# Spinorial idempotents in the Clifford algebras of Minkowski spaces

Wojciech Kopczyński
*Institute of Theoretical Physics, Warsaw University, Hoża 69, 00-681 Warsaw, Poland*

(Received 13 April 1988; accepted for publication 24 August 1988)

Algebraic spinor spaces determined by spinorial idempotents are considered in the case of two- and four-dimensional, real and complexified Minkowski spaces. The form of the idempotents and their classification is given.

## I. INTRODUCTION

A spinor space is conventionally defined as a space in which a representation of a Clifford algebra acts irreducibly and equivalence of spinor spaces is understood in the sense of the representation theory. It follows that if the Clifford algebra is simple (for instance, in the even-dimensional case), all spinor spaces are equivalent.

There is, however, another approach to spinors. The algebraic spinor spaces are defined as minimal left ideals of a Clifford algebra. They provide a representation space for the Clifford algebra and the groups Pin and Spin related to that algebra. In the mathematical literature this approach to spinors was put forward by Chevalley[1] and in the physical literature by Sommerfeld[2] and Sauter.[3] For more information about the history of various approaches to spinors consult Budinich and Trautman.[4] It was also pointed by Banks, Dothan, and Horn[5] that the "ideal" approach may be useful in the description of fermionic generations.

For algebraic spinor spaces, as it was emphasized by Graf,[6] there is another notion of equivalence than that from the representation theory. The algebraic spinor spaces are equivalent if there exists a Clifford automorphism (i.e., an element of the orthogonal group) that maps one space onto the other. Nonequivalent spinor spaces might in principle correspond to different fermions. Motivated by this idea, I consider in this paper a step in the direction of a classification of algebraic spinor spaces. Namely, a classification of spinorial idempotents generating algebraic spinor spaces is given. The classification is based on topological properties of the space of orbits of the orthogonal group.

In Sec. II, I discuss some general definitions. In Sec. III, I consider spinorial idempotents, which determine algebraic spinor spaces, in the two-dimensional Minkowski space for the real and complexified cases. Section IV is devoted to spinorial idempotents in the four-dimensional Minkowski space. In Secs. IV A and IV B, I consider the two real cases corresponding to the signatures $(+,+,+,-)$ and $(+,-,-,-)$, respectively, and in Sec. IV C, the complexified case is considered.

## II. GENERAL NOTIONS

Let $V$ be a real vector space and $g$ a bilinear symmetric form on $V$. We consider a two-sided ideal $J_g$ in the tensor algebra $T(V)$ generated by the elements of the form $u\otimes u - g(u,u)$, where $u\in V$. The Clifford algebra $C(V,g)$ is the quotient algebra $T(V)/J_g$. The natural projection of the tensor algebra onto $C(V,g)$ when restricted to skew tensors is one-to-one. Therefore we have a natural linear space isomorphism between the exterior algebra and the Clifford algebra, $\chi\colon \Lambda V\to C(V,g)$ and the direct sum decomposition

$$C(V,g)=\bigoplus_{p=0}^{n}\chi(\overset{p}{\Lambda}V),\quad n=\dim V.$$

The Clifford algebra is $\mathbb{Z}_2$ graded, $C(V,g)=C^+(V,g)\oplus C^-(V,g)$, where

$$C^+(V,g)=\bigoplus_{p=0}^{[n/2]}\chi(\overset{2p}{\Lambda}V),$$

$$C^-(V,g)=\bigoplus_{p=0}^{[(n-1)/2]}\chi(\overset{2p+1}{\Lambda}V).$$

We write $|s|=0$ if $s\in C^+$ and $|s|=1$ if $s\in C^-$.

The Clifford automorphism $f$ of $C(V,g)$ is an automorphism in the category of algebras with unity such that there exists a linear automorphism $f'$ of $V$ and the following diagram is commutative:

$$\begin{array}{ccc} V & \xrightarrow{f'} & V \\ \downarrow & & \downarrow \\ C(V,g) & \xrightarrow{f} & C(V,g) \end{array}$$

The vertical arrows denote the natural inclusion. One can easily show that $f'$ is an orthogonal transformation. Conversely, for each $f'\in\mathbb{O}(V,g)$ there exists a unique Clifford automorphism $f$; so we can identify the group of Clifford automorphisms with the orthogonal group, $f\in\mathbb{O}(V,g)$. The Clifford automorphisms have the following form:

$$f_s(c)=(-1)^{|s|\,|c|}scs^{-1},$$

where $s$ is an element of $C(V,g)$ of the form

$$s=u_i\cdots u_k,\quad u_i\in V,\quad g(u_i,u_i)=\pm 1.$$

Such elements constitute the group $\mathrm{Pin}(V,g)$. We have the epimorphism $\mathrm{Pin}(V,g)\ni s\to f_s\in\mathbb{O}(V,g)$, its kernel is $\mathbb{Z}_2$. We shall be interested mainly in the connected component of unity $\mathbb{SO}^\uparrow(V,g)$ and its counterimage $\mathrm{Spin}_0(V,g)$.

A spinorial idempotent $p$ is an element of $C(V,g)$ such that $p^2=p$ and the rank of the linear map $C(V,g)\ni c\to cp\in C(V,g)$ is minimal nonzero. A spinorial idempotent $p$ generates a minimal left ideal (spinor space)

$$\mathscr{L}_p=C(V,g)p=\{c\in C(V,g):\ cp=p\}.$$

Any minimal left ideal is generated by a spinorial idempotent, but such an idempotent is not, in general, uniquely determined by the ideal.

The spinorial idempotents $p$ and $p'$ are equivalent, $p\sim p'$,

 0022-2488/89/020243-06$02.50 © 1989 American Institute of Physics 

if there exists $f \in \mathbb{O}(V,g)$ such that $fp = p'$. They are strongly equivalent, $p \approx p'$, if moreover $f \in \mathbb{SO}^{\uparrow}(V,g)$.

We shall consider also the complexified Clifford algebra $C(V,g)^{\mathbb{C}}$, the complex algebraic spinor spaces, and complex spinorial idempotents; the equivalence of spinorial idempotents will be defined under the action of *real* orthogonal groups.

## III. SPINORIAL IDEMPOTENTS IN TWO-DIMENSIONAL MINKOWSKI SPACE

Let $(k,l)$ be a null basis of two-dimensional Minkowski space $V(1,1)$: $g(k,k) = g(l,l) = 0, g(k,l) = \frac{1}{2}$. It generates a basis

$$1,\ k,\ l,\ (kl - lk)/2$$

of the Clifford algebra $C(1,1)$. We look for solutions of the equation $p^2 = p$ such that the rank of the map $c \mapsto cp$ is 1 or, equivalently, $p \neq 0,1$. The solutions have the form

$$p = \tfrac{1}{2} + bk + cl + d[(kl - lk)/2],$$

where the parameters $b,c,d$ are constrained by

$$bc = \tfrac{1}{4}(1 - d^2).$$

First, we consider the strong equivalence of real spinorial idempotents. The action of the special orthochronous Lorentz group $\mathbb{SO}^{\uparrow}(1,1)$ is given by $k \mapsto e^{\psi}k$, $l \mapsto e^{-\psi}l$. It acts on the topological space $M_{\mathbb{R}}(1,1)$ of real spinorial idempotents and $M_{\mathbb{R}}(1,1)/\mathbb{SO}^{\uparrow}(1,1)$ is the set of strongly nonequivalent classes of spinorial idempotents. Let us note that we have the following invariants under the action of $\mathbb{SO}^{\uparrow}(1,1)$:

$$d,\quad \operatorname{sgn} b,\quad \operatorname{sgn} c.$$

Considering the quotient topology on $M_{\mathbb{R}}(1,1)/\mathbb{SO}^{\uparrow}(1,1)$ leads to an interesting classification of spinorial idempotents.

*(1) Continuous idempotents:* These are idempotents for which $d \neq \mp 1$. The orbits of $\mathbb{SO}^{\uparrow}(1,1)$ are grouped into six branches. A branch is given, if we specify one of the open intervals $(-\infty,-1)$, $(-1,+1)$, $(+1,+\infty)$ to which belongs the invariant $d$ and sgn $b$ (or sgn $c$—both are related). Each orbit of this type is one-dimensional.

*(2) Bridge-type idempotents:* These are idempotents for which $d = \mp 1$ and $b,c$ do not vanish simultaneously. There are eight orbits of this type, each is one-dimensional. Each open neighborhood of such an orbit contains orbits belonging to at least two branches of type (1).

*(3) Multicross-type idempotents:* These are

$$p = \tfrac{1}{2} \mp \tfrac{1}{2}(kl - lk)$$

and lead to left ideals used by Chevalley[1] for the case of the neutral signature. Let us note that there is only one element of an orbit of this type. Each neighborhood of an orbit of this type contains orbits belonging to four branches of type (1) and four orbits of type (2).

In order to consider $M_{\mathbb{R}}(1,1)/\mathbb{O}(1,1)$ we have to take into account the action of a space reflection $k \leftrightarrow l$ and of the space-time reflection $k \mapsto -k$, $l \mapsto -l$ on $M_{\mathbb{R}}(1,1)/\mathbb{SO}^{\uparrow}(1,1)$. On Fig. 1, the space reflection and the space-time reflection are represented as reflections with respect to the vertical and horizontal symmetry axes, respectively. Generically an orbit of $\mathbb{O}(1,1)$ consists of four orbits of $\mathbb{SO}^{\uparrow}(1,1)$. There are two exceptional orbits of $\mathbb{O}(1,1)$ that consist of two orbits of $\mathbb{SO}(1,1)$. These are the orbit of multicross-type idempotents and the orbit for which $d = 0$. The last orbit is of boundary type in the topological space $M_{\mathbb{R}}(1,1)/\mathbb{O}(1,1)$. Therefore we can distinguish in the internal branches the continuous idempotents of *boundary type.*

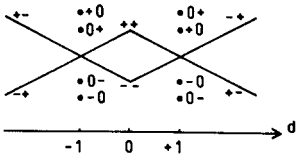



FIG. 1. The orbit space $M_{\mathbb{R}}(1,1)/\mathbb{SO}^{\uparrow}(1,1)$. The pair sgn$(b)$, sgn$(c)$ is indicated.

Next we consider the space $M_{\mathbb{C}}(2)$ of complex idempotents. They are given by the same equations as real ones, the parameters $b,c,d$ now being complex. The space $M_{\mathbb{C}}(2)/\mathbb{SO}(1,1)$ consists of (1) one three-dimensional sheet of continuous idempotent orbits $(d \neq \mp 1)$, (2) four one-dimensional circles of orbits of bridge-type idempotents $(d = \mp 1$; $b, c$ do not vanish simultaneously), and (3) two multicross-type idempotents $(d = \mp 1, b = c = 0)$. The orbits of continuous type and of bridge type are one-dimensional, whereas orbits of multicross type consist of single points. Reflections distinguish the complex idempotents of boundary type for which $d = 0$.

## IV. SPINORIAL IDEMPOTENTS IN FOUR-DIMENSIONAL MINKOWSKI SPACES

We shall look for spinorial idempotents in the real Clifford algebras $C(1,3)$ (signature $+,-,-,-$), $C(3,1)$ (signature $+,+,+,-$), and in their complexification $C(1,3)^{\mathbb{C}}$.

Let $(e_\mu)$ be an orthonormal basis of the Minkowski space and let

$$e_5 = e_0e_1e_2e_3 = -(1/4!)\eta^{\mu\nu\rho\sigma}e_\mu e_\nu e_\rho e_\sigma,$$

where $\eta^{\mu\nu\rho\sigma}$ is skew and normalized by the condition $\eta^{0123} = -1$. An idempotent $p$ can be decomposed as follows:

$$p = a + u^\mu e_\mu + \tfrac{1}{2}f^{\mu\nu}e_\mu e_\nu + v^\mu e_\mu e_5 + be_5.$$

The condition $p^2 = p$ leads to the following set of equations:

$$\alpha^2 - b^2 + u^2 + v^2 - \tfrac{1}{2}f_{\mu\nu}f^{\mu\nu} = \tfrac{1}{4}, \tag{1}$$

$$\tfrac{1}{2}f_{\mu\nu}{}^*f^{\mu\nu} = -2\alpha b, \tag{2}$$

$${}^*f^{\mu\nu}v_\nu = \alpha u^\mu, \tag{3}$$

$${}^*f^{\mu\nu}u_\nu = -\alpha v^\mu, \tag{4}$$

$$af^{\mu\nu} + b\,{}^*f^{\mu\nu} = -\eta^{\mu\nu\rho\alpha}u_\rho v_\sigma, \tag{5}$$

where $\alpha = a - \frac{1}{2}$ and ${}^*f_{\mu\nu} = \frac{1}{2}\eta_{\mu\nu\rho\sigma}f^{\rho\sigma}$. Dualization of (5) gives

$$-bf^{\mu\nu} + \alpha\,{}^*f^{\mu\nu} = u^\mu v^\nu - v^\mu u^\nu \tag{5'}$$

and both (5) and (5′) lead to

$$(\alpha^2 + b^2)f^{\mu\nu} = -b(u^\mu v^\nu - v^\mu u^\nu) - \alpha\eta^{\mu\nu\rho\sigma}u_\rho v_\sigma. \quad (6)$$

Assume that $\alpha \neq 0$, $\alpha^2 + b^2 \neq 0$. Under the second assumption we can eliminate $f_{\mu\nu}$. Rejecting the trivial solutions $p = 0,1$, Eqs. (1)–(5) are satisfied if and only if the vectors $u$, $v$ satisfy the constraints

$$uv = 0, \quad (7)$$

$$u^2 = v^2 = \alpha^2 + b^2, \quad (8)$$

and moreover $\alpha^2 = \frac{1}{16}$. The last condition leads either to

$$\alpha = \tfrac{1}{4} \quad \text{and} \quad a = \tfrac{3}{4}$$

or to

$$\alpha = -\tfrac{1}{4} \quad \text{and} \quad a = \tfrac{1}{4}. \quad (9)$$

The case (9) corresponds to idempotents of the minimal rank 1 if Eqs. (7) and (8) have solutions.

If the signature is (1,3) there are no real solutions of (7) and (8).

### A. Signature (3,1)

In the case of the signature (3,1) the real solutions of (7) and (8) exist and give all minimal idempotents of $C(3,1) \cong \mathbb{R}(4)$. Each orbit of the action of $\mathrm{SO}^\uparrow(3,1)$ on the space of spinorial idempotents $M_{\mathbf{R}}(3,1)$ is five-dimensional. The space of orbits $M_{\mathbf{R}}(3,1)/\mathrm{SO}^\uparrow(3,1)$ is one-dimensional, the unique invariant is the parameter $b$. The unique invariant under the action of the full Lorentz group is $|b|$ and once more we can distinguish idempotents of the boundary type for which $b = 0$.

### B. Signature (1,3)

In the case of signature (1,3) we have to look for idempotents of rank 2 of $C(1,3) \cong \mathbb{H}(2)$ for which $a = \frac{1}{2}$ and $\alpha = 0$. Since $\alpha = 0$, the bivector $f^{\mu\nu}$ is simple. If moreover $b \neq 0$, then we get from (5)

$$f^{\mu\nu} = -(1/b)(u^\mu v^\nu - v^\mu u^\nu). \quad (10)$$

Equations (2)–(4) are satisfied identically and (1) gives the biquadratic equation for $b$,

$$b^4 - \beta b^2 + \gamma = 0, \quad (11)$$

where $\beta = u^2 + v^2 - \frac{1}{4}$ and $\gamma = u^2v^2 - (uv)^2$. In the case $\gamma > 0$, (11) could have solutions under the conditions $\beta > 0$ and $\beta^2 > 4\gamma$. But $\gamma > 0$ implies that both vectors $u$ and $v$ are spacelike, therefore $\beta > 0$ cannot be satisfied.

*Types $II_T^+$ and $II_T^-$:* In the case $\gamma < 0$, i.e.,

$$u^2v^2 - (uv)^2 < 0, \quad (12)$$

we have two solutions $\pm b$ of (11). In this case the vectors $u,v$ span a two-dimensional timelike surface. We have two disjoint types: $\mathrm{II}_T^+$ for which $b > 0$ and $\mathrm{II}_T^-$ for which $b < 0$. Equation (12) restricts the values of the scalar products $u^2, v^2, uv$ to the exterior of a cone in three-dimensional space; it can be written as

$$t^2 - r^2 < 0, \quad (12')$$

where we have introduced the new parameters:

$$t = u^2 + v^2$$

and $r,\phi$ defined by

$$u^2 - v^2 + 2i(uv) = re^{i\phi}.$$

If we turn around the cone following a continuous closed curve, the vectors $u,v$ become $-u, -v$ and they determine a different orbit of $\mathrm{SO}^\uparrow(1,3)$ than that determined by $u,v$. The space of orbits in both types is a $\mathbb{Z}_2$ bundle over exterior of the cone. As parameters in $\mathrm{II}_T^+$ and $\mathrm{II}_T^-$ we can take $t$, $r$ satisfying the inequality (12′) and the angle $\phi$, $0 \leqslant \phi < 4\pi$. The dimension of the orbit is 5.

*Types $I_{T,ext}^+$ and $I_{T,ext}^-$:* In the case $\gamma = 0$ the solutions $\pm b \neq 0$ of (11) exist under the condition $\beta > 0$. Therefore the vectors $u,v$ span a one-dimensional timelike subspace and the parameters $t$, $r$ satisfy the restrictions

$$t = r > \tfrac{1}{4}. \quad (13)$$

So $r > \frac{1}{4}$ and $0 \leqslant \phi < 4\pi$ are parameters on the space of orbits for types $\mathrm{I}_{T,\mathrm{ext}}^+$ (for which $b > 0$) and $\mathrm{I}_{T,\mathrm{ext}}^-$ (for which $b < 0$). Note that $f^{\mu\nu} = 0$ according to (10). The dimension of the orbit is 3.

Consider now the case $b = 0$. This implies $u \| v$. At first assume that the subspace spanned by $u$ and $v$ is one-dimensional and let $n$ be a nonzero vector in this subspace. Then, from (2)–(4), $f^{\mu\nu}$ has the form

$$f^{\mu\nu} = w^\mu n^\nu - n^\mu w^\nu.$$

It remains to solve (1), which becomes

$$2(nw)^2 - 2n^2w^2 = \tfrac{1}{4} - t. \quad (14)$$

*Types $I_{T,int}$ and $I_{T,circle}$:* If the one-dimensional subspace spanned by $u$ and $v$ is timelike, we can determine $n^\mu$ uniquely, requiring that it is future oriented and normalized to 1. Then $w^\mu$ can be determined by the requirement $wn = 0$. Then, Eq. (14) can be solved if

$$0 < t = r < \tfrac{1}{4} \quad (15)$$

and if

$$t = r = \tfrac{1}{4}. \quad (16)$$

In the limiting case (16) $f^{\mu\nu} = 0$. The range of the parameter $\phi$ is form 0 to $4\pi$. We have two types: $\mathrm{I}_{T,\mathrm{int}}$ corresponding to the case (15) and $\mathrm{I}_{T,\mathrm{circle}}$ corresponding to the case (16). The dimension of the orbit in the type $\mathrm{I}_{T,\mathrm{int}}$ is 5 and in the type $\mathrm{I}_{T,\mathrm{circle}}$ is 3.

*Type $I_s$:* If the one-dimensional subspace spanned by the vectors $u$ and $v$ is spacelike, we can determine $w^\mu$ uniquely requiring that $wn = 0$ and $w$ is future oriented and normalized to 1. Then $f^{\mu\nu}$ determines uniquely $n^\mu$. We have in this case

$$-t = r > 0, \quad (17)$$

and the range of $\phi$ is from 0 to $4\pi$ since the triads $n,u,v$ and $n, -u, -v$ determine different orbits. The dimension of the orbit is 5.

*Types $I_N^\uparrow$ and $I_N^\downarrow$:* If the one-dimensional subspace spanned by $u$ and $v$ is null, we can determine $n$ uniquely requiring that $n$ be future oriented and $u + iv = zn$, where $|z| = 1$. Then, if we impose the condition $w^2 = 0$, the vector $w$ will be chosen uniquely. From (14) we get $(nw)^2 = \frac{1}{8}$. We have two disjoint types: $\mathrm{I}_N^\uparrow$ for which $w$ is future oriented and

$I_N^{\downarrow}$ for which $w$ is past oriented. The parameter is $z \in U(1)$ and the dimension of the orbit is 5.

*Type 0:* The last type, for which $u^\mu = v^\nu = b = 0$. The bivector $f^{\mu\nu}$ is simple and satisfies

$$-\tfrac{1}{2} f^{\mu\nu} f_{\mu\nu} = \tfrac{1}{4},$$

so it spans a two-dimensional timelike surface. There is only one orbit of this type. The dimension of the orbit is 4.

Figure 2 displays all types and relations between them. An arrow on this diagram means that there exists a sequence converging to a type pointed by it. We do not draw an arrow if it is possible to pass from one type to another with the help of arrows already indicated.

### C. Complex idempotents

In the complex case $C(1,3)^{\mathbb{C}} \cong \mathbb{C}(4)$ and the rank of minimal idempotents is 1, so $a = \frac{1}{4}$. Therefore Eqs. (6)–(8) give idempotents for which $b \neq \pm i/4$. Let us discuss this generic case.

We shall introduce the real vectors $\pi, \tau, \rho, \sigma$ defined by

$$u = (\pi + \sigma)/2 + i[(\rho - \tau)/2]$$

$$v = (\rho + \tau)/2 + i[(\pi - \sigma)/2].$$

Equations (7) and (8) now take the form

$$\pi^2 = \tau^2, \quad \pi\tau = 0, \quad \rho^2 = \sigma^2, \quad \rho\sigma = 0,$$

$$\tfrac{1}{2}(\pi\sigma + \rho\tau) = m, \quad \tfrac{1}{2}(\pi\rho - \sigma\tau) = n,$$

where

$$b^2 + \tfrac{1}{16} = m + in.$$

The numbers $\pi^2$, $\tau^2$, $m$, $n$, and

$$x = \tfrac{1}{2}(\pi\sigma - \rho\tau), \quad y = \tfrac{1}{2}(\pi\rho + \sigma\tau),$$

parametrize all nontrivial scalar products. In the discussion below we shall also use the polar coordinates $r_1$, $\phi_1$, $r_2$, $\phi_2$ defined by

$$m + in = r_1 \exp(i\phi_1), \quad x + iy = r_2 \exp(i\phi_2).$$

The scalar products of the vectors $\pi,\tau,\rho,\sigma$ belong to the space with signature (1,3) if and only if $\pi^2 \leqslant 0$, $\rho^2 \leqslant 0$, and

$$\Delta = \begin{vmatrix} \pi^2 & 0 & \pi\rho & \pi\sigma \\ 0 & \pi^2 & \tau\rho & \tau\sigma \\ \pi\rho & \tau\rho & \rho^2 & 0 \\ \pi\sigma & \tau\sigma & 0 & \rho^2 \end{vmatrix} \leqslant 0. \tag{18}$$

The determinant $\Delta$ can be factorized as

$$\Delta = (\pi^2\rho^2 - (r_1 - r_2)^2)(\pi^2\rho^2 - (r_1 + r_2)^2)$$

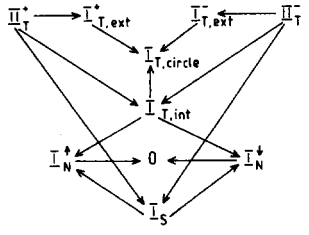



FIG. 2. Types of spinorial idempotents in the Clifford algebra $C(1,3)$ and relations between them.

so (18) can be written as

$$(r_1 - r_2)^2 \leqslant \pi^2\rho^2 \leqslant (r_1 + r_2)^2.$$

If $\pi^2 < 0$, we can introduce the orthogonal projections $\rho_\perp$, $\sigma_\perp$ of the vectors $\rho$, $\sigma$ onto the surface orthogonal to the surface spanned by $\pi$ and $\tau$. Their scalar products are

$$\begin{aligned} \pi^2\rho_\perp{}^2 &= Q + 2r_1r_2\cos(\phi_1 + \phi_2), \\ \pi^2\sigma_\perp{}^2 &= Q - 2r_1r_2\cos(\phi_1 + \phi_2), \\ \pi^2\rho_\perp\sigma_\perp &= -2r_1r_2\sin(\phi_1 + \phi_2), \end{aligned} \tag{19}$$

where $Q = \pi^2\rho^2 - r_1{}^2 - r_2{}^2$ satisfies $-2r_1r_2 \leqslant Q \leqslant 2r_1r_2$. If $\rho^2 < 0$, we can introduce orthogonal projections $\pi_\perp$, $\tau_\perp$ of the vectors $\pi,\tau$ onto the surface orthogonal to the surface spanned by $\rho$ and $\sigma$ and their scalar products are

$$\begin{aligned} \rho^2\pi_\perp{}^2 &= Q - 2r_1r_2\cos(\phi_1 - \phi_2), \\ \rho^2\tau_\perp{}^2 &= Q + 2r_1r_2\cos(\phi_1 - \phi_2), \\ \rho^2\pi_\perp\tau_\perp &= 2r_1r_2\sin(\phi_1 - \phi_2). \end{aligned} \tag{20}$$

*Types $IV^+$ and $IV^-$:* The space spanned by the vectors $\pi$, $\rho$, $\sigma$, and $\tau$ is four-dimensional. The parameters Re $b$, Im $b$, $\pi^2$, $\rho^2$, $x$, and $y$ satisfy the restrictions $\pi^2 < 0$, $\rho^2 < 0$, $r_1 > 0$, $r_2 > 0$, and $(r_1 - r_2)^2 < \pi^2\rho^2 < (r_1 + r_2)^2$. The orbits corresponding to different space-time orientations of the basis $(\pi,\tau,\rho,\sigma)$ form disjoint subsets in the topology of $M_{\mathbb{C}}(4)/\mathrm{SO}^{\uparrow}(1,3)$. Thus we have subtypes $IV^+$ and $IV^-$. It can be seen from (19) and (20) that the continuous change of $\phi_1$ or $\phi_2$ from 0 to $2\pi$ leads to the change of all vectors $\pi$, $\tau$, $\rho$, and $\sigma$. Thus there is no further subdivision of the type IV. To describe fully the configuration we need a $\mathbb{Z}_2$ bundle over the space of the parameters $b$, $x + iy$, i.e., over $(\mathbb{C} - \{-i/4, +i/4\}) \times (\mathbb{C} - \{0\})$, which is nontrivial if we turn $b$ once around $-i/4$ or $+i/4$ or if we turn $x + iy$ once around 0. A point of $M_{\mathbb{C}}(4)/\mathrm{SO}^{\uparrow}(1,3)$ of type $IV^+$ or $IV^-$ corresponds to a point of this bundle and to the values of the parameters $\pi^2$ and $\rho^2$. The dimension of the orbit is 6.

*Type $III_S$:* The space spanned by $\pi,\tau,\rho,\sigma$ is three-dimensional spacelike. The parameters $b$, $x$, $y$, and $\pi^2$ satisfy the restrictions $\pi^2 < 0$, $r_1 > 0$, and $\rho^2$ is determined by $\pi^2\rho^2 = (r_1 + r_2)^2$. The bundle structure remains the same as in type IV. The dimension of the orbit is 6.

*Type $III_T$:* The space spanned by $\pi$, $\tau$, $\rho$, $\sigma$ is three-dimensional timelike. The parameters $b$, $x$, $y$, and $\pi^2 + \rho^2$ satisfy the restrictions $\pi^2 + \rho^2 < 0$, $r_1 > 0$, and $r_2 > 0$. The squares $\pi^2$ and $\rho^2$ satisfy the condition $\pi^2\rho^2 = (r_1 - r_2)^2$. The bundle structure remains as in type IV. The dimension of the orbit is 6.

*Type $II_T$:* The space spanned by $\pi$, $\tau$, $\rho$, $\sigma$ is two-dimensional timelike. Here $r_1 = r_2 > 0$ and $\pi^2 = \rho^2 = 0$. We now have a $\mathbb{Z}_2$ bundle over the space of the parameters $b$ and $\phi_2$, i.e., over $(\mathbb{C}\{-i/4, +i/4\}) \times \mathbb{U}(1)$. The dimension of the orbit is 5.

*Types $III_N^{\pm}$, $III_N^{-}$, and $II_S$:* They correspond to $r_2 = 0$, $r_1 > 0$, which implies $\pi^2\rho^2 = r_1{}^2$. It follows from (19) that the space spanned by $\rho_\perp$ and $\sigma_\perp$ is either one-dimensional null (type $III_N$) or both vectors $\rho_\perp$ and $\sigma_\perp$ vanish (type $II_S$). Since there are two null directions in the space orthogonal to that spanned by $\pi$ and $\tau$, type $III_N$ decomposes on subtypes

$III_N^+$ and $III_N^-$. If we choose a vector $k$ so that $\rho_\perp + i\sigma_\perp = k\exp(i\psi)$, the angle $\psi$ is the only one invariant constructed out of $\rho_\perp$ and $\sigma_\perp$. The parameters in types $III_N^+$ and $III_N^-$ are $b$, $\pi^2 < 0$, and the angle $\psi$; the dimension of the orbit is 6. In type $II_s$ the parameters are $b$ and $\pi^2 < 0$; the dimension of the orbit is 5.

Consider now the case $b = \pm i/4$. Equation (5) now gives only one part of $f^{\mu\nu}$: self-dual or anti-self-dual. We have

$$b = -i/4, \quad \overset{+}{f}{}^{\mu\nu} = -2i(u^\mu v^\nu - v^\mu u^\nu), \tag{21}$$

or

$$b = +i/4, \quad \overset{-}{f}{}^{\mu\nu} = 2i(u^\mu v^\nu - v^\mu u^\nu). \tag{22}$$

Equations (6), (3), and (4) lead to

$$u^2 = v^2 = uv = 0,$$

which is the limiting case of (7) and (8). Thus we can use the same parametrization of the vectors $u$ and $v$ as in the generic case; $b = \pm i/4$ leads to $r_1 = 0$ and $\pi^2\rho^2 = r_2^2$.

We shall use a null basis $(k,l,m,\bar{m})$ such that $kl = \frac{1}{2}$, $m\bar{m} = -\frac{1}{2}$ and all other products vanish; the real vectors $k$ and $l$ are future oriented and the space-time orientation of the null basis is established by

$$\eta_{\mu\nu\rho\sigma} k^\mu l^\nu m^\rho \bar{m}^\sigma = i/4.$$

If $\mu_1 = \sqrt{-\pi^2} \neq 0$, the vector $m$ can be uniquely determined by

$$m = (1/2\mu_1)(\pi + i\tau).$$

We shall also use the notation $\mu_2 = \sqrt{-\rho^2}$. In the case when the space spanned by $\pi, \tau, \rho, \sigma$ is three-dimensional and null, we can choose the vectors $k$ and $l$ uniquely requiring that $\rho_\perp + i\sigma_\perp = k\exp(i\psi)$ or $\rho_\perp + i\sigma_\perp = l\exp(i\psi)$. This leads either to

$$\begin{aligned} u &= \tfrac{1}{2}k\exp(-i\psi) + (\mu_1 - \mu_2\exp(i\phi_2))\bar{m}, \\ v &= \tfrac{1}{2}k\exp(-i\psi) + i(\mu_1 + \mu_2\exp(i\phi_2))\bar{m}, \end{aligned} \tag{23}$$

or to

$$\begin{aligned} u &= \tfrac{1}{2}l\exp(-i\psi) + (\mu_1 - \mu_2\exp(i\phi_2))\bar{m}, \\ v &= \tfrac{1}{2}l\exp(-i\psi) + i(\mu_1 + \mu_2\exp(i\phi_2))\bar{m}. \end{aligned} \tag{24}$$

Since $k \wedge \bar{m}$ is self-dual and $l \wedge \bar{m}$ is anti-self-dual, (23) corresponds to the case (21) and (24) to the case (22). Consider $b = -i/4$. Anti-self-dual bivectors have a basis

$$k \wedge l - m \wedge \bar{m}, \quad k \wedge m, \quad l \wedge \bar{m}.$$

Substituting $\overset{-}{f}{}^{\mu\nu}$, expressed as a linear combination of this basis, into (3) and (4) and solving these equations, we obtain

$$\begin{aligned} \overset{-}{f}{}^{\mu\nu} = &-\tfrac{1}{2}(k^\mu l^\nu - l^\mu k^\nu - m^\mu \bar{m}^\nu + \bar{m}^\mu m^\nu) \\ &+ 2i\mu_2\exp(i(\psi + \phi_2))(l^\mu \bar{m}^\nu - \bar{m}^\mu l^\nu). \end{aligned} \tag{25}$$

Consider $b = i/4$. The self-dual bivector $\overset{+}{f}{}^{\mu\nu}$ can be written as a combination of

$$k \wedge l - m \wedge \bar{m}, \quad k \wedge \bar{m}, \quad l \wedge m.$$

From (3) and (4) we obtain

$$\begin{aligned} \overset{+}{f}{}^{\mu\nu} = &\tfrac{1}{2}(k^\mu l^\nu - l^\mu k^\nu + m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu) \\ &+ 2i\mu_2\exp(i(\psi + \phi_2))(k^\mu \bar{m}^\nu - \bar{m}^\mu k^\nu). \end{aligned} \tag{26}$$

Equations (1) and (2) are in both cases satisfied identically.

*Types $III_N$ - and $III_N$ +:* The space spanned by $\pi, \tau, \rho, \sigma$ is three-dimensional null and $b = -i/4$ (type $III_N$ -) or $b = i/4$ (type $III_N$ +). The parameters are $x, y, \pi^2 + \rho^2 < 0$ and the angle $\psi$. The squares $\pi^2$, $\rho^2$ satisfy the condition $\pi^2\rho^2 = r_2^2$. If $\pi^2 < 0$ the solution is given: for the type $III_N$ - by (21), (23), and (25), whereas for the type $III_N$ + by (22), (24), and (26). If the space spanned by $\pi, \tau$ is null (so $\pi^2 = 0$), we can use $\rho\sigma$ instead of $\pi, \tau$ in the definition of the basis vector $m$. This leads to replacements $u \leftrightarrow v$ and $0 = \mu_1 \leftrightarrow \mu_2$ in (23) and (24) and to the change of the overall sign in (25) and (26). The dimension of the orbit is 6.

If $r_2 > 0$ but the space spanned by $\pi, \tau, \rho, \sigma$ is two-dimensional spacelike (3) and (4) cannot be satisfied.

*Types $II^1_{S--}$, $II^1_{S-+}$, $II^1_{S+-}$, $II^1_{S++}$, and $II^0_{S--}$, $II^0_{S-+}$, $II^0_{S+-}$, $II^0_{S++}$:* The space spanned by $\pi, \tau, \rho, \sigma$ is two-dimensional spacelike and $b = -i/4$ (types $II_{S--}$ and $II_{S-+}$) or $b = i/4$ (types $II_{S+-}$ and $II_{S++}$). This can be achieved only if $\rho = \sigma = 0$ (types $II_{S--}$ and $II_{S+-}$) or $\pi = \tau = 0$ (types $II_{S-+}$ and $II_{S++}$). For the types $II_{S--}$ and $II_{S+-}$ we have

$$v = iu = i\mu_1\bar{m}$$

and for the types $II_{S-+}$ and $II_{S++}$

$$u = iv = i\mu_2\bar{m}.$$

Equations (3) and (4) give for the types $II_{S--}$ and $II_{S-+}$ (the upper sign applies for the first of these types):

$$\begin{aligned} \overset{-}{f}{}^{\mu\nu} = &\mp\tfrac{1}{2}(k^\mu l^\nu - l^\mu k^\nu - m^\mu \bar{m}^\nu + \bar{m}^\mu m^\nu) \\ &+ \delta(l^\mu \bar{m}^\nu - \bar{m}^\mu l^\nu) \end{aligned} \tag{27}$$

and for the types $II_{S+-}$ and $II_{S++}$

$$\begin{aligned} \overset{+}{f}{}^{\mu\nu} = &\pm\tfrac{1}{2}(k^\mu l^\nu - l^\mu k^\nu + m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu) \\ &+ \delta(k^\mu \bar{m}^\nu - \bar{m}^\mu k^\nu), \end{aligned} \tag{28}$$

where $\delta$ is arbitrary. Equations (1) and (2) are satisfied identically. If $\phi \neq 0$ (subtypes $II^1_S$) we can, performing a Lorentz boost in the plane spanned by $k$ and $l$, normalize it. So for the types $II^1_S$ the parameters are either $\pi^2 < 0$ (then $\rho^2 = 0$) or $\rho^2 < 0$ (then $\pi^2 = 0$) and the parameter $\delta \in U(1)$; the dimension of the orbit is 6. For the types $II^0_S$ (i.e., if $\delta = 0$) there is only one parameter: either $\pi^2 < 0$ or $\rho^2 < 0$; the dimension of the orbit is 5.

*Types $I_{N--}$, $I_{N-+}$, $I_{N+-}$, and $I_{N++}$:* The space spanned by $\pi, \tau, \rho, \sigma$ is one-dimensional null. From (3) and (4) both vectors $u, v \neq 0$, so we can fix the basis vector $k$ so that $u = k\exp(-i\psi)$. Equations (3) and (4) can be satisfied if and only if $v = iu$ (types $I_{N--}$ and $I_{N+-}$) or $u = iv$ (types $I_{N-+}$ and $I_{N++}$). We can choose the null basis so that (27) and (28) hold with $\delta = 0$; (27) holds for the types $I_{N--}$, $I_{N-+}$, (28) holds for the types $I_{N+-}$ and $I_{N++}$

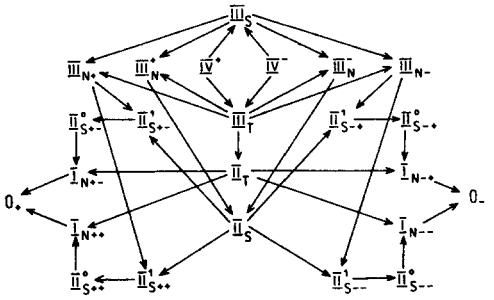


FIG. 3. Types of spinorial idempotents in the Clifford algebra $C(1,3)^{\mathbb{C}}$ and relations between them.

(respecting the signs in both cases). The angle $\psi$ is the single parameter. The dimension of the orbit is 5.

*Types $0_-$ and $0_+$:* The vectors $u$ and $v$ vanish. Equations (1) and (2) remain to be solved. They lead to $\bar{f}^{\mu\nu}\bar{f}_{\mu\nu} = -\frac{1}{4}$ (type $0_-$ if $b = -i/4$ so $\overset{+}{f}{}^{\mu\nu} = 0$) or $\overset{+}{f}{}^{\mu\nu}\overset{+}{f}_{\mu\nu} = -\frac{1}{4}$ (type $0_+$ if $b = i/4$ so $\bar{f}^{\mu\nu} = 0$). A convenient choice of the null tetrad leads in the case $0_-$ to

$$\bar{f}^{\mu\nu} = -\tfrac{1}{2}(k^\mu l^\nu - l^\mu k^\nu - m^\mu \bar{m}^\nu + \bar{m}^\mu m^\nu)$$

and in the case $0_+$ to

$$\overset{+}{f}{}^{\mu\nu} = -\tfrac{1}{2}(k^\mu l^\nu - l^\mu k^\nu + m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu).$$

There are no continuous parameters in both types $0_-$ and $0_+$. The dimension of the orbit is 4.

Figure 3 shows the relations between various types of the complex idempotents.

## ACKNOWLEDGMENT

The research reported in this paper was partially supported by the Polish Research Project CPBP 01.03.

[1]C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U.P., New York, 1954).
[2]A. Sommerfeld, *Atombau und Spektrallinien* (Vieweg, Braunschweig, 1951).
[3]F. Sauter, Z. Phys. **63**, 803 (1930); **64**, 295 (1930).
[4]P. Budinich and A. Trautman, *The Spinorial Chessboard* (Springer, Berlin, 1988).
[5]T. Banks, Y. Dothan, and D. Horn, Phys. Lett. B **117**, 413 (1982).
[6]W. Graf, Ann. Inst. H. Poincaré A **29**, 85 (1978).

# The Clifford algebra of an infinite-dimensional space

G. P. Wene
*Division of Mathematics, Computer Science, and Systems Design, The University of Texas at San Antonio, San Antonio, Texas 78285*

(Received 9 December 1987; accepted for publication 31 August 1988)

The Clifford algebra of a real infinite-dimensional vector space is constructed. This algebra contains all finite- and infinite-dimensional Clifford algebras over the real and complex numbers. Some of the mathematical properties of these algebras are discussed; the automorphisms and involutions are specifically addressed. The special case with metric $g_{\nu\mu} = \delta_{\nu\mu}$, $\nu, \mu = 1,2,...$, follows as a corollary to the more general results.

## I. INTRODUCTION

The spaces that appear in quantum mechanics usually have a countably infinite number of complex coordinates. As Dirac[1] noted, a real space with a countably infinite number of coordinates is the more elementary concept. A complex space can be looked upon as a real space in which a pairing of coordinates has been introduced, each pair being considered as a complex number.

This fundamental concept is illustrated by Shale and Stinespring[2] who, beginning with a real pre-Hilbert space $M$, construct an infinite-dimensional Clifford algebra $\mathscr{C}_0(M)$ over the complex numbers. They define a state of $\mathscr{C}_0(M)$ as a linear functional $E$ on $\mathscr{C}_0(M)$ such that $E(1) = 1$ and $E(u^*u) \geqslant 0$, for all $u$ in $\mathscr{C}_0(M)$. The algebra $\mathscr{C}_0(M)$ then has a unique central state $E_0$ [$E_0(uv) = E_0(vu)$], for all $u, v$ in $\mathscr{C}_0(M)$, that is invariant under the orthogonal group of $M$.

Plymen[3] begins with an infinite-dimensional Hilbert space and constructs two real infinite-dimensional Clifford algebras. We present his construction in considerable detail because it illustrates the fundamental nature of the infinite-dimensional real Clifford algebras.

Let $\mathscr{H}$ be an infinite-dimensional real Hilbert space with orthonormal basis $h_i$, $i = 1,2,...$. Let $\mathscr{A}$ be the canonical anticommutation relation (CAR) algebra over $\mathscr{H}$, and let $J$ be a complex structure in $\mathscr{H}$. Let $\mathscr{H}_j$ and $\pi_j$ be the Fock space and representation determined by $J$. Let $F$ be $\mathscr{H}_j$ as a real Hilbert space; i.e., we restrict scalars from the complex numbers to the real numbers and take the inner product in $F$ to be the real part of the inner product in $\mathscr{H}_j$.

Let $U$ be an open set in $\mathscr{H}$ and let $f$: $U \to F$ be a differentiable map such that its Fréchet derivative is nuclear for all $x \in U$. The Dirac operator $Pf(x)$, for $x \in U$, is given by

$$Pf(x) = \sum_{i=1}^{\infty} c_n f_n(x), \tag{1}$$

where $f_n(x)$ is the $n$th partial derivative of $f$ at $x$ evaluated at $h_n$. The $c_i$, $i = 1,2,...$, are given by

$$c_i = \pi_j(e_i), \tag{2}$$

and satisfy

$$c_i c_j + c_j c_i = 2\delta_{ij}. \tag{3}$$

An orthonormal basis for $F$ can be chosen in which each $c_i$ is an infinite matrix that is symmetric and orthogonal. The operator $(Pf)^2$ acts as Gross's[4] Laplacian.

We will show that a real infinite-dimensional Clifford algebra is the fundamental structure in the sense that any real or complex, finite-dimensional or not, Clifford algebra can be embedded in it.

The interplay between infinite-dimensional real Clifford algebras and complex Clifford algebras is further shown in Shale and Stinespring[5] where they discover the infinite-dimensional spinors relative to $SO(H)$; de la Harpe[6] constructs the universal covering group $\mathrm{Spin}(\mathscr{H})_\infty$ of $SO(\mathscr{H})_1$, and Plymen[7] constructs the spin representation of $\mathrm{Spin}(\mathscr{H})_\infty$. Carey and O'Brien[8] continue the discussion of the group of automorphisms of the infinite-dimensional Clifford algebra. The interested reader is referred to the references of the cited literature for a more complete set of references.

We begin with a real, countably infinite-dimensional vector space and will assume sufficient structure on this space in order to generate a Clifford algebra. The mathematical properties of this algebra are discussed, with rather arbitrary limitations being imposed to keep the paper relatively short. There is extensive literature on finite-dimensional Clifford algebras (see the listing in Salingaros and Wene[9]).

## II. THE CONSTRUCTION

Let $\Lambda$ be a well-ordered, countably infinite set. Proceeding as in Chevalley,[10] we will denote by $\mathscr{C}$ the vector space over the field $\mathbf{R}$ of real numbers spanned by all finite products $e_{\lambda_1} e_{\lambda_2} \cdots e_{\lambda_n}$, where $\lambda_i \in \Lambda$ and $\lambda_i < \lambda_{i+1}$. The dimension of $\mathscr{C}$ over $\mathbf{R}$ will be countably infinite. An associative product is defined in $\mathscr{C}$ by

$$e_i e_j + e_j e_i = 2\delta_{ij}\alpha_i, \quad i, j \in \Lambda, \tag{4}$$

where $\alpha_i = -1$, 0, or 1. Hence we must add an identity element to $\mathscr{C}$.

The set $\Lambda$ can be written as the union of sets $M$, $Z$, and $P$, where

$$M = \{\lambda \in \Lambda | e_\lambda^2 = -1\}, \tag{5}$$

$$Z = \{\lambda \in \Lambda | e_\lambda^2 = 0\}, \tag{6}$$

and

$$P = \{\lambda \in \Lambda | e_\lambda^2 = 1\}. \tag{7}$$

Lemmas 1 and 2 exploit the fact that the set $\Lambda$ has infinite cardinality.

 0022-2488/89/020249-03$02.50 © 1989 American Institute of Physics 

*Lemma 1:* If one of the sets $M$ and $P$ is infinite, then there exist elements $p_i, q_i$, $i = 1,2,...$, of the Clifford algebra $\mathscr{C}$ such that for all $i, j = 1,2,...$,

$$p_i^2 = 1, \tag{8}$$

$$q_i^2 = -1, \tag{9}$$

$$p_i p_j + p_j p_i = 2\delta_{ij} 1, \tag{10}$$

$$q_i q_j + q_j q_i = 2\delta_{ij} 1, \tag{11}$$

and

$$p_i q_j + q_j p_i = 0. \tag{12}$$

*Proof:* If the set $M$ of all $i \in \Lambda$ such that $e_i^2 = -1$ is an infinite subset of the well-ordered set $\Lambda$, we can put the elements of $M$ into a one-to-one correspondence with the natural numbers 1,2,... . Define

$$p_i = e_2 e_4 e_{4+2i}, \tag{13}$$

and

$$q_i = e_{2i-1}, \tag{14}$$

for all $e_i$, $i = 1,2,...$ .

If the set $P$ of all $i \in \Lambda$ such that $e_i^2 = 1$ is an infinite subset of the well-ordered set $\Lambda$, we can put the elements of $P$ into a one-to-one correspondence with the material 1,2,... . Define

$$p_i = e_{2i-1}, \tag{15}$$

and

$$q_i = e_2 e_4 e_{4+2i}, \tag{16}$$

for all $e_i$, $i = 1,2,...$ .

*Lemma 2:* Let $p_i$, $i = 1,2,...$, be a set of elements of $\mathscr{C}$ such that

$$p_i p_j + p_j p_i = 2\delta_{ij} 1, \quad i,j = 1,2,... . \tag{17}$$

Then $\mathscr{C}$ contains elements $c, f_i, g_i$, $i = 1,2,...$, such that

$$c^2 = -1, \tag{18}$$

$$cf_i = f_i c \quad \text{and} \quad cg_i = g_i c, \tag{19}$$

$$f_i f_j + f_j f_i = 2\delta_{ij} 1, \tag{20}$$

$$g_i g_j + g_j g_i = -2\delta_{ij} 1, \tag{21}$$

and

$$f_i g_j + g_j f_i = 0, \tag{22}$$

for all $i, j = 1,2,...$ .

*Proof:* Define

$$c = p_1 p_2, \tag{23}$$

$$f_i = p_{2i+1}, \tag{24}$$

and

$$g_i = p_4 p_6 p_{6+2i}. \tag{25}$$

The metric of a Clifford algebra with generators $e_i$, $i = 1,2,...$, is the mapping $g_{\mu\nu} = \frac{1}{2}(e_\nu e_\mu + e_\mu e_\nu) = \frac{1}{2}\delta_{\mu\nu}\alpha_\nu$, $\mu\nu = 1,2,...$ . If $\lambda = M \cup Z \cup P$ is finite we will write $A^{p,m,q}$ for the Clifford algebra with the metric determined by $|P| = p$, $|M| = m$, and $|Z| = q$. If $Z = \emptyset$, we write $A^{p,m}$ in place of $A^{p,m,o}$. Most of the literature discusses the algebras $A^{p,m}$; the algebras $A^{p,m,q}$ are discussed in Ablamowicz.[11,12]

**Theorem 3:** If the set $M \cup P$ is infinite, then the real Clifford algebra $\mathscr{C}$ contains all finite-dimensional Clifford algebras $A^{p,m}$ over the field of real numbers and those over the field of complex numbers.

*Proof:* By Lemma 1, $\mathscr{C}$ contains a set of elements $p_i$, $i = 1,2,..., p_i^2 = 1$. Construct elements $c, f_i$, and $g_i$, $i = 1,2,...$, as in Lemma 2.

Let $F = \{f_i | i = 1,2,...\}$ and $G = \{g_i | i = 1,2,...\}$. Then any set $W$ of $m$ elements of $G$ and $p$ elements of $F$ under the product of $\mathscr{C}$ will generate the real Clifford algebra $A^{p,m}$.

Since the element $c$ of Lemma 2 can play the role of a complex unit, the set $W \cup \{c\}$ will generate the complex Clifford algebra $A^{p,m}$.

*Corollary 4:* The Clifford algebra associated with the metric $g_{\nu\mu} = \delta_{\nu\mu}$, $\nu,\mu = 1,2,...$, contains all Clifford algebras $A^{p,m}$, $p,m = 0,1,2,...$ .

*Corollary 5:* If each of the sets $M \cup P$ and $Z$ are infinite, then the real Clifford algebra $\mathscr{C}$ contains all finite-dimensional Clifford algebras $A^{p,m,q}$ (real and complex).

*Proof:* Let $M \cup P$ be as in Theorem 3 and $Z$ infinite. Define elements $c, f_i$, and $g_i$ as before. Let $H$ be the set of elements $\{z_i | i \in Z\}$. If $H_q$ is a set of $q$ elements of $H$, then the set $W \cup H_q$ generates the real algebra $A^{p,m,q}$. The complex algebra will be generated by $W \cup H_q \cup \{c\}$.

## III. THE IDEAL STRUCTURE

The ideal structure of the algebras $A^{p,m}$ is well known (see, for example, Chevalley[10]). We can easily determine the ideal structure of the Clifford algebra of an infinite-dimensional vector space.

**Theorem 6:** Let $\Lambda = P \cup M \cup U$ be infinite. The algebra $\mathscr{C}$ is simple if and only if $Z = \emptyset$.

*Proof:* Suppose $Z$ is not the empty set. Then there is some $e_0 \in Z$, $e_0^2 = 0$, and the algebra $e_0 \mathscr{C}$ is an ideal in $\mathscr{C}$.

Now if $Z = \emptyset$, then any element $a$ of $\mathscr{C}$ is an element of some $A^{p,m}$. If $A^{p,m}$ is not simple, that is, $p + m = 2r + 1$, then (without constructing a new basis) either $A^{p,m+1}$ or $A^{p+1,m}$ is also in $\mathscr{C}$ and contains $A^{p,m}$. But then $p + m + 1$ is even and the ideal generated by $a$ contains 1 (see Chevalley,[10] p. 47).

Compare our Theorem 5 with Lemma 1 of Shale and Stinespring.[2]

An ideal $I$ in an algebra $\mathscr{A}$ is *nil* if for each $x \in I$, $x^{n(x)} = 0$, for some positive integer $n(x)$. The *nil radical* $\mathscr{R}$ of an algebra is the sum of all nil ideals. We get as a corollary to Theorem 5 that the algebra $\mathscr{C}$ has a nil ideal if and only if $Z \neq \emptyset$. There are many additional things that can be said about the nil radical; we end our current discussion with the following theorem.

**Theorem 7:** Let $\Lambda = M \cup P \cup Z$ be infinite and $Z \neq \emptyset$. Then $\mathscr{C}$ can be written as a vector space direct sum

$$\mathscr{C} = \mathscr{A} \oplus \mathscr{R}, \tag{26}$$

where $\mathscr{R}$ is the nil radical of $\mathscr{C}$ and $\mathscr{A}$ the Clifford algebra generated by $\{e_i | i \in P \cup M\}$ is isomorphic to the factor algebra $\mathscr{C}/\mathscr{R}$. If $|Z| = n$, then $\mathscr{R}^{n+1} = 0$.

*Proof:* Let $\mathscr{A}$ be the Clifford algebra generated by the set

$\{e_i | i \in M \cup P\}$. The radical $\mathscr{R}$ is the ideal in $\mathscr{C}$ generated by the set $\{e_i | i \in Z\}$. Then

$$\mathscr{C} = \mathscr{A} \oplus \mathscr{R} \tag{27}$$

as a vector space. The map

$$a + \mathscr{R} \to a \tag{28}$$

is an isomorphism of $\mathscr{C}/\mathscr{R}$ and $\mathscr{A}$. Clearly if $|Z| = n$, $\mathscr{R}^{n+1} = 0$.

## IV. AUTOMORPHISMS AND INVOLUTIONS

Let $\mathscr{U}$ be an algebra over the field $K$. A vector space isomorphism $a$: $\mathscr{U} \to \mathscr{U}$ is an *automorphism* of $\mathscr{U}$ if

$$\alpha(xy) = \alpha(x)\alpha(y), \tag{29}$$

for all $x,y$ in $\mathscr{U}$.

Let $\mathscr{C}$ be the Clifford algebra generated by the elements $e_i$, $i \in \Lambda$, $\Lambda = M \cup Z \cup P$ and the set $M \cup P$ is infinite.

Denote by $\mathscr{C}^{\#}$ the group of invertible elements of $\mathscr{C}$. If $c \in \mathscr{C}^{\#}$ then the map $\alpha$: $\mathscr{C} \to \mathscr{C}$ defined by $\alpha(x) = cxc^{-1}$ is an (inner) automorphism of $\mathscr{C}$. The group of inner automorphism is isomorphic to the factor group

$$\mathscr{C}^{\#}/\text{center}(\mathscr{C}^{\#}).$$

We determine all the automorphism of the real Clifford algebra $A^{p,m}$ contained in $\mathscr{C}$ and generated by a finite set of $e_i$, $i \in M \cup P$. For a discussion of the automorphisms of an infinite-dimensional Clifford algebra see Refs. 2, 3, 5–8.

**Theorem 8:** Let $A^{p,m}$ be a real Clifford algebra generated by a finite set of elements $e_i$, $i \in M \cup P$. If $\rho$ is an automorphism of $A^{p,m}$, then there exists an element $c \in \mathscr{C}^{\#}$ such that $\rho(x) = cxc^{-1}$ for all $x$ in $A^{p,m}$.

*Proof:* If $A^{p,m}$ is simple, then by the Noether–Skolem theorem (Herstein,[13] p. 99) all automorphisms are inner and $c$ is an element of $A^{p,m}$. The term $A^{p,m}$ is not simple if and only if $p + m = 2r + 1$ and $2(-1)^r(-1)^q$ is a square in $\mathbf{R}$ (Chevalley,[10] p. 47).

Since the set $M \cup P$ is infinite we can construct a simple algebra $\mathscr{A}$ that is isomorphic to either $A^{p+1,q}$ or $A^{p,q+1}$. Let $\Theta$: $\mathscr{A} \to \mathscr{A}$ be defined by linearity and

$$\Theta(p_1 e_{p+q+1} p_2) = \rho(p_1) e_{p+q+1} \rho(p_2), \tag{30}$$

where $p_1$ and $p_2$ are products of the $e_i$'s, $i = 1,2,\ldots,p+q$. Then $\Theta$ is an automorphism of $\mathscr{A}$ that fixes $A^{p,q}$ setwise. Since $\mathscr{A}$ is simple, $\Theta$ is inner.

A linear transformation (superscript $\sigma$) of an algebra $\mathscr{U}$ is an *involution* if $(a^\sigma)^\sigma = a$ and $(ab)^\sigma = b^\sigma a^\sigma$ for all $a$, $b$ in $\mathscr{U}$.

The main involution * in a Clifford algebra is defined by linearity and

$$e_i^* = e_i, \quad i = 1,2,\ldots, \tag{31}$$

$$(e_{i_1} e_{i_2} \cdots e_{i_n})^* = e_{i_n} e_{i_{n-1}} \cdots e_{i_2} e_{i_1}. \tag{32}$$

The main involution * in the algebra $\mathscr{C}$ when restricted to $A^{p,q}$ induces the main involution in $A^{p,q}$.

**Theorem 9:** Let $A^{p,m}$ be a real Clifford algebra generated by a finite set $e_i$, $i \in M \cup P$. If $\sigma$ is an involution in $A^{p,m}$, then there exists an involution $\Theta$ in $\mathscr{C}$ such that $\Theta$ restricted to $A^{p,m}$ is $\sigma$.

*Proof:* Let $F$ be the subset of $M \cup P$ such that $e_i$, $i \in F$ generates $A^{p,m}$.

Extend $\sigma$ to all of $\mathscr{C}$ by linearity and, for each product $\mu$ of $e_i$'s, $i \notin F$,

$$\sigma(p_1 u p_2) = p_2^\sigma u^* p_1^\sigma, \tag{33}$$

where * is the main involution in the algebra generated by the $e_i$'s, $i \notin F$, and $p_1$ and $p_2 \in A^{p,m}$.

## V. CONCLUSION

The Clifford algebra $\mathscr{C}$ associated with an infinite-dimensional real vector space is a suitable background for the mathematical study of finite-dimensional real and complex Clifford algebras. If $A^{p,q}$ is a real Clifford algebra contained in $\mathscr{C}$, then any automorphism of $A^{p,q}$ is induced by an inner automorphism of $\mathscr{C}$. Similarly, any involution in $A^{p,q}$ is induced by an involution in $\mathscr{C}$.


[1] A. M. Dirac, *Spinors in Hilbert Spaces* (Plenum, New York, 1974).
[2] D. Shale and W. F. Stinespring, "States of the Clifford algebra," Ann. Math. **80**, 365 (1964).
[3] R. J. Plymen, "The Laplacian and Dirac operator in infinitely many variables," Compositio Math. **41**, 137 (1980).
[4] L. Gross, "Potential theory on Hilbert space," J. Funct. Anal. **1**, 123 (1967).
[5] D. Shale and W. F. Stinespring, "Spinor representations of infinite orthogonal groups," J. Math. Mech. **14**, 315 (1965).
[6] P. de La Harpe, "The Clifford algebra and the spinor group of a Hilbert space," Compositio Math. **25**, 245 (1972).
[7] R. J. Plymen, "Spinors in Hilbert space," Math. Proc. Cambridge Philos. Soc. **80**, 337 (1976).
[8] A. L. Carey and D. M. O'Brien, "Automorphism of the infinite dimensional Clifford algebra and the Atiyah–Singer Mod 2 index," Topology **22**, 437 (1983).
[9] N. A. Salingaros and G. P. Wene, "The Clifford algebra of differential forms," Acta Appl. Math. **4**, 271 (1985).
[10] C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U.P., New York, 1954).
[11] R. Ablamowicz, "Structure of spin groups associated with degenerate Clifford algebras," J. Math. Phys. **27**, 1 (1986).
[12] R. Ablamowicz, "Deformation and contraction in Clifford algebras," J. Math. Phys. **27**, 423 (1986).
[13] I. N. Herstein, *Noncommutative Rings*, Mathematical Association of America, Carus Mathematical Monographs, No. 15 (Math. Assoc. Am., Buffalo, 1975).

# Small oscillations of regular polytopes in $d \geqslant 4$ space dimensions

A. Eastaugh[a] and P. van Nieuwenhuizen
*State University of New York at Stony Brook, Stony Brook, New York 11794-3840*

(Received 12 July 1988; accepted for publication 7 September 1988)

Using the theory of finite groups, the spectrum of a regular "tetrahedron" in four and more space dimensions is analyzed.

## I. INTRODUCTION

The theory of finite groups is useful for the analysis of small oscillations of molecules when one is only interested in certain qualitative properties of the spectrum, such as degeneracies. In particular, when one does not know the details of the interactions between the atoms but only their symmetry properties, or when the interactions are too complicated to allow a direct solution of the Lagrange equations, group theory is a useful tool for analyzing the spectrum. The standard examples treated in textbooks of classical mechanics are the $H_2O$ and $NH_3$ molecules, whose symmetries under rotations and reflections form the Klein group of order 4 and the dihedral group of order 6. However, more amusing are the applications of group theory to the five Platonic solids: the tetrahedron (simplex), hexahedron (cube) and its dual the octahedron, and the icosahedron (20-face) together with its dual the dodecahedron (12-face). Their symmetry groups are $S_4$, $S_4 + \sigma S_4$, and $A_5 + \sigma A_5$, respectively, where $\sigma$ denotes the space inversion. (The dual is obtained by drawing lines from the center to the vertices and erecting planes orthogonal to these lines at the vertices.)

In this paper we shall analyze the spectrum of the simplest polytope in four space dimensions, the hypersimplex, by means of finite group theory. We do this out of curiosity and not because there are obvious physical applications (see, however, Ref. 1).

Polytopes are the objects that come after polyhedra in the series

$$\text{point, segment, polygon, polyhedron, polytope.} \tag{1.1}$$

In four space dimensions there are six regular polytopes but in $d \geqslant 5$ space dimensions there are only three regular polytopes: the hypersimplex (which is self-dual) and the hypercube which is dual to a solid which may be called the dual hypercube. One may visualize the hypersimplex in $d$ dimensions as obtained by adding a point "somewhere above" the hypersimplex in $(d-1)$ dimensions and drawing all edges from this extra point to the vertices. In four space dimensions, the hypersimplex contains therefore $V = 5$ vertices, $E = 6 + 4 = 10$ edges, $F = 4 + \binom{4}{2} = 10$ faces, and $S = 5$ tetrahedra, in agreement with Euler's formula $V - E + F - S = 0$. Since it is self-dual, $V = S$.

The hypercube is obtained by transporting the hypercube in one lower dimension parallel to itself, and drawing edges between corresponding vertices. In $d$ space dimensions the hypercube has $2^d$ vertices with coordinates $(\pm\frac{1}{2}, \pm\frac{1}{2}, \ldots, \pm\frac{1}{2})$, and it is bounded by $2d$ regular polytopes of maximal lower dimension $(d-1)$. Hence the dual hypercube contains $2d$ vertices and $2^d$ $(d-1)$-dimensional polytopes. Euler's formula for the hypercube in $d$ space dimensions reads

$$2^d - \binom{d}{1}2^{d-1} + \binom{d}{2}2^{d-2} - \cdots \mp 2d = (2-1)^d \mp 1 = \begin{cases} 0 & (d \text{ even}), \\ 2 & (d \text{ odd}), \end{cases} \tag{1.2}$$

while for the hypersimplex one has

$$(d+1) - \binom{d+1}{2} + \binom{d+1}{3} - \cdots \mp \binom{d+1}{d} = 1 - (1-1)^{d+1} \mp 1 = \begin{cases} 0 & (d \text{ even}), \\ 2 & (d \text{ odd}). \end{cases} \tag{1.3}$$

The three "exceptional" regular polytopes in four dimensions are

$$\begin{gathered} \text{the self-dual 24-tope with 24 vertices,} \\ \text{the 120-tope with 600 vertices,} \\ \text{the 600-tope with 120 vertices.} \end{gathered} \tag{1.4}$$

It is easy to prove that there are only five Platonic solids. At each vertex $q$ faces meet $(q \geqslant 3)$ and each face is a polygon with $p$ vertices $(p \geqslant 3)$. If the angle between two edges bordering a face is $\alpha$, one has $q\alpha < 2\pi$. (If $q\alpha = 2\pi$, the $q$ faces would lie in a plane.) The $p$ edges of a face are parallel to $p$ vectors through the origin, and it follows that $p(\pi - \alpha) = 2\pi$. Eliminating $\alpha$ from these two relations, one obtains the inequality

$$1/p + 1/q > \tfrac{1}{2}. \tag{1.5}$$

The only solutions for $(p,q)$ are (3,3), (3,4), and (4,3), and (3,5) and (5,3).

To see that there are six polytopes in four space dimensions, and only three polytopes in more than four space dimensions, is less easy.[2] The general procedure for constructing polytopes in $d$ space dimensions consists of piecing together suitable polytopes in $d-1$ dimensions. Not all $(d-1)$-dimensional polytopes are suitable, e.g., for $d = 3$, the $p$-gon is suitable only if $p < 6$, as we have seen. Let us piece together the three-dimensional polyhedra $(p,q)$ to construct the four-dimensional polytopes. Just as one may construct a polyhedron by starting in one dimension lower and putting there $q$ polygons around a vertex but with a deficit angle $2\pi - q\alpha$, which, when closed, moves the whole figure into three dimensions and then yields a polyhedron, also here one may start with $r$ polyhedra around an edge in three dimensions with a deficit angle $2\pi - r\beta$, and by closing this

[a] Current address: The Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637.

 0022-2488/89/020252-09$02.50 © 1989 American Institute of Physics 

angle the whole figure moves into four space dimensions and yields a polytope. Then $r\beta < 2\pi$ and $r > 2$. [Alternatively, one may start directly in four space dimensions and consider an $S^2$ erected at the middle of the edge and orthogonal to the edge. Then the intersection of this $S^2$ with the $(d-1)$ polytope (with $d=4$ here) yields a closed sequence of arcs, and again one obtains $r\beta < 2\pi$.] Consider now the dihedral angle $\beta$ between two adjacent faces of $(p,q)$. This immediately rules out the icosahedron (3,5) as a possible building block since it has $\beta > 120°$.

The only solutions to the constraints on $\beta$ and $r$ are $(p,q,r) = (3,3,3)$, $(3,3,4)$, $(3,3,5)$, $(3,4,3)$, $(4,3,3)$, and $(5,3,3)$. This is easily seen by measuring the angle $\beta$ for each of the polyhedra. More analytically, one may use that for any polyhedron $(p,q)$, one has for the dihedral angle the formula $\sin\theta/2 = (\cos\pi/q)/(\sin\pi/p)$. Hence for a tetrahedron $\beta \sim 70°$, for the octohedron $\theta \sim 109°$, for the cube $\beta = 90°$, for the dodecahedron $\beta \sim 116°$, while for the icosahedron $\beta \sim 138°$. This formula is most easily checked (and thus proved) by considering each case separately. (In particular, use the facts that an octahedron is cut in half by a square and that an icosahedron has a cap of five equilateral triangles, and then cut from the dodecahedron a pyramid whose four vertices are four nearest-neighbor vertices of the dodecahedron.)

We thus know that there are at most six polytopes in four dimensions. The (3,3,3) is the hypersimplex and extends to (3,3,...,3) for $d$ dimensions. The (4,3,3) is the hypercube and extends to (4,3,...,3) in $d$ dimensions. Its dual is (3,3,...,4). These polytopes obviously exist since their construction is trivial.

The (3,4,3) is the (self-dual) 24-tope consisting of 24 octahedrons (and 24 vertices since it is self-dual). It can be constructed, nontrivially, from the hypercube (4,3,3) by joining the midpoints of the 24 faces of the (4,3,3). That the result is a regular polytope is not at all obvious. (For example, joining the midpoints of the edges of a cube in three space dimensions does not yield a polyhedron, but rather an object that contains squares and triangles, as one may verify). Obviously, taking the midpoints of the faces of the eight cubes of the hypercube will give eight octahedra, but it is by no means obvious that 16 more octahedra are formed. However, it is not too difficult to see where these extra 16 come from.

Consider the hypercube as being obtained by parallel-transporting a cube "on the right" to a cube "on the left." Then the eight vertices on the right go to the eight vertices on the left. Each of these 16 vertices leads to an octahedron in the following way (and that yields then the remaining 16 octahedra). Consider a vertex $V$, say, of the right cube. It is common to three faces. Take the middle of these three faces; these will be three of the six vertices of the sought-after octahedron. The other three vertices of the desired octahedron are the middles of those three faces that are obtained by parallel transporting the three edges that meet at vertex $V$.

The (3,3,5) and its dual (5,3,3) can also be constructed but it is not so simple to prove that they are regular.[2]

When an $N$-atom molecule in its equilibrium position is invariant under a set of rotations and reflections, these symmetries form a finite group $G$. Denoting the position, equilibrium position, and the deviation of atom $j$ in $d$-dimensional space by

$$\mathbf{r}_j = \mathbf{r}_j^0 + \mathbf{s}_j, \tag{1.6}$$

one can plot the deviations $\mathbf{s}_j$ of the atoms as a $dN$-dimensional vector in a $dN$-dimensional vector space $R^{dN}$,

$$\mathbf{R} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N\}. \tag{1.7}$$

Thus $\mathbf{R}(t)$ gives the deviation of the whole molecule at time $t$. A normal mode corresponds to the motion $\mathbf{R}(t) = \mathbf{R}_0 \cos(\omega t + \alpha)$, and the general motion is a linear combination of the $dN$ normal modes. Among the $dN$ degrees of freedom there are, in $d$ space dimensions, $d$ translations and $\frac{1}{2}d(d-1)$ rotations, while the remaining modes are genuine normal modes. For example, in $d=4$ space dimensions there are ten genuine normal vibrations for the hypersimplex. A normal mode with frequency $\omega$ is mapped by a symmetry $g$ in $G$ into another normal mode *with the same frequency*. Hence $R^{dN}$ splits into subspaces $S(\omega_1), S(\omega_2), ...$, which are invariant under $G$ and any vector in $S(\omega_j)$ corresponds to a motion of the molecule that is a linear combination of normal modes with the same frequency $\omega_j$, and that is a harmonic motion in $S(\omega_j)$. The invariant subspaces $S(\omega_j)$ are the carrier spaces of representations of $G$, because any element $g$ in $G$ maps a vector in $S(\omega_j)$ into another in vector in $S(\omega_j)$. In general, the subspaces $S(\omega_j)$ will be irreducible representations (irreps) of $G$ but if one chooses the interactions between the atoms in a special way, it may happen that two irreps correspond to the same frequency. We shall not consider such special cases.

It is physically clear that the translations and rotations,

$$\mathbf{r}_j = \mathbf{r}^0{}_j + \mathbf{a}, \quad r_j^k = r_j^{0,k} + (b^k{}_l - b^l{}_k) r_j^{0,l}, \tag{1.8}$$

form invariant subspaces under $G$. (They need not be irreps: for example, for the $H_2O$ molecule all irreps are one dimensional.) One can find out how many times a given irrep of $G$ is contained in $R^{dN}$ by using characters. To explain this, we briefly review some elements of finite group theory.

A given group $G$ splits up into $r$ classes ($gag^{-1}$ with $g$ arbitrary). If there are $r$ classes there are $r$ irreps. A character $\chi$ is the trace of the matrices that form a representation of $G$. The characters are class functions, not only group functions, because $\chi(gag^{-1}) = \chi(a)$. The order of $G$ (i.e., the number of elements it contains) is related to the dimensions $d_s$ $(s = 1,...,r)$ of the irreps by

$$\text{order } G = \sum_{s=1}^{r} (d_s)^2. \tag{1.9}$$

Usually, there are only a few one-dimensional irreps of $G$. The number of them is given by the ratio of the order of $G$ and the order of the commutator subgroup $C(G)$

$$\text{number of one-dimensional irreps} = (\text{order } G)/\text{order } C(G). \tag{1.10}$$

The commutator subgroup $C(G)$ is generated by all elements $aba^{-1}b^{-1}$. It is an invariant subgroup (because $gaba^{-1}b^{-1}g^{-1} = a'b'a'^{-1}b'^{-1}$ with $a' = gag^{-1}$ and $b' = gbg^{-1}$). Any invariant subgroup consists of entire classes, so to determine $C(G)$ we evaluate $aba^{-1}b^{-1}$ for a

few elements $a$ and $b$, and each time the result lies in class $C$, the whole class $C$ is part of $C(G)$.

A most useful formula states that the characters are orthogonal functions of the classes

$$\sum_{g\in G}\chi^i(g)\chi^j(g)=\sum_{\text{classes } C^k}\chi^i(C^k)\chi^j(C^k)\times\text{order } C^k=\delta^{ij}\text{ order } G. \quad (1.11)$$

The symbol "order $C^k$" is the number of elements in class $C^k$ $(k=1,...,r)$, which is a divisor of the order of $G$. This allows a quick check on the enumeration of classes. Another check is the theorem that for any subgroup $H$ of $G$, in particular $C(G)$, the order of $H$ is a divisor of the order of $G$ (Lagrange).

Conceptually important is Cayley's theorem, which states that every group $G$ is a (proper or improper) subgroup of the symmetric group $S_N$ with $N=$ order $G$. ($S_N$ is the group of permutations of $N$ objects and is of order $N!$) Thus every group element of a group $G$ can be written as a product of permutation cycles. For example, the Abelian Klein group with elements $e$, $a$, $b$, $ab$ satisfying $a^2=b^2=e$, consists of the cycles $e$, (12)(34), (13)(24), and (14)(23). The conjugation relation $a'=gag^{-1}$ can then simply be interpreted as changing the *names* of the cycles according to the recipe $g$. [For example, (123)(345)(321) equals (145) and is obtained from (345) by changing the name of 3 into 1.] It follows that the classes of $S_N$ contain with each element $a$ also its inverse $a^{-1}$. Hence for $S_N$ one can generate $C(G)$ by multiplying all elements of a given class (and that for all classes).

## II. THE CHARACTERS OF $S_5$ DERIVED BY ELEMENTARY MEANS

The regular simplex in four space dimensions has five vertices, whose Cartesian coordinates we take at

$$(2,0,0,0),\quad (0,2,0,0,),\quad (0,0,2,0),\quad (0,0,0,2),\quad (\tau,\tau,\tau,\tau), \quad (2.1)$$

where $\tau=\frac{1}{2}(1-\sqrt{5})$ is the "golden ratio," satisfying $\tau^2-\tau-1=0$. The center is then at $\alpha(\tau,\tau,\tau,\tau)$ with $\alpha=\tau/(\tau+2)$. From the center to the five vertices one has the vectors $\mathbf{e}_1,...,\mathbf{e}_5$ with

$$(\tau+2)^{-1}(\tau+3,-\tau-1,-\tau-1,-\tau-1)$$
and three others; (2.2)
$$(\tau+2)^{-1}(2\tau,2\tau,2\tau,2\tau),$$

The oscillations have $5\times4=20$ degrees of freedom, of which four are translations and six are rotations. Thus there are ten genuine normal modes.

The isometries of the simplex in equilibrium permute the vertices. In fact, pair exchange of two vertices is an isometry so that the full isometry group is $S_5$. To see that pair exchange is an isometry, we consider the two vertices at (2,0,0,0) and (0,2,0,0), and choose as a basis in $R^4$ the vectors $(2,-2,0,0)$, $(0,0,2,2)$, $(0,0,2,-2)$, and $(1-\alpha)(\tau,\tau,\tau,\tau)$, i.e., the vectors $\mathbf{e}_1-\mathbf{e}_2$, $\mathbf{e}_3+\mathbf{e}_4+\tau\mathbf{e}_5$, $\mathbf{e}_3-\mathbf{e}_4$, and $\mathbf{e}_5$. On this basis the isometry takes the form

$$M(\text{pair exchange})=\begin{pmatrix}-1&&&\\&1&&\\&&1&\\&&&1\end{pmatrix} \quad (2.3)$$

and thus corresponds to a reflection. It follows that the rotations correspond to the alternating group $A_5$ and the pure reflections to $S_5-A_5$.

Since we must eventually reduce the 20-dimensional rep of $S_5$ that is produced by the small oscillations, into irreps of $S_5$, we will first deduce the characters of $S_5$. We can then use the orthogonality relation of the characters for the reduction.

To find the one-dimensional irreps, we need the commutator subgroup, as explained before. The commutator subgroup $C(G)$ of $S_N$ contains the products of all elements of a given class. Thus it contains all even permutations, and thus

$$C(S_5)=A_5. \quad (2.4)$$

It follows that there are only two one-dimensional irreps. It is easy to identify them: one is the unit representation itself, while the other is $+1$ on $A_5$ and $-1$ on the odd elements of $S_5$.

The number of irreps is equal to the number of classes. There are seven classes, namely the unit element, all pair exchanges (12), (23), etc., the three-cycles (123), etc., and so on. Denoting a permutation by the number of vertices it permutes among themselves, we have the following seven classes:

$$\begin{array}{lllllll} 1^5, & 1^32, & 1^23, & 1\ 4, & 5, & 1\ 2^2, & 2\ 3 \\ (1), & (10), & (20), & (30), & (24), & (15), & (20) \end{array} \quad (2.5)$$

Below each class we have written its order, and as a check one may note that they are divisors of the order of $S_5$ (which is 120). There are thus seven irreps, of which two are one dimensional. To determine the dimensionality of the other five irreps we use (1.9). We obtain the Diophantine equation

$$120=1+1+x^2+y^2+z^2+t^2+u^2. \quad (2.6)$$

If we order such that $x\leqslant y\leqslant z\leqslant t\leqslant u$, it is clear that $u\leqslant10$. Since $x=y=z=t=2$ is not a solution, $t\geqslant3$ so that $u\leqslant9$. Continuing in this way we find only one solution

$$x=y=4,\quad z=t=5,\quad u=6. \quad (2.7)$$

We will now identify these irreps by various elementary methods.

First of all, the isometries of the hypersimplex as an object in $R^4$ form a discrete subgroup of O(4); in fact, we already saw that the rotations are elements of SO(4). Hence there are at least two four-dimensional representations of $S_5$, as matrices $M$ in O(4) and as $M$ times det $M$. The character $\chi^4$ of the $4\times4$ matrix representation of $S_5$ as an isometry group in $R^4$ can be explicitly evaluated on the seven classes. For the pair exchanges we already obtained $\chi^4=2$, while on the unit class it equals $+4$. On the three-cycles (123), etc., we find its value by noting that the matrix that cyclically permutes $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ but keeps $\mathbf{e}_4$ and $\mathbf{e}_5$ fixed, is a rotation

in the three-volume $(\alpha,\beta,\gamma,0)$ around the axis (1,1,1,0) over an angle $2\pi/3$. Hence

$$M=\begin{pmatrix} -\frac{1}{2} & 1/2\sqrt{3} & \vdots & 0 \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} & \vdots & \\ \cdots & \cdots & \vdots & \cdots \\ & & \vdots & 1\ 0 \\ 0 & & \vdots & \\ & & \vdots & 0\ 1 \end{pmatrix},$$

$$\operatorname{tr} M(\text{three-cycles}) = 1. \tag{2.8}$$

The permutation (12)(34) can be represented on the basis $\mathbf{e}_1+\mathbf{e}_2$, $\mathbf{e}_1-\mathbf{e}_2$, $\mathbf{e}_3+\mathbf{e}_4$, $\mathbf{e}_3-\mathbf{e}_4$ as a diagonal matrix with entries $+1$, $-1$, $+1$, $-1$. Hence $\chi^4$ vanishes on the class $12^2$. For the class 23 with element (123)(45), we choose the basis $\mathbf{e}_1-\mathbf{e}_2$, $\mathbf{e}_2-\mathbf{e}_3$, $\mathbf{e}_4$, $\mathbf{e}_5$. Since $\mathbf{e}_2-\mathbf{e}_3$ goes over into $\mathbf{e}_3-\mathbf{e}_1=-(\mathbf{e}_1-\mathbf{e}_2)-(\mathbf{e}_2-\mathbf{e}_3)$, the matrix representation reads

$$M=\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \operatorname{tr} M(23\text{-cycles}) = -1. \tag{2.9}$$

That leaves us with the characters for the classes 14 and 5. The class 14 contains the permutation (1234) which maps $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4$ cyclically into each other, hence $\chi^4(14)=0$. On the other hand, the permutation (12345) can also be written down on the basis $\mathbf{e}_1,\ldots,\mathbf{e}_4$, but since $\mathbf{e}_1+\cdots+\mathbf{e}_5=0$, the vertex $\mathbf{e}_4$ is mapped into $\mathbf{e}_5$ which is equal to $-\mathbf{e}_1-\mathbf{e}_2-\mathbf{e}_3-\mathbf{e}_4$. Hence

$$M=\begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad \chi^4(\text{five-cycles}) = -1. \tag{2.10}$$

Summarizing, the character of $S_5$ as the isometry group in $R^4$ has the following value on the classes:

| class | $1^5$ | $1^32$ | $1^23$ | 14 | 5 | $12^2$ | 23 |
|---|---|---|---|---|---|---|---|
| order | (1) | (10) | (20) | (30) | (24) | (15) | (20). |
| character | 4 | 2 | 1 | 0 | $-1$ | 0 | $-1$ |

(2.11)

In principle this character could be a sum of irreducible characters. In that case we would find

$$\sum_{\text{classes } C^k} \chi^4(C^k)\chi^4(C^k) \text{ order } C^k = \left(\sum n_j^2\right) \text{order } G. \tag{2.12}$$

However, in this case we find for this sum

$$16\times1+4\times10+1\times20+1\times24+1\times20=120. \tag{2.13}$$

Thus the four-dimensional representation of $S_5$ as the isometry group in $R^4$ is irreducible.

The other four-dimensional representation can now at once be written down; it is the product of the nontrivial one-dimensional representation with $\chi^4$. This is the representation $M\det M$, whose character $\operatorname{tr} M\det M$ is obviously again an irreducible representation. To distinguish them we will denote (2.11) by $\chi^4_{(1)}$ and this irrep by $\chi^4_{(2)}$.

The six-dimensional representation of $S_5$ can also readily be identified. It should only be nonzero on $A_5$ (the classes with even permutations), because otherwise one could construct another six-dimensional irrep (by multiplication with the nontrivial one-dimensional irrep). By orthogonality one could further narrow it down. However, we can directly obtain it by noting that O(4) has the six-dimensional irrep $A_{ij}=-A_{ji}$ $(i,j=1,4)$, which suggests that this representation as a representation of $S_5$ might also be an irrep. On the unit class the trace of the six-dimensional rep is obviously 6. The isometry (12) keeps $\mathbf{e}_3$, $\mathbf{e}_4$, and $\mathbf{e}_5$ fixed. Hence

$$\begin{aligned} &A_{12}\to -A_{12}, \quad A_{1k}+A_{2k}\to A_{1k}+A_{2k}, \quad k=3,4, \\ &A_{34}\to A_{34}, \quad A_{1k}-A_{2k}\to -(A_{1k}-A_{2k}), \quad k=3,4. \end{aligned} \tag{2.14}$$

Clearly the trace of this matrix representation is zero, as expected (see above). The isometry (123) acts on the six components of $A_{ij}$ as follows:

$$\begin{aligned} &A_{12}\to A_{23}, \quad A_{23}\to -A_{13}, \quad A_{13}\to -A_{12}, \\ &A_{14}\to A_{24}, \quad A_{24}\to A_{34}, \quad A_{34}\to A_{14}. \end{aligned} \tag{2.15}$$

Hence $\chi^6$ equals zero on the class with three-cycles. On the class with (12)(34) one finds that

$$\begin{aligned} &A_{12}\to -A_{12}, \quad A_{1\pm2,3+4}\to \pm A_{1\pm2,3+4}, \\ &A_{34}\to -A_{34}, \quad A_{1\pm2,3-4}\to \mp A_{1\pm2,3-4}, \end{aligned} \tag{2.16}$$

where $A_{1-2,3+4}=A_{13}-A_{23}+A_{14}-A_{24}$. Hence $\chi^6$ on $12^2$ equals $-2$. Finally, for the class with (12345) we have $\mathbf{e}_4\to -\mathbf{e}_1-\mathbf{e}_2-\mathbf{e}_3-\mathbf{e}_4$. Hence

$$\begin{aligned} &A_{12}\to A_{23}, \quad A_{13}\to A_{24}, \\ &A_{23}\to A_{34}, \quad A_{24}\to A_{13}+A_{23}-A_{34}, \\ &A_{34}\to A_{14}+A_{24}+A_{34}, \quad A_{14}\to A_{12}-A_{23}-A_{24}. \end{aligned} \tag{2.17}$$

Clearly, on this class $\chi^6$ takes on the value $+1$.

Summarizing, we get for the six-dimensional representation of $S_5$,

| $1^5$ | $1^32$ | $1^23$ | 14 | 5 | $12^2$ | 23 |
|---|---|---|---|---|---|---|
| (1) | (10) | (20) | (30) | (24) | (15) | (20). |
| 6 | 0 | 0 | 0 | 1 | $-2$ | 0 |

(2.18)

This representation is indeed irreducible, and as a check one may verify that its character is orthogonal to the characters previously obtained.

We are left with the problem of finding the two five-dimensional characters. A natural guess would be to consider the five-dimensional matrix representation of $S_5$ as the permutation group of the vertices. Its character is the number of vertices that are left fixed. One has

$$\begin{array}{ccccccc} 1^5 & 1^32 & 1^23 & 14 & 5 & 12^2 & 23 \\ (1) & (10) & (20) & (30) & (24) & (15) & (20). \\ 5 & 3 & 2 & 1 & 0 & 1 & 0 \end{array} \quad (2.19)$$

However, this representation is not irreducible since $\Sigma\chi(g) \times\chi(g) = 240$. Using the orthogonality relations it is easy to see that the unit trace is contained once in it. Subtracting it, one is left with the character $\chi^4_{(1)}$ that we determined before.

We can, however, also consider the action of $S_5$ on the ten edges. Each edge can be written as $(ij)$ with $i<j=1,5$. The character is the number of edges that is kept fixed. Hence

$$\begin{array}{ccccccc} 1^5 & 1^32 & 1^23 & 14 & 5 & 12^2 & 23 \\ (1) & (10) & (20) & (30) & (24) & (15) & (20). \\ 10 & 4 & 1 & 0 & 0 & 2 & 1 \end{array} \quad (2.20)$$

From the orthogonality relations we get

$$100 + 16\times 10 + 1\times 20 + 4\times 15 + 1\times 20 = 3\times 120, \quad (2.21)$$

so that it contains three irreps ($\Sigma n_j^2 = 3$). The unit trace is contained once. Thus one of the four-dimensional irreps must be contained in it. As one may verify, $\chi^4_{(1)}$ is contained in it. Hence one is left with the following five-dimensional irrep:

$$\begin{array}{ccccccc} 1^5 & 1^32 & 1^23 & 14 & 5 & 12^2 & 23 \\ (1) & (10) & (20) & (30) & (24) & (15) & (20). \\ 5 & 1 & -1 & -1 & 0 & 1 & 1 \end{array} \quad (2.22)$$

One may directly verify that it is irreducible by using the orthogonality relations.

The other five-dimensional irrep is now at once found, by multiplication with the nontrivial one-dimensional irrep. The complete set of characters is given in Table I, and agrees with the standard results in the literature.[3]

## III. THE SPECTRUM OF THE OSCILLATING HYPERSIMPLEX

To each symmetry of a molecule corresponds a mapping of the linear vector space of deviations of the atoms from their equilibrium positions. Let $g$ be a symmetry (rotation or reflection) that maps the equilibrium position $\mathbf{r}_j^0$ of atom $j$ to the equilibrium position $\mathbf{r}_k^0$ of atom $k$. Then under $g$ the atoms are permuted according to the element $\pi_g$ of a permutation group $\pi$ of the $N$ atoms that is a proper or improper subgroup of the symmetric group $S_N$,

$$g\mathbf{r}_j^0 = \mathbf{r}_k^0, \quad \pi_g j = k. \quad (3.1)$$

Thus both $\pi_g$ and $g$ form a representation of the symmetry group $G$, but $\pi$ acts on the set of $N$ atoms while $g$ acts on the linear vector space $R^d$. Given a position $\mathbf{r}_j = \mathbf{r}_j^0 + \mathbf{s}_j$ of atom $j$, where $\mathbf{s}_j$ is its deviation, which we take small, then $g\mathbf{r}_j$ will be a vector near the equilibrium position $\mathbf{r}_k^0$ of atom $k$. Hence

$$g\mathbf{r}_j = \mathbf{r}_k^0 + \mathbf{s}(g)_k. \quad (3.2)$$

In other words, given a deviation $\mathbf{s}_j$ of atom $j$ the operator $g$ leads to a corresponding deviation of atom $k$. We can take all deviations $\mathbf{s}_j$ ($j = 1,N$) of the atoms together, and construct a $dN$-dimensional vector $\mathbf{R}$,

$$\mathbf{R} = \{\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_N\} \in R^{dN}, \quad (3.3)$$

which gives the deviation of the whole molecule. We can also define a projection operator $P_j$ that selects the *deviation* $\mathbf{s}_j$ and replaces it by the *position* $\mathbf{r}_j = \mathbf{r}_j^0 + \mathbf{s}_j$ of atom $j$,

$$P_j\mathbf{R} = \mathbf{r}_j = \mathbf{r}_j^0 + \mathbf{s}_j. \quad (3.4)$$

Note that $P_j$ is not a linear operator in $R^{dN}$ because the origin O of $R^{dN}$ is mapped to the equilibrium positions of the atoms

$$P_j\mathrm{O} = \mathbf{r}_j^0. \quad (3.5)$$

The group elements $g$ map deviations of one atom to deviations of another atom as we have discussed. It is therefore natural to investigate how the group $G$ acts on the vectors $\mathbf{R}$ that constitute the deviation of the whole molecule. Let $g\mathbf{r}_j$ be equal to $\mathbf{r}_k$. Then we can write $\mathbf{r}_j$ in terms of the deviation vectors $\mathbf{R}$ of the whole molecule as $\mathbf{r}_j = P_j\mathbf{R}$. From the set of deviations $\mathbf{s}_j$ of all atoms one obtains a new set of deviations $\mathbf{s}(g)_j$ after acting with $g$. These new deviations constitute a new deviation vector $\mathbf{R}'$ of the whole molecule. The new deviation at atom $k$ is then given by $\mathbf{s}'_k = g\mathbf{r}_j - \mathbf{r}_k^0$. We can view the new position vector $\mathbf{r}'_k = \mathbf{r}_k^0 + \mathbf{s}'_k$ of atom $k$ as coming from $\mathbf{R}'$ according to the usual recipe: $\mathbf{r}'_k = P_k\mathbf{R}'$. Let the transformation $\mathbf{R}\to\mathbf{R}'$, which is due to the group element $g$, be denoted by

$$\mathbf{R}' = A_g\mathbf{R}. \quad (3.6)$$

Then we can rewrite the $d$-dimensional linear transformation $g\mathbf{r}_j = \mathbf{r}_k$ as follows:

$$gP_j\mathbf{R} = P_kA_g\mathbf{R}. \quad (3.7)$$

Or, since $k = \pi_g j$, we obtain

$$P_{\pi_g j}A_g\mathbf{R} = gP_j\mathbf{R}. \quad (3.8)$$

The mapping $A_g$ is a linear operator in $R^{dN}$. (Note that if $\mathbf{R}$ is equal to the origin O then

$$P_{\pi_g j}A_g\mathrm{O} = g\mathbf{r}_j^0 = \mathbf{r}_k^0 = P_k\mathrm{O} = P_{\pi_g j}\mathrm{O} \quad (3.9)$$

so that $A_g\mathrm{O} = \mathrm{O}$.) In fact, $A_g$ forms an (in general reducible) representation of $G$ in $R^{dN}$. Hence, we now have three representations for $G$:

(i) the permutation representation $\pi$ acting on the atoms,
(ii) the linear operator $g$ acting in $R^d$, (3.10)
(iii) the linear operator $A_g$ acting in $R^{dN}$.

TABLE I. Characters and classes of the hyper simplex.

| Class/irrep | $\chi^1_{(1)}$ | $\chi^1_{(2)}$ | $\chi^4_{(1)}$ | $\chi^4_{(2)}$ | $\chi^5_{(1)}$ | $\chi^5_{(2)}$ | $\chi^6$ | tr $A_g$ |
|---|---|---|---|---|---|---|---|---|
| $1^5$(1) | 1 | 1 | 4 | 4 | 5 | 5 | 6 | 20 |
| $1^32$ (10) | 1 | −1 | 2 | −2 | 1 | −1 | 0 | 6 |
| $1^23$ (20) | 1 | 1 | 1 | 1 | −1 | −1 | 0 | 2 |
| 14 (30) | 1 | −1 | 0 | 0 | −1 | 1 | 0 | 0 |
| 5 (24) | 1 | 1 | −1 | −1 | 0 | 0 | 1 | 0 |
| $12^2$ (15) | 1 | 1 | 0 | 0 | 1 | 1 | −2 | 0 |
| 23 (20) | 1 | −1 | −1 | 1 | 1 | −1 | 0 | 0 |

As discussed in Sec. I, the linear vector space $R^{dN}$ splits up into subspaces $S(\omega_1)$, $S(\omega_2)$,..., which are invariant under $G$. If there is no accidental degeneracy, each $S(\omega_j)$ will be the carrier space of an irrep of $G$. To find out how often a particular irrep of $G$ is contained in $R^{dN}$ (Ref. 4) we need the character of $A_g$, and the characters of the irreps of $G$. Using the orthogonality relations of the characters we can then decompose the character of $A_g$ as follows:

$$\operatorname{tr} A_g = \sum_{j=1}^{r} n_j \chi^j(g), \tag{3.11}$$

where $n_j$ are non-negative integers that give the number of times a given irrep with character $\chi^j$ is contained in $A_g$. Moreover, since we know how $g$ acts inside a given $S(\omega_j)$, we will also be able to deduce how the normal modes with a given frequency $\omega_j$ transform into each other under the symmetry operation $g$, and from that knowledge we will try to identify the various normal modes.

To obtain the character of $A_g$ we proceed as follows. The linear vector space $R^{dN}$ is the direct sum of $N$ $d$-dimensional linear vector spaces $R^1$, $R^2$,...,$R^N$, one for each atom. Since $A_g$ maps a deviation vector $\mathbf{s}_j$ in $R^j$ to a deviation vector $\mathbf{s}'_k$ in $R^k$, only those subspaces $R^j$ can contribute to the trace of $A_g$ that are not permuted but held fixed. Inside such a given subspace $R^j$, $A_g$ acts just like $g$ acts in the physical $d$-dimensional space. Hence

$$\operatorname{tr} A_g = N_g \operatorname{tr} g, \tag{3.12}$$

where $N_g$ is the number of atoms that is held fixed by $g$, and $\operatorname{tr} g$ is the character of the matrix representation of $g$ in the $d$-dimensional space. In our case, we already established the identification

$$\operatorname{tr} g = \chi^4_{(1)}. \tag{3.13}$$

Hence the character of the representation $A_g$ of $G$ on the deviation space of the molecule is given by

$$\operatorname{tr} A_g = N_g \chi^4_{(1)}. \tag{3.14}$$

The result for $\operatorname{tr} A_g$ is given in Table I, where of course also $\chi^4_{(1)}$ is given, and where $N_g$ can be found as the superscript $p$ in the notation $1^p2^q3^r$, etc., for the classes as products of cycles.

From the orthogonality relations in (1.11) we can find the non-negative numbers $n_j$ in (3.11). Since $\Sigma(\operatorname{tr} A_g)^2 = 7\times 120$, it follows that $\Sigma n_j^2 = 7$. On the other hand, we also know that the dimension of the representation $A_g$ is equal to $5\times 4 = 20$. Hence

$$\sum n_j d_j = 20, \quad \sum n_j^2 = 7, \tag{3.15}$$

where $d_j$ can be 1, 4, 5, or 6, and $n_j = 0$, 1, or 2. Using the orthogonality relations, we find the following multiplicities of the irreps contained in $A_g$:

$$n(\chi^1_{(1)}) = 1, \quad n(\chi^1_{(2)}) = 0, \quad n(\chi^4_{(1)}) = 2, \quad n(\chi^4_{(2)}) = 0,$$
$$n(\chi^5_{(1)}) = 1, \quad n(\chi^5_{(2)}) = 0, \quad n(\chi^6) = 1. \tag{3.16}$$

Among these irreps, the translations and rotations are still present, and we will now remove them.

The translations and rotations form invariant subspaces. In fact, under an infinitesimal translation $\mathbf{r}_j = \mathbf{r}_j^0 + \mathbf{a}$, the subspaces $R^j$ are not permuted and the deviations $\mathbf{a}$ inside each $d$-dimensional subspace $\mathbf{R}^j$ are transformed into $g\mathbf{a}$. Hence the $d$-dimensional subspace of $R^{dN}$ corresponding to the translations is transformed by $g$ just as $g$ acts in $R^d$. Thus the translations correspond to the character $\chi^4_{(1)}$.

In this case, the subspace of the translation vectors in $R^{dN}$ forms an irrep of $G$. This need not always be the case. For example, for the $H_2O$ molecule, $G$ is Abelian and all irreps are one dimensional, so that the translations form a reducible rep.

The $\frac{1}{2}d(d-1)$ infinitesimal rotations in $R^d$ are of the form

$$r^k \to r^k + \epsilon^{klm_1\cdots m_{d-2}} b_{m_1\cdots m_{d-2}} r^l. \tag{3.17}$$

In particular, under an infinitesimal rotation, atom $j$ is displaced from its equilibrium position $\mathbf{r}_j^0$ as follows:

$$r_j^{0,k} \to r_j^{0,k} + \epsilon^{klm_1\cdots m_{d-2}} b_{m_1\cdots m_{d-2}} r_j^{0,l} \equiv r_j^{0,k} + s(b)_j^k. \tag{3.18}$$

The deviations $\mathbf{s}(b)_j$ again form a vector $\mathbf{R}(b)$ in $R^{dN}$ that describes the rotation of the whole molecule, and $A_g$ acts on these vectors $\mathbf{R}(b)$ as follows:

$$\begin{aligned} &P_{\pi_g j} A_g \mathbf{R}(b) = g P_j \mathbf{R}(b), \\ &P_j \mathbf{R}(b) = \mathbf{r}_j^0 + \epsilon b \mathbf{r}_j^0, \\ &g P_j \mathbf{R}(b) = g\mathbf{r}_j^0 + \det g \epsilon (gb)(g\mathbf{r}_j^0) \\ &\qquad = \mathbf{r}_k^0 + \epsilon(\det g)(gb)\mathbf{r}_k^0 \\ &\qquad = P_k \mathbf{R}\{(\det g)(gb)\} \\ &\qquad = P_{\pi_g j} \mathbf{R}\{(\det g)(gb)\}. \end{aligned} \tag{3.19}$$

Hence

$$A_g \mathbf{R}(b) = \mathbf{R}\{(\det g)(gb)\}. \tag{3.20}$$

Thus $g$ acts on the labels $b_{ij} = -b_{ji} \equiv \epsilon_{ij}^{\ m_1\cdots m_{d-2}} b_{m_1\cdots m_{d-2}}$ as the antisymmetric tensor representation times $\det g$. The antisymmetric tensor representation was denoted by $\chi^6$; it was irreducible and vanished on the classes with odd permutations. Hence the six-dimensional subspace of $R^{dN}$ corresponding to the rotational deviations transforms under $G$ as an irrep with character $\chi^6$.

After subtracting the translations and rotations, we are left with a ten-dimensional reducible representation, which contains the irreps corresponding to the characters $\chi^1_{(1)}$, $\chi^4_{(1)}$, and $\chi^5_{(1)}$ each once. Thus we have one normal mode with frequency $\omega_{\mathrm{I}}$, four normal modes with frequency $\omega_{\mathrm{II}}$, and five normal modes with frequency $\omega_{\mathrm{III}}$. The singlet normal mode with frequency $\omega_{\mathrm{I}}$ is mapped by all symmetries of $S_5$ into itself. Clearly, this is the dilational mode in which all five atoms move harmonically, radially, and in phase. To identify the quartet and quintet of modes, we compare with the spectrum of the tetrahedron in three space dimensions.

A similar analysis for the tetrahedron yields $4\times 3 - 3 - 3 = 6$ genuine normal modes, which split into a singlet, a doublet, and a triplet. The doublet consists of two normal modes that are transformed into a linear combination of them under the action of $g$ in $G$, where $G$ now is $S_4$. The characters of these two-dimensional and three-dimensional irreps of $S_4$ are given by

$$\begin{array}{lccccc} \text{class} & 1^4 & 1^2 2 & 13 & 4 & 2^2 \\ \text{order} & (1) & (6) & (8) & (6) & (3) \\ \chi^{(2)} & 2 & 0 & -1 & 0 & 2 \\ \chi^{(3)} & 3 & 1 & 0 & -1 & -1 \end{array} . \tag{3.21}$$

Hence the normal modes of the doublet are invariant under the Klein group, which is in this case the set of rotations around an axis connecting the middles of two opposite edges [corresponding to the cycles (12)(34), etc.]. The triplet forms the same irrep as the translations, so that this triplet seems similar to the quartet of the hypersimplex.

To get a clearer picture, we consider instead of the tetrahedron the ammonia molecule $NH_3$, and consider what happens if we let the N atom become similar in mass and interactions to the three H atoms. The symmetry group of the $NH_3$ molecule is $D_3 = S_3$, the symmetry group of a triangle in a plane, and has order 6. There are of course again six genuine normal modes, which, however, now fall into two singlets and two doublets. Clearly, one singlet is the dilational mode, while the other singlet is the motion with the N atom moving radially towards and from the center while the three H atoms make counteroscillations, (a "pumper motion"). One of the doublets is due to a similar motion but with the role of the N atom taken over by the H atoms. Although there are three such motions, because there are three H atoms, their sum vanishes by symmetry arguments, so that these motions form a doublet. (From the fact that the sum of the three "H pumpers" vanishes, it also follows that the N atom moves perpendicular to the vector connecting it to the center.) If the N atom becomes an H atom, the "N pumper" and the doublet of H pumpers are expected to fuse into a triplet of "tetrahedron pumpers." The other doublet of the $NH_3$ molecule must then remain a doublet of the tetrahedron. This suggests that this doublet is a motion of, roughly, atom 1 and 2 along the edge connecting them, and atom 3 and 4 along the edge connecting them. There are three pairs of opposite edges for the tetrahedron, but the sum of these three vibrations vanishes again due to symmetry considerations, so that these "Klein modes" indeed form a doublet.

One can write down these motions in Cartesian coordinates. Denoting by **a,b,c,d** the equilibrium positions of the four atoms, one has $\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}=0$; hence $\mathbf{a}^2=\mathbf{b}^2=\mathbf{c}^2=\mathbf{d}^2=1$ and $\mathbf{a}\cdot\mathbf{b}=-\frac{1}{3}$, etc. The translational deviations can be written as

$$\begin{aligned} &\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}\to\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}+\epsilon\mathbf{T}_a\,(\text{or } \mathbf{T}_b,\mathbf{T}_c,\mathbf{T}_d), \\ &\mathbf{T}_a=\{\mathbf{a},\mathbf{a},\mathbf{a},\mathbf{a}\},\quad \text{idem } \mathbf{T}_b,\mathbf{T}_c,\mathbf{T}_d, \end{aligned} \tag{3.22}$$

with the constraint $\mathbf{T}_a+\mathbf{T}_b+\mathbf{T}_c+\mathbf{T}_d=0$, while the rotational deviations are given by

$$\begin{aligned} &\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}\to\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}+\epsilon\mathbf{R}_a\ \ (\text{or } \mathbf{R}_b,\mathbf{R}_c,\mathbf{R}_d), \\ &\mathbf{R}_a=\{\mathbf{a}\wedge\mathbf{a},\mathbf{a}\wedge\mathbf{b},\mathbf{a}\wedge\mathbf{c},\mathbf{a}\wedge\mathbf{d}\},\quad \text{idem } \mathbf{R}_b,\mathbf{R}_c,\mathbf{R}_d, \end{aligned} \tag{3.23}$$

with the constraint $\mathbf{R}_a+\mathbf{R}_b+\mathbf{R}_c+\mathbf{R}_d=0$. Hence there are three translational and three rotational modes.

The dilational mode of the tetrahedron is given by

$$\mathbf{D}=\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}. \tag{3.24}$$

For the "pumper mode" with atom $A$ moving along the radius vector **a** connecting $A$ to the origin O, any other atom $B$ moves in the $OAB$ plane. Hence $\mathbf{P}_a=\{\mathbf{a},\alpha\mathbf{a}+\beta\mathbf{b},\alpha\mathbf{a}+\beta\mathbf{c},\alpha\mathbf{a}+\beta\mathbf{d}\}$. To fix $\alpha$ and $\beta$ we use the orthogonality of the modes belonging to different frequencies. According to the Lagrange equations, $(M^i{}_j\omega^2{}_{\rm I}-V^i{}_j)J^j_{\rm I}=0$, so that for two normal modes with frequencies $\omega_{\rm I}$ and $\omega_{\rm II}$ one has the orthogonality relation

$$\sum_{i,j=1}^{3N} M^i{}_j s^i_{\rm I} s^j_{\rm II}(\omega^2_{\rm I}-\omega^2_{\rm II})=0, \tag{3.25}$$

where $M_i$ is the mass matrix, which is the unit matrix in our case.

Hence the pumper modes must be orthogonal to the translational, rotational, and dilatational modes. Orthogonality to the translational modes (conservation of momentum) gives $1+3\alpha-\beta=0$, and orthogonality to the dilatational mode (the center of mass is at rest) leads to $1+3\beta-\alpha=0$. Hence, $\alpha=\beta=\frac{1}{2}$ and the pumper modes of the tetrahedron are

$$\begin{aligned} \mathbf{P}_a&=\{\mathbf{a},-\tfrac{1}{2}(\mathbf{a}+\mathbf{b}),-\tfrac{1}{2}(\mathbf{a}+\mathbf{c}),-\tfrac{1}{2}(\mathbf{a}+\mathbf{d})\}, \\ \mathbf{P}_b&=\{-\tfrac{1}{2}(\mathbf{b}+\mathbf{a}),\mathbf{b},-\tfrac{1}{2}(\mathbf{b}+\mathbf{c}),-\tfrac{1}{2}(\mathbf{b}+\mathbf{d})\}, \\ \mathbf{P}_c&=\{-\tfrac{1}{2}(\mathbf{c}+\mathbf{a}),-\tfrac{1}{2}(\mathbf{c}+\mathbf{b}),\mathbf{c},-\tfrac{1}{2}(\mathbf{c}+\mathbf{d})\}, \\ \mathbf{P}_d&=\{-\tfrac{1}{2}(\mathbf{d}+\mathbf{a}),-\tfrac{1}{2}(\mathbf{d}+\mathbf{b}),-\tfrac{1}{2}(\mathbf{d}+\mathbf{c}),\mathbf{d}\}. \end{aligned} \tag{3.26}$$

Obviously their sum vanishes so that only three are independent.

For the $NH_3$ atom, the N pumper must be orthogonal to the H pumpers. In the limit that the N atom becomes an H atom, one finds that $(\mathbf{P}_a)^2=(\mathbf{P}_b)^2=2$ and $\mathbf{P}_a\cdot\mathbf{P}_b=-\frac{2}{3}$. Hence the N-pumper mode is given by $\mathbf{P}_a$, while the H-pumper mode for the atom at **b** is given by $\mathbf{P}'_b=\mathbf{P}_b+\frac{1}{3}\mathbf{P}_a$. From the explicit expressions for $\mathbf{P}_a$ and $\mathbf{P}_b$ it follows that the N atom moves perpendicularly to its radius vector **a**, as we already anticipated. It is also clear that the $\mathbf{P}'_b$, $\mathbf{P}'_c$, and $\mathbf{P}'_d$ modes form a doublet for the $NH_3$ molecule since their sum vanishes.

To identify the other doublet of the $NH_3$ molecule, the doublet which remains a doublet of the tetrahedron, it suffices to find three modes that are orthogonal to all previous modes and whose sum vanishes. We begin with the ansatz

$$\mathbf{K}(12,34)=\{\mathbf{a}-\mathbf{b},\mathbf{b}-\mathbf{a},\mathbf{c}-\mathbf{d},\mathbf{d}-\mathbf{c}\} \tag{3.27}$$

and a similar expression for $\mathbf{K}(13,24)$ and $\mathbf{K}(14,23)$. Clearly, these modes are orthogonal to the **T** and **R** vectors. For the inner product with **D** one finds $\frac{16}{3}$, so that after subtraction of the dilatational mode one obtains

$$\begin{aligned} \mathbf{K}(12,34)'&=\mathbf{K}(12,34)-\tfrac{4}{3}\mathbf{D} \\ &=\{-\tfrac{1}{3}\mathbf{a}-\mathbf{b},-\tfrac{1}{3}\mathbf{b}-\mathbf{a},-\tfrac{1}{3}\mathbf{c}-\mathbf{a},-\tfrac{1}{3}\mathbf{d}-\mathbf{b}\}. \end{aligned} \tag{3.28}$$

This mode is already orthogonal to all **P**'s, so that **K**' are the three Klein modes, whose sum indeed vanishes

$$\begin{aligned} \mathbf{K}(13,24)'&=\{-\tfrac{1}{3}\mathbf{a}-\mathbf{c},-\tfrac{1}{3}\mathbf{b}-\mathbf{d},-\tfrac{1}{3}\mathbf{c}-\mathbf{a},-\tfrac{1}{3}\mathbf{d}-\mathbf{b}\}, \\ \mathbf{K}(14,23)'&=\{-\tfrac{1}{3}\mathbf{a}-\mathbf{d},-\tfrac{1}{3}\mathbf{b}-\mathbf{c},-\tfrac{1}{3}\mathbf{c}-\mathbf{b},-\tfrac{1}{3}\mathbf{d}-\mathbf{a}\}. \end{aligned} \tag{3.29}$$

Geometrically, the motion of these modes, say of **K** (12,34) is such that the angle $\phi$ between **a** and **b** and the angle $\phi$ between **c** and **d** both increase and decrease harmonically and in phase, while the lengths of $\mathbf{r}_a$, $\mathbf{r}_b$, $\mathbf{r}_c$, and $\mathbf{r}_d$ do not change but stay equal to $(\mathbf{r}^0_a)^2$. In other words, one can view

this Klein mode as follows. Connect the middles of $\mathbf{r}_1^0 - \mathbf{r}_2^0$ and $\mathbf{r}_3^0 - \mathbf{r}_4^0$ by a vector $\mathbf{v}$, and erect a plane $P$ at $\frac{1}{2}\mathbf{v}$, orthogonal to $\mathbf{v}$. Thus atoms 1 and 2 lie on one side of the plane, and atoms 3 and 4 on the other side. Then the four atoms move towards and away from $P$, and when they move towards $P$, all four move away from each other (and from $\mathbf{v}$), while when they move away from $P$, they approach the line $\mathbf{v}$.

Returning to the hypersimplex, we now readily identify the quartet as describing the five "pumper motions" where one atom ($A$) moves radially and the other four atoms ($B$) make countermotions in the plane through the origin, atom $A$ and an atom $B$. The sum of these five vibrations vanishes; hence the pumpers form a quartet. In fact, as one makes a one-to-one correspondence between a pumper mode and the vector from the origin to atom $A$, the action of $g$ on the five pumper modes is the same as the action of $g$ on the five vectors from the origin to the edges. In general, for a hypersimplex in $d$ space dimensions, the pumper modes always correspond to the representation of $g$ in $R^d$.

To identify the quintet of normal modes, we first analyze the hypersimplex in $d = 5$ and $d = 6$ space dimensions. Using the same techniques as before, and employing the character tables for $S_6$ and $S_7$ of Ref. 3, we find the following results.

*$d=5$:* The group $G = S_6$ has order 720 and 11 classes. There are pairs of irreps of dimensions 1, 5, again 5, 9, 10, and one irrep of dimension 16. Further $A_g$ has $\Sigma n_j^2 = 7$ and contains $\chi^1_{(1)}$ once (the dilation), $\chi^5_{(1)}$ twice (the translations and the pumpers), $\chi^{10}_{(1)}$ once (the rotations), and finally $\chi_{(1)}$ once.

*$d=6$:* The group $S_7$ has order 5040, and 15 classes. There are pairs of irreps of dimension 1, 6, 14, again 14, 15, 21, 35, and one irrep of dimension 20. Now $A_g$ has $\Sigma n_i^2 = 7$ and contains $\chi^1_{(1)}$ once, $\chi^6_{(1)}$ twice, $\chi^{15}_{(1)}$ once, and $\chi^{14}_{(1)}$ once.

The characters $\chi^9_{(1)}$ in $d = 5$, and $\chi^{14}_{(1)}$ in $d = 6$, as well as the characters $\chi^5_{(1)}$ in $d = 4$ and $\chi^2_{(1)}$ in $d = 3$, correspond to the normal modes we want to identify. The sequence 2,5,9,14 for $d = 3,4,5,6$ can easily be extended to all values of $d$. We now proceed as in the case of the tetrahedron, and make an ansatz, which we then complete by requiring orthogonality to all other modes.

If $\mathbf{a}^2 = 1$, one has in $d$ space dimensions that $\mathbf{a}\cdot\mathbf{b} = -1/(n-1)$ where $n = d+1$ is the number of atoms. The pumper modes are then $d$-dimensional vectors; for example

$$\mathbf{P}_a = \{\mathbf{a}, -(\mathbf{a}+\mathbf{b})/(n-2)), -((\mathbf{a}+\mathbf{c})/(n-2)),\ldots\}. \tag{3.30}$$

They satisfy $\mathbf{P}^2{}_a = n/(n-2)$ and $\mathbf{P}_a\cdot\mathbf{P}_b = -n/((n-1)(n-2))$. From now on we consider the case $d = 4$.

The pumper modes $\mathbf{P}_a$, $\mathbf{P}_b$, $\mathbf{P}_c$, $\mathbf{P}_d$, and $\mathbf{P}_e$, whose sum equals zero, can be made orthogonal. The resulting modes are

$$\begin{aligned}&\mathbf{P}_a,\quad \mathbf{P}_b + \tfrac{1}{4}\mathbf{P}_a,\quad \mathbf{P}_a,\quad \mathbf{P}_c + \tfrac{1}{3}(\mathbf{P}_a + \mathbf{P}_b),\\ &\mathbf{P}_d + \tfrac{1}{2}(\mathbf{P}_a + \mathbf{P}_b + \mathbf{P}_c).\end{aligned} \tag{3.31}$$

As an ansatz for the Klein modes we consider $\mathbf{K}(12,34) = \{\mathbf{a}-\mathbf{b},\mathbf{b}-\mathbf{a},\mathbf{c}-\mathbf{d},\mathbf{d}-\mathbf{c},0\}$. There are ten such expressions because $\mathbf{K}(12,34) + \mathbf{K}(13,24) + \mathbf{K}(14,23) = 0$. We will therefore have to explain how this decuplet is reduced to a quintet.

Using that $\mathbf{K}\cdot\mathbf{D} = 5$ and $\mathbf{D}^2 = 5$, we find that the mode orthogonal to $\mathbf{D}$ is given by

$$\mathbf{K}(12,34)' = \{-\mathbf{b}, -\mathbf{a}, -\mathbf{d}, -\mathbf{c}, -\mathbf{e}\}. \tag{3.32}$$

Next we orthogonalize it to the pumper modes. Straightforward but slightly tedious algebra yields

$$\begin{aligned}\mathbf{K}(12,34)'' &= \mathbf{K}(12,34)' - \mathbf{P}_a - \mathbf{P}_b - \mathbf{P}_c - \mathbf{P}_d\\ &= \mathbf{K}(12,34)' + \mathbf{P}_e\\ &= \{\tfrac{1}{3}\mathbf{a} - \mathbf{b} - \tfrac{1}{3}\mathbf{e}, -\tfrac{1}{3}\mathbf{b} - \mathbf{a} - \tfrac{1}{3}\mathbf{e},\\ &\quad -\tfrac{1}{3}\mathbf{c} - \mathbf{d} - \tfrac{1}{3}\mathbf{e}, -\tfrac{1}{3}\mathbf{d} - \mathbf{c} - \tfrac{1}{3}\mathbf{e}, 0\}.\end{aligned} \tag{3.33}$$

It is clear that $\mathbf{K}''$ in (3.33) is orthogonal to the translations ($\mathbf{T}_a$ and $\mathbf{T}_e$ are enough to check), dilatations, and pumpers ($\mathbf{P}_a$ is enough to check). The rotational deviations of a vector $\mathbf{v}$ can be written as $v^i \to v^i + (a^i b^j - b^i a^j)v^j$ and nine similar expressions, of which only six are independent. The rotational deviations of the molecule are then

$$\begin{aligned}\mathbf{R}_{ab} &= \{-\tfrac{1}{4}\mathbf{a} - \mathbf{b}, \mathbf{a} + \tfrac{1}{4}\mathbf{b},\\ &\quad -\tfrac{1}{4}(\mathbf{a}-\mathbf{b}), -\tfrac{1}{4}(\mathbf{a}-\mathbf{b}), -\tfrac{1}{4}(\mathbf{a}-\mathbf{b})\},\\ \mathbf{R}_{ac} &= \{-\tfrac{1}{4}\mathbf{a} - \mathbf{c}, -\tfrac{1}{4}(\mathbf{a}-\mathbf{c}), \mathbf{a}\\ &\quad + \tfrac{1}{4}\mathbf{c}, -\tfrac{1}{4}(\mathbf{a}-\mathbf{c}), -\tfrac{1}{4}(\mathbf{a}-\mathbf{c})\},\\ \mathbf{R}_{de} &= \{-\tfrac{1}{4}(\mathbf{d}-\mathbf{e}), -\tfrac{1}{4}(\mathbf{d}-\mathbf{e}),\\ &\quad -\tfrac{1}{4}(\mathbf{d}-\mathbf{e}), -\tfrac{1}{4}\mathbf{d} - \mathbf{e}, \mathbf{d} + \tfrac{1}{4}\mathbf{e}\}.\end{aligned} \tag{3.34}$$

It is straightforward to verify that $\mathbf{R}_{ab}$ is orthogonal to $\mathbf{T}_a$, $\mathbf{T}_e$, $\mathbf{D}$, $\mathbf{P}_a$, and $\mathbf{P}_e$ (use $\mathbf{P}_a = \{\frac{5}{3}\mathbf{a},0,0,0\} - \frac{1}{3}\mathbf{T}_a - \frac{1}{3}\mathbf{D}$) and to $\mathbf{K}''$ in (3.33). Also $\mathbf{R}_{ac}$ and $\mathbf{R}_{ae}$ are orthogonal to $\mathbf{K}''$. We have therefore identified the Klein modes in $d = 4$. In order to interpret them, we shift the origin to the center of mass of the (1234) system by defining $\mathbf{a}' = \mathbf{a} + \frac{1}{4}\mathbf{e}$, $b'\mathbf{b} + \frac{1}{4}\mathbf{e}$, etc., such that $\mathbf{a}' + \mathbf{b}' + \mathbf{c}' + \mathbf{d}' = 0$. We then obtain

$$\begin{aligned}\mathbf{K}(12,34)'' = \{&-\tfrac{1}{3}a' - b', -\tfrac{1}{3}b' - a', -\tfrac{1}{3}c' - d',\\ &-\tfrac{1}{3}d' - c', 0\}.\end{aligned} \tag{3.35}$$

These are just the Klein modes of the (1234) system with atom 5 at rest. Similarly the other Klein modes in $d = 4$ can be interpreted as Klein modes in $d = 3$ of four atoms in their center of mass frame, the fifth atom being a spectator.

## IV. CONCLUSIONS

In $d$ space dimensions the small oscillations of a hypersimplex have only three nonzero frequencies: the dilatational mode (a singlet), the pumper modes (degeneracy $d$), and the Klein modes [degeneracy $\frac{1}{2}d(d-1) - 1$]. In the pumper modes, one atom ($A$) moves to and away from the center while the other $d$ atoms make counter motions such that an atom $B$ moves in the $OAB$ plane. If the displacement of atom $A$ is $\mathbf{a}$, the displacement of atom $B$ is $-(\mathbf{a}+\mathbf{b})/(d-1)$,

and it is clear that the sum of the deviations of the $d$ $B$ atoms cancels the deviation of atom $A$.

In the Klein modes, two pairs of atoms, (1,2) and (3,4), oscillate roughly as follows: 1 and 2 oscillate harmonically along the edge (12) connecting them, and also 3 and 4 along the edge (34); moreover, when 1 and 2 (and 3 and 4) move apart, the edges (12) and (34) approach each other, while when 1 and 2 (and 3 and 4) come together, the edges (12) and (34) move apart. The edges (12) and (34) remain orthogonal. In $d = 4$, there are 15 such doublets of pairs, but since the sum of the three Klein modes in $d = 3$ vanishes $[\mathbf{K}(12,34) + \mathbf{K}(13,24) + \mathbf{K}(14,23) = 0]$, we can restrict our attentions to ten modes. These are still linearly dependent, for example, $\mathbf{K}(12,34)'' + \mathbf{K}(12,35)'' + \mathbf{K}(12,45)'' + \mathbf{K}(13,45)'' + \mathbf{K}(23,45)'' = 0$, and this leaves five independent Klein modes.


## ACKNOWLEDGMENTS

It is a pleasure to thank C. N. Yang for exhibiting in a clear and simple way the Klein modes of the tetrahedron, and C. H. Sah for useful discussions concerning polytopes.



[1] G. Venkatearaman and D. Sahoo, Contemp. Phys. **26**, 579 (1985).

[2] H. S. M. Coxeter, *Regular Polytopes* (MacMillan, New York, 1963), 2nd ed., Chap. 7. A simpler account can be found in *Introduction to Geometry* (Wiley, New York, 1961).

[3] F. D. Murnaghan, *The Theory of Group Representations* (Dover, New York, 1963), p. 142.

[4] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955). J. S. Lomont, *Applications of Finite Group Theory* (Academic, New York, 1959); H. Freudenthal, course in finite group theory at Utrecht University, 1962 (unpublished).

# Fierz identities in ten-dimensional superspace

Seichi Naito, Masahiro Nishimoto, and Tohru Hirokane[a)]
*Department of Physics, Osaka City University, Sumiyoshiku, Osaka, Japan*

(Received 12 April 1988; accepted for publication 21 September 1988)

All SO(1,9)-irreducible bases composed of tensors and tensor–spinors, which span the whole space up to 16th-order polynomials in $\theta$, $\theta_j$ ($j=1$–16) being Majorana–Weyl spinors of SO(1,9), are explicitly constructed. Any superfield $\varphi(x^m,\theta)$ in ten-dimensional superspace can be expanded by these bases, expansion coefficients being ordinary tensor and tensor–spinor fields. Then explicit formulas are given for $\varphi(x,\theta)\Phi(x,\theta)$ ($\Phi$ being another superfield) up to sixth order in $\theta$ in terms of ordinary fields, while those for $(\partial/\partial\bar{\theta})\varphi(x,\theta)$ and $\Gamma^m\theta\,\partial_m\varphi(x,\theta)$ are given up to eighth order in $\theta$.

## I. INTRODUCTION AND PRELIMINARIES

Historically four-dimensional supersymmetry[1,2] has attracted much attention since fermions are related with bosons by supercharge operators, which generate translations in four-dimensional superspace. If supersymmetry is respected by physical vacuum, there should exist mass degeneracy between bosons and fermions. However, our real world has no such degeneracy, but nevertheless there is no candidate for the Goldstino particle[2] that is present when global supersymmetry is spontaneously broken. Absence of Goldstino can be explained by the super-Higgs mechanism,[2] provided that we have a gauge theory having local supersymmetry. Since energy-momentum operators are constructed from supercharge operators, we are led to local translations in space-time, so that gravity appears naturally in such gauge theories (i.e., supergravity theories[3,4]). Furthermore, it has been found that nice properties such as vanishing anomalies and counterterms at one loop level[5] exist in extended supergravity theories that are arrived at by using dimensional reduction from a ten-dimensional supergravity, and as a bonus we have Yang–Mills particles as well as gravitons in four-dimensional space-time. On the other hand, the superstring model has long been known to be consistent only in ten-dimensional space-time, and Yang–Mills (gravitational) fields appear as excited modes of open (closed) strings. Moreover, Green and Schwarz[6] have proposed a superstring Lagrangian, where a superstring moves in ten-dimensional superspace. Thus ten-dimensional superspace has been found from two approaches (i.e., supergravity[4] and superstring[7]) by physicists trying to unify Yang–Mills interactions with gravitational ones. Therefore we believe that the superfield in ten-dimensional superspace $(x^m,\theta)$ will play an important role in various possible unified theories of all interactions. However, it is necessary to transform proposed equations[8,9] for *super*fields into those for fields. For this purpose, we must expand superfields in terms of $\theta$, in such a way that coefficients turn out to be fields that are irreducible under SO(1,9) transformations. Therefore, we must explicitly construct all SO(1,9)-irreducible bases composed of tensors and tensor–spinors that span the whole space up to 16th order of polynomials in $\theta$, and they are given in this section.

We consider the ten-dimensional superspace

$$z^M=(x^m,\theta_j), \tag{1.1}$$

where space-time coordinate $x^m$ (supercoordinate $\theta_j$) transforms as a ten-dimensional vector (16-dimensional Majorana–Weyl spinor) representation of SO(1,9), so that vector (spinor) indices $m$ ($j$) run over $1+9$ (16) values. In investigating the spinor representation of SO(1,9) it is useful to introduce $\Gamma^m$ matrices satisfying the Clifford algebra

$$\Gamma^{m_1}\Gamma^{m_2}+\Gamma^{m_2}\Gamma^{m_1}=2\eta^{m_1m_2}, \tag{1.2}$$

where

$$\eta^{m_1m_2}=\begin{cases}1, & \text{for } m_1=m_2=0,\\ -1, & \text{for } m_1=m_2=1\text{–}9,\\ 0, & \text{otherwise.}\end{cases} \tag{1.3}$$

Then the Majorana–Weyl property of $\theta_j$ is characterized by the Majorana condition

$$\theta=C\bar{\theta}^T \quad (\text{i.e., } \bar{\theta}=-\theta^TC^{-1}) \tag{1.4}$$

and the Weyl (or chirality) condition

$$\Gamma^{11}\theta=\pm\theta, \tag{1.5}$$

where $\Gamma^{11}\equiv\Gamma^0\Gamma^1\Gamma^2\Gamma^3\Gamma^4\Gamma^5\Gamma^6\Gamma^7\Gamma^8\Gamma^9$ and $C$ is the (antisymmetric) charge conjugation matrix satisfying

$$C^{-1}\Gamma^mC=-(\Gamma^m)^T, \tag{1.6}$$

where $(\Gamma^m)^T$ is the transposed matrix of $\Gamma^m$. The antisymmetrization operator $[1,2,\ldots,n]$ and the symmetrization one $\{1,2,\ldots,n\}$ are defined, respectively, by

$$[1,2,\ldots,n]=\frac{1}{n!}\sum_P\delta_PP \tag{1.7}$$

and

$$\{1,2,\ldots,n\}=\frac{1}{n!}\sum_P P, \tag{1.8}$$

where $P$ represents any permutation among $n$ indices $(1,2,\ldots,n)$ and

$$\delta_P=\begin{cases}1, & \text{for even } P,\\ -1, & \text{for odd } P.\end{cases} \tag{1.9}$$

When calculating products of some $\Gamma$ matrices, it is quite useful to introduce

$$\Gamma^{m_1m_2\cdots m_n}\equiv[m_1,m_2,\cdots,m_n]\Gamma^{m_1}\Gamma^{m_2}\cdots\Gamma^{m_n}$$
$$\equiv\Gamma^{[m_1}\Gamma^{m_2}\cdots\Gamma^{m_n]}$$

[a)] On leave of absence from Department of Physics, Kanazawa University, Kanazawa 920, Japan.

 0022-2488/89/020261-13$02.50 © 1989 American Institute of Physics 

and use the formula[8]

$$\Gamma^{m}\Gamma^{m_1\cdots m_n}\equiv\Gamma^{m\,m_1\cdots m_n}+n\eta^{m[m_1}\Gamma^{m_2\cdots m_n]}. \tag{1.10}$$

The Fierz identity having two $\theta$'s is easily derived by using the Majorana–Weyl property of $\theta$ and is given by[8]

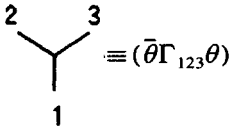
$$\theta_{j_1}\bar{\theta}_{j_2}=-\frac{1}{96}\sum_{m_1m_2m_3}(\bar{\theta}\Gamma_{m_1m_2m_3}\theta)(\Gamma^{m_3m_2m_1})_{j_1j_2}. \tag{1.11}$$

In (1.11), $\theta_{j_1}\bar{\theta}_{j_2}$ with only spinor indices $j_1$ and $j_2$ is called open, while $(\bar{\theta}\Gamma_{m_1m_2m_3}\theta)$ with only vector indices $m_1$–$m_3$ is called closed. For simplicity of notation, vector indices $m_i$ in (1.11) are denoted by $i$, so that (1.11) is simply expressed by

$$\theta\cdot\bar{\theta}=\tfrac{1}{96}(\bar{\theta}\Gamma_{123}\theta)\Gamma^{123}, \tag{1.12}$$

where these and future dummy indices are to be summed over. These *notations i for vector indices $m_i$ will be adopted throughout this paper* and they should not be confused with definite values $i$ among 0,1,...,9.

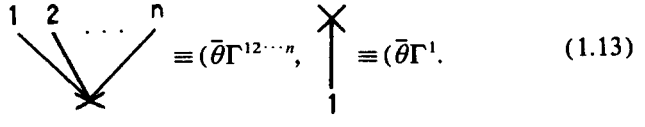
First, we introduce diagrammatical expressions for the $\theta$ polynomials, diagrams being composed of the following building blocks:




$\equiv(\bar{\theta}\Gamma_{123}\theta)$, which is called $Y$,

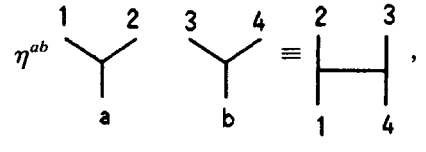
$\times\equiv(\bar{\theta}$, which is called $X$,

and

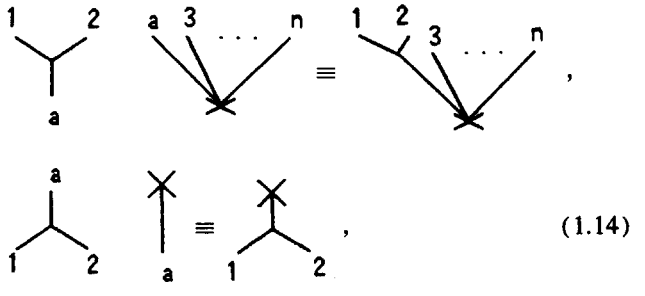



$\equiv(\bar{\theta}\Gamma^{12\cdots n}$, $\equiv(\bar{\theta}\Gamma^{1}$. (1.13)

These blocks can be connected by taking an $(a,b)$ trace



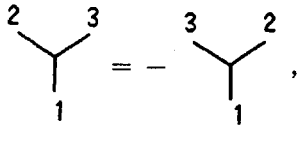

which is called $H$, and a $\Gamma^a$ trace

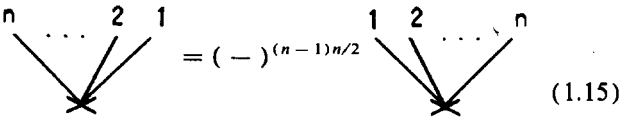



(1.14)

and connected lines will be called links.

We can easily find the "mirror reflection"

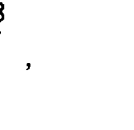



$=(-)^{(n-1)n/2}$ (1.15)

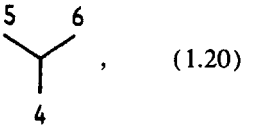
We explicitly construct all SO(1,9)-irreducible bases composed of tensors and tensor–spinors that span the whole space up to 16th-order polynomials in $\theta$. They can be compactly expressed diagrammatically as follows (note that notations 1,2,... are used to represent $m_1,m_2,...$):

$(\bar{\theta}\equiv\times$, (1.16)

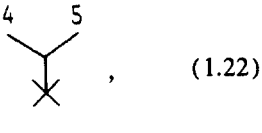
$(\theta^2)_{123}\equiv$  , (1.17)

$(\bar{\theta}^3)_{12}\equiv$  , (1.18)

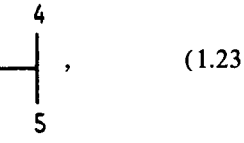
$(\theta^4)_{12,34}\equiv(\theta^4)_{34,12}\equiv\frac{3}{4}$  , (1.19)

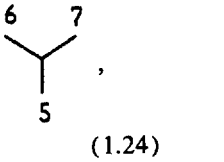
$(\theta^4)_{12345,6}\equiv[1,2,3,4,5]$  , (1.20)

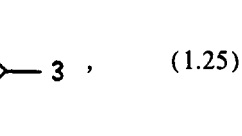
$(\bar{\theta}^5)_{12,3}\equiv\frac{3}{4}$  , (1.21)

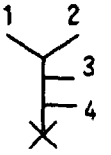
$(\bar{\theta}^5)_{12345}\equiv[1,2,3,4,5]$  , (1.22)

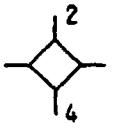
$(\theta^6)_{124,3,5}\equiv(\theta^6)_{124,5,3}$

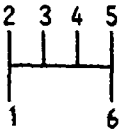
$\equiv\frac{5}{6}[1,2,4]$  , (1.23)

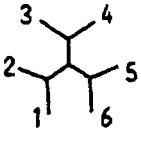
$(\theta^6)_{12567,34}\equiv-\frac{3}{8}[1,2,5,6,7]$  , (1.24)

$(\bar{\theta}^7)_{1,2,3}\equiv(\bar{\theta}^7)_{\{1,2,3\}}\equiv$  , (1.25)

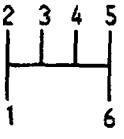
$$(\bar{\theta}^{7})_{124,3} \equiv -\frac{1}{4}[1,2,4]$$  , (1.26)

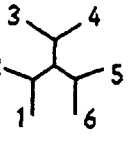
$$(\theta^{8})_{1,2,3,4} \equiv (\theta^{8})_{\{1,2,3,4\}} \equiv$$  , (1.27)

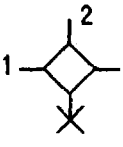
$$(\theta^{8})_{2345,1,6} \equiv (\theta^{8})_{2345,6,1} \equiv [2,3,4,5]$$ 

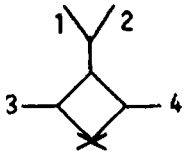
$$\equiv \frac{1}{2}[2,3,4,5]$$  , (1.28)

$$(\theta^{8})_{124,356} \equiv (\theta^{8})_{356,124}$$

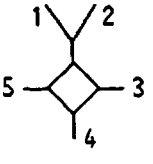
$$\equiv -\frac{3}{16}[1,2,4][3,5,6]$$ 

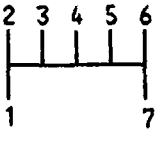
$$\equiv -\frac{3}{16}[1,2,4][3,5,6]$$  , (1.29)

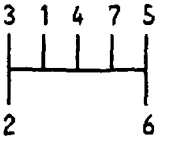
$$(\bar{\theta}^{9})_{1,2,3} \equiv (\bar{\theta}^{9})_{\{1,2,3\}} \equiv$$  , (1.30)

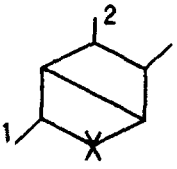
$$(\bar{\theta}^{9})_{123,4} \equiv \frac{1}{2}[1,2,3]$$  , (1.31)

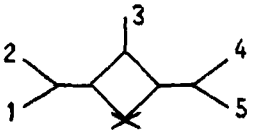
$$(\theta^{10})_{123,4,5} \equiv (\theta^{10})_{123,5,4}$$

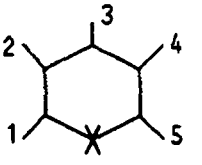
$$\equiv \frac{1}{3}[1,2,3]$$  , (1.32)

$$(\theta^{10})_{23456,17} \equiv \frac{1}{2}[2,3,4,5,6]$$ 

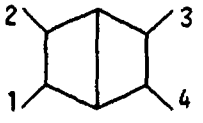
$$\equiv \frac{1}{8}[2,3,4,5,6]$$  , (1.33)

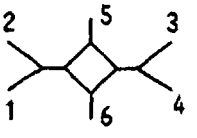
$$(\bar{\theta}^{11})_{12,3} \equiv \frac{1}{2}[1,2]$$  , (1.34)

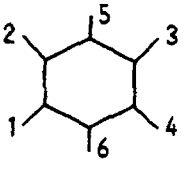
$$(\bar{\theta}^{11})_{12345} \equiv [1,2,3,4,5]$$ 

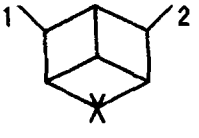
$$\equiv 4[1,2,3,4,5]$$  , (1.35)

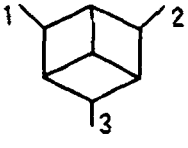
$$(\theta^{12})_{13,24} \equiv (\theta^{12})_{24,13}$$

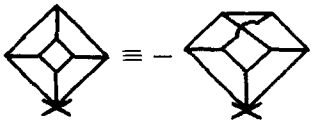
$$\equiv \frac{1}{2}[1,3]$$  , (1.36)

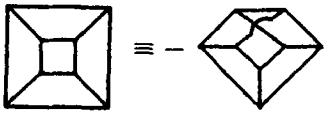
$$(\theta^{12})_{12345,6} \equiv \frac{1}{2}[1,2,3,4,5]$$ 

$$\equiv 2[1,2,3,4,5]$$  , (1.37)

$$(\bar{\theta}^{13})_{12} \equiv$$  , (1.38)

$$(\theta^{14})_{123} \equiv$$  , (1.39)

$$(\bar{\theta}^{15}) \equiv$$  , (1.40)

$$(\theta^{16}) \equiv$$  . (1.41)

In (1.16)–(1.41), the notation

$$(\theta^n)_{m_{(1,1)}\cdots m_{(g_1,1)},\ldots,m_{(1,f_1)}\cdots m_{(g_{f_1},f_1)}} \tag{1.42}$$

explicitly shows that (1.42) is completely antisymmetric among blocks (of vector indices) divided by commas, i.e., we have

$$[m_{(1,s)},m_{(2,s)},\ldots,m_{(g_s,s)}] \quad (\text{for } 1\leqslant s\leqslant f_1). \tag{1.43}$$

Furthermore, we have traceless properties

$$\eta^{ab}(\theta^{2n})_{\ldots a\ldots b\ldots}=0 \quad (\text{for } n=1\text{–}8), \tag{1.44}$$

and $\Gamma$-traceless properties

$$\Gamma^a(\theta^{2n+1})_{\ldots a\ldots}=0 \quad (\text{for } n=2\text{–}7). \tag{1.45}$$

By using $g_1,g_2,\ldots,g_{f_1}$ in (1.42), we construct a Young diagram, the $s$th column being composed of $g_s$ boxes. Suppose that $f_i$ boxes exist in the $i$th row of the thus obtained Young diagram. Then, (1.42) for even (odd) $n$ are tensors (tensor–spinors) of SO(1,9)-irreducible representations $(f_1,f_2,f_3,f_4,f_5)$ $[(f_1+\frac{1}{2},f_2+\frac{1}{2},f_3+\frac{1}{2},f_4+\frac{1}{2},f_5+\frac{1}{2})]$.[10,11]

We can prove that our irreducible tensors and tensor–spinors in fact satisfy the following "antisymmetric properties":

$$(\theta^4)_{[12,3]4}=0, \tag{1.19'}$$
$$(\theta^4)_{[12345,6]}=0, \tag{1.20'}$$
$$(\bar\theta^5)_{[12,3]}=0, \tag{1.21'}$$
$$(\theta^6)_{[124,3]5}=0, \tag{1.23'}$$
$$(\theta^6)_{[12567,3]4}=0, \tag{1.24'}$$
$$(\theta^6)_{1[2567,34]}=0, \tag{1.24''}$$
$$(\bar\theta^7)_{[124,3]}=0, \tag{1.26'}$$
$$(\theta^8)_{[2345,6],1}=(0), \tag{1.28'}$$
$$(\theta^8)_{[124,3]56}=0, \tag{1.29'}$$
$$(\theta^8)_{1[24,56]3}=0, \tag{1.29''}$$
$$(\bar\theta^9)_{[123,4]}=0, \tag{1.31'}$$
$$(\theta^{10})_{[123,4]5}=0, \tag{1.32'}$$
$$(\theta^{10})_{[23456,1]7}=0, \tag{1.33'}$$
$$(\bar\theta^{11})_{[12,3]}=0, \tag{1.34'}$$
$$(\theta^{12})_{[13,2]4}=0, \tag{1.36'}$$
$$(\theta^{12})_{[12345,6]}=0 \tag{1.37'}$$

In Sec. II, the superfield $\varphi(x,\theta)$ is expanded with respect to $\theta$ by using irreducible tensors and tensor–spinors so that coefficients turn out to be tensor fields and tensor–spinor fields. Then, explicit formulas for $\varphi(x,\theta)\Phi(x,\theta)$ ($\Phi$ being another superfield) are given up to sixth order in $\theta$, while those for $(\partial/\partial\bar\theta)\varphi(x,\theta)$, $\Gamma^m\theta(\partial/\partial x^m)\varphi(x,\theta)$ are given up to eighth order in $\theta$, in terms of tensor and tensor–spinor fields. In the Appendix, we give various Fierz identities necessary in deriving formulas in Sec. II. In Sec. III, we summarize and discuss our results.

## II. SUPERFIELDS

Superfields in superspace $(x^m,\theta_j)$ are useful in constructing physical models having supersymmetry. However, it is necessary to transform proposed equations for superfields[8,9] into those for fields. For this purpose, we must transform operations such as

$$\varphi^2(x,\theta),\frac{\partial}{\partial\bar\theta}\varphi(x,\theta),\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta),$$

for superfields $\varphi(x,\theta)$'s into those in various fields in $x$ space, which appear as coefficients when superfields $\varphi(x,\theta)$'s are expanded in $\theta$. These analyses in four dimensions have been carried out completely by Salam and Strathdee,[2] while those in ten dimensions are lacking. (This situation is unfortunate, since ten dimensions are used in both ten-dimensional supergravity and the superstring model, which are possible candidates for a unified theory of all interactions including gravity.) In this section, we carry out these analyses which are extremely tedious but straightforward. Since expansions in $\theta$ can be treated similarly for any superfield, we give those for any scalar superfields $\varphi(x,\theta)$ only *as examples*:

$$\begin{aligned}\varphi(x,\theta)=&\,t(x)+\bar\theta\psi(x)+(\theta^2)_{123}t^{123}(x)+(\bar\theta^3)_{12}\psi^{12}(x)+(\theta^4)_{12,34}t^{12,34}(x)+(\theta^4)_{12345,6}t^{12345,6}(x)\\&+(\bar\theta^5)_{12,3}\psi^{12,3}+(\bar\theta^5)_{12345}\psi^{12345}(x)+(\theta^6)_{123,4,5}t^{123,4,5}(x)+(\theta^6)_{12345,67}t^{12345,67}(x)\\&+(\bar\theta^7)_{1,2,3}\psi^{1,2,3}(x)+(\bar\theta^7)_{123,4}\psi^{123,4}(x)+(\theta^8)_{1,2,3,4}t^{1,2,3,4}(x)+(\theta^8)_{1234,5,6}t^{1234,5,6}(x)\\&+(\theta^8)_{123,456}t^{123,456}(x)+(\bar\theta^9)_{1,2,3}\lambda^{1,2,3}(x)+(\bar\theta^9)_{123,4}\lambda^{123,4}(x)\\&+(\theta^{10})_{123,4,5}f^{123,4,5}(x)+(\theta^{10})_{12345,67}f^{12345,67}(x)+(\bar\theta^{11})_{12,3}\lambda^{12,3}(x)+(\bar\theta^{11})_{12345}\lambda^{12345}(x)\\&+(\theta^{12})_{12,34}f^{12,34}(x)+(\theta^{12})_{12345,6}f^{12345,6}(x)+(\bar\theta^{13})_{12}\lambda^{12}(x)+(\theta^{14})_{123}f^{123}(x)\\&+(\bar\theta^{15})\lambda(x)+(\theta^{16})f(x),\end{aligned} \tag{2.1}$$

where fields $t\ldots(x)$ and $f\ldots(x)$ will be called tensor fields, while $\psi\ldots(x)$ and $\lambda\ldots(x)$ will be called tensor–spinor fields.

In order to derive formulas for transforming the products of superfields into those of tensor (–spinor) fields, we introduce a new superfield $\Phi(x,\theta)$ obtained from (4.1) by replacing all (super-, tensor, and tensor–spinor) fields denoted by lowercase letters with those denoted by capital letters. When calculating the product $\varphi(x,\theta)\cdot\Phi(x,\theta)$, all tensor and tensor–spinor fields in $\varphi(x,\theta)$ and $\Phi(x,\theta)$ are assumed to satisfy the same antisymmetric properties as those for corresponding $(\theta^n)$-irreducible bases, but they are not assumed to be traceless and $\Gamma$ traceless, respectively. In order to analyze $(\theta^n)$ polynomials of the product $\varphi(x,\theta)\cdot\Phi(x,\theta)$, which are denoted by $[\varphi\Phi]_n$, we need

Fierz identities expressing tensor products of two irreducible bases $\theta^i\theta^{n-i}$ $(i=1\text{–}n-1)$ in terms of irreducible bases $\theta^n$. These Fierz identities are lacking in the literature, so we give them in the Appendix. In the following, we give formulas $[\varphi\cdot\Phi]_n$ up to sixth order in $\theta$, while previous authors have given them only up to $n=3$. In order to express the results compactly, we introduce the notation $(\varphi\leftrightarrow\Phi)$ for those terms obtained from all other preceding terms by the operation

$$\begin{aligned} &t_{\dots}(x)\leftrightarrow T_{\dots}(x),\quad f_{\dots}(x)\leftrightarrow F_{\dots}(x),\\ &\psi_{\dots}(x)\leftrightarrow\Psi_{\dots}(x),\quad \lambda_{\dots}(x)\leftrightarrow\Lambda_{\dots}(x). \end{aligned} \tag{2.2}$$

The results obtained are as follows:

$$[\varphi\cdot\Phi]_0 = t(x)T(x), \tag{2.3}$$

$$[\varphi\cdot\Phi]_1 = (\bar\theta t(x)\Psi(x) + (\varphi\leftrightarrow\Phi), \tag{2.4}$$

$$[\varphi\cdot\Phi]_2 = (\theta^2)_{123}(t(x)T^{123}(x) + \tfrac{1}{192}\bar\psi^c(x)\Gamma^{123}\Psi(x)) + (\varphi\leftrightarrow\Phi), \tag{2.5}$$

where we have used

$$\bar\theta\cdot\psi(x)\equiv\bar\psi^c(x)\cdot\theta$$

[with $\psi^c(x)\equiv C\bar\psi^T(x)$],

$$[\psi\cdot\Phi]_3 = (\bar\theta^3)_{12}(t(x)\Psi^{12}(x) + \tfrac{1}{2}\Gamma_a\psi(x)T^{12a}(x)) + (\varphi\leftrightarrow\Phi), \tag{2.6}$$

$$\begin{aligned}[\varphi\cdot\Phi]_4 = {}&(\theta^4)_{12,34}\begin{pmatrix} t(x)T^{12,34}(x) + \frac{1}{24}\bar\psi^c(x)\Gamma^{12}\Psi^{34}(x)\\ + t_a{}^{12}(x)T^{a34}(x)\end{pmatrix}\\ &+(\theta^4)_{12345,6}\begin{pmatrix} t(x)T^{12345,6}(x) + \frac{5}{2}t^{123}(x)T^{456}(x)\\ -\frac{5}{96}\bar\psi^c(x)\Gamma^{3456}\Psi^{12}(x)\end{pmatrix} + (\varphi\leftrightarrow\Phi),\end{aligned} \tag{2.7}$$

$$\begin{aligned}[\varphi\cdot\Phi]_5 = {}&(\bar\theta^5)_{12,3}\begin{pmatrix} t(x)\Psi^{12,3}(x) - \frac{6}{5}\Psi^{3a}(x)t_a{}^{12}(x)\\ -\frac{2}{5}\Gamma_{ab}\Psi^{3a}(x)t^{12b}(x) + \frac{2}{5}\Gamma_{ab}\Psi^{12}(x)t^{3ab}(x)\\ +\frac{2}{5}t^{12,3a}(x)\Gamma_a\Psi(x) + \frac{2}{15}\Gamma_{abc}\Psi(x)t^{12abc,3}(x)\end{pmatrix}\\ &+(\bar\theta^5)_{12345}\begin{pmatrix} t(x)\Psi^{12345}(x) + \Psi^{45}(x)t^{123}(x)\\ +\frac{1}{10}\Gamma_a\Psi(x)t^{12345,a}(x)\end{pmatrix} + (\varphi\leftrightarrow\Phi),\end{aligned} \tag{2.8}$$

$$\begin{aligned}[\varphi\cdot\Phi]_6 = {}&(\theta^6)_{123,4,5}\begin{pmatrix} t(x)T^{123,4,5} + \frac{1}{240}\bar\psi^{c\,a4}(x)\Gamma_a\Gamma^{123}\Gamma_b\Psi^{b\,5}(x)\\ +\frac{3}{100}\bar\psi^{c\,a5}(x)\Gamma_a\Gamma^{12}\Psi^{34}(x) - \frac{1}{50}\bar\psi^{c\,a4}(x)\Gamma^{123}\Psi_a{}^5(x)\\ +\frac{3}{20}\bar\psi^c(x)\Gamma^{34}\Psi^{12,5}(x) + \frac{1}{80}\bar\psi^c(x)\Gamma^{123}\Gamma_a\Psi^{a4,5}(x)\\ -\frac{108}{25}t_a{}^{34}(x)T^{12,a5}(x) - \frac{9}{25}t^{123}(x)T_a{}^{4,a5}(x)\\ -\frac{36}{25}t^{ab\,123,4}(x)T_{ab}{}^5(x) - \frac{3}{10}\bar\psi^{c\,14}(x)\Gamma^2\Psi^{35}(x)\end{pmatrix}\\ &+(\theta^6)_{12345,67}\begin{pmatrix} t(x)T^{12345,67}(x) - 8T_a{}^{1234,7}(x)t^{a56}(x)\\ +\frac{8}{3}T^{a1234,}{}_a(x)t^{567}(x) - \frac{8}{5}T^{12345,}{}_a(x)t^{a67}(x)\\ -\frac{1}{36}\bar\psi^c(x)\Gamma^{1237}\Psi^{45,6}(x) - \frac{1}{6}\bar\psi^c(x)\Gamma^{67}\Psi^{12345}(x)\\ +\frac{5}{36}\bar\psi^c(x)\Gamma^{567}\Gamma_a\Psi^{a1234}(x) - 4t^{123}(x)T^{45,67}(x)\\ -\frac{1}{18}\bar\psi^c{}_a{}^6(x)\Gamma^{12345}\Psi^{a7}(x) + \frac{1}{4}\bar\psi^{c\,12}(x)\Gamma^{367}\Psi^{45}(x)\\ -\frac{1}{72}\bar\psi^{c\,67}(x)\Gamma^{1234}\Gamma_a\Psi^{a5}(x) - \frac{1}{27}\bar\psi^{c\,16}(x)\Gamma^{234}\Psi^{57}(x)\\ +\frac{5}{216}\bar\Psi^{c\,56}(x)\Gamma^{1234}\Gamma_a\psi^{a7}(x)\end{pmatrix} + (\varphi\leftrightarrow\Phi)\end{aligned} \tag{2.9}$$

As we can see from $[\varphi\cdot\Phi]_0$–$[\varphi\cdot\Phi]_6$, tensor and tensor–spinor fields for $\varphi(x,\theta)\Phi(x,\theta)$ do not necessarily satisfy "antisymmetric properties" and traceless or $\Gamma$-traceless conditions. But GL(10) projections specified by certain Young tableaux easily transform these fields into those satisfying "antisymmetric properties." The product $\varphi(x,\theta)\Phi(x,\theta)$ being expanded by tensor or tensor–spinor fields obtained after this projection is denoted by $\phi(x,\theta)$. Then, if we calculate $\varphi(x,\theta)\phi(x,\theta)$, we obtain $\varphi^2(x,\theta)\Phi(x,\theta)$. In this manner, we can calculate arbitrary polynomials of superfields. (There are no essential difficulties in deriving expressions for $[\varphi\cdot\Phi]_n$ $(n=7\text{–}12)$. However, calculations are laborious and expressions are extremely lengthy, so we do not give explicit expressions for them in this paper.)

In the following, $(\theta^n)$ polynomials for $\bar{\epsilon}(\partial/\partial\bar{\theta})\varphi(x,\theta)$ and $\bar{\epsilon}\Gamma^m\theta(\partial/\partial x^m)\varphi(x,\theta)$ are denoted, respectively, by $[\bar{\epsilon}(\partial/\partial\bar{\theta})\varphi(x,\theta)]_n$ and $[\bar{\epsilon}\Gamma^m\theta(\partial/\partial x^m)\varphi(x,\theta)]_n$ ($\epsilon$ being an arbitrary Majorana–Weyl spinor), and they are explicitly given up to $n=8$;

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_0=\bar{\epsilon}\psi(x), \tag{2.10}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_0=0, \tag{2.10'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_1=2(\bar{\theta}\Gamma_{123}\epsilon)\times t^{123}(x), \tag{2.11}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_1=-(\bar{\theta}\Gamma^a\epsilon)\times\partial_a t(x), \tag{2.11'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_2=3(\theta^2)_{123}\times\bar{\epsilon}\Gamma^1\psi^{23}(x), \tag{2.12}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_2=\frac{1}{96}(\theta^2)_{123}\times\bar{\epsilon}\Gamma^a\Gamma^{123}\partial_a\psi(x), \tag{2.12'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_3=3(\bar{\theta}^3)_{34}\Gamma_{12}\epsilon\times t^{12,34}(x)-(\bar{\theta}^3)_{56}\Gamma_{1234}\epsilon\times t^{12345,6}(x), \tag{2.13}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_3=-\frac{1}{2}(\bar{\theta}^3)_{12}\Gamma_b\Gamma^a\epsilon\times\partial_a t^{12b}(x), \tag{2.13'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_4=-\frac{5}{2}(\theta^4)_{12,34}\bar{\epsilon}\Gamma^1\psi^{34,2}(x)-\frac{25}{16}(\theta^4)_{12345,6}\times\bar{\epsilon}\Gamma^{123}\psi^{45,6}(x)+5(\theta^4)_{12345,6}\times\bar{\epsilon}\Gamma^6\psi^{12345}(x), \tag{2.14}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_4=\frac{1}{24}(\theta^4)_{12,34}\times\bar{\epsilon}\Gamma^a\Gamma^{12}\partial_a\psi^{34}(x)+\frac{5}{192}(\theta^4)_{12345,6}\times\bar{\epsilon}\Gamma^a\Gamma^{1234}\partial_a\psi^{56}(x), \tag{2.14'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_5=\frac{20}{3}(\bar{\theta}^5)_{12,5}\Gamma_{34}\epsilon\times t^{123,4,5}(x)-\frac{1}{2}(\bar{\theta}^5)_{67,5}\Gamma_{1234}\epsilon\times t^{12345,67}(x)-\frac{3}{4}(\bar{\theta}^5)_{12345}\Gamma_{67}\epsilon\times t^{12345,67}(x), \tag{2.15}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_5=-\frac{2}{5}(\bar{\theta}^5)_{12,3}\Gamma_4\Gamma^a\epsilon\times\partial_a t^{12,34}(x)-\frac{2}{15}(\bar{\theta}^5)_{12,6}\Gamma_{345}\Gamma^a\epsilon\times\partial_a t^{12345,6}(x)-\frac{1}{10}(\bar{\theta}^5)_{12345}\Gamma_6\Gamma^a\epsilon\times\partial_a t^{12345,6}(x), \tag{2.15'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_6=\frac{54}{5}(\theta^6)_{123,4,5}\times\bar{\epsilon}\Gamma^{12}\psi^{3,4,5}(x)-\frac{169}{200}(\theta^6)_{123,4,5}\times\bar{\epsilon}\Gamma^4\psi^{123,5}(x)+\frac{95}{27}(\theta^6)_{12345,67}\times\bar{\epsilon}\Gamma^{123}\psi^{456,7}(x), \tag{2.16}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_6=\frac{9}{80}(\theta^6)_{123,4,5}\times\bar{\epsilon}\Gamma^a\Gamma^{34}\partial_a\psi^{12,5}(x)+\frac{5}{192}(\theta^6)_{12345,67}\times\bar{\epsilon}\Gamma^a\Gamma^{1234}\partial_a\psi^{67,5}(x)-\frac{1}{6}(\theta^6)_{12345,67}\times\bar{\epsilon}\Gamma^a\Gamma^{67}\partial_a\psi^{12345}(x), \tag{2.16'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_7=8(\bar{\theta}^7)_{1,2,3}\Gamma_4\epsilon\times t^{1,2,3,4}(x)+6(\bar{\theta}^7)_{123,4}\Gamma_{56}\epsilon\times t^{123,456}(x)+\frac{80}{7}(\bar{\theta}^7)_{123,4}\Gamma_{56}\epsilon\times t^{1235,4,6}(x)-\frac{4}{21}(\bar{\theta}^7)_{1,2,3}\Gamma_{456}\epsilon\times t^{4561,2,3}(x), \tag{2.17}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_7=\frac{5}{168}(\bar{\theta}^7)_{1,4,5}\Gamma_{23}\Gamma^a\epsilon\times\partial_a t^{123,4,5}(x)+\frac{10}{21}(\bar{\theta}^7)_{123,4}\Gamma_5\Gamma^a\epsilon\times\partial_a t^{123,4,5}(x)+\frac{5}{14}(\bar{\theta}^7)_{124,5}\Gamma_3\Gamma^a\epsilon\times\partial_a t^{123,4,5}(x)-\frac{9}{28}(\bar{\theta}^7)_{671,2}\Gamma_{345}\Gamma^a\epsilon\times\partial_a t^{12345,67}(x)-\frac{3}{14}(\bar{\theta}^7)_{123,6}\Gamma_{457}\Gamma^a\epsilon\times\partial_a t^{12345,67}(x), \tag{2.17'}$$

$$\left[\bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\varphi(x,\theta)\right]_8=2(\theta^8)_{1,2,3,4}\times\bar{\epsilon}\Gamma^4\psi^{1,2,3}(x)-9(\theta^8)_{1234,5,6}\times\bar{\epsilon}\Gamma^{123}\psi^{4,5,6}(x)-19(\theta^8)_{1234,5,6}\times\bar{\epsilon}\Gamma^{12}\psi^{345,6}(x)-14(\theta^8)_{123,456}\times\bar{\epsilon}\Gamma^{56}\psi^{123,4}(x), \tag{2.18}$$

$$\left[\bar{\epsilon}\Gamma^m\theta\frac{\partial}{\partial x^m}\varphi(x,\theta)\right]_8=-\frac{1}{16}(\theta^8)_{1,2,3,4}\times\bar{\epsilon}\Gamma^a\Gamma^4\partial_a\psi^{1,2,3}(x)$$

$$-\frac{9}{32}(\theta^8)_{1234,5,6}\times\bar{\epsilon}\Gamma^a\Gamma^{123}\partial_a\psi^{4,5,6}(x)$$
$$+\frac{3}{40}(\theta^8)_{1235,6,4}\times\bar{\epsilon}\Gamma^a\Gamma^{56}\partial_a\psi^{123,4}(x)$$
$$+\frac{1}{16}(\theta^8)_{123,456}\times\bar{\epsilon}\Gamma^a\Gamma^{56}\partial_a\psi^{123,4}(x), \qquad (2.18')$$

where we have used various Fierz identities given in the Appendix.

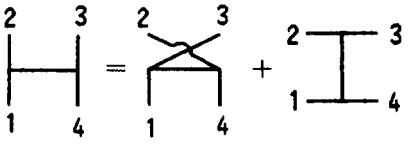
## III. SUMMARY AND DISCUSSION

Superfields in superspace $(x^m,\theta_j)$ are useful in constructing physical models having supersymmetry. However, it is necessary to transform proposed equations[8,9] for *superfields* into those for fields. For this purpose, we must expand superfields in terms of $\theta$, in such ways that coefficients turn out to be fields that are irreducible under SO(1,9) transformations. Therefore, we have constructed all SO(1,9)-irreducible bases (1.16)–(1.41) that span the whole space up to 16th order of polynomials in $\theta$, and satisfy antisymmetric properties (1.43) and (1.19')–(1.37'). With the help of expansion (2.1), we must transform operations such as $\varphi^2(x,\theta)$, $(\partial/\partial\bar{\theta})\varphi(x,\theta)$, $\Gamma^m\theta(\partial/\partial x^m)\varphi(x,\theta)$, for superfields $\varphi(x,\theta)$ into those in various fields in $x$ space. We have carried out these analyses and have obtained (2.3)–(2.19'), where it was necessary to have Fierz identities (A1)–(A35) expressing products of two SO(1,9)-irreducible bases $\theta^i\theta^{n-i}$ in terms of SO(1,9)-irreducible bases $\theta^n$. Furthermore, we have given the most general Fierz identities $Y_nX$, $Y_{n+1}$ for $2\leqslant n\leqslant 3$ in (A36)–(A39). Although our formulas are still not exhaustive, we have given formulas (A20)–(A35) for calculating $Y_nX$ and $Y_{n+1}$ for $4\leqslant n\leqslant 7$.

In this paper, we have proposed using diagrams that have one-to-one correspondence with $(\theta^n)$ polynomials. Building blocks for these diagrams are given by (1.13) and these blocks can be connected by contracting operators such as shown in (1.14), where connected lines are called links. We have found that representing $(\theta^n)$ polynomial by using diagrams are quite useful, since we can easily find out the existence of the following structures in $(\theta^n)$ polynomials.

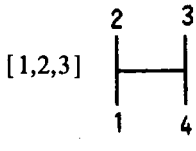
(1) We have mirror reflection (1.15).

(2) "Antisymmetric properties" (1.19') are expressed by




(3.1)

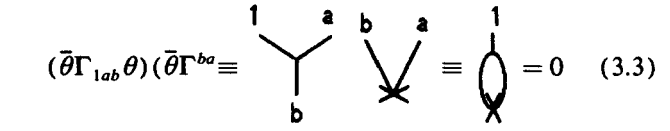
and




$= 0,$ (3.2)

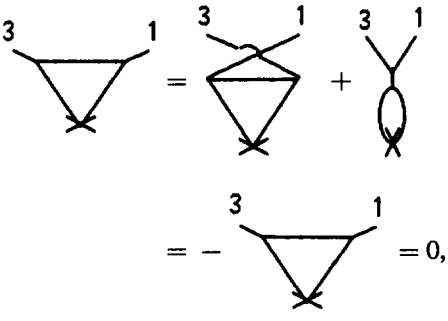
so that (3.1) will be called a "crossing relation," while (3.2) will be named "erased $H$."

(3) Any loop diagram composed of two or three links does not exist, and these facts are called the "vanishing root"




(3.3)

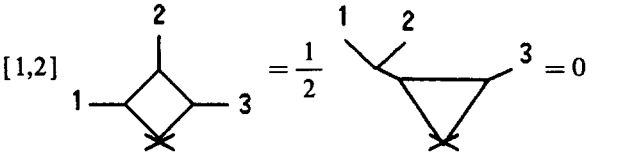
and "vanishing triangle"




(3.4)

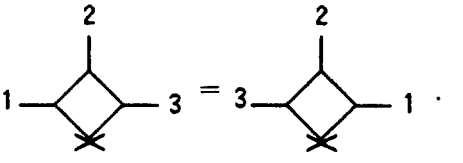
where we have used (3.1) and (3.3).

(4) We have complete symmetry among space-time indices for any loop diagram composed of four links, and this fact is called "completely symmetric square";




(3.5)

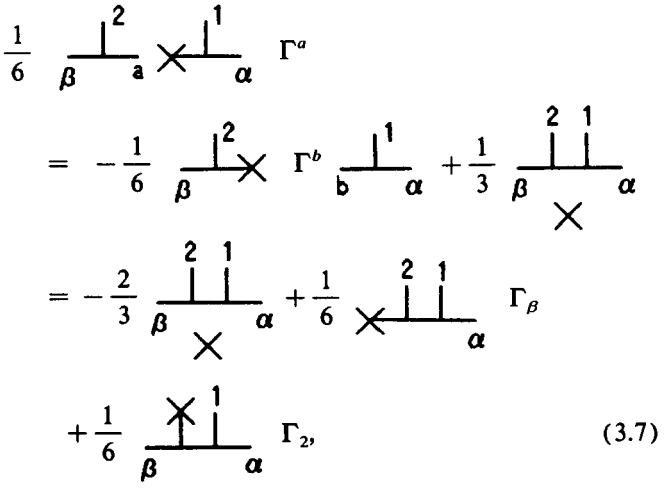
and




(3.6)

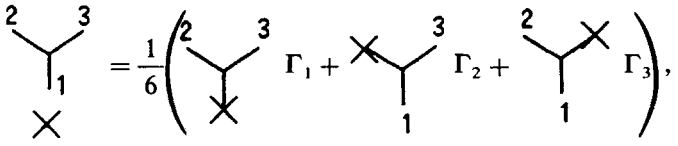
In (3.5) we have used "crossing relation" (3.1) and then "vanishing triangle" (3.4), while "mirror reflection" (1.15) has been used in (3.6).

(5) Any loop diagram composed of five, six, etc., links can be reduced to loop diagram with four links, i.e., a completely symmetric square, which can be concluded in the following way. First, we find




(3.7)

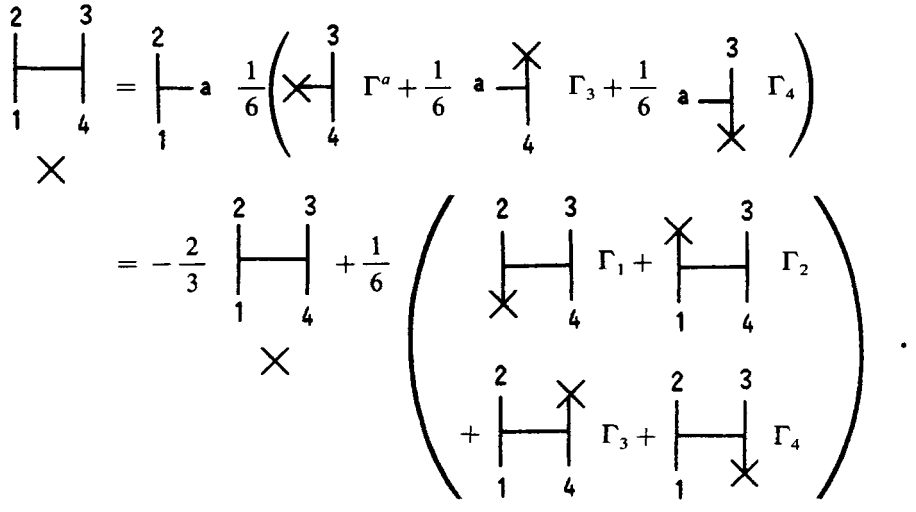
in the last step of which we have used



(3.8)

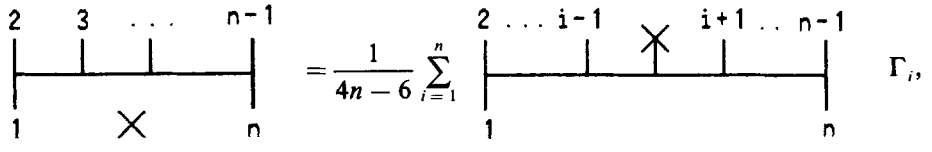
which will be called the "bonding formula," since the right- (left-) hand side of (3.8) is connected (disconnected).

With the help of (3.7) and (3.8), we have a new bonding formula



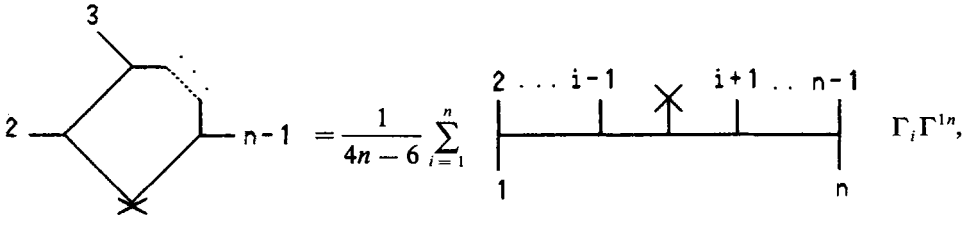
(3.9)

A similar technique leads to the "general bonding formula"



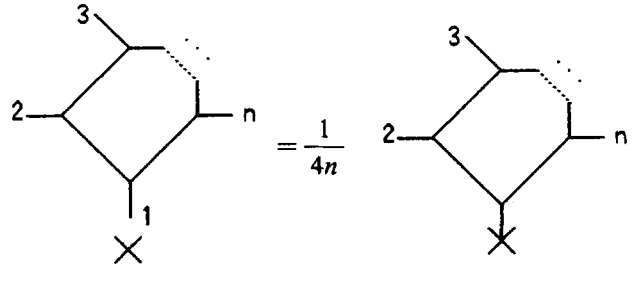
(3.10)

so that we obtain



(3.11)

and



+ (other $(n-1)$ terms obtained by cyclic permutations $1\to 2\to 3\to\cdots\to n\to 1$). (3.12)

With the help of the above five properties, we have proved that our irreducible tensors and tensor–spinors in fact satisfy the "antisymmetric properties" (1.43) and (1.19′)–(1.37′) in addition to the "traceless and Γ-traceless conditions" (1.44) and (1.45).

Our diagrammatical method makes it easy to divide given diagrams into parts, so that Fierz identities for the original diagrams can be obtained by successively applying Fierz identities among parts, the final result being independent of the order of application because of the associative law. [See (A11)–(A14).] Thus we have found that Fierz identities of $(\theta^n)$ polynomials for $2 \leqslant n \leqslant N$ are useful in calculating those of $(\theta^{N+1})$ polynomials.


## ACKNOWLEDGMENT

We are grateful to Mr. K. Osada for useful discussions during the early stages of this work.


## APPENDIX: FIERZ IDENTITIES

We can derive any Fierz identity as follows. First, we notice that any $(\theta^n)$ polynomial can be expanded in terms of SO(1,9)-irreducible bases (1.16)–(1.41), satisfying antisymmetric properties (1.19′)–(1.37′) that lead to miscellaneous "antisymmetric properties" in such a way that

$$0=(\theta^4)_{[12345,6]}$$
$$=\tfrac{1}{3}(\theta^4)_{1234[5,6]}+\tfrac{2}{3}(\theta^4)_{56[123,4]}$$
$$=\tfrac{1}{2}(\theta^4)_{123[45,6]}-\tfrac{1}{2}(\theta^4)_{456[12,3]}$$
$$=\tfrac{1}{6}(\theta^4)_{12345,6}-\tfrac{5}{6}(\theta^4)_{6[1234,5]}.$$

All of these antisymmetric properties show that there exist various linear relations among irreducible tensors and tensor–spinors. Therefore, these antisymmetric properties play essential roles, when we express any $(\theta^n)$ polynomials by sums of *linearly independent* terms, each of which is the product of $\eta_{m_1m_2}$'s, $\Gamma_{m_1m_2\cdots}$, irreducible tensors or tensor–spinors and constants (i.e., coefficients). These coefficients are determined by equating traces or Γ traces of both sides in the above expansions, where we must take account of (1.44) and (1.45). We have carried out these analyses and we give some Fierz identities necessary for the calculations in Sec. II.

### 1. Fierz identities necessary for calculating $\varphi(x,\theta)\Phi(x,\theta)$ and $\Gamma^m\theta\,\partial_m\varphi(x,\theta)$

Fierz identities expressing products of two SO(1,9)-irreducible bases $\theta^i\theta^{n-i}$ in terms of SO(1,9)-irreducible bases $\theta^n$ are given in the following. (Hereafter, $[S_1,S_2,\ldots,S_n]$ and $\{S_1,S_2,\ldots,S_n\}$ denote, respectively, antisymmetrization and symmetrization operators among $n$ sets, each set being composed of indices that appear in $S_i$ and are operated by $S_i$.) For example,

$$\{[1,2],[3,4]\}T_{1234}\equiv\tfrac{1}{2}(T_{[12][34]}+T_{[34][12]})$$

and

$$[[1,2],[3,4]]T_{1234}\equiv\tfrac{1}{2}(T_{[12][34]}-T_{[34][12]}),$$
$$\theta\cdot\bar\theta=\tfrac{1}{96}(\bar\theta\Gamma_{123}\theta)\Gamma^{123}, \tag{A1}$$
$$(\theta^2)_{123}(\bar\theta=\tfrac{1}{2}(\bar\theta^3)_{[12}\Gamma_{3]}\quad(\text{called } YX), \tag{A2}$$
$$(\theta^2)_{123}(\theta^2)_{456}\quad(\text{called } Y_2)=2[4,5,6]\eta_{4[1}(\theta^4)_{23],56}+5(\theta^4)_{123[45,6]}, \tag{A3}$$
$$(\theta^4)_{12,34}(\bar\theta=\tfrac{2}{5}\{[1,2],[3,4]\}(\bar\theta^5)_{12,3}\Gamma_4, \tag{A4}$$
$$(\theta^4)_{12345,6}(\bar\theta=\tfrac{1}{12}(\bar\theta^5)_{12345}\Gamma_6+\tfrac{1}{12}(\bar\theta^5)_{6[1234}\Gamma_{5]}+\tfrac{2}{15}[1,2,3,4,5](\bar\theta^5)_{12,6}\Gamma_{345}, \tag{A5}$$
$$(\theta^2)_{123}(\bar\theta^3)_{45}=[1,2,3][4,5]\begin{pmatrix}\tfrac{2}{5}(\bar\theta^5)_{12,4}\Gamma_{35}+\tfrac{2}{5}(\bar\theta^5)_{45,1}\Gamma_{23}\\ +\tfrac{6}{5}\eta_{14}(\bar\theta^5)_{23,5}+(\bar\theta^5)_{12345}\end{pmatrix}, \tag{A6}$$
$$(\theta^4)_{12,34}(\theta^2)_{567}=\{[1,2],[3,4]\}[5,6,7]\begin{pmatrix}-\tfrac{108}{25}\eta_{15}(\theta^6)_{346,7,2}-\tfrac{9}{25}\eta_{13}(\theta^6)_{567,2,4}\\ -\tfrac{8}{3}(\theta^6)_{56712,34}+\tfrac{8}{3}(\theta^6)_{56731,24}\end{pmatrix}, \tag{A7}$$
$$(\theta^4)_{12345,6}(\theta^2)_{789}=[1,2,3,4,5][7,8,9]\begin{pmatrix}\tfrac{8}{3}\eta_{16}(\theta^6)_{23457,89}-\tfrac{8}{3}\eta_{17}(\theta^6)_{23456,89}\\ -8\eta_{17}(\theta^6)_{23458,96}-\tfrac{16}{15}\eta_{67}(\theta^6)_{12345,89}\\ -\tfrac{36}{25}\eta_{17}\eta_{28}(\theta^6)_{345,9,6}\end{pmatrix}, \tag{A8}$$
$$(\theta^6)_{124,3,5}(\bar\theta=[1,2,4]\{3,5\}\begin{pmatrix}-\tfrac{5}{168}(\bar\theta^7)_{3,5,1}\Gamma_{24}-\tfrac{10}{21}(\bar\theta^7)_{124,3}\Gamma_5\\ -\tfrac{5}{14}(\bar\theta^7)_{123,5}\Gamma_4\end{pmatrix}, \tag{A9}$$
$$(\theta^6)_{12567,34}(\bar\theta=[1,2,5,6,7][3,4]\begin{pmatrix}\tfrac{9}{28}(\bar\theta^7)_{341,2}\Gamma_{567}+\tfrac{3}{14}(\bar\theta^7)_{125,3}\Gamma_{467}\\ -\tfrac{3}{7}\eta_{13}(\bar\theta^7)_{256,4}\Gamma_7\end{pmatrix}, \tag{A10}$$

$$(\bar\theta^5)_{12345}(\theta^2)_{678}=[(\theta^4)_{12345,a}(\bar\theta\Gamma^a)(\theta^2)_{678}$$
$$=(((\theta^4)_{12345,a}(\theta^2)_{678})(\bar\theta)\Gamma^a=[1,2,3,4,5][6,7,8]$$
$$\times\begin{pmatrix}-\frac{18}{35}(\bar\theta^7)_{671,2}\Gamma_{3458}+\frac{4}{35}(\bar\theta^7)_{678,1}\Gamma_{2345}\\ -\frac{16}{35}(\bar\theta^7)_{123,6}\Gamma_{7845}+\frac{40}{7}\eta_{16}\eta_{27}(\bar\theta^7)_{345,8}\\ +\frac{30}{7}\eta_{16}(\bar\theta^7)_{782,3}\Gamma_{45}-\frac{40}{7}\eta_{16}(\bar\theta^7)_{234,7}\Gamma_{85}\end{pmatrix}, \tag{A11}$$

where we have used (1.22) together with (1.20), "associative law," $(\theta^6)$-Fierz identity (A8) and finally $(\theta^7)$-Fierz identities (A9) and (A10),

$$(\theta^2)_{123}(\bar\theta^5)_{45,6}=(\theta^2)_{123}[(\theta^4)_{45,6a}(\bar\theta\Gamma^a]=(((\theta^2)_{123}(\theta^4)_{45,6a})(\bar\theta)\Gamma^a$$
$$=[1,2,3][4,5]\begin{pmatrix}\frac{3}{7}\eta_{64}(\bar\theta^7)_{123,5}-\frac{45}{14}\eta_{61}(\bar\theta^7)_{234,5}\\ -\frac{135}{56}\eta_{14}(\bar\theta^7)_{236,5}+\frac{45}{56}\eta_{14}(\bar\theta^7)_{235,6}\\ +\frac{9}{56}\eta_{14}(\bar\theta^7)_{2,5,6}\Gamma_3+\frac{3}{108}(\bar\theta^7)_{1,4,6}\Gamma_{235}\\ +\frac{1}{7}(\bar\theta^7)_{123,4}\Gamma_{56}-\frac{1}{7}(\bar\theta^7)_{123,6}\Gamma_{45}\\ +\frac{15}{56}(\bar\theta^7)_{126,4}\Gamma_{35}+\frac{75}{56}(\bar\theta^7)_{124,6}\Gamma_{35}\\ +\frac{15}{28}(\bar\theta^7)_{456,1}\Gamma_{23}-\frac{45}{28}(\bar\theta^7)_{451,6}\Gamma_{23}\\ +\frac{15}{14}(\bar\theta^7)_{124,5}\Gamma_{36}\end{pmatrix}, \tag{A12}$$

where we have used (1.21) together with (1.19), the associative law, the $(\theta^6)$-Fierz identity (A7), and finally the $(\theta^7)$-Fierz identities (A9) and (A10).

Similarly we can derive

$$(\theta^4)_{12345,6}(\bar\theta^3)_{78}=(\theta^4)_{12345,6}[(\theta^2)_{78a}(\bar\theta\Gamma^a]=(((\theta^4)_{12345,6}(\theta^2)_{78a})(\bar\theta)\Gamma^a$$
$$=[1,2,3,4,5][7,8]\begin{pmatrix}\frac{8}{7}\eta_{17}\eta_{28}(\bar\theta^7)_{345,6}+\frac{1}{35}\eta_{17}(\bar\theta^7)_{8,6,2}\Gamma_{345}\\ -\frac{8}{7}\eta_{17}(\bar\theta^7)_{234,8}\Gamma_{56}-\frac{3}{7}\eta_{17}(\bar\theta^7)_{236,8}\Gamma_{45}\\ +\frac{57}{35}\eta_{17}(\bar\theta^7)_{238,6}\Gamma_{45}+\frac{48}{35}\eta_{17}(\bar\theta^7)_{234,6}\Gamma_{58}\\ -\frac{48}{35}\eta_{16}(\bar\theta^7)_{782,3}\Gamma_{45}+\frac{32}{35}\eta_{16}(\bar\theta^7)_{234,7}\Gamma_{85}\\ -\frac{36}{35}\eta_{67}(\bar\theta^7)_{123,8}\Gamma_{45}\\ -\frac{1}{7}(\bar\theta^7)_{786,1}\Gamma_{2345}-\frac{1}{7}(\bar\theta^7)_{126,7}\Gamma_{3458}\\ -\frac{4}{35}(\bar\theta^7)_{781,2}\Gamma_{3456}-\frac{4}{35}(\bar\theta^7)_{123,7}\Gamma_{4568}\\ -\frac{1}{35}(\bar\theta^7)_{781,6}\Gamma_{2345}+\frac{3}{35}(\bar\theta^7)_{127,6}\Gamma_{3458}\\ +\frac{4}{35}(\bar\theta^7)_{123,6}\Gamma_{4578}\end{pmatrix} \tag{A13}$$

and

$$(\theta^4)_{12,34}(\bar\theta^3)_{56}=(\theta^4)_{12,34}[(\theta^2)_{56a}(\bar\theta\Gamma^a]=(((\theta^4)_{12,34}(\theta^2)_{56a})(\bar\theta)\Gamma^a$$
$$=\{[1,2],[3,4]\}[5,6]\begin{pmatrix}\frac{3}{140}\eta_{13}(\bar\theta^7)_{2,4,5}\Gamma_6-\frac{6}{35}\eta_{35}(\bar\theta^7)_{4,6,1}\Gamma_2\\ -\frac{9}{7}\eta_{35}(\bar\theta^7)_{412,6}+\frac{27}{7}\eta_{35}(\bar\theta^7)_{612,4}\\ +\frac{9}{7}\eta_{13}(\bar\theta^7)_{256,4}\\ -\frac{6}{7}(\bar\theta^7)_{125,6}\Gamma_{34}+\frac{3}{7}(\bar\theta^7)_{123,5}\Gamma_{46}\\ -\frac{9}{7}(\bar\theta^7)_{125,3}\Gamma_{46}+\frac{6}{7}(\bar\theta^7)_{561,3}\Gamma_{24}\end{pmatrix}, \tag{A14}$$

$$(\theta^6)_{123,4,5}(\theta^2)_{678}=[1,2,3][6,7,8]\{4,5\}$$
$$\times\begin{pmatrix}\frac{5}{28}\eta_{16}\eta_{27}(\theta^8)_{3,8,4,5}-3\eta_{56}(\theta^8)_{1237,8,4}\\ +\frac{5}{21}\eta_{45}(\theta^8)_{678,123}-\frac{20}{7}\eta_{46}(\theta^8)_{785,123}\\ +\frac{10}{21}\eta_{14}(\theta^8)_{678,235}-\frac{5}{3}\eta_{15}(\theta^8)_{6782,3,4}\\ -6\eta_{16}(\theta^8)_{2357,8,4}-\frac{15}{4}\eta_{16}(\theta^8)_{7823,4,5}\\ -\frac{15}{7}\eta_{16}(\theta^8)_{785,234}\end{pmatrix}, \tag{A15}$$

$$(\theta^6)_{12345,67}(\theta^2)_{abc} = [1,2,3,4,5][6,7][a,b,c]$$

$$\times\begin{pmatrix} -\frac{3}{14}\eta_{16}\eta_{27}(\theta^8)_{345,abc} - \frac{27}{20}\eta_{6a}\eta_{1b}(\theta^8)_{2345,7,c} \\ +\frac{18}{7}\eta_{16}\eta_{2a}(\theta^8)_{345,bc7} - \frac{9}{5}\eta_{16}\eta_{2a}(\theta^8)_{345b,c,7} \\ -\frac{45}{14}\eta_{1a}\eta_{2b}(\theta^8)_{345,c67} - \frac{9}{5}\eta_{1a}\eta_{2b}(\theta^8)_{3456,7,c} \end{pmatrix}, \tag{A16}$$

$$(\theta^4)_{12,34}(\theta^4)_{56,78} = \{[1,2],[3,4]\}\{[5,6],[7,8]\}$$

$$\times\begin{pmatrix} -\frac{9}{280}\eta_{13}\eta_{57}(\theta^8)_{2,4,6,8} + \frac{9}{35}\eta_{15}\eta_{37}(\theta^8)_{2,6.4,8} \\ -\frac{36}{5}\eta_{15}(\theta^8)_{2673,4,8} - \frac{18}{5}\eta_{15}(\theta^8)_{3478,2,6} \\ -\frac{72}{7}\eta_{15}(\theta^8)_{273,486} + \frac{72}{7}\eta_{15}(\theta^8)_{263,478} \\ -\frac{54}{7}\eta_{13}(\theta^8)_{257,864} \end{pmatrix}, \tag{A17}$$

$$(\theta^4)_{12,34}(\theta^4)_{56789,a} = \{[1,2],[3,4]\}[5,6,7,8,9]$$

$$\times\begin{pmatrix} \frac{12}{7}\eta_{51}\eta_{63}(\theta^8)_{789,a24} + \frac{10}{7}\eta_{15}\eta_{26}(\theta^8)_{789,34a} \\ +\frac{9}{20}\eta_{a5}\eta_{13}(\theta^8)_{6789,2,4} - \frac{9}{5}\eta_{a1}\eta_{53}(\theta^8)_{6789,2,4} \\ -\frac{9}{10}\eta_{13}\eta_{25}(\theta^8)_{6789,4,a} + \frac{12}{5}\eta_{15}\eta_{26}(\theta^8)_{7893,4,a} \\ +\frac{12}{5}\eta_{a5}\eta_{36}(\theta^8)_{7891,2,4} \\ +\frac{12}{5}\eta_{15}\eta_{36}(\theta^8)_{7892,4,a} \end{pmatrix}, \tag{A18}$$

$$(\theta^4)_{01234,a}(\theta^4)_{56789,b} = \{a[0,1,2,3,4],b[5,6,7,8,9]\}$$

$$\times\begin{pmatrix} -\frac{3}{35}\eta_{05}\eta_{16}\eta_{27}\eta_{38}(\theta^8)_{a,b,4,9} \\ +\frac{16}{7}\eta_{ab}\eta_{05}\eta_{16}(\theta^8)_{234,789} + \frac{16}{7}\eta_{05}\eta_{6a}\eta_{1b}(\theta^8)_{234,789} \\ -\frac{16}{7}\eta_{05}\eta_{6b}\eta_{1a}(\theta^8)_{234,789} - \frac{48}{7}\eta_{05}\eta_{16}\eta_{7b}(\theta^8)_{234,89a} \\ -\frac{12}{7}\eta_{05}\eta_{16}\eta_{27}(\theta^8)_{34a,89b} + \frac{36}{7}\eta_{05}\eta_{16}\eta_{27}(\theta^8)_{34b,89a} \\ +\frac{24}{5}\eta_{05}\eta_{16}\eta_{7b}(\theta^8)_{2348,9,a} \\ +\frac{24}{5}\eta_{05}\eta_{16}\eta_{7a}(\theta^8)_{2348,9,b} \\ +\frac{12}{5}\eta_{05}\eta_{16}\eta_{27}(\theta^8)_{34a8,9,b} \\ -\frac{36}{5}\eta_{05}\eta_{16}\eta_{27}(\theta^8)_{34b8,9,a} \end{pmatrix}, \tag{A19}$$

where we have used the operation defined by

$$\{a[0,1,2,3,4],b[5,6,7,8,9]\}T_{a01234b56789} \equiv \tfrac{1}{2}(T_{a[01234]b[56789]} + T_{b[56789]a[01234]}),$$

$$(\theta^8)_{1,2,3,4}(\bar{\theta} = \tfrac{1}{4}(\bar{\theta}^9)_{\{1,2,3}\Gamma_{4\}}, \tag{A20}$$

$$(\theta^8)_{1234,5,6}(\bar{\theta} = [1,2,3,4]\{5,6\}\begin{pmatrix} -\frac{2}{63}(\bar{\theta}^9)_{123,5}\Gamma_{46} + \frac{1}{42}(\bar{\theta}^9)_{125,6}\Gamma_{34} \\ +\frac{1}{42}\eta_{15}(\bar{\theta}^9)_{234,6} - \frac{1}{168}(\bar{\theta}^9)_{5,6,1}\Gamma_{234} \end{pmatrix}, \tag{A21}$$

$$(\theta^8)_{123,456}(\bar{\theta} = \{[1,2,3],[4,5,6]\}(\tfrac{1}{60}(\bar{\theta}^9)_{123,4}\Gamma_{56} + \tfrac{1}{40}(\bar{\theta}^9)_{124,5}\Gamma_{63}), \tag{A22}$$

$$(\theta^8)_{1,2,3,4}(\theta^2)_{567} = \{1,2,3,4\}[5,6,7](-\tfrac{3}{4}\eta_{12}(\theta^{10})_{567,3,4} + \tfrac{63}{10}\eta_{15}(\theta^{10})_{672,3,4}), \tag{A23}$$

$$(\theta^8)_{1234,5,6}(\theta^2)_{789} = [1,2,3,4]\{5,6\}[7,8,9]$$

$$\times\begin{pmatrix} -\frac{9}{20}\eta_{17}\eta_{28}(\theta^{10})_{345,9,6} + \frac{9}{20}\eta_{17}\eta_{28}(\theta^{10})_{349,5,6} \\ +\frac{9}{20}\eta_{15}\eta_{27}(\theta^{10})_{893,4,6} + \frac{3}{10}\eta_{57}\eta_{18}(\theta^{10})_{234,9,6} \\ -\frac{32}{7}\eta_{57}(\theta^{10})_{89612,34} + \frac{32}{7}\eta_{57}(\theta^{10})_{89123,46} \\ -\frac{40}{21}\eta_{15}(\theta^{10})_{78923,46} - \frac{8}{21}\eta_{15}(\theta^{10})_{78962,34} \\ +\frac{32}{7}\eta_{17}(\theta^{10})_{89523,46} + \frac{16}{21}\eta_{56}(\theta^{10})_{78912,34} \end{pmatrix}, \tag{A24}$$

$$(\theta^8)_{123,456}(\theta^2)_{789}=\{[1,2,3],[4,5,6]\}[7,8,9]$$

$$\times\begin{pmatrix} -\frac{3}{200}\eta_{14}\eta_{25}(\theta^{10})_{789,6,3}+\frac{63}{200}\eta_{14}\eta_{27}(\theta^{10})_{895,6,3} \\ -\frac{21}{100}\eta_{17}\eta_{28}(\theta^{10})_{456,9,3}+\frac{63}{200}\eta_{17}\eta_{48}(\theta^{10})_{235,6,9} \\ -\frac{24}{5}\eta_{17}(\theta^{10})_{89245,63}-\frac{12}{5}\eta_{17}(\theta^{10})_{89456,23} \\ -\frac{12}{5}\eta_{17}(\theta^{10})_{89234,56} \\ +\frac{8}{5}\eta_{14}(\theta^{10})_{23789,56}-\frac{8}{5}\eta_{14}(\theta^{10})_{78952,36}\end{pmatrix}, \tag{A25}$$

$$(\theta^{10})_{123,4,5}(\bar\theta=[1,2,3]\{4,5\}\begin{pmatrix}\frac{2}{105}(\bar\theta^{11})_{14,5}\Gamma_{23}-\frac{1}{35}(\bar\theta^{11})_{12,4}\Gamma_{35} \\ -\frac{1}{63}\eta_{14}(\bar\theta^{11})_{23,5}\end{pmatrix}, \tag{A26}$$

$$(\theta^{10})_{12345,67}(\bar\theta=[1,2,3,4,5][6,7]\begin{pmatrix}\frac{1}{320}\eta_{16}(\bar\theta^{11})_{23,7}\Gamma_{45}-\frac{1}{1056}(\bar\theta^{11})_{12345}\Gamma_{67} \\ -\frac{1}{528}(\bar\theta^{11})_{61234}\Gamma_{57}-\frac{1}{1056}(\bar\theta^{11})_{67123}\Gamma_{45} \\ -\frac{1}{960}(\bar\theta^{11})_{12,6}\Gamma_{3457}+\frac{1}{960}(\bar\theta^{11})_{67,1}\Gamma_{2345}\end{pmatrix}, \tag{A27}$$

$$(\theta^{10})_{123,4,5}(\theta^2)_{678}=[1,2,3]\{4,5\}[6,7,8]$$

$$\times\begin{pmatrix}\frac{2}{105}\eta_{45}\eta_{16}(\theta^{12})_{78,23}-\frac{4}{135}\eta_{64}\eta_{15}(\theta^{12})_{23,78} \\ +\frac{64}{945}\eta_{16}\eta_{24}(\theta^{12})_{53,78}+\frac{8}{35}\eta_{16}\eta_{47}(\theta^{12})_{85,23} \\ +\frac{16}{105}\eta_{16}\eta_{27}(\theta^{12})_{84,53} \\ -\frac{4}{77}\eta_{46}(\theta^{12})_{12378,5}-\frac{2}{77}\eta_{46}(\theta^{12})_{12357,8} \\ -\frac{3}{77}\eta_{16}(\theta^{12})_{78234,5}+\frac{1}{154}\eta_{45}(\theta^{12})_{67812,3} \\ +\frac{13}{693}\eta_{14}(\theta^{12})_{67823,5}-\frac{4}{693}\eta_{14}(\theta^{12})_{67852,3}\end{pmatrix} \tag{A28}$$

$$(\theta^{10})_{12345,67}(\theta^2)_{abc}=[1,2,3,4,5][6,7][a,b,c]$$

$$\times\begin{pmatrix}\frac{1}{48}\eta_{1a}\eta_{2b}\eta_{3c}(\theta^{12})_{45,67}-\frac{1}{40}\eta_{16}\eta_{2a}\eta_{3b}(\theta^{12})_{c7,45} \\ +\frac{1}{240}\eta_{16}\eta_{27}\eta_{3a}(\theta^{12})_{bc,45} \\ +\frac{1}{44}\eta_{6a}\eta_{7b}(\theta^{12})_{12345,c}+\frac{45}{352}\eta_{1a}\eta_{6b}(\theta^{12})_{c2345,7} \\ -\frac{25}{352}\eta_{1a}\eta_{6b}(\theta^{12})_{72345,c}-\frac{15}{88}\eta_{1a}\eta_{2b}(\theta^{12})_{c3456,7} \\ +\frac{5}{88}\eta_{1a}\eta_{2b}(\theta^{12})_{34567,c}-\frac{15}{88}\eta_{16}\eta_{2a}(\theta^{12})_{bc345,7} \\ +\frac{15}{88}\eta_{16}\eta_{2a}(\theta^{12})_{7345b,c}+\frac{15}{176}\eta_{16}\eta_{7a}(\theta^{12})_{2345b,c} \\ +\frac{5}{88}\eta_{16}\eta_{27}(\theta^{12})_{abc34,5}\end{pmatrix} \tag{A29}$$

$$(\theta^{12})_{12,34}(\bar\theta=\{[1,2],[3,4]\}(-\tfrac{1}{132}(\bar\theta^{13})_{12}\Gamma_{34}+\tfrac{1}{132}(\bar\theta^{13})_{31}\Gamma_{24}), \tag{A30}$$

$$(\theta^{12})_{12345,6}(\bar\theta=[1,2,3,4,5]\begin{pmatrix}-\frac{1}{540}(\bar\theta^{13})_{12}\Gamma_{3456}-\frac{1}{540}(\bar\theta^{13})_{61}\Gamma_{2345} \\ -\frac{1}{180}\eta_{16}(\bar\theta^{13})_{23}\Gamma_{45}\end{pmatrix}, \tag{A31}$$

$$(\theta^{12})_{12,34}(\theta^2)_{567}=\{[1,2],[3,4]\}[5,6,7]\begin{pmatrix}\frac{1}{308}\eta_{13}\eta_{24}(\theta^{14})_{567}+\frac{6}{77}\eta_{15}\eta_{26}(\theta^{14})_{734} \\ +\frac{6}{77}\eta_{15}\eta_{36}(\theta^{14})_{247}+\frac{3}{308}\eta_{15}\eta_{23}(\theta^{14})_{467} \\ +\frac{9}{308}\eta_{15}\eta_{23}(\theta^{14})_{467}-\frac{3}{154}\eta_{13}\eta_{25}(\theta^{14})_{467}\end{pmatrix}, \tag{A32}$$

$$(\theta^{12})_{12345,6}(\theta^2)_{789}=[1,2,3,4,5][7,8,9]\begin{pmatrix}\frac{1}{35}\eta_{17}\eta_{28}\eta_{39}(\theta^{14})_{456}-\frac{1}{35}\eta_{16}\eta_{27}\eta_{38}(\theta^{14})_{459} \\ +\frac{1}{35}\eta_{67}\eta_{18}\eta_{29}(\theta^{14})_{345}\end{pmatrix}. \tag{A33}$$

$$(\theta^{14})_{123}(\bar\theta=\tfrac{1}{720}(\bar\theta^{15})\Gamma_{123}, \tag{A34}$$

$$(\theta^{14})_{123}(\theta^2)_{456}=-\tfrac{1}{120}[1,2,3][4,5,6]\eta_{14}\eta_{25}\eta_{36}(\theta^{16}). \tag{A35}$$

## 2. Fierz identities necessary in calculating $(\partial/\partial\bar\theta)\varphi(x,\theta)$

In addition to (A1)–(A3), we give

$$(\theta^2)_{123}(\theta^2)_{456}(\bar\theta\quad(\text{called}\quad Y_2X)=\{[1,2,3],[4,5,6]\}\begin{pmatrix}\frac{1}{2}(\bar\theta^5)_{12345}\Gamma_6+\frac{4}{5}\eta_{14}(\bar\theta^5)_{23,5}\Gamma_6 \\ +\frac{2}{5}(\bar\theta^5)_{45,1}\Gamma_{236}\end{pmatrix}, \tag{A36}$$

$$(\theta^2)_{123}(\theta^2)_{456}(\theta^2)_{789}\ (\text{called } Y_3) = \{[1,2,3],[4,5,6],[7,8,9]\}\begin{pmatrix} 48\eta_{14}(\theta^6)_{78952,36} - \frac{324}{25}\eta_{17}\eta_{24}(\theta^6)_{895,6,3} \\ -\frac{54}{25}\eta_{14}\eta_{25}(\theta^6)_{789,3,6} \end{pmatrix}. \tag{A37}$$

$$(\theta^2)_{123}(\theta^2)_{456}(\theta^2)_{789}(\bar{\theta}\quad (\text{called } Y_3X) = \{[1,2,3],[4,5,6],[7,8,9]\}$$

$$\times\begin{pmatrix} \frac{27}{7}\eta_{14}\eta_{25}(\bar{\theta}^7)_{378,6}\Gamma_9 + \frac{9}{140}\eta_{14}\eta_{25}(\bar{\theta}^7)_{3,6,7}\Gamma_{89} \\ +\frac{27}{7}\eta_{14}\eta_{27}(\bar{\theta}^7)_{389,5}\Gamma_6 - \frac{81}{35}\eta_{14}\eta_{27}(\bar{\theta}^7)_{589,3}\Gamma_6 \\ -\frac{27}{7}\eta_{14}\eta_{27}(\bar{\theta}^7)_{563,8}\Gamma_9 + \frac{27}{5}\eta_{14}\eta_{27}(\bar{\theta}^7)_{568,3}\Gamma_9 \\ -\frac{9}{35}\eta_{14}\eta_{27}(\bar{\theta}^7)_{3,5,8}\Gamma_{69} \\ -\frac{27}{7}\eta_{14}(\bar{\theta}^7)_{237,5}\Gamma_{689} - \frac{27}{35}\eta_{14}(\bar{\theta}^7)_{235,7}\Gamma_{689} \\ +\frac{36}{35}\eta_{14}(\bar{\theta}^7)_{789,2}\Gamma_{356} - \frac{108}{35}\eta_{14}(\bar{\theta}^7)_{237,8}\Gamma_{569} \\ +\frac{54}{35}\eta_{14}(\bar{\theta}^7)_{278,5}\Gamma_{369} \end{pmatrix}, \tag{A38}$$

and

$$(\theta^2)_{123}(\theta^2)_{456}(\theta^2)_{789}(\theta^2)_{abc}\quad (\text{called } Y_4) = 4!\{[1,2,3],[4,5,6],[7,8,9],[a,b,c]\}$$

$$\times\begin{pmatrix} -\frac{9}{140}\eta_{14}\eta_{27}\eta_{8a}\eta_{5b}(\theta^8)_{3,6,9,c} - \frac{9}{560}\eta_{14}\eta_{25}\eta_{7a}\eta_{8b}(\theta^8)_{3,6,9,c} \\ +\frac{27}{14}\eta_{14}\eta_{25}\eta_{a7}(\theta^8)_{bc3,896} + \frac{27}{40}\eta_{14}\eta_{25}\eta_{a7}(\theta^8)_{bc89,3,6} \\ +\frac{9}{5}\eta_{14}\eta_{25}\eta_{37}(\theta^8)_{abc8,9,6} + \frac{6}{7}\eta_{14}\eta_{27}\eta_{85}(\theta^8)_{936,abc} \\ -\frac{9}{5}\eta_{14}\eta_{27}\eta_{3a}(\theta^8)_{895b,c,6} - \frac{27}{7}\eta_{14}\eta_{27}\eta_{5a}(\theta^8)_{389,6bc} \\ +\frac{27}{7}\eta_{14}\eta_{27}\eta_{5a}(\theta^8)_{3bc,689} - \frac{27}{10}\eta_{14}\eta_{27}\eta_{5a}(\theta^8)_{bc89,3,6} \end{pmatrix}. \tag{A39}$$

Suppose we have already obtained the Fierz identity for $Y_n$. Then, the Fierz identity for $Y_nX$ [$Y_{n+1}$] (for $n = 4$–7) can be derived, provided we have Fierz identities that express products of irreducible ($\theta^{2n}$) tensors with ($\bar{\theta}$ [$(\theta^2)_{123}$] in terms of irreducible ($\theta^{2n+1}$) tensor–spinors [irreducible ($\theta^{2n+2}$) tensors], and these necessary Fierz identities have already been given by (A21)–(A35). Fierz identities for $Y_n$ and $Y_nX$ have a maximum number of vector indices among all ($\theta^{2n}$)- and ($\theta^{2n+1}$)- Fierz identities. Taking traces of $Y_n$ or $\Gamma$ traces of $Y_nX$ gives any Fierz identity with fewer numbers of vector indices, so that missing formulas in subsection 1 for $\theta^i\theta^{n-i}\to\theta^n$ ($n = 9$–16) can be derived by using the following procedures. We first transform $\theta^i\theta^{n-i}$ into $\theta^n$ polynomials with the help of (1.16)–(1.41), and the thus-obtained $\theta^n$ polynomials are expressed by $\theta^n$ bases by using above-mentioned Fierz identities.


[1]Yu. A. Gel'fand and E. P. Likhtman, JETP Lett. **13**, 323 (1971); D. V. Volkov and V. P. Akulov, Phys. Lett. B **46**, 109 (1973); Teor. Mat. Fiz. (USSR) **18**, 39 (1974); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).

[2]For reviews on supersymmetry, see P. Fayet and S. Ferrara, Phys. Rep. C **32**, 249 (1977); A. Salam and J. Strathdee, Fortschr. Phys. **26**, 57 (1978); J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton U.P., Princeton, NJ, 1983); S. J. Gates Jr., M. T. Grisaru, M. Rocek, and W. Siegel, *Superspace or One Thousand and One Lessons in Supersymmetry* (Benjamin/Cummings, Reading, PA, 1983).

[3]For a four-dimensional supergravity, see D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976); S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976). For an 11-dimensional supergravity, see E. Cremmer, B. Julia, and J. Scherk, Phys. Lett. B **76**, 409 (1978), together with S. Naito, K. Osada, and T. Fukui, Phys. Rev. D **34**, 536 (1986); **35**, 1536(E) (1987).

[4]For reviews on supergravity, see P. van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981); E. S. Fradkin and A. A. Tseytlin, *ibid.* **119**, 235 (1985); M. F. Sohnius, *ibid.* **128**, 39 (1985); P. West, *Introduction to Supersymmetry and Supergravity* (World Scientific, Singapore, 1986).

[5]M. Duff, "Anti-symmetric tensors and supergravity" in *Superspace and Supergravity*, edited by S. Hawking and M. Rocek (Cambridge U.P., Cambridge, 1981); P. van Nieuwenhuizen, "Quantum (super)gravity" in *Quantum Gravity 2*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1981).

[6]M. B. Green and J. H. Schwarz, Phys. Lett. B **136**, 367 (1984).

[7]For reviews on superstrings, see J. H. Schwarz, *Superstrings* (World Scientific, Singapore, 1985); M. B. Green and D. J. Gross, *Unified String Theories* (World Scientific, Singapore, 1986); M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U.P., Cambridge, 1987).

[8]B. E. W. Nilsson, Nucl. Phys. B **188**, 176 (1981); E. Bergshoeff, M. de Roo, B. de Wit, and P. van Nieuwenhuizen, *ibid.* **195**, 97 (1982).

[9]R. D'Auria, P. Fre, and A. J. DaSilva, Nucl. Phys. B **196**, 205 (1982).

[10]H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, NJ, 1946); Z-X. Wan, *Lie Algebras* (Pergamon, New York, 1975).

[11]E. Bergshoeff and M. de Roo, Phys. Lett. B **112**, 53 (1982). A misprint for the SO(1,9) representation for ($\theta^8$) bases (1.28) exists in Ref. 11.

# Coherent states for the noncompact supergroups Osp(2/2N,R)

A. B. Balantekin
*Department of Physics, University of Wisconsin, Madison, Wisconsin 53706*
H. A. Schmitt and P. Halse
*Department of Physics, University of Arizona, Tucson, Arizona 85721*

(Received 12 April 1988; accepted for publication 5 October 1988)

An analysis of the coherent states for the noncompact supergroup Osp(2/2N,R) is presented. In contrast to Osp(1/2N,R), both typical and atypical representations have to be considered. The measure of integration, in general, for Osp(2/2N,R) coherent states is calculated; it is then used to construct the decomposition of unity for the special case of Osp(2/2,R). It is found, however, that the typical and atypical representations of Osp(2/2,R) have to be treated separately. It is verified that the coherent states for Osp(2/2,R) are "closest to classical" in the sense of Perelomov.

## I. INTRODUCTION

Coherent states have long been used to provide a natural link between classical and quantal physics,[1] often providing better insight to a problem. Path integrals offer an alternative to the Schrödinger and Heisenberg formalisms.[2] The use of group theoretical techniques to study the symmetry properties of physical systems have likewise aided our understanding. Recently, there has been some interest in unifying these ideas by studying path integrals over coherent states for both compact[3] and noncompact[4] groups.

In the past few years there has been great interest in the study of supersymmetries, and the use of supergroups has now entered many fields of physics. Within the interacting boson–fermion model[5] (IBFM), the compact supergroups U(6/n) (Ref. 6) have been used to study nuclear spectra and compact superalgebras have recently been employed in statistical mechanics to study many fermion systems.[7] The noncompact supergroups Osp(p/2N,R) have been employed in general superfield theories.[8] Among the simplest of these are Osp(2/2N,R) and their analysis would form a useful first step in understanding the Osp(p/2N,R) supergroups. An essential ingredient for any path integral construction is the measure of integration, and until now no systematic construction for the noncompact orthosymplectic supergroups has been given.

In an earlier paper,[9] the coherent states for the noncompact supergroup Osp(1/2N,R) were studied, and a general method for calculating the corresponding measure of integration was presented giving in detail various results for Osp(1/2,R) coherent states. We present an analysis of the coherent states for Osp(2/2N,R) supergroups, which turn out to have a more interesting structure than those for Osp(1/2N,R). One of the more interesting features of Osp(2/2N,R) is the vanishing of its quadratic Casimir operator for some nontrivial representations: the atypical representations that must then be treated separately.

In Sec. II, the superalgebra Osp(2/2,R) is defined, and its coherent states are introduced. These coherent states are shown to take on a much simpler form for the atypical representations, and can in fact be handled using the techniques of Ref. 9. The measure of integration is calculated for the general Osp(2/2N,R) supergroup in Sec. III, by studying the transformation properties of the supercoset variables. Section IV contains a summary of our results and a discussion of possible extensions and applications. We also include two appendices containing some of the more technical results for the representation theory and coherent states for Osp(2/2,R). In Appendix A, the representation theory for Osp(2/2,R) is worked out in detail; Appendix B contains the matrix elements of the various generators.

## II. COHERENT STATES FOR THE SUPERGROUP Osp(2/2,R)

The noncompact supergroup Osp(2/2,R) is generated by eight elements.[10] The four even generators close under commutation [see Eq. (2.2a) below] and generate the subgroup

$$\mathrm{Sp}(2,R)\times\mathrm{SO}(2)\,. \tag{2.1}$$

The superalgebra Osp(2/2,R) is defined by the following commutation and anticommutation relations[11]:

$$[K_0,K_\pm]=\pm K_\pm\,,\quad [K_+,K_-]=-2K_0\,,$$
$$[B,K_\pm]=[B,K_0]=0\,, \tag{2.2a}$$
$$[K_0,V_\pm]=\pm\tfrac{1}{2}V_\pm\,,\quad [K_0,W_\pm]=\pm\tfrac{1}{2}W_\pm\,, \tag{2.2b}$$
$$[K_\pm,V_\pm]=[K_\pm,W_\pm]=0,\quad [K_\pm,V_\mp]=\mp V_\pm\,,$$
$$[K_\pm,W_\mp]=\mp W_\pm\,, \tag{2.2c}$$
$$[B,V_\pm]=\tfrac{1}{2}V_\pm\,,\quad [B,W_\pm]=-\tfrac{1}{2}W_\pm\,, \tag{2.2d}$$
$$\{V_\pm,V_\pm\}=\{V_\pm,V_\mp\}=\{W_\pm,W_\pm\}$$
$$=\{W_\pm,W_\mp\}=0\,, \tag{2.2e}$$
$$\{V_\pm,W_\pm\}=K_\pm\,,\quad \{V_\pm,W_\mp\}=K_0\mp B\,. \tag{2.2f}$$

We first define the lowest weight state for an Osp(2/2,R) representation as that state that is annihilated by all the lowering operators

$$|\tau\,q;\tau\,\tau\,q\rangle\,. \tag{2.3}$$

In Eq. (2.3), the first two quantum numbers label the Osp(2/2,R) representation, while the next three labels are the Sp(2,R) quantum number, the corresponding third component, and the SO(2) quantum number, respectively. In complete analogy with coherent states for SU(1,1) (Ref.

 0022-2488/89/020274-06$02.50 © 1989 American Institute of Physics 

12) and Osp(1/2,$R$) (Ref. 9), we may then define the coherent states for the supergroup Osp(2/2,$R$) as

$$|\theta,\chi,\alpha\rangle = \mathcal{N}\exp(\alpha K_+ + \theta V_+ + \chi W_+)|\tau\, q;\tau\,\tau\, q\rangle\,, \tag{2.4}$$

where $\alpha$ is a complex variable and $\theta$ and $\chi$ are Grassmann variables.[13] The factor $\mathcal{N}$ is introduced to ensure that the coherent state is normalized to unity. This normalization factor can be conveniently expressed in terms of the superdeterminant of

$$M(\omega,\bar{\omega}';\xi,\bar{\xi}) = \begin{bmatrix} 1 & \xi \\ \bar{\xi}' & 1-\omega\bar{\omega}' \end{bmatrix}. \tag{2.5}$$

Explicitly, we find

$$\mathcal{N}^{-2} = [\mathrm{Sdet}\, M(\beta,\bar{\beta};\chi,\bar{\chi})]^{\tau+q}[\mathrm{Sdet}\, N(\sigma,\bar{\sigma};\theta,\bar{\theta})]^{\tau-q}, \tag{2.6a}$$

where

$$\beta = \alpha - \tfrac{1}{2}\theta\chi \quad \text{and} \quad \sigma = \alpha - \tfrac{1}{2}\chi\theta\,. \tag{2.6b}$$

The superdeterminant of a graded matrix

$$G = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{2.7a}$$

is given by

$$\mathrm{Sdet}\, G = \det(A - BD^{-1}C)/\det D\,. \tag{2.7b}$$

We show in Appendix A that for $\tau = q\ (=-q)$, the generator $V_+$ ($W_+$) annihilates the state $|\tau\, q;\,\tau\,\tau\, q\rangle$. Thus for the $\tau = -q$ representation, we may simplify Eq. (2.4) to

$$|\theta\sigma\rangle = \mathcal{N}\exp(\sigma K_+ + \theta V_+)|\tau\, q;\,\tau,\tau\, q = -\tau\rangle\,, \tag{2.4'}$$

and the normalization Eq. (2.5) reduces to

$$\mathcal{N}^{-2} = [\mathrm{Sdet}\, M(\sigma,\bar{\sigma}';\theta,\bar{\theta}')]^{2\tau}\,. \tag{2.6a'}$$

Likewise, for the atypical representation $\tau = q$, we find

$$|\theta,\beta\rangle = \mathcal{N}\exp(\beta K_+ + \chi W_+)|\tau\, q;\,\tau\,\tau\, q = \tau\rangle\,, \tag{2.4''}$$

and

$$\mathcal{N}^{-2} = [\mathrm{Sdet}\, M(\beta,\bar{\beta}';\chi,\bar{\chi}')]^{2\tau}\,. \tag{2.6a''}$$

The variables $\sigma$ and $\beta$ are defined as in Eq. (2.6b). In order to calculate the decomposition of unity, we will need to perform integrations over the Grassmann variables. The integration over Grassmann variables is only defined up to a normalization,[13] which we take for convenience as

$$\int d\bar{\theta}\, d\theta(1,\theta,\bar{\theta}) = 0\,, \tag{2.8a}$$

$$\int d\bar{\theta}\, d\theta\, \theta\bar{\theta} = 1\,. \tag{2.8b}$$

It is now possible to calculate the decomposition of unity for the Osp(2/2,$R$) coherent states. By using the integration measure derived in Sec. III, we can explicitly show [by suppressing the Osp(2/2,$R$) labels]

$$\begin{aligned}\int d\mu|\theta,\chi,\alpha\rangle\langle\theta,\chi,\alpha| &= \sum_{m=0}^{\infty}|\tau\,\tau+m\,q\rangle\langle\tau\,\tau+m\,q| \\ &+ \sum_{m=0}^{\infty}|\tau+1\,\tau+1+m\,q\rangle\langle\tau+1\,\tau+1+m\,q| \\ &+ \sum_{m=0}^{\infty}\left|\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,q-\frac{1}{2}\right\rangle \\ &\times\left\langle\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,q-\frac{1}{2}\right| \\ &+ \sum_{m=0}^{\infty}\left|\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,q+\frac{1}{2}\right\rangle \\ &\times\left\langle\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,q+\frac{1}{2}\right| = 1\end{aligned} \tag{2.9a}$$

with

$$d\mu = \frac{(\tau+q)(\tau-q)}{2\tau\pi}\, d\bar{\theta}\, d\theta\, d\bar{\chi}\, d\chi\, d\alpha^*\, d\alpha\,. \tag{2.9b}$$

Equation (2.9b) clearly indicates that the cases $\tau = \pm q$ must be handled separately. However, we note that from Eqs. (2.4') and (2.4''), we can use the techniques developed for Osp(1/2,$R$) coherent states. We merely quote the results here:

$$\begin{aligned}\int d\mu_{\tau=-q}|\theta,\sigma\rangle\langle\theta,\sigma| &= \sum_{m=0}^{\infty}|\tau\,\tau+m\,-\tau\rangle\langle\tau\,\tau+m\,-\tau| \\ &+ \sum_{m=0}^{\infty}\left|\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,-\tau+\frac{1}{2}\right\rangle \\ &\times\left\langle\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,-\tau+\frac{1}{2}\right| = 1\,,\end{aligned} \tag{2.10a}$$

with

$$d\mu_{\tau=-q} = (2/\pi)d\bar{\theta}\, d\theta\, d\bar{\sigma}\, d\sigma\ \mathrm{Sdet}\, M(\sigma,\bar{\sigma};\theta,\bar{\theta})\,. \tag{2.10b}$$

Similarly, if $\tau = q$, then

$$\begin{aligned}\int d\mu_{\tau=q}|\chi,\beta\rangle\langle\chi,\beta| &= \sum_{m=0}^{\infty}|\tau\,\tau+m\,\tau\rangle\langle\tau\,\tau+m\,\tau| \\ &+ \sum_{m=0}^{\infty}\left|\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,\tau-\frac{1}{2}\right\rangle \\ &\times\left|\tau+\frac{1}{2}\,\tau+\frac{1}{2}+m\,\tau-\frac{1}{2}\right| = 1\end{aligned} \tag{2.11a}$$

and

$$d\mu_{\tau=q} = \frac{2}{\pi}\, d\bar{\chi}\, d\chi\, d\bar{\beta}\, d\beta\ \mathrm{Sdet}\, M(\beta,\beta^*;\chi,\bar{\chi})\,. \tag{2.11b}$$

In Ref. 14, it was shown that coherent states constructed from lowest weight states are "closest to classical" in the sense that they minimize the dispersion

$$\Delta = \langle C_2\rangle - g^{ij}\langle X_i\rangle\langle X_j\rangle\,. \tag{2.12}$$

In Eq. (2.12), $g^{ij}$ is the Cartan–Killing metric, the $X_i$ are generators of the supergroup Osp(2/2,$R$), and $C_2 = g^{ij}X_iX_j$ is the quadratic Casimir operator of Osp(2/2,$R$), where a sum over repeated indices is to be understood. By using the expectation values of the generators from Appendix B, it is possible to show that the coherent states as defined in Eq. (2.4) minimize

$$\Delta = \langle C_2\rangle - \langle K_0\rangle\langle K_0\rangle + \tfrac{1}{2}\langle K_+\rangle\langle K_-\rangle + \tfrac{1}{2}\langle K_-\rangle\langle K_+\rangle + \langle B\rangle\langle B\rangle - \tfrac{1}{2}\langle V_-\rangle\langle W_+\rangle - \tfrac{1}{2}\langle W_-\rangle\langle V_+\rangle + \tfrac{1}{2}\langle V_+\rangle\langle W_-\rangle + \tfrac{1}{2}\langle W_+\rangle\langle V_-\rangle , \quad (2.13a)$$

where the quadratic Casimir $C_2$ is given by

$$C_2 = K_0^2 - \tfrac{1}{2}K_+K_- - \tfrac{1}{2}K_-K_+ - B^2 + \tfrac{1}{2}V_-W_+ + \tfrac{1}{2}W_-V_+ - \tfrac{1}{2}V_+W_- - \tfrac{1}{2}W_+V_- . \quad (2.13b)$$

It is interesting that the dispersion $\Delta$ vanishes for the typical as well as the atypical representations.

## III. INTEGRATION MEASURE FOR Osp(2/2N,R)

In this section we consider the construction of coherent states for the supergroups Osp(2/2N,R) and the calculation of their measure of integration. Since we have treated Osp(1/2N,R) coherent states in detail in Ref. 9, we only sketch the calculations here, and the reader is referred there for more details. We introduce the boson creation and annihilation operators $b_j^\dagger$, $b_i$ and fermion creation and annihilation operators $a_j^\dagger$, $a_i$. These operators satisfy the standard commutation or anticommutation relations

$$[b_i,b_j] = 0, \quad [b_i,b_j^\dagger] = \delta_{ij}, \quad [b_i,a_j] = 0, \quad [b_i,a_j^\dagger] = 0, \quad \{a_i,a_j^\dagger\} = \delta_{i,j}, \quad \{a_i,a_j\} = 0 . \quad (3.1)$$

For Osp(2/2N,R), we need $N$ creation/annihilation operators in the bosonic space, and only one creation/annihilation operator in the fermion space. The operators are then transformed among themselves with the most general mapping that preserves Hermiticity,

$$b_i \to \tilde{b}_i = u_{ij}b_j + v_{ij}b_j^\dagger + \theta_{1i}a + \theta_{2i}a^\dagger , \quad (3.2a)$$

$$b_i^\dagger \to \tilde{b}_i^\dagger = u_{ij}^* b_j^\dagger + v_{ij}^* b_j - \bar{\theta}_{1i}a^\dagger - \bar{\theta}_{2i}a , \quad (3.2b)$$

$$a \to \tilde{a} = \lambda_1 a + \lambda_2 a^\dagger + \chi_{1i}b_i + \chi_{2i}b_i^\dagger , \quad (3.2c)$$

$$a^\dagger \to \tilde{a}^\dagger = \lambda_1^* a^\dagger + \lambda_2^* a + \bar{\chi}_{1i}b_i^\dagger + \bar{\chi}_{2i}b_i . \quad (3.2d)$$

Since we want the transformation to be canonical, we require the transformed (tilded) operators to satisfy the same relations as the untransformed operators. This leads to the following set of independent equations:

$$u_{il}v_{jl} - v_{il}u_{jl} - \theta_{2i}\theta_{1j} - \theta_{1i}\theta_{2j} = 0 , \quad (3.3a)$$

$$u_{il}u_{jl}^* - v_{il}v_{jl}^* + \theta_{1i}\bar{\theta}_{1j} + \theta_{2i}\bar{\theta}_{2j} = \delta_{ij} , \quad (3.3b)$$

$$u_{il}\chi_{2l} - v_{il}\chi_{1l} + \theta_{1i}\lambda_2 + \theta_{2i}\lambda_1 = 0 , \quad (3.3c)$$

$$u_{il}\chi_{1l} - v_{il}\chi_{2l} + \theta_{1i}\lambda_1^* + \theta_{2i}\lambda_2^* = 0 , \quad (3.3d)$$

$$\lambda_1\lambda_1^* + \lambda_2\lambda_2^* + \chi_{1i}\bar{\chi}_{1i} - \chi_{2i}\bar{\chi}_{2i} = 1 , \quad (3.3e)$$

$$\lambda_1\lambda_2 + \chi_{li}\chi_{2i} = 0 . \quad (3.3f)$$

Here the $u_{ij}$, $v_{ij}$, $\lambda_1$, and $\lambda_2$ are complex numbers, while $\theta_{1i}$, $\bar{\theta}_{2i}$, $\chi_{1i}$, and $\bar{\chi}_{2i}$ are Grassmann (anticommuting) numbers. It will be convenient to express the above conditions [Eqs. (3.3a)–(3.3f)] in matrix form. Therefore, $u_{ij}$ will be interpreted as the $ij$th entry of the $N\times N$ matrix $u$, and so on. Next we introduce the supermatrix

$$R = \begin{bmatrix} u & v & \theta_1 & \theta_2 \\ v^* & u^* & -\bar{\theta}_2 & -\bar{\theta}_1 \\ \chi_1 & \chi_2 & \lambda_1 & \lambda_2 \\ \bar{\chi}_2 & \bar{\chi}_1 & \lambda_2^* & \lambda_1^* \end{bmatrix} . \quad (3.4)$$

It is not hard to show then that the transformation in Eq. (3.2) is canonical if

$$RHR^{\rm ST} = H , \quad (3.5)$$

where ST indicates the supertranspose of the matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{\rm ST} = \begin{bmatrix} A^T & C^T \\ -B^T & D^T \end{bmatrix} . \quad (3.6)$$

The matrix $H$ in Eq. (3.5) is given by

$$H = \begin{bmatrix} 0 & I & 0 & 0 \\ -I & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} , \quad (3.7)$$

where $I$ is the $N\times N$ unit matrix. It follows from Eq. (3.5) that the matrix $R$ is an element of the supergroup Osp(2/2N,R). We can now put the transformation equations into matrix form:

$$\begin{bmatrix} \tilde{b}^\dagger \\ \tilde{b} \\ \tilde{a}^\dagger \\ \tilde{a} \end{bmatrix} = R \begin{bmatrix} b^\dagger \\ b \\ a^\dagger \\ a \end{bmatrix} . \quad (3.8)$$

It is clear from Eq. (3.8) that this transformation is a generalized Bogoliubov transformation for a mixed system of bosons and fermions.

We find it convenient to introduce the following operator:

$$T = \exp\left[\tfrac{1}{2}Z_{ij}b_i^\dagger b_j^\dagger\right]\exp\left[\tfrac{1}{2}\psi_i a b_i^\dagger\right]\exp\left[\tfrac{1}{2}\eta_i a^\dagger b_i^\dagger\right] , \quad (3.9)$$

with $Z$ an $N\times N$ symmetric matrix, and $\psi$ and $\eta$ $N\times 1$ Grassmann column vectors. It is then easy to verify that

$$Tb_iT^{-1} = b_i - Z_{ij}b_j^\dagger - \tfrac{1}{2}\eta_i a^\dagger - \tfrac{1}{2}\psi_i a . \quad (3.10a)$$

Thus it follows that we can write the transformation of $b_i$ as

$$\tilde{b}_i = u_{ij}Tb_iT^{-1} , \quad (3.10b)$$

$$Z_{jk} = -u_{jl}^{-1}v_{lk}, \quad \eta_j = -2u_{jl}^{-1}\theta_{2l}, \quad \psi_j = -2u_{jl}^{-1}\theta_{1l} . \quad (3.10c)$$

We then define the Osp(2/2N,R) coherent state as

$$|Z,\psi,\eta\rangle = T|0\rangle . \quad (3.11)$$

It then follows that the quasiparticle operators annihilating the coherent state are

$$b_i - Z_{ij}b_j^\dagger - \tfrac{1}{2}\eta_i a^\dagger - \tfrac{1}{2}\psi_i a . \quad (3.12)$$

In Eqs. (3.9)–(3.12), $i,j = 1-N$, and a summation over repeated indices is implied. For the coherent state to be left invariant, we need $Z$, $\eta$, and $\psi$ to transform as

$$Z \to (u^\dagger + Zv^\dagger - \tfrac{1}{2}\eta\bar{\theta}_2^T - \tfrac{1}{2}\psi\bar{\theta}_1^T)^{-1} \times (v^T + Zu^T + \tfrac{1}{2}\eta\theta_1^T + \tfrac{1}{2}\psi\theta_2^T) , \quad (3.13a)$$

$$\psi \to (u^\dagger + Zv^\dagger - \tfrac{1}{2}\eta\bar{\theta}_2^T - \tfrac{1}{2}\psi\bar{\theta}_1^T)^{-1} \times (\psi\lambda_1^* + \eta\lambda_2^* - 2Z\bar{\chi}_2^T - 2\bar{\chi}_1^T) , \quad (3.13b)$$

$$\eta \to (u^\dagger + Zv^\dagger - \tfrac{1}{2}\eta\bar{\theta}_2^T - \tfrac{1}{2}\psi\bar{\theta}_1^T)^{-1}$$
$$\times(\eta\lambda_1^* + \psi\lambda_2^* - Z\chi_1^T - 2\chi_2^T)\,. \tag{3.13c}$$

By defining $\Xi = [Z, \frac{1}{2}\eta, \frac{1}{2}\psi]$, we may express the transformation of $Z$, $\psi$ and $\eta$ as

$$\Xi \to \Xi' = (A + \Xi C)^{-1}(B + \Xi D) \tag{3.14a}$$

with

$$A = u^\dagger, \quad B = [v^T - \chi_2^T - \bar{\chi}_1^T]\,, \tag{3.14b}$$

$$C = \begin{bmatrix} v^\dagger \\ -\bar{\theta}_2^T \\ -\bar{\theta}_1^T \end{bmatrix}, \quad D = \begin{bmatrix} u^T & -\chi_1^T & -\bar{\chi}_2^T \\ \theta_1^T & \lambda_1 & \lambda_2^* \\ \theta_2^T & \lambda_2 & \lambda_1^* \end{bmatrix}. \tag{3.14c}$$

Now, by following the methods of Ref. 9, we find the following set of relations, which allow us to express one set of variables in terms of the other:

$$u = UB, \quad B^\dagger = B, \quad UU^\dagger = 1\,, \tag{3.15a}$$

$$\chi_1^T = \tfrac{1}{2}(1 - Z^*Z)^{-1}[\bar{\psi} - Z\eta, \bar{\eta} - Z\psi]\begin{bmatrix}\lambda_1 \\ \lambda_2\end{bmatrix}, \tag{3.15b}$$

$$\chi_2^T = \tfrac{1}{2}(1 - ZZ^*)^{-1}[\eta - Z\bar{\psi}, \psi - Z\bar{\eta}]\begin{bmatrix}\lambda_1 \\ \lambda_2\end{bmatrix}, \tag{3.15c}$$

$$B^T B = (1 - ZZ^\dagger + \tfrac{1}{4}\bar{\psi}\psi^T + \tfrac{1}{4}\eta\bar{\eta}^T)^{-1}\,, \tag{3.15d}$$

$$\lambda_1^*\lambda_1 + \lambda_2^*\lambda_2 + \chi_1\bar{\chi}_1^T - \chi_2\bar{\chi}_2^T = 1\,. \tag{3.15e}$$

The invariant measure $\mu(\Xi)$ is defined by

$$\mu(\Xi) = \mu(\Xi')|J(\Xi', \Xi)| = \mu(0)|J(0, \Xi)|\,, \tag{3.16}$$

where $J(\Xi', \Xi)$ is the Jacobian of the transformation that sends $\Xi$ to $\Xi'$. We finally conclude that

$$[\det(A - BD^{-1}C)]^{2(N-2)}|J(0,\Xi)| = (\mathrm{Sdet}\, D)^{2N}. \tag{3.17}$$

Hence the invariant measure in the decomposition of unity is

$$d\mu(Z,\psi,\eta) \propto d\theta\, d\bar{\theta}\, d\chi\, d\bar{\chi}\, d^{2N}Z \times \frac{(\mathrm{Sdet}\, D)^{2N}}{[\det(A - BD^{-1}C)]^{2(N-2)}}\,. \tag{3.18}$$

For the case $N = 1$, this can be shown to be consistent with Eq. (2.9b).

## IV. CONCLUSION

In this paper we have continued our treatment of coherent states for the Osp($p/2N,R$)-type noncompact supergroups by investigating those for Osp($2/2N,R$). It should also be noted that there is no difficulty in extending this treatment to compact versions of these Osp($p/2N,R$) supergroups. Since Osp($p/2m,R$) supergroups have been used in generalized superfield theories,[8] the study of coherent states, in general, and the measure of integration in particular, should prove useful. The supergroup Osp($2/2,R$) is clearly among the simplest of the Osp($p/2m,R$) supergroups. While we have not found any physical applications for Osp($2/2,R$) superalgebras, it appears that Osp($4/2,R$) superalgebras may be useful in the study of nuclear structure.[15] In Ref. 15, it was shown that Osp($4/2,R$) has two possibly useful subgroup chains:

$$\mathrm{Osp}(4/2,R) \supset \mathrm{Sp}(2,R) \times \mathrm{SO}(4)\,, \tag{4.1a}$$

$$\mathrm{Osp}(4/2,R) \supset \mathrm{U}(1/2) \cong \mathrm{SU}(1/2) \times \mathrm{U}(1)\,. \tag{4.1b}$$

The superalgebra $\mathrm{SU}(1/2) \cong \mathrm{Spl}(2,1)$ is a compact superalgebra whose representations have been studied by several authors,[16] and, as we noted earlier, their coherent states could be studied by the techniques presented in this paper. Finally, it might be useful to study the coherent states for Osp($4/2,R$), although the difficulty of handling such supergroups increases rapidly. A start towards this end is now underway, beginning with an investigation of the representations for Osp($4/2,R$) superalgebras.[17]


## ACKNOWLEDGMENTS

The authors would like to acknowledge useful discussions with A. Bincer. HAS and PH would like to thank the members of the Theoretical Nuclear Physics group at University of Wisconsin for hospitality during their stay.

The research of ABB was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, in part by an National Science Foundation Grant No. PHY-8610851, and in part by a Presidential Young Investigator Award administered by the U.S. National Science Foundation. The research of HAS and PH was supported in part by the U.S. National Science Foundation Grant No. PHY-8603883.


## APPENDIX A: REPRESENTATION THEORY FOR THE SUPERALGEBRA Osp(2/2,R)

We construct the irreducible representations of Osp($2/2,R$) by acting with the raising operators on the lowest weight state (LWS), that state that is annihilated by all the lowering operators. We choose the first and second components of the weight to be the eigenvalues of $K_0$ and $B$, respectively. From the commutation and anticommutation relations [Eq. (2.2)], it is then clear that the lowering operators are $V_-$, $W_-$, and $K_-$, whereas the raising operators are $V_+$, $W_+$, and $K_+$. The Osp($2/2,R$) representations can be labeled by the $\mathrm{Sp}(2,R) \times \mathrm{SO}(2)$ quantum numbers of the LWS, the state vectors being further specified by including the labels $k(k-1)$ and $M_k$ (the eigenvalues of $K^2$ and $K_0$) and $Q$ [the SO(2) label].

By using the anticommutation relations of Eq. (2.2), it can be shown that there are four independent combinations of the supercaret raising operators, 1, $V_+$, $W_+$, and $V_+W_+$. Thus each Osp($2/2,R$) representation decomposes into at most four representations of the subgroup. Upon evaluating the norms of these states, we find that all four independent multiplets are not always realized; the conditions for the existence of any multiplet can be expressed in terms of the Osp($2/2,R$) quantum numbers $\tau$ and $q$. The decomposition is (the conditions for the existence of each multiplet are given below the corresponding labels)

$$(\tau\, q)\to(\tau)(q)\oplus(\tau+\tfrac{1}{2})(q+\tfrac{1}{2})\oplus(\tau+\tfrac{1}{2})(q-\tfrac{1}{2})\oplus(\tau+1)(q)$$
$$\tau-q\neq 0 \qquad \tau+q\neq 0 \qquad \tau-q\neq 0,\ \tau+q\neq 0. \tag{A1}$$

We denote the $Sp(2,R)\times SO(2)$ LWS by

$$|\tau\, q;\tau'\,\tau'\,q'\rangle\equiv|\tau\, q;\tau'\,q'\rangle\,, \tag{A2}$$

where the first two quantum numbers of the ket on the right-hand side of Eq. (A2) label the $Osp(2/2,R)$ representation, while the second set denote those for $Sp(2,R)\times SO(2)$. The four orthogonalized state vectors obtained from $Osp(2/2,R)$ on reduction to $Sp(2,R)\times SO(2)$, are

$$|\tau\, q;\tau\, q\rangle=|\phi_0\rangle\,, \tag{A3a}$$

$$|\tau\, q;\tau+\tfrac{1}{2}\,b+\tfrac{1}{2}\rangle=\sqrt{[1/2(\tau-q)]}\,V_+|\phi_0\rangle\,, \tag{A3b}$$

$$|\tau\, q;\tau+\tfrac{1}{2}\,b-\tfrac{1}{2}\rangle=\sqrt{[1/2(\tau+q)]}\,W_+|\phi_0\rangle\,, \tag{A3c}$$

$$|\tau\, q;\tau+1\,b\rangle=\sqrt{[1/4\tau(2\tau+1)(\tau+q)(\tau-q)]}\times\{\tau V_+W_+-(\tau+q)\}|\phi_0\rangle\,. \tag{A3d}$$

It is possible to use the raising operator $K_+$ to ladder within a representation, with the result

$$|\tau\, q;k\,k+m\,q'\rangle=\sqrt{[\Gamma(2k)/\Gamma(m+1)\Gamma(2k+m)]}\,K_+^m\,|\tau\, q;k\,k\,q'\rangle\,, \tag{A4}$$

where the square root factor ensures that the state remains normalized. Finally, by using the commutation and anticommutations relations, one can find the action of the operators on the various multiplets. In Eq. (A5), we suppress the $Osp(2/2,R)$ representation labels and obtain the following results:

$$K_\pm\,|\tau\,\tau+m\,q\rangle=\sqrt{(\tau+m\pm\tau)(\tau+m\mp\tau\pm1)}\,|\tau\,\tau+m\pm1\,q\rangle\,, \tag{A5a}$$

$$V_\pm\,|\tau\,\tau+m\,q\rangle=\alpha\sqrt{\tau+m\pm\tau}\,|\tau+\tfrac{1}{2}\,\tau+m\pm\tfrac{1}{2}\,q+\tfrac{1}{2}\rangle\,, \tag{A5b}$$

$$W_\pm\,|\tau\,\tau+m\,q\rangle=\beta\sqrt{\tau+m\pm\tau}\,|\tau+\tfrac{1}{2}\,\tau+m\pm\tfrac{1}{2}\,q-\tfrac{1}{2}\rangle\,, \tag{A5c}$$

$$V_\pm\,|\tau+\tfrac{1}{2}\,\tau+\tfrac{1}{2}+m\,q+\tfrac{1}{2}\rangle=0\,, \tag{A5d}$$

$$W_\pm\,|\tau+\tfrac{1}{2}\,\tau+\tfrac{1}{2}+m\,q+\tfrac{1}{2}\rangle=\gamma\sqrt{\tau+\tfrac{1}{2}+m\mp\tau\pm\tfrac{1}{2}}\,|\tau\,\tau+\tfrac{1}{2}+m\pm\tfrac{1}{2}\,q\rangle+\delta\sqrt{\tau+\tfrac{1}{2}+m\pm\tau\pm\tfrac{1}{2}}\,|\tau+1\,\tau+\tfrac{1}{2}+m\pm\tfrac{1}{2}q\rangle\,, \tag{A5e}$$

$$V_\pm\,|\tau+\tfrac{1}{2}\,\tau+\tfrac{1}{2}+m\,q-\tfrac{1}{2}\rangle=\rho\sqrt{\tau+\tfrac{1}{2}+m\mp\tau\pm\tfrac{1}{2}}\,|\tau\,\tau+\tfrac{1}{2}+m\pm\tfrac{1}{2}\,q\rangle+\sigma\sqrt{\tau+\tfrac{1}{2}+m\pm\tau\pm\tfrac{1}{2}}\,|\tau+1\,\tau+\tfrac{1}{2}+m\pm\tfrac{1}{2}\,q\rangle\,, \tag{A5f}$$

$$W_\pm\,|\tau+\tfrac{1}{2}\,\tau+\tfrac{1}{2}+m\,q-\tfrac{1}{2}\rangle=0\,, \tag{A5g}$$

$$V_\pm\,|\tau+1\,\tau+1+m\,q\rangle=\kappa\sqrt{\tau+1+m\mp\tau}\,|\tau+\tfrac{1}{2}\,\tau+1+m\pm\tfrac{1}{2}\,q+\tfrac{1}{2}\rangle\,, \tag{A5h}$$

$$W_\pm\,|\tau+1\,\tau+1+m\,q\rangle=\omega\sqrt{\tau+1+m\mp\tau}\,|\tau+\tfrac{1}{2}\,\tau+1+m\pm\tfrac{1}{2}\,q-\tfrac{1}{2}\rangle\,. \tag{A5i}$$

The complex numbers $\alpha,\beta,\ldots$ may depend on $\tau$ and $b$ but not on $m$, a result of using the Wigner–Eckart theorem. There are three different representations that we must consider separately, and these are discussed next.

(i) *The $\tau=q$ representation*: The coefficients then satisfy

$$\alpha=\gamma=\delta=\sigma=\kappa=\omega=0 \quad\text{and}\quad \beta\rho=1\,. \tag{A6}$$

(ii) *The $\tau=-q$ representation*: For this case the coefficients satisfy

$$\beta=\rho=\delta=\sigma=\kappa=\omega=0 \quad\text{and}\quad \alpha\gamma=1\,. \tag{A7}$$

(iii) If $\tau\neq\pm q$ then all four multiplets occur, and we have a representation iff the following relations hold among the coefficients:

$$\gamma=\frac{1}{\alpha}\frac{\tau-q}{2\tau},\quad \rho=\frac{1}{\beta}\frac{\tau+q}{2\tau},\quad \sigma=-\frac{\alpha\delta}{\beta},\quad \kappa=\frac{1}{\delta}\frac{\tau+q}{2\tau},\quad \omega=\beta\frac{1}{\delta\alpha}\frac{\tau-q}{2\tau}, \tag{A8a}$$

and

$$\alpha^2=\frac{\tau-q}{2\tau},\quad \beta^2=\frac{\tau+q}{2\tau},\quad \sigma^2=\frac{\tau-q}{2\tau}. \tag{A8b}$$

## APPENDIX B: MATRIX ELEMENTS FOR THE GENERATORS OF Osp(2/2,R)

In this Appendix we present the matrix elements for the generators of the supergroup $Osp(2/2,R)$. These are presented in a form where the simplification for the atypical representations are easily seen. For convenience, we define two supermatrices $M_1$ and $M_2$ by

$$M_1=M(\beta,\bar\beta;\chi,\bar\chi)\,, \tag{B1a}$$

$$M_2=M(\sigma,\bar\sigma;\theta,\bar\theta)\,. \tag{B1b}$$

We may then write the matrix elements of the various generators of the superalgebra in the following compact form:

$$\frac{\langle\theta'\chi'\alpha'|K_+|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle}=\bar\beta(\tau+b)\,\mathrm{Sdet}\,M_1+\bar\sigma'(\tau-b)\,\mathrm{Sdet}\,M_2\,, \tag{B2a}$$

$$\frac{\langle\theta'\chi'\alpha'|K_-|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle}=\beta(\tau+b)\,\mathrm{Sdet}\,M_1+\sigma(\tau-b)\,\mathrm{Sdet}\,M_2\,, \tag{B2b}$$

$$\frac{\langle\theta'\chi'\alpha'|K_0|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle}=\frac{\tau+b}{2}(1+\beta\bar\beta')\,\mathrm{Sdet}\,M_1+\frac{\tau-b}{2}(1+\sigma\bar\sigma')\,\mathrm{Sdet}\,M_2\,, \tag{B2c}$$

$$\frac{\langle\theta'\chi'\alpha'|V_+|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle} = -\bar{\theta}'(\tau-b)\mathrm{Sdet}\,M_2 - \chi\bar{\beta}'(\tau+b)\mathrm{Sdet}\,M_1\,, \tag{B2d}$$

$$\frac{\langle\theta'\chi'\alpha'|W_-|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle} = -\theta(\tau-b)\mathrm{Sdet}\,M_2 - \bar{\chi}'(\tau+b)\beta\,\mathrm{Sdet}\,M_1\,, \tag{B2e}$$

$$\frac{\langle\theta'\chi'\alpha'|V_-|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle} = -\bar{\theta}'(\tau-b)\sigma\,\mathrm{Sdet}\,M_2 - \chi(\tau+b)\mathrm{Sdet}\,M_1\,, \tag{B2f}$$

$$\frac{\langle\theta'\chi'\alpha'|W_+|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle} = -\theta\bar{\sigma}'(\tau-b)\mathrm{Sdet}\,M_2 - \bar{\chi}'(\tau+b)\mathrm{Sdet}\,M_1\,, \tag{B2g}$$

$$\frac{\langle\theta'\chi'\alpha'|B|\theta\chi\alpha\rangle}{\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle} = b + \frac{\chi\bar{\chi}'}{2}\mathrm{Sdet}\,M_2 - \frac{\theta\bar{\theta}'}{2}\mathrm{Sdet}\,M_1\,. \tag{B2h}$$

Of course the coherent states for Osp(2/2,$R$) are not orthogonal, and the overlap is conveniently expressed in terms of the superdeterminants of the matrices in Eq. (B1):

$$\begin{aligned}\langle\theta'\chi'\alpha'|\theta\chi\alpha\rangle &= (\mathrm{Sdet}\,M(\beta,\bar{\beta}';\chi,\bar{\chi}'))^{\tau+q}(\mathrm{Sdet}\,M(\sigma,\bar{\sigma}';\theta,\bar{\theta}'))^{\tau-q} \\ &= (\mathrm{Sdet}\,M(\beta,\bar{\beta};\chi,\bar{\chi}))^{-(\tau+q)/2} \\ &\quad\times(\mathrm{Sdet}\,M(\sigma,\bar{\sigma};\theta,\bar{\theta}))^{-(\tau-q)/2} \\ &= (\mathrm{Sdet}\,M(\beta',\bar{\beta}';\chi',\bar{\chi}'))^{-(\tau+q)/2} \\ &\quad\times(\mathrm{Sdet}\,M(\sigma',\bar{\sigma}';\theta',\bar{\theta}'))^{-(\tau-q)/2}\,. \end{aligned} \tag{B3}$$


[1] P. Carruthers and M. M. Nieto, Rev. Mod. Phys. **46**, 411 (1968); L. G. Yaffe, *ibid.* **54**, 407 (1982); K. T. Hecht, *The Vector Coherent State Methods and Its Application to Problems of Symmetries* (Springer, Berlin, 1987); D. J. Rowe, "Some recent advances in coherent state theory and its application to nuclear collective motion," in *Phase Space Applications to Nuclear Dynamics*, edited by M. di Toro, W. Nuremberg, M. Rosina, and S. Stringari (World Scientific, Singapore, 1986), p. 546.

[2] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[3] H. Kuratsuji and T. Suzuki, J. Math. Phys. **21**, 472 (1980).

[4] C. C. Gerry and S. Silverman, J. Math. Phys. **23**, 1995 (1982).

[5] F. Iachello and O. Scholten, Phys. Rev. Lett. **43**, 679 (1979).

[6] A. B. Balantekin, I. Bars, and F. Iachello, Nucl. Phys. A **370**, 284 (1981); J. Jolie, K. Heyde, P. Van Isacker, and A. Frank, *ibid.* **466**, 1 (1987).

[7] A. Montorsi, M. Rasetti, and A. I. Solomon, Phys. Rev. Lett. **59**, 2243 (1987).

[8] A. Salam and J. Strathdee, Nucl. Phys. B **79**, 477 (1974); M. Günaydin and N. P. Warner, *ibid.* **272**, 99 (1986).

[9] A. B. Balantekin, H. A. Schmitt, and B. R. Barrett, J. Math. Phys. **29**, 1634 (1988).

[10] V. Rittenberg, *VI International Colloquium on Group Theoretical Methods in Physics*, edited by P. Kramer and A. Rieckers (Springer, Berlin, 1978).

[11] A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985); N. Baake and P. Reinicke, J. Math. Phys. **27**, 859 (1986).

[12] A. Perelomov, Commun. Math. Phys. **26**, 222 (1972).

[13] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).

[14] A. Perelomov, Sov. J. Nucl. Phys. **29**, 867 (1979).

[15] H. A. Schmitt, P. Halse, A. B. Balantekin, and B. R. Barrett, Phys. Lett. B **210**, 1 (1988).

[16] M. Scheunert, W. Nahm, and V. J. Rittenberg, J. Math. Phys. **18**, 155 (1977).

[17] H. A. Schmitt, P. Halse, A. B. Balantekin, and B. R. Barrett, submitted to J. Math. Phys.

# Continuous subgroups of the generalized Schrödinger groups

L. F. Barannik and W. I. Fushchich
*Institute of Mathematics, Academy of Sciences of the Ukrainian SSR, Repin Street 3, Kiev, USSR*

(Received 12 August 1987; accepted for publication 31 August 1988)

Some general results on the subalgebras of the Lie algebra $\mathrm{ASch}(n)$ of the generalized Schrödinger group $\mathrm{Sch}(n)$ and on the subalgebras of the Lie algebra $\widetilde{\mathrm{ASch}}(n)$ of the generalized extended Schrödinger group $\widetilde{\mathrm{Sch}}(n)$ have been obtained. The subalgebra structure of the algebras $\mathrm{ASch}(n)$ and $\widetilde{\mathrm{ASch}}(n)$ are studied with respect to inner automorphisms of the groups $\mathrm{Sch}(n)$ and $\widetilde{\mathrm{Sch}}(n)$, respectively. The maximal Abelian subalgebras and the one-dimensional subalgebras of the algebras $\mathrm{ASch}(n)$ and $\widetilde{\mathrm{ASch}}(n)$ have been explicitly found. The full classification of the subalgebras of the algebras $\mathrm{ASch}(3)$, $\widetilde{\mathrm{ASch}}(3)$, which are nonconjugate to the subalgebras of $\mathrm{ASch}(2)$, $\widetilde{\mathrm{ASch}}(2)$, respectively, has been carried out.

## I. INTRODUCTION

To construct exact solutions of both linear and nonlinear Schrödinger and heat equations it is important to know the subgroup structure of the extended Schrödinger group $\widetilde{\mathrm{Sch}}(3)$ (see Ref. 1). Other important applications of subgroup structure of this group were discussed in Refs. 2 and 3. It is natural to generalize the notions of the three-dimensional Schrödinger group for the case of arbitrary $n$-dimensional Euclidean space and to solve the problem of subgroup classification for these generalized groups. If we restrict ourselves by continuous subgroups, then the problem will be reduced to classification of subalgebras of correspondent Lie algebras. This classification was realized for $n = 1$ in Ref. 4 and for $n = 2$ in Ref. 2.

In the present paper we study subalgebra structure of both the Lie algebra $\mathrm{ASch}(n)$ of the Schrödinger group $\mathrm{Sch}(n)$ and the Lie algebra $\widetilde{\mathrm{ASch}}(n)$ of the extended Schrödinger group $\widetilde{\mathrm{Sch}}(n)$ with respect to inner automorphisms of the group $\mathrm{Sch}(n)$ and the group $\widetilde{\mathrm{Sch}}(n)$, respectively. This paper is a continuation of investigations that were carried out in Refs. 5–9. The applied general method of Patera, Winternitz, and Zassenhaus[10] gets further development for classes of groups under consideration.

In Sec. II we give definitions of the generalized Schrödinger groups and algebras and introduce some other concepts and basis notation used in the whole paper. In Sec. III, completely reducible subalgebras of the algebra $\mathrm{AO}(n) \oplus \mathrm{ASL}(2,R)$ are derived, and all subalgebras of this algebra are described for $n = 3$. In Sec. IV a number of general results about splitting subalgebras of the algebra $\mathrm{ASch}(n)$ are obtained. Abelian subalgebras of the extended Schrödinger algebra $\widetilde{\mathrm{ASch}}(n)$ are described in Sec. V. Classification of subalgebras of the algebras $\mathrm{ASch}(3)$ and $\widetilde{\mathrm{ASch}}(3)$ is carried out in Sec. VI. The conclusions are summarized in Sec. VII.

## II. DEFINITIONS OF SCHRÖDINGER GROUPS AND ALGEBRAS. MAIN NOTATION

Let $R$ be the real number field, $R^n$ an arithmetical $n$-dimensional Euclidean space, and AG the Lie algebra of the Lie group G. The Schrödinger group $\mathrm{Sch}(n)$ is the multiplicative group of matrices

$$\begin{pmatrix} W & \mathbf{v} & \mathbf{a} \\ 0 & \alpha & \beta \\ 0 & \gamma & \delta \end{pmatrix},$$

where $W \in \mathrm{O}(n)$, $\mathbf{a}, \mathbf{v} \in R^n$, and $\alpha\delta - \beta\gamma = 1$ ($\alpha, \beta, \gamma, \delta \in R$). If $\alpha = \delta = 1$, $\gamma = 0$, we obtain matrices that are elements of the Galilei group $\mathrm{G}(n)$. If at the same time $\beta = 0$, we have elements of the isochronous Galilei group $\mathrm{G}^0(n)$. Besides, the Schrödinger group $\mathrm{Sch}(n)$ can be realized as the transformation group

$$\mathbf{x} \to \frac{W\mathbf{x} + t\,\mathbf{v} + \mathbf{a}}{\gamma t + \delta}, \quad t \to \frac{\alpha t + \beta}{\gamma t + \delta},$$

where $t$ is time and $\mathbf{x}$ is a variable vector of the space $R^n$.

The Lie algebra $\mathrm{ASch}(n)$ of the group $\mathrm{Sch}(n)$ consists of real matrices

$$\begin{pmatrix} X & \mathbf{v} & \mathbf{a} \\ 0 & \alpha & \beta \\ 0 & \gamma & -\alpha \end{pmatrix},$$

where $X \in \mathrm{AO}(n)$, $\alpha, \beta, \gamma \in R$, and $\mathbf{a}, \mathbf{v} \in R^n$. Let $I_{ab}$ be a matrix of degree $n + 2$ having unity at the intersection of the $a$th line and the $b$ th column and zeros at the other places ($a$, $b = 1,\ldots,n+2$). Then the basis of the algebra $\mathrm{ASch}(n)$ is formed by the matrices

$$J_{ab} = I_{ab} - I_{ba}, \quad G_a = I_{a,n+1}, \quad P_a = I_{a,n+2},$$
$$D = -I_{n+1,n+1} + I_{n+2,n+2},$$
$$S = -I_{n+2,n+1}, \quad T = I_{n+1,n+2}$$

($a < b$, $a,b = 1,\ldots,n$). They satisfy the following commutation relations:

$$[J_{ab},J_{cd}] = \delta_{ad}J_{bc} + \delta_{bc}J_{ad} - \delta_{ac}J_{bd} - \delta_{bd}J_{ac};$$
$$[P_a,J_{bc}] = \delta_{ab}P_c - \delta_{ac}P_b; \quad [P_a,P_b] = 0;$$
$$[G_a,J_{bc}] = \delta_{ab}G_c - \delta_{ac}G_b; \quad [G_a,G_b] = 0;$$
$$[G_a,P_b] = 0; \quad [D,J_{ab}] = [S,J_{ab}] = [T,J_{ab}] = 0;$$
$$[D,P_a] = -P_a, \quad [D,G_a] = G_a;$$
$$[S,P_a] = G_a, \quad [S,G_a] = 0;$$
$$[T,P_a] = 0, \quad [T,G_a] = -P_a;$$
$$[D,S] = 2S, \quad [D,T] = -2T, \quad [T,S] = D$$
$$(a,b,c,d = 1,2,\ldots,n).$$

 0022-2488/89/020280-11$02.50 © 1989 American Institute of Physics 

The extended Schrödinger algebra $\widetilde{\mathrm{ASch}}(n)$ is obtained from the algebra $\mathrm{ASch}(n)$ by adding the central element $M$, and, moreover, $[G_a,P_b]=\delta_{ab}M$ and other commutation relations do not change. The factor algebra $\widetilde{\mathrm{ASch}}(n)/\langle M\rangle$ is identified with $\mathrm{ASch}(n)$. We shall denote the generators of algebras $\mathrm{ASch}(n)$ and $\widetilde{\mathrm{ASch}}(n)$ by the same symbols.

The algebra $\widetilde{\mathrm{AG}}^0(n)=\mathrm{AO}(n)\oplus\langle M,P_1,...,P_n,G_1,...,G_n\rangle$ is called the extended isochronous Galilei algebra, and the algebra $\mathrm{AG}^0(n)=\widetilde{\mathrm{AG}}^0(n)/\langle M\rangle$ is called the isochronous Galilei algebra.

Since the Lie algebra $L=\langle M,P_1,...,P_n,G_1,....G_n\rangle$ is nilpotent, $L$ is a Lie algebra of some connected and simply connected nilpotent Lie group $H$. As $H$ is an exponential group, any of its elements can be denoted as $\exp(\theta M)\times\exp(\mathbf{vG}+\mathbf{aP})$, where $\theta\in R$, $\mathbf{vG}=v_1G_1+\cdots+v_nG_n$, and $\mathbf{aP}=a_1P_1+\cdots+a_nP_n$ $(a_i,v_i\in R,\ i=1,...,n)$. The multiplication law is derived by the Campbell–Hausdorf formula. Let

$$\Delta=\begin{pmatrix} W & 0 & 0\\ 0 & \alpha & \beta\\ 0 & \gamma & \delta\end{pmatrix}$$

be an element of $\mathrm{O}(n)\times\mathrm{SL}(2,R)$. It is not difficult to show that in $\mathrm{Sch}(n)$ we have

$$\Delta\cdot\exp(\mathbf{vG}+\mathbf{aP})=\exp((\delta W\mathbf{v}-\gamma W\mathbf{a})\mathbf{G}+(-\beta W\mathbf{v}+\alpha W\mathbf{a})\mathbf{P})\cdot\Delta. \tag{1}$$

An arbitrary element of the group $\widetilde{\mathrm{Sch}}(n)$ has the form

$$\exp(\theta M)\cdot\exp(\mathbf{vG}+\mathbf{aP})\cdot\Delta.$$

By definition, $\exp(\theta M)\cdot\Delta=\Delta\cdot\exp(\theta M)$, and the equality (1) holds true for $\Delta\cdot\exp(\mathbf{vG}+\mathbf{aP})$. Using these equalities and multiplication laws in $H$ and $\mathrm{O}(n)\times\mathrm{SL}(2,R)$ we shall establish multiplication in $\widetilde{\mathrm{Sch}}(n)$ in the usual way. Evidently, $\widetilde{\mathrm{Sch}}(n)=H\lambda(\mathrm{O}(n)\times\mathrm{SL}(2,R))$.

Subalgebras $L_1$ and $L_2$ of the algebra $\widetilde{\mathrm{ASch}}(n)$ are called $\widetilde{\mathrm{Sch}}(n)$ conjugated if $gL_1g^{-1}=L_2$ for some element $g\in\widetilde{\mathrm{Sch}}(n)$. Mapping $\varphi_g: X\to gXg^{-1}$, $X\in\widetilde{\mathrm{ASch}}(n)$, is called an automorphism corresponding to the element $g$. If $g=\mathrm{diag}[W,1,1]$, where $W\in\mathrm{O}(n)$, then $\varphi_g$ is called an $\mathrm{O}(n)$ automorphism corresponding to the matrix $W$. We shall identify the automorphism $\varphi_g$ with the element $g$.

Henceforth we shall use the following notations: $\langle X_1,...,X_s\rangle$ is a vector space or Lie algebra over $R$ with the generators $X_1,...,X_s$; $V[k,l]=\langle G_k,...,G_l\rangle$ $(k\leqslant l)$ is a Euclidean space having the orthonormal basis $G_k,...,G_l$, $V[k]=V[k,k]$; $W[k,l]=\langle P_k,...,P_l\rangle$ $(k\leqslant l)$, $W[k]=W[k,k]$; $\mathfrak{M}[r,t]=\langle M,P_r,...,P_t,G_r,...,G_t\rangle$ $(r\leqslant t)$, $\mathfrak{M}[r]=\mathfrak{M}[r,r]$, $\overline{\mathfrak{M}}[r,t]=\mathfrak{M}[r,t]/\langle M\rangle$; $\pi,\omega,\tau,\epsilon$, and $\xi$ are projections of the algebras $\widetilde{\mathrm{ASch}}(n)$ and $\mathrm{ASch}(n)$ onto $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$, $\mathrm{AO}(n)$, $\mathrm{ASL}(2,R)$, $V[1,n]$, and $W[1,n]$, respectively.

Let $U$ be a subspace of $\overline{\mathfrak{M}}[1,n]$ and $\widehat{F}$ be a subalgebra of $\mathrm{ASch}(n)$ such that $\pi(\widehat{F})=F$. The notation $\widehat{F}+U$ means that $[F,U]\subset U$ and $\widehat{F}\cap\overline{\mathfrak{M}}[1,n]\subset U$. Considering algebras $(\widehat{F}+U_1),...,(\widehat{F}+U_s)$ we shall use the notation $\widehat{F}\colon U_1,...,U_s$. In the case of the algebra $\widetilde{\mathrm{ASch}}(n)$ this notation has the same meaning.

Let $L$ be the direct sum of Lie algebras $L_1,...,L_s$, $K$ a Lie subalgebra of $L$, and $\pi_i$ the projection of $L$ onto $L_i$. If $\pi_i(K)=L_i$ for all $i=1,...,s$, then $K$ is called the subdirect sum of algebras $L_1,...,L_s$. In this case we shall use the notation $K=L_1\dot{+}\cdots\dot{+}L_s$. The subdirect sum of modules over a Lie algebra is defined in a similar way.

## III. ON THE SUBALGEBRAS OF THE ALGEBRA $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$

In this section a number of auxiliary results to be used in following sections are obtained.

*Lemma 3.1:* Subalgebras of the algebra $\mathrm{ASL}(2,R)$ are exhausted with respect to $\mathrm{SL}(2,R$ conjugation by the following algebras: $0,\langle D\rangle$, $\langle T\rangle$, $\langle S+T\rangle$, $\langle D,T\rangle$, $\mathrm{ASL}(2,R)$. The written algebras are not conjugated mutually.

Later on, when we speak about subalgebras of the algebra $\mathrm{ASL}(2,R)$ we shall mean the subalgebras given by Lemma 3.1.

By direct calculations we are convinced that the normalizer of $\langle D\rangle$ in the group $\mathrm{SL}(2,R)$ consists of matrices

$$\begin{pmatrix} 0 & \alpha\\ -\alpha^{-1} & 0\end{pmatrix},\quad \begin{pmatrix}\sigma & 0\\ 0 & \alpha^{-1}\end{pmatrix},$$

where $\alpha\in R$, $\alpha\neq 0$. The normalizer of $\langle T\rangle$ and the normalizer of $\langle D,T\rangle$ in the group $\mathrm{SL}(2R)$ consists of matrices $\pm\exp(\theta_1D)\cdot\exp(\theta_2T)$, where $\theta_1,\theta_2\in R$. The normalizer of $\langle S+T\rangle$ coincides with the group

$$\mathrm{SO}(2)=\left\{\begin{pmatrix}\cos\varphi & \sin\varphi\\ -\sin\varphi & \cos\varphi\end{pmatrix}\middle|\ \varphi\in R\right\}.$$

*Proposition 3.1:* Let $\mathrm{AH}(n)$ be the Cartan subalgebra of the algebra $\mathrm{AO}(n)$. Up to conjugacy under $\mathrm{O}(n)\times\mathrm{SL}(2,R)$ the algebra $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$ has two maximal soluable subalgebras $\mathrm{AH}(n)\oplus\langle S+T\rangle$, $\mathrm{AH}(n)\oplus\langle D,T\rangle$.

Proposition 3.1 follows immediately from Lemma 3.1 and the fact that $\mathrm{AO}(n)$ has, with respect to $\mathrm{O}(n)$ conjugation, the only maximal solvable subalgebra $\mathrm{AH}(n)$.

*Proposition 3.2:* Up to conjugacy under $\mathrm{O}(n)\times\mathrm{SL}(2,R)$ the algebra $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$ has the following subalgebras: (i) $F\oplus K$, where $F\subset\mathrm{AO}(n)$, $K\subset\mathrm{ASL}(2,R)$; (ii) $F\oplus\langle X+Y\rangle$, where $F\oplus\langle X\rangle\subset\mathrm{AO}(n)$, $Y\in\mathrm{ASL}(2,R)$; and (iii) $\langle X+D\rangle\dot{+}(F\oplus\langle T\rangle)$, where $F\oplus\langle X\rangle\subset\mathrm{AO}(n)$.

Proposition 3.2 is proved by the Goursat twist method.[11]

*Corollary:* Subalgebras of the algebra $\mathrm{AO}(3)\oplus\mathrm{ASL}(2,R)$ are exhausted with respect to $\mathrm{O}(3)\times\mathrm{SL}(2,R)$ conjugation by the following algebras:

$0$; $\langle J_{12}\rangle$; $\langle D\rangle$; $\langle T\rangle$; $\langle S+T\rangle$; $\langle J_{12}+\alpha D\rangle$ $(\alpha>0)$;
$\langle J_{12}+T\rangle$; $\langle J_{12}+\alpha(S+T)\rangle$ $(\alpha>0)$; $\langle J_{12}+\alpha D,T\rangle$ $(\alpha>0)$;
$\langle D,T\rangle$; $\langle J_{12},D\rangle$; $\langle J_{12},T\rangle$; $\langle J_{12},S+T\rangle$; $\langle J_{12},D,T\rangle$;
$\mathrm{AO}(3)$; $\mathrm{ASL}(2,R)$; $\langle J_{12}\rangle \oplus \mathrm{ASL}(2,R)$; $\mathrm{AO}(3)\oplus\langle D\rangle$;
$\mathrm{AO}(3)\oplus\langle T\rangle$; $\mathrm{AO}(3)\oplus\langle S+T\rangle$; $\mathrm{AO}(3)\oplus\langle D,T\rangle$; $\mathrm{AO}(3)\oplus\mathrm{ASL}(2,R)$.

The written algebras are not conjugated mutually.

The space $\mathfrak{M}[1,n]$ can be considered as an exact module over the Lie algebra $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$. Let $L$ be a subalgebra of this algebra. If $\mathfrak{M}[1,n]$ is a completely reducible $L$ module, then the algebra $L$ will be called completely reducible.

**Theorem 3.1:** A subalgebra $L$ of the algebra $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$ is completely reducible if and only if $\tau(L)$ does not coincide with $\langle T\rangle$ and $\langle D,T\rangle$.

*Proof:* If $\tau(L)=0$, then $L$ is a completely reducible algebra. If $\tau(L)=\langle D,T\rangle$, then $L=L_1\oplus L_2$, where $L_1\subset\mathrm{AO}(n)$, $L_2=\langle X+D,T\rangle$, $X\in\mathrm{AO}(n)$. Since the algebra $L_2$ is solvable and non-Abelian, then $L$ is not a completely reducible algebra.[12] Let $\tau(L)=\mathrm{ASL}(2,R)$. Since direct decomposition of $F\subset\mathrm{AO}(n)$ can be realized through every ideal, and since every subalgebra of the algebra $\mathrm{AO}(n)$ is compact and $\mathrm{ASL}(2,R)$ is not compact, then $L=\omega(L)+\tau(L)$. That is why[12] $L$ is completely reducible.

Let us assume that $\tau(L)=\langle D\rangle$. Since $[D,P_a]=-P_a$, $[D,G_a]=G_a$, then $\mathfrak{M}[1,n]$ can be decomposed into a direct sum of $L$-irreducible spaces. Consequently $L$ is a completely reducible algebra.

As $[S+T,P_a]=G_a$ and $[S+T,G_a]=-P_a$, then the skew-symmetric matrix

$$\begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix}$$

corresponds to the operator $S+T$ in a basis $P_1,\ldots,P_n$, $G_1,\ldots,G_n$ of the space $\mathfrak{M}[1,n]$. Hence it follows that if $\tau(L)=\langle S+T\rangle$, then in the basis mentioned above every element of an algebra $L$ is represented by a skew-symmetric matrix of degree $2n$, and that is why $L$ is a completely reducible algebra.

Let $\tau(L)=\langle T\rangle$, and $V[k,l]$ be an irreducible $\omega(L)$ module. Evidently $V[k,l]+W[k,l]$ is an $L$ module. Since by Lemma 4.2 of Ref. 9 this module can not be decomposed into a direct sum of irreducible $L$ modules, an algebra $L$ is not completely reducible. The theorem is proved.

## IV. ON THE STRUCTURE OF SPLITTING SUBALGEBRAS OF THE SCHRÖDINGER ALGEBRA

The aim of this section is to study up to conjugation the subspaces of the space $\mathfrak{M}[1,n]$ invariant under subalgebras of the algebra $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$. The main results are Theorems 4.1 and 4.2.

Let $F$ be a subalgebra of $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$, and $\hat{F}$ be a subalgebra of the algebra $\mathrm{ASch}(n)$ such that $\pi(\hat{F})=F$. If algebra $\hat{F}$ is $\mathrm{Sch}(n)$ conjugated to the algebra $F\dot{+}\mathfrak{N}$, where $\mathfrak{N}$ is an $F$-invariant subspace of the space $\mathfrak{M}[1,n]$, then $\hat{F}$ is called a splitting in the algebra $\mathrm{ASch}(n)$. The notion of a splitting subalgebra of the algebra $\widetilde{\mathrm{ASch}}(n)$ is defined in an analogous way. If every subalgebra $\hat{F}$ is a splitting, we shall say that $F$ has only splitting extensions in the algebra $\mathrm{ASch}(n)$ [resp. in the algebra $\widetilde{\mathrm{ASch}}(n)$].

We shall find all subalgebras $F$, which possess only splitting extensions.

Let

$$J(a,b)=J_{2a-1,2a}+\cdots+J_{2b-1,2b}\quad(a\leqslant b),\quad J(a)=J(a,a),$$

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

$$X=S+T+\alpha_1 J_{12}+\cdots+\alpha_t J_{2t-1,2t},\quad 0\leqslant\alpha_1\cdots\leqslant\alpha_t,$$

$$Y_{2a-1}=G_{2a-1}+P_{2a},$$

$$Y_{2a}=G_{2a}-P_{2a-1},\quad Z_{2a-1}=G_{2a-1}-P_{2a},$$

$$Z_{2a}=G_{2a}+P_{2a-1},\quad \mathfrak{L}_a=\langle Y_{2a-1},Y_{2a}\rangle,$$

$$\mathfrak{N}_a=\langle Z_{2a-1},Z_{2a}\rangle.$$

Obviously, $\mathfrak{L}_a+\mathfrak{N}_a=\mathfrak{M}[2a-1,2a]$. If $1\leqslant a\leqslant t$, then

$$\begin{aligned}
&[X,Y_{2a-1}]=-(\alpha_a-1)Y_{2a},\\
&[X,Y_{2a}]=(\alpha_a-1)Y_{2a-1},\\
&[X,Z_{2a-1}]=-(\alpha_a+1)Z_{2a},\\
&[X,Z_{2a}]=(\alpha_a+1)Z_{2a-1}.
\end{aligned}\tag{2}$$

Thus $(\alpha_a-1)J$ is the matrix of ad $X$ in the basis $Y_{2a-1}$, $Y_{2a}$ of the space $\mathfrak{L}_a$, and $(\alpha_a+1)J$ is the matrix of ad $X$ in the basis $Z_{2a-1}$, $Z_{2a}$ of the space $\mathfrak{N}_a$ $(1\leqslant a\leqslant t)$. If $\alpha_a=0$, we obtain a matrix corresponding to $\mathrm{ad}(S+T)$.

Let $\alpha_a\neq 0$, $\alpha_a\neq 1$. The $\langle X\rangle$ module $\mathfrak{N}_a$ is called an elementary module of the first kind, and the $\langle X\rangle$ module $\mathfrak{N}_a$ is called an elementary module of the second kind. A subdirect sum of elementary modules of the first kind is called a module of the first kind, and a subdirect sum of elementary modules of the second kind is called a module of the second kind.

*Lemma 4.1:* Let $C$ be a matrix obtained from the identity matrix of degree $n$ as a result of fulfilling a permutation over its columns

$$\begin{pmatrix} 2k-1 & 2k & 2l-1 & 2l \\ 2l & 2l-1 & 2k & 2k-1 \end{pmatrix}\quad(k<l),$$

followed by the multiplication on $(-1)$ columns which have number $2k$ and $2l$. The $\mathrm{O}(n)$ automorphism $\varphi$ of the algebra $\mathrm{ASch}(n)$ which corresponds to the matrix $C$ has the following properties:

(1) $\varphi(J_{2d-1,2d})=J_{2d-1,2d}$, if $d\neq k$, $d\neq l$;
$\varphi(J_{2k-1,2k})=J_{2l-1,2l}$, $\varphi(J_{2l-1,2l})=J_{2k-1,2k}$;

(2) $\varphi(G_{2k-1})=G_{2l}$, $\varphi(G_{2k})=-G_{2l-1}$,
$\varphi(G_{2l-1})=G_{2k}$, $\varphi(G_{2l})=-G_{2k-1}$;

(3) $\varphi(\mathfrak{L}_k)=\mathfrak{L}_l$, $\varphi(\mathfrak{L}_l)=\mathfrak{L}_k$,

$\varphi(\mathfrak{N}_k)=\mathfrak{N}_l$, $\varphi(\mathfrak{N}_l)=\mathfrak{N}_k$.

*Proof:* For simplicity we can take $n=4$ and

$$C=\begin{pmatrix}0 & -J\\ -J & 0\end{pmatrix}.$$

Then

$$C(\alpha J_{12}+\beta J_{34})C^{-1}=\beta J_{12}+\alpha J_{34},$$

$$C\cdot\begin{pmatrix}y_1\\ y_2\\ y_3\\ y_4\end{pmatrix}=\begin{pmatrix}-y_4\\ y_3\\ -y_2\\ y_1\end{pmatrix}.$$

Using the last equality we conclude that $\varphi(G_1)=G_4$, $\varphi(G_2)=-G_3$, $\varphi(G_3)=G_2$, and $\varphi(G_4)=-G_1$. The lemma is proved.

*Lemma 4.2:* Letting $n\geqslant 4$, $1\leqslant q\leqslant[n/2]-1$, and $E_a$ be the identity matrix of degree $a$,

$$C_1(\lambda)=\begin{pmatrix}\frac{1}{\sqrt{1+\lambda^2}} & \frac{\lambda}{\sqrt{1+\lambda^2}}\\ \frac{\lambda}{\sqrt{1+\lambda^2}} & \frac{-1}{\sqrt{1+\lambda^2}}\end{pmatrix}\otimes E_2,$$

$$C_k(\lambda)=\begin{pmatrix}\frac{1}{\sqrt{1+\lambda^2}} & 0 & \frac{\lambda}{\sqrt{1+\lambda^2}}\\ 0 & E_{k-1} & 0\\ \frac{\lambda}{\sqrt{1+\lambda^2}} & 0 & \frac{-1}{\sqrt{1+\lambda^2}}\end{pmatrix}\otimes E_2 \quad (k\geqslant 2);$$

$$\Delta(1,k;\lambda)=\operatorname{diag}[C_k(\lambda),E_{n-2(k+1)}],$$

if $2(k+1)<n$,

$$\Delta(1,k;\lambda)=C_k(\lambda), \quad \text{if } 2(k+1)=n;$$

$$\Delta(q,k;\lambda)=\operatorname{diag}[E_{2q-2},C_k(\lambda),E_{n-2(k+q)}],$$

if $q>1$, $2(k+q)<n$,

$$\Delta(q,k;\lambda)=\operatorname{diag}[E_{2q-2},C_k(\lambda)],$$

if $q>1$, $2(k+q)=n$;

and $\varphi(q,k;\lambda)$ is an $O(n)$ automorphism of the algebra $\mathrm{ASch}(n)$ which corresponds to a matrix $\Delta(q,k;\lambda)$. Then

$$\varphi(q,k;\lambda)(J(q,q+k))=J(q,q+k),$$

$$\varphi(q,k;\lambda)(G_{2q-1}+\lambda G_{2(q+k)-1})=\sqrt{1+\lambda^2}\cdot G_{2q-1},$$

$$\varphi(q,k;\lambda)(G_{2q}+\lambda G_{2(q+k)})=\sqrt{1+\lambda^2}G_{2q}.$$

*Proof:* We may restrict ourselves only to the case when $n=4$, $q=1$, $k=1$. Since

$$C_1(\lambda)\cdot\begin{pmatrix}X' & 0\\ 0 & X'\end{pmatrix}=\begin{pmatrix}X' & 0\\ 0 & X'\end{pmatrix}\cdot C_1(\lambda),$$

for every matrix $X'$ of degree 2, and

$$C_1(\lambda)\cdot\begin{pmatrix}y_1\\ y_2\\ \lambda y_1\\ \lambda y_2\end{pmatrix}=\sqrt{1+\lambda^2}\begin{pmatrix}y_1\\ y_2\\ 0\\ 0\end{pmatrix},$$

then

$$\varphi(1,1;\lambda)(J(1,2))=J(1,2),$$

$$\varphi(1,1;\lambda)(G_1+\lambda G_3)=\sqrt{1+\lambda^2}G_1,$$

$$\varphi(1,1;\lambda)(G_2+\lambda G_4)=\sqrt{1+\lambda^2}G_2.$$

The lemma is proved.

*Proposition 4.1:* Let $X=S+T+\alpha J(k,l)$ where $\alpha>0$, $\alpha\neq 1$. If $U$ is an $\langle X\rangle$ submodule of the first (the second) kind of the module $\overline{\mathfrak{M}}[2k-1,2l]$, then $U$ is conjugated to the module

$$\sum_{a=k}^{t}\mathfrak{L}_a\Big(\sum_{a=k}^{t}\mathfrak{N}_a\Big) \quad (t\leqslant l).$$

*Proof:* Let us assume that $U$ is a module of the first kind. By Lemma 4.1 we shall suppose that a projection of $U$ onto $\mathfrak{L}_k$ differs from 0. As

$$\exp(\theta J_{2a-1,2a})(\gamma_a Y_{2a-1}+\delta_a Y_{2a})\exp(-\theta J_{2a-1,2a})$$
$$=(\gamma_a\cos\theta+\delta_a\sin\theta)Y_{2a-1}$$
$$+(\delta_a\cos\theta-\gamma_a\sin\theta)Y_{2a},$$

putting $\delta_a\cos\theta-\gamma_a\sin\theta=0$, we may assume that if a projection of an element $Y\in U$ onto $\mathfrak{L}_a$ is equal to $\gamma_a Y_{2a-1}+\delta_a Y_{2a}$, then $\delta_a=0$. Hence it follows that $U$ has the element

$$Y=Y_{2k-1}+\lambda_{k+1}Y_{2k+1}+\cdots+\lambda_l Y_{2l-1}$$
$$=(G_{2k-1}+\lambda_{k+1}G_{2k+1}+\cdots+\lambda_l G_{2l-1})$$
$$+(P_{2k}+\lambda_{k+1}P_{2k+2}+\cdots+\lambda_l P_{2l}).$$

In view of Lemma 4.2, for some $O(n)$ automorphism $\varphi=\varphi(k,1;\mu_1)\cdot\varphi(k,2;\mu_2)\cdot\cdots\cdot\varphi(k,l-k;\mu_{l-k})$ of the algebra $\mathrm{ASch}(n)$ the following equalities hold true: $\varphi(X)=X$, $\varphi(Y)=\gamma(G_{2k-1}+P_{2k})$ ($\gamma\in R$, $\gamma\neq 0$). Therefore we may assume that $Y_{2k-1}\in U$. Then $Y_{2k}\in U$, and thus $\mathfrak{L}_k\subset U$. Using induction we conclude that $U=\Sigma\mathfrak{L}_a$.

The case when $U$ is a module of the second kind is treated similarly. The proposition is proved.

**Theorem 4.1:** Let $F$ be a subalgebra of the algebra $\mathrm{AO}(n)\oplus\mathrm{ASL}(2,R)$. Then $F$ has only splitting extensions in $\mathrm{ASch}(n)$ if and only if one of the following conditions is satisfied: (1) $D\in\tau(F)$; (2) $\tau(F)=\langle S+T\rangle$ and $F$ is not conjugated to $\langle J_{12}+S+T\rangle+K$, where $K$ is a subalgebra of the algebra $\langle J_{ab}\,|\,a,b=3,\ldots,n\rangle$; (3) $\tau(F)\subset\langle T\rangle$ and $\omega(F)$ is not conjugated to any subalgebra of the algebra $\mathrm{AO}(n-1)$; or (4) $\tau(F)=0$ and $\omega(F)$ is a semisimple algebra.

*Proof:* Let $D\in\tau(F)$. If $\tau(F)=\mathrm{ASL}(2,R)$, then by Theorem 3.1 $F$ is a completely reducible algebra. Since in this case $F$ annuls only zero subspace in $\overline{\mathfrak{M}}[1,n]$, then by Proposition 2.1 of Ref. 9 the algebra $F$ has only splitting extensions in $\mathrm{ASch}(n)$. If $\tau(F)=\langle D,T\rangle$, then $T\in F$. Algebra $F/\langle T\rangle$ acts completely reducible in $\overline{\mathfrak{M}}[1,n]$ and annuls only zero subspace in this space. From this, using Proposition 2.1 and Lemma 3.1,[9] we conclude that $F$ has only splitting extensions in $\mathrm{ASch}(n)$. At the same time the case $\tau(F)=\langle D\rangle$ is considered.

Let $\tau(F)=\langle S+T\rangle$. If $S+T\in F$, then $F$ annuls only zero subspace in $\overline{\mathfrak{M}}[1,n]$. Because of Theorem 3.1 the algebra $F$ is completely reducible; then by Proposition 2.1 of Ref. 9 any algebra $\hat{F}$ such that $\pi(\hat{F})=F$ is splitting. If $F$ has a nonsplitting extension $\hat{F}$ in the algebra $\mathrm{ASch}(n)$, then $F$ con-

tains $X = S + T + \alpha_1 J_{12} + \cdots + \alpha_t J_{2t-1,2t}$, where $0 < \alpha_1 \leqslant \cdots \leqslant \alpha_t$. We may assume that projections of other basis elements of the algebra $F$ onto $\langle S + T \rangle$ are equal to 0. In view of Proposition 2.1 of Ref. 9 the algebra $F$ annuls in $\overline{\mathfrak{M}}[1,n]$ a certain nonzero subspace $U$. It follows from this and formula (2) that $U \subset \langle Y_1, Y_2, \ldots, Y_{2k} \rangle$ and $X = S + T + J(1,t)$ $(k \leqslant t)$ or

$$X = S + T + J(1,k) + \beta_{k+1} J(k+1) + \cdots + \beta_t J(t) \quad (t > k),$$

where $\beta_{k+1} > 0, \ldots, \beta_t > 0$, $\beta_{k+1} \neq 1, \ldots, \beta_t \neq 1$. Arguing as in the proof of Proposition 4.1 we obtain that $Y_1 \in U$ up to conjugacy. Hence it follows that $F = \langle S + T + J_{12} \rangle \dotplus K$, where $K \subset \langle J_{ab} \mid a,b = 3,\ldots,n \rangle$. By lemma 2.1 of Ref. 9 the algebra $\hat{F}$, which is obtained from $F$ by replacing $S + T + J_{12}$ by $S + T + J_{12} + Y_1$, is nonsplitting.

Let $\tau(F) = \langle T \rangle$, $F_1 = \omega(F)$, and $\hat{F}$ be a subalgebra of the algebra $\mathrm{ASch}(n)$ such that $\pi(\hat{F}) = F$. If $F_1$ is not conjugated to a subalgebra of the algebra $\mathrm{AO}(n-1)$, then by Proposition 2.1 and Lemma 3.1 of Ref. 9 an algebra $\hat{F}$ is splitting. If $F_1$ is conjugated to a subalgebra of the algebra $\mathrm{AO}(n-1)$, then $F = \langle X \rangle \oplus F_2$, where $X \neq 0$, and $\langle X \rangle$ and $F_2$ are subalgebras of the algebra $\mathrm{AO}(n-1) \oplus \langle T \rangle$. An algebra $F_2 \dotplus \langle P_n X + G_n \rangle$ is nonsplitting.

The case $\tau(F) = 0$ is considered in Refs. 5 and 7. The theorem is proved.

*Proposition 4.2:* A subalgebra $F$ of the algebra $\mathrm{AO}(n) \oplus \mathrm{ASL}(2,R)$ possesses only splitable extensions in $\widetilde{\mathrm{ASch}}(n)$ if and only if $F$ is a semisimple algebra.

The proof of Proposition 4.2 is similar to the proof of Theorem 4.1.

Let $\Gamma: X \to X$ be the trivial representation of a subalgebra $F$ of the algebra $\mathrm{AO}(n)$. Then $\Gamma$ is $\mathrm{O}(n)$ equivalent to $\mathrm{diag}[\Gamma_1, \ldots, \Gamma_m]$, where $\Gamma_i$ is an irreducible subrepresentation $(i = 1,\ldots,m)$. It is well known that if representations $\Delta$ and $\Delta'$ of Lie algebra $L$ by skew-symmetric matrices are equivalent over $R$, then $C\Delta(X)C^{-1} = \Delta'(X)$ for some orthogonal matrix $C$ $(X \in L)$, hence we conclude that if $\Gamma_i$ and $\Gamma_j$ are equivalent representations, then we can assume that for every $X \in F$ the equality $\Gamma_i(X) = \Gamma_j(X)$ takes place. Uniting equivalent nonzero irreducible subrepresentations we shall get nonzero disjunctive primary subrepresentations $\Delta_1, \ldots, \Delta_q$ of the representation $\Gamma$. An algebra

$$K_i = \{\mathrm{diag}[0,\ldots,\Delta_i(X),\ldots,0] \mid X \in F\} \quad (1 \leqslant i \leqslant q)$$

is called a primary part of the algebra $F$. Evidently $F$ is a subdirect sum of its primary parts.

We shall say that the splitting subalgebra $\hat{F}$ of the algebra $\mathrm{ASch}(n)$ or of the algebra $\widetilde{\mathrm{ASch}}(n)$ has a splitting factor algebra in the case $\pi(\hat{F}) = F_1 \oplus F_2$, where $F_1 \subset \mathrm{AO}(n)$, $F_2 \subset \mathrm{ASL}(2,R)$. If this condition does not hold, then the factor algebra $\pi(\hat{F})$ of an algebra $\hat{F}$ is called nonsplitting.

**Theorem 4.2**: Let $K_1, K_2, \ldots, K_q$ be primary parts of the nonzero subalgebra $L'$ of the algebra $\mathrm{AO}(n)$, $L''$ be a subalgebra of the algebra $\mathrm{ASL}(2,R)$ differing from $\langle S + T \rangle$, and $L$ be a subdirect sum of $L'$ and $L''$. If $U$ is a subspace of $\overline{\mathfrak{M}}[1,n]$, being invariant under $L$, then $U = U_1 \oplus \cdots \oplus U_q \oplus \tilde{U}$, where $U_i = [K_i, U] = [K_j, U_i]; [L'', U_i] \subset U_i$; $[K_j, U_i] = 0$ in the case $j \neq i$; $[K_i, \tilde{U}] = 0$, $[L'', \tilde{U}] \subset \tilde{U}$ $(i,j = 1,\ldots,q)$.

*Proof:* If $L'' = \mathrm{ASL}(2,R)$, then $L'' \subset L$. Therefore from $[L,U] \subset U$ it follows that $[L'', U] \subset U$. Since $\overline{\mathfrak{M}}[a]$ is invariant under $\mathrm{ASL}(2,R)$ for any $a$, $1 \leqslant a \leqslant n$, then each subspace $U_i = [K_i, U]$ is invariant under this algebra. Let $\tilde{U}$ be a maximal subspace of the space $U$ annulled by $L'$, $U' = [L', U]$. Since $L'$ is a completely reducible algebra, $U = U' \oplus \tilde{U}$ and $[L', U'] = U'$. Applying Lemma 3.1 of Ref. 9 we conclude that $U' = U_1 \oplus \cdots \oplus U_q$, where $U_i = [K_i, U] = [K_i, U_i]$ $(i = 1,\ldots,q)$.

Let $L'' = \langle T, D \rangle$. Since $\langle T, D \rangle$ is a non-Abelian solvable algebra and every subalgebra of the algebra $\mathrm{AO}(n)$ is reductive, then applying the Goursat twist method[11] we obtain that $T \in L$. Therefore it is enough to consider the case $L'' = \langle D \rangle$. By Lemma 4.2 of Ref. 9, $[D,U] \subset U$, it follows that $[D, U_i] \subset U_i$, $[D, \tilde{U}] \subset \tilde{U}$ $(i = 1,\ldots,q)$.

The case $L'' = \langle T \rangle$ is considered in Refs. 5 and 7. The theorem is proved.

Because of Theorem 4.2, the study of splitting subalgebras $\hat{F}$ of the algebra $\mathrm{ASch}(n)$, for which $\tau(\hat{F}) \neq \langle S + T \rangle$, is reduced to the study of splitting subalgebras $\hat{K}$ of the algebra $\mathrm{ASch}(n)$ having the splitting factor algebra $\pi(\hat{K})$ and zero or primary projection onto $\mathrm{AO}(n)$. Such subalgebras have been described in Ref. 13.

*Proposition 4.3:* Nonzero subspaces of the space $\overline{\mathfrak{M}}[1,n]$ invariant under $\langle S + T \rangle$ are exhausted with respect to $\mathrm{O}(n)$ conjugation by the following spaces: $\overline{\mathfrak{M}}[1,d]$ $(d = 1,\ldots,n)$; $U_q$ $(q = 1,\ldots,[n/2])$, $U_m + \overline{\mathfrak{M}}[2m+1,t]$ $(m = 1,\ldots,[(n-1)/2]; t = 2m+1,\ldots,n)$, where $U_q$ is a subdirect sum of $V[1,2q]$ and $W[1,2q]$ having zero intersections with these spaces. If

$$\{G_j + \alpha_{1j} P_1 + \cdots + \alpha_{2qj} P_{2q} \mid j = 1,\ldots,2q\}$$

is a basis of $U_q$, then with respect to $\mathrm{O}(2q)$ conjugation a matrix $(\alpha_{kj})$ $(k,j = 1,\ldots,2q)$ coincides with $\mathrm{diag}[\Gamma(\lambda_1),\ldots,\Gamma(\lambda_q)]$, where $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_q \leqslant 1$ and

$$\Gamma(\lambda) = \begin{pmatrix} 0 & \lambda \\ -\lambda^{-1} & 0 \end{pmatrix}.$$

The numbers $\lambda_1, \ldots, \lambda_q$ are defined by the space $U_q$ uniquely.

Proposition 4.3 is proved along with Proposition 2.4 and Theorem 3.4 in Ref. 13.

*Proposition 4.4:* Let

$$\Lambda_b(a) = \langle P_1 + \lambda_1 P_{a+1}, \ldots, P_b + \lambda_b P_{a+b} \rangle,$$

where $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_b$, $b \leqslant a$, $a + b \leqslant n$. A subalgebra $\hat{F}$ of the algebra $\mathrm{ASch}(n)$ such that $\omega(\hat{F}) = 0$, $D \in \tau(\hat{F})$, is $\mathrm{Sch}(n)$ conjugated to $L \dotplus U$, where $L \subset \mathrm{ASL}(2,R)$ and $U$ is a subspace of the space $\overline{\mathfrak{M}}[1,n]$. Let $U \neq 0$. If $L = \mathrm{ASL}(2,R)$, then $U$ is conjugated to $\overline{\mathfrak{M}}[1,d]$ $(1 \leqslant d \leqslant n)$. If $L = \langle D, T \rangle$, then $U$ is conjugated to one of the following spaces: $W[1,d]$, $\overline{\mathfrak{M}}[1,d]$, $(1 \leqslant d \leqslant n)$; $V[1,d] + W[1,t]$ $(1 \leqslant d \leqslant n-1, d+1 \leqslant t \leqslant n)$. If $L = \langle D \rangle$, then $U$ is conjugated to one of the following spaces:

$W[1,d]$, $\overline{\mathfrak{M}}[1,d]$ $(1\leqslant d\leqslant n)$; $V[1,d]+W[d+1,d+t]$ $(1\leqslant d\leqslant[n/2];\ d\leqslant t\leqslant n-d)$;

$V[1,d]+W[1,c]+W[d+1,d+t]$

$(1\leqslant d\leqslant n-1;\ 1\leqslant c\leqslant d;\ d-c\leqslant t\leqslant n-d,$ if $c\neq d;\ 1\leqslant t\leqslant n-d,$ if $c=d)$;

$V[1,d]+\Lambda_d(d)$ $(1\leqslant d\,[n/2])$;

$V[1,d]+\Lambda_t(d)+W[t+1,d]$ $(2\leqslant d\leqslant n-1;\ 1\leqslant t\leqslant\min\{d-1,n-d\})$;

$V[1,d]+\Lambda_t(d)+W[d+t+1,d+t+s]$

$(1\leqslant d\leqslant[n/2];\ 1\leqslant t\leqslant\min\{d,n-d-1\};\ 1\leqslant s\leqslant n-d-t;\ s+t\geqslant d)$;

$V[1,d]+\Lambda_t(d)+W[t+1,b]+W[d+t+1,d+t+s]$

$(2\leqslant d\leqslant n-2;\ 1\leqslant t\leqslant\min\{d-1,n-d-1\};\ t+1\leqslant b\leqslant d;\ 1\leqslant s\leqslant n-d-t;\ b+s\geqslant d)$.

The proof of Proposition 4.4 is similar to the proof of Theorem 3.3.[13]

Using Theorem 4.2 to investigte splitting subalgebras with nonsplitting factor algebra, it is enough to consider the algebras $\hat{F}\subset\mathrm{ASch}(n)$ for which $\tau(\hat{F})=\langle S+T\rangle$ and $\tau(\hat{F})\not\subset\hat{F}$. In this case $\pi(\hat{F})=F'\oplus\langle X\rangle$, where $F'$ is a subalgebra of $\mathrm{AO}(n)$ and $X=S+T+\alpha_1J_{12}+\cdots+\alpha_kJ_{2k-1,2k}$. We may suppose that $0<\alpha_1\leqslant\cdots\leqslant\alpha_k$. Henceforth we shall discuss subspaces of the space $\overline{\mathfrak{M}}[1,n]$ that are invariant under $X$.

*Lemma 4.3:* Let $1\leqslant a,b\leqslant k$. Then $\mathfrak{L}_a\cong\mathfrak{L}_b$ if and only if $\alpha_a=\alpha_b$ or $\alpha_a+\alpha_b=2$; $\mathfrak{N}_a\cong\mathfrak{N}_b$ if and only if $\alpha_a=\alpha_b$; $\mathfrak{L}_a\cong\mathfrak{N}_b$ if and only if $\alpha_a=2+\alpha_b$ $(a\neq b)$. Modules $\mathfrak{L}_a$ and $\mathfrak{N}_a$ are not isomorphic.

*Proof:* The matrices $\lambda J$, $\mu J$ are similar if and only if $\lambda^2=\mu^2$. It follows that $\mathfrak{L}_a\cong\mathfrak{L}_b$ if and only if $(\alpha_a-1)^2=(\alpha_b-1)^2$. In the case $\alpha_a-\alpha_b\neq0$, $\alpha_a+\alpha_b=2$.

If $\mathfrak{N}_a\cong\mathfrak{N}_b$, then $(\alpha_a+1)^2=(\alpha_b+1)^2$, whence $2(\alpha_a-\alpha_b)=-(\alpha_a-\alpha_b)(\alpha_a+\alpha_b)$. In the case $\alpha_a-\alpha_b\neq0$, $2=-(\alpha_a+\alpha_b)$. But this contradicts the fact that $\alpha_a,\alpha_b>0$.

Let $\mathfrak{L}_a\cong\mathfrak{N}_b$. Then $(\alpha_a-1)^2=(\alpha_b+1)^2$, whence $2(\alpha_a+\alpha_b)=(\alpha_a-\alpha_b)(\alpha_a+\alpha_b)$. Thus if $a\neq b$, then $\alpha_a-\alpha_b=2$. The lemma is proved.

Let us remark that if $\alpha_a\neq1$, then the $\langle X\rangle$ modules $\mathfrak{L}_a$ and $\mathfrak{N}_a$ are irreducible, and any $\langle X\rangle$ submodule of the module $\overline{\mathfrak{M}}[1,n]$ is completely reducible.

*Proposition 4.5:* Let

$$X=S+T+\sum_{i=1}^{s}\beta_iJ(k_{i-1}+1,k_i),$$

where $s\geqslant2$, $k_0=0$, $\beta_i>0$, $\beta_i\neq1$, $\beta_i\neq\beta_j$ if $i\neq j$. There exists an indecomposable $\langle X\rangle$ submodule with nonzero projections onto $\overline{\mathfrak{M}}[1,2k_1]$, $\overline{\mathfrak{M}}[2k_1+1,\ 2k_2]$,..., $\overline{\mathfrak{M}}[2k_{s-1}+1,2k_s]$ of the $\langle X\rangle$ module $\overline{\mathfrak{M}}[1,2k_s]$ if and only if $s=2$ and one of the following conditions is satisfied: (i) $\beta_1=2+\beta_2$; (ii) $\beta_2=2+\beta_1$; (iii) $\beta_1+\beta_2=2$. If $U$ is a demanded indecomposable $\langle X\rangle$ module and $U_i$ is the projection of $U$ onto $\overline{\mathfrak{M}}[2k_{i-1}+1,2k_i]$ $(i=1,2)$, then in case (i) $U_1$ is a module of the first kind and $U_2$ is a module of the second kind; in the case (ii) $U_1$ is a module of the second kind and $U_2$ is a module of the first kind; and in case (iii) $U_1$ and $U_2$ are modules of the first kind.

*Proof:* By Lemma 3.1 of Ref. 9, in the $\langle X\rangle$ module $\overline{\mathfrak{M}}[1,2k_s]$ there exists an indecomposable submodule demanded if and only if the $\langle X\rangle$ modules $\overline{\mathfrak{M}}[1,2k_1]$, $\overline{\mathfrak{M}}[2k_1+1,\ 2k_2]$,..., $\overline{\mathfrak{M}}[2k_{s-1}+1,2k_s]$ have isomorphic composition factors. If $\mathfrak{L}_{k_i}\cong\mathfrak{L}_{k_j}$, *and* $\mathfrak{L}_{k_j}\cong\mathfrak{L}_{k_r}$, then by Lemma 4.3 $\beta_i+\beta_j=2$ and $\beta_j+\beta_r=2$. From this it follows that $\beta_i=\beta_r$, and that is why $i=r$. If $\mathfrak{L}_{k_i}\cong\mathfrak{N}_{k_j}$ and $\mathfrak{N}_{k_j}\cong\mathfrak{N}_{k_r}$, then $\beta_i=2+\beta_j$ and $\beta_r=2+\beta_j$, whence $i=r$. Thus $s\leqslant2$ and one of the following conditions is satisfied: (1) $\beta_1=2+\beta_2$; (2) $\beta_2=2+\beta_1$; (3) $0<\beta_1<2,\beta_2=2-\beta_1$. Statements about the kinds of projections follow from Lemma 4.3. The proposition is proved.

*Proposition 4.6:* Let $X=S+T+\beta J(1,k)$ $(\beta>0)$. In the $\langle X\rangle$ module $\overline{\mathfrak{M}}[1,n]$ there exists an indecomposable $\langle X\rangle$ submodule with nonzero projections onto $\overline{\mathfrak{M}}[1,2k]$ and $\overline{\mathfrak{M}}[2k+1,n]$ if and only if $\beta=2$. If $U$ is such a submodule and $U_1$ is the projection $U$ onto $\overline{\mathfrak{M}}[1,2k]$, then $U_1$ is a module of the first kind.

## V. ABELIAN SUBALGEBRAS OF THE EXTENDED SCHRÖDINGER ALGEBRA

The main results of this section are Theorem 5.1 and its two corollaries.

Let us use the following notation:

$$X_t=\alpha_1J_{12}+\alpha_2J_{34}+\cdots+\alpha_tJ_{2t-1,2t},$$

where $\alpha_1=1$, $0<\alpha_2\leqslant\cdots\leqslant\alpha_t\leqslant1$ if $t\geqslant2$;

$\mathrm{AH}(0)=0$,

$\mathrm{AH}(2d)=\mathrm{AH}(2d+1)=\langle J_{12},J_{34},\ldots,J_{2d-1,2d}\rangle$;

$\Delta_0[r,t]=\langle G_r+\alpha_rP_r,\ldots,G_t+\alpha_tP_t\rangle$,

$\Delta[r,t]=\Delta_0[r,t]+\langle M\rangle$,

where $r\leqslant t$, $\alpha_r\leqslant\cdots\leqslant\alpha_t$, $\alpha_r=0$, and $\alpha_t=1$ if $\alpha_t\neq0$;

$\Pi(a,b)=\langle Y_{2a-1},Y_{2a+1},\ldots,Y_{2b-1}\rangle$ $(a\leqslant b)$.

We recall that $Y_{2c-1}=G_{2c-1}+P_{2c}$ and $Y_{2c}=G_{2c}-P_{2c-1}$.

The algebra $\mathrm{AH}(n)$ is a maximal Abelian subalgebra of the algebra $\mathrm{AO}(n)$. It is well known that any maximal Abelian subalgebra of the algebra $\mathrm{AO}(n)$ is conjugated $\mathrm{AH}(n)$ with respect to inner automorphisms of the algebra $\mathrm{AO}(n)$. Henceforth when speaking about Abelian subalgebras of the algebra $\mathrm{AO}(n)$ we shall mean subalgebras of the algebra $\mathrm{AH}(n)$.

*Lemma 5.1:* Let $L$ be an Abelian subalgebra of the algebra $\langle J(a,b)+S+T\rangle\oplus\mathfrak{M}[2a-1,2b]$ such that its projection onto $\langle J(a,b)+S+T\rangle$ is nonzero and its projection

onto $\langle M\rangle$ is equal to 0. Then $L$ is conjugated to one of the following algebras:

$$\langle J(a,b)+S+T+\alpha Y_{2b-1}\rangle \quad (\alpha\geqslant 0);$$

$$\Pi(a,c)\oplus\langle J(a,b)+S+T+\alpha Y_{2b-1}\rangle \quad (\alpha\geqslant 0,\ c\leqslant b).$$

The written algebras are pairwise nonconjugated.

*Proof:* The maximal subspace of the space $\mathfrak{M}[2a-1,2b]$ anulled by $\langle J(a,b)+S+T\rangle$ and having zero projection onto $\langle M\rangle$ coincides with

$$\sum_{c=a}^{b}\mathfrak{L}_c.$$

Let $U=L\cap\mathfrak{M}[2a-1,2b]$. By the same arguments as in the proof of Proposition 4.1 we can establish that if $U\neq 0$, then $U$ contains $Y_{2a-1}$. As $[Y_{2a-1},Y_{2a}]=-2M$, so $U=\langle Y_{2a-1}\rangle+U^1$, where $U^1$ is a subspace of the space

$$\sum_{c=a+1}^{b}\mathfrak{L}_c.$$

Continuing these arguments we obtain that $U=\Pi(a,c)$ $(c\leqslant b)$ and $L$ contains $J(a,b)+S+T+\alpha Y_{2b-1}$ $(\alpha\geqslant 0)$. The lemma is proved.

**Theorem 5.1:** Let $L$ be a nonzero Abelian subalgebra of the algebra $\mathrm{ASch}(n)$. If $\tau(L)=\langle D\rangle$, then $L$ is conjugated to the subdirect sum $L_1\dotplus L_2\dotplus L_3$ of algebras $L_1, L_2, L_3$, where $L_1\subset\mathrm{AH}(2d)$, $L_2=\langle D\rangle$, $L_3\subset\langle M\rangle$ $(0\leqslant d\leqslant[n/2]$. If $\tau(L)=\langle T\rangle$, then $L$ is conjugated to $L_1\dotplus L_2\dotplus L_3\dotplus L_4$, where $L_1\subset\mathrm{AH}(2d)$, $L_2=\langle T+\alpha G_{2d+1}\rangle$ $(\alpha\in\{0,1\})$, $L_3=0$ or $L_3=W[r,t]$, $L_4\subset\langle M\rangle$ $(0\leqslant d\leqslant[n/2]; r=2d+1$ if $\alpha=0$, $2d+1\leqslant n$; $r=2d+2$ if $\alpha=1$, $2d+2\leqslant n)$. If $\tau(L)=\langle S+T\rangle$, then $L$ is conjugated to $L_1\dotplus L_2\dotplus L_3\dotplus L_4$, where $L_1\subset\langle M\rangle$, $L_2\subset\mathrm{AH}(2d)$ $(0\leqslant d\leqslant[n/2])$, and the algebras $L_3$ and $L_4$ satisfy one of the following conditions:

(1) $L_3=\langle S+T\rangle$, $L_4=0$;

(2) $L_3=\langle J(d+1,t)+S+T+\alpha Y_{2t-1}\rangle$, $L_4=0$ $(\alpha>0)$;

(3) $L_3=\langle J(d+1,t)+S+T+\alpha Y_{2t-1}\rangle$, $L_4=\Pi(d+1,s)$ $(s\leqslant t;\ \alpha\geqslant 0)$.

If $L\subset\tilde{\mathrm{AG}}^0(n)$, then $L$ is conjugated to $L_1\dotplus L_2\dotplus L_3\dotplus L_4$, where $L_1\subset\mathrm{AH}(2d)$, $L_2=0$ or $L_2=\Delta_0[2d+1,s]$, $L_3=0$ or $L_3=W[k,l]$, $L_4=0$ or $L_4=\langle M\rangle$ $(0\leqslant d\leqslant[n/2]$; $k=s+1$ if $L_2\neq 0$; $k=2d+1$ if $L_2=0$; $l\leqslant n)$.

*Proof:* If $\tau(L)=\langle D\rangle$, then by Theorem 4.1 the algebra $L$ is conjugated to the algebra $U+F$, where $U\subset\mathfrak{M}[1,n]$ and $F\subset\mathrm{AH}(n)\oplus\langle D,M\rangle$. Since $D$ annuls only $\langle M\rangle$ in $\mathfrak{M}[1,n]$ and by Theorem 4.2 $[D,U]\subset U$, then $U\subset\langle M\rangle$. Thus $L$ is conjugated to some subalgebra of the algebra $\mathrm{AH}(2d)\oplus\langle D,M\rangle$ $(0\leqslant d\leqslant[n/2])$.

If $\tau(L)=\langle T\rangle$, then by Theorem 4.1 the algebra $L$ is conjugated to the algebra $U+F$ satisfying one of the following conditions: $U\subset\mathfrak{M}[1,n]$ and $F$ is a subalgebra of $\mathrm{AH}(n)+\langle M,T\rangle$; or $U\subset\mathfrak{M}[1,2d]$ and $F$ is a subalgebra of $\mathrm{AH}(2d)+\mathfrak{M}[2d+1,n]+\langle T\rangle$ $(d\geqslant 1)$. Let us consider the last case. Let us suppose that the projection $K$ of the algebra $F$ onto $\mathrm{AO}(n)$ is not conjugated to any subalgebra of the algebra $\mathrm{AH}(2d-2)$. Since $K$ annuls only the zero subspace of $V[1,2d]$, then $U\subset\langle M\rangle$. Therefore we shall assume that $U=0$. As $[T,G_a]=-P_a$, so by Witt's mapping theorem[14] $\epsilon(F)=0$, or $\epsilon(F)=\langle G_{2d+1}\rangle$. Since

$$\exp(\theta T)(T+\alpha G_{2d+1}+\beta P_{2d+1})\exp(-\theta T)$$
$$=T+\alpha G_{2d+1}+(\beta-\theta\alpha)P_{2d+1}$$

and

$$\exp(\lambda D)(T+\alpha G_{2d+1})\exp(-\lambda D)$$
$$=\exp(-2\lambda)(T+\alpha\exp(3\lambda)\cdot G_{2d+1}),$$

then if $\epsilon(F)\neq 0$, the projection of $F$ onto $\langle T\rangle\oplus\mathfrak{M}[2d+1,n]$ contains $T+G_{2d+1}$. In this case, by Witt's theorem $\xi(F)$ coincides with 0 or $W[2d+2,t]$. If $\epsilon(F)=0$, then $\xi(F)=0$ or $\xi(F)=W[2d+1,t]$.

Let $\tau(L)=\langle S+T\rangle$. If $S+T\in L$, then $\epsilon(L)=0$ and $\xi(L)=0$. If $S+T\notin L$, then an algebra $L$ contains

$$Y=S+T+\sum_{a=1}^{[n/2]}\alpha_a J_{2a-1,2a}+\sum_{i=1}^{n}(\beta_i G_i+\gamma_i P_i)+\delta M.$$

We shall suppose that projections of the at rest basis elements of the algebra $L$ onto $\langle S+T\rangle$ are equal to zero, and $\alpha_a\geqslant 0$ for all $a$. If $\alpha_c\neq 1$, then $\langle S+T+\alpha_c J_{2c-1,2c}\rangle$ is a completely reducible algebra of linear transformations of the vector space $\overline{\mathfrak{M}}[2c-1,2c]$ and annuls only the zero subspace of this space, whence by Proposition 2.1 of Ref. 9 we conclude that the projection of $L$ onto $\overline{\mathfrak{M}}[2c-1,2c]$ is equal to zero. Therefore we may assume that

$$Y=J(d+1,t)+S+T+\sum_{i=2d+1}^{2t}(\beta_i G_i+\gamma_i P_i).$$

From Proposition 2.1 of Ref. 9 it also follows that

$$\sum_{i=2d+1}^{2t}(\beta_i G_i+\gamma_i P_i)\in\sum_{j=d+1}^{t}\mathfrak{L}_j.$$

Applying Theorem 4.1 and Lemma 5.1 we conclude that, with respect to the conjugation $\omega(L)\subset\mathrm{AH}(2d)+J(d+1,t)\rangle$,

$$Y=J(d+1,t)+S+T+\alpha Y_{2t-1},$$

and $L\cap\overline{\mathfrak{M}}[1,n]=0$ or $L\cap\overline{\mathfrak{M}}[1,n]=\Pi(d+1,s)$ $(\alpha\geqslant 0;$ $s\leqslant t)$.

Let us assume that $L\subset\tilde{\mathrm{AG}}^0(n)$. By Theorem 2 of Ref. 7 the algebra $L$ is conjugated to an algebra $U+F$, which satisfies one of the following conditions: $U\subset\mathfrak{M}[1,n]$ and $F$ is a subalgebra of $\mathrm{AH}(n)+\langle M\rangle$; or $U\subset\mathfrak{M}[1,2d]$ and $F$ is a subalgebra of $\mathrm{AH}(2d)+\mathfrak{M}[2d+1,n]$ $(1\leqslant d\leqslant[n-1/2])$. Let us restrict ourselves to the last case. Let the projection $K$ of the algebra $F$ onto $\mathrm{AO}(n)$ be not conjugated to any subalgebra of the algebra $\mathrm{AH}(2d-2)$. Since $K$ annuls only the zero subspace of the space $\overline{\mathfrak{M}}[1,2d]$, $U\subset\langle M\rangle$. Therefore we suppose that $U=0$.

Let $N$ be the projection of $F$ onto $\mathfrak{M}[2d+1,n]$ and $\epsilon(N)=V[2d+1,2d+q]$. By Witt's mapping theorem[14] the algebra $N$ is a subdirect sum of the algebras $N_1, N_2, N_3$, where $N_1\subset\overline{\mathfrak{M}}[2d+1,2d+q]$ (as a space), $N_2=0$ or $N_2=W[2d+q+1,t]$, and $N_3\subset\langle M\rangle$. Let

$Z_i = G_i + \beta_{2d+1,i}P_{2d+1} + \cdots + \beta_{2d+q,i}P_{2d+q}$ $(i = (2d+1),...,(2d+q))$,

$N_1 = \langle Z_{2d+1},...,Z_{2d+q}\rangle$. Evidently $[Z_i,Z_j] = (\beta_{ij} - \beta_{ji})M$. Since $N_1$ is an Abelian algebra, $\beta_{ij} = \beta_{ji}$. Hence it follows that the matrix $B = (\beta_{ij})$ is symmetric. Therefore there exists a matrix $Q \in O(q)$ such that $QBQ^{-1} = \operatorname{diag}[\lambda_1,...,\lambda_q]$. From this it follows that with respect to automorphisms from the group $O(2d)\times O(q)\times O(n-2d-q)$ we may assume that $Z_{2d+j} = G_{2d+j} + \lambda_j P_{2d+j}$ $(j = 1,...,q)$, where $\lambda_1 \leqslant \cdots \leqslant \lambda_q$. Applying the automorphism $\exp(\lambda_1 T)$ we obtain the generators $G_{2d+j} + \mu_j P_{2d+j}$ $(j = 1,...,q)$, where $\mu_1 = 0$, $0 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_q$. If $\mu_q > 0$, then $\mu_q = \exp(-2\theta)$. Obviously

$$\exp(\theta D)(G_{2d+j} + \mu_j P_{2d+j})\exp(-\theta D) = \exp\theta(G_{2d+j} + \mu_j \exp(-2\theta)P_{2d+j}).$$

Therefore if $\mu_q > 0$, we may suppose that $\mu_q = 1$. This proves that the algebra $N_1$ is conjugated to $\Delta_0[2d+1,2d+q]$. The theorem is proved.

*Corollary 1:* The maximal Abelian subalgebras of the algebra $\widetilde{\mathrm{ASch}}(n)$ are exhausted with respect to the $\widetilde{\mathrm{Sch}}(n)$ conjugation by the following algebras:

$\mathrm{AH}(n) \oplus \langle T,M\rangle$ $(n \equiv 0 \pmod 2)$; $\mathrm{AH}(n) \oplus \langle S+T,M\rangle$;

$\mathrm{AH}(n) \oplus \langle D,M\rangle$; $\mathrm{AH}(n-1) \oplus \langle G_n + T,M\rangle$ $(n \equiv 1 \pmod 2)$;

$\mathrm{AH}(2d) \oplus \Delta[2d+1,n]$ $(d = 0,1,...,[(n-1)/2])$;

$\mathrm{AH}(2d) \oplus \Delta[2d+1,t] \oplus W[t+1,n]$ $(d = 0,1,...,[(n-2)/2]; t = 2d+1,...,n-1)$;

$\mathrm{AH}(2d) \oplus \langle T,M\rangle \oplus W[2d+1,n]$ $(d = 0,1,...,[(n-1)/2])$;

$\mathrm{AH}(2d) \oplus \langle G_{2d+1} + T\rangle \oplus W[2d+2,n] \oplus \langle M\rangle$ $(d = 0,1,...,[(n-2)/2])$;

$\mathrm{AH}(2d) \oplus \langle J(d+1,r) + S + T\rangle \oplus \langle M\rangle \oplus \Pi(d+1,r)$ $(d = 0,1,...,[(n-2)/2]; r = d+1,...,[n/2])$.

*Corollary 2:* Let $\alpha, \beta \in R$, $\alpha > 0$, $\beta > 0$; $t = 1,...,[n/2]$; $s = 1,...,[(n-1)/2]$; $n \geqslant 3$. One-dimensional subalgebras of the algebra $\widetilde{\mathrm{ASch}}(n)$ are exhausted with respect to the $\widetilde{\mathrm{Sch}}(n)$ conjugation by the following algebras:

$\langle D\rangle$; $\langle T\rangle$; $\langle S+T\rangle$; $\langle M\rangle$; $\langle D+\alpha M\rangle$; $\langle T \pm M\rangle$;

$\langle S+T \pm \alpha M\rangle$; $\langle P_1\rangle$; $\langle G_1 + P_2\rangle$; $\langle G_1 + T\rangle$; $\langle X_t\rangle$; $\langle X_t + \alpha D\rangle$;

$\langle X_t + \alpha D + \beta M\rangle$; $\langle X_t + T\rangle$; $\langle X_t + T \pm \alpha M\rangle$; $\langle X_t + \alpha(S+T)\rangle$; $\langle X_t + \alpha M\rangle$;

$\langle X_t + \alpha(S+T) \pm \beta M\rangle$; $\langle X_s + P_{2s+1}\rangle$; $\langle X_r + G_{2r+1} + \alpha P_{2r+2}\rangle$ $(r = 1,...,[(n-2)/2])$;

$\langle X_s + T + \alpha G_{2s+1}\rangle$; $\langle X_t + S + T + \alpha(G_1 + P_2)\rangle$.

*Remark:* One-dimensional subalgebras of the algebra $\mathrm{ASch}(n)$ are exhausted with respect to the $\mathrm{Sch}(n)$ conjugation by one-dimensional subalgebras of the algebra $\widetilde{\mathrm{ASch}}(n)$ whose generators do not contain $\lambda M$ as an addend $(\lambda \neq 0)$.

**Theorem 5.2:** Let $L$ be a nonzero Abelian subalgebra of the algebra $\mathrm{ASch}(n)$. If $\tau(L) = \langle D\rangle$, then $L$ is conjugated to a subdirect sum of $\langle D\rangle$ and the subalgebra of the algebra $\mathrm{AH}(2d)$ $(0 \leqslant d \leqslant [n/2])$. If $\tau(L) = \langle T\rangle$, then $L$ is conjugated to $L_1 \dotplus L_2 \dotplus L_3$, where $L_1 \subset \mathrm{AH}(2d)$, $L_2 = \langle T + \alpha G_{2d+1}\rangle$, and $L_3$ is one of the following algebras:

$0$; $W[2d+2,t]$; $\langle P_{2d+1} + \lambda P_{2d+2}\rangle + \gamma W[2d+2] + \delta W[2d+3,t]$ $(0 \leqslant d \leqslant [n/2]; t \leqslant n; \alpha,\gamma,\delta \in \{0,1\}; \lambda \leqslant 0)$.

If $\tau(L) = \langle S+T\rangle$, then $L$ is conjugated to $L_1 \dotplus L_2 \dotplus L_3$, where $L_1 \subset \mathrm{AH}(2d)$ $(0 \leqslant d \leqslant [n/2])$ and the algebras $L_2$, $L_3$ satisfy one of the following conditions: (1) $L_2 = \langle S+T\rangle$ and $L_3 = 0$; or (2) $L_2 = \langle J(d+1,t) + S + T + \alpha Y_{2t-1}\rangle$ $(\alpha \geqslant 0)$ and $L_3$ is a subalgebra of the algebra

$$\sum_{a=d+1}^{t} \dotplus \mathfrak{L}_\alpha.$$

If $L \subset \mathrm{AG}^0(n)$, then $L$ is conjugated to $L_1 \dotplus L_2$, where $L_1 \subset \mathrm{AH}(2d)$ and $L_2 \subset \mathfrak{M}[2d+1,n]$ $(0 \leqslant d \leqslant [n/2])$.

The theorem is proved along the same lines as Theorem 5.1.

*Corollary:* The maximal Abelian subalgebras of the algebra $\mathrm{ASch}(n)$ are exhausted with respect to the $\mathrm{Sch}(n)$ conjugation by the following algebras:

$\mathrm{AH}(n) \oplus \langle D\rangle$; $\mathrm{AH}(n) \oplus \langle S+T\rangle$; $\mathrm{AH}(n) \oplus \langle T\rangle$ $[n \equiv 0 \pmod 2]$;

$\mathrm{AH}(2d) \oplus \langle T\rangle \oplus W[2d+1,n]$ $(d = 0,1,...,[(n-1)/2])$;

$\mathrm{AH}(2d) \oplus \mathfrak{M}[2d+1,n]$ $(d = 0,1,...,[(n-1)/2])$;

$\mathrm{AH}(2d) \oplus \langle G_{2d+1} + T\rangle + W[2d+1,n]$ $(d = 0,1,...[(n-1)/2])$;

$$\mathrm{AH}(2d) \oplus \langle J(d+1,r) + S + T\rangle \oplus \sum_{a=d+1}^{r} \dotplus \mathfrak{L}_a \quad (d = 0,1,...[(n-2)/2];\ r = d+1,...,[n/2]).$$

## VI. CLASSIFICATION OF SUBALGEBRAS OF THE ALGEBRAS ASch(3) AND $\widetilde{\mathrm{ASch}}(3)$

In this section we make use of the previous results to provide a classification of all subalgebras of the algebras $\mathrm{ASch}(3)$ and $\widetilde{\mathrm{ASch}}(3)$.

Let $\tilde{AG}(3) = (AO(3) \oplus \langle T \rangle) \dot{+} \mathfrak{M}[1,3]$ and $AG(3) = \tilde{AG}(3)/\langle M \rangle$. Subalgebras of the algebras $AG(3)$ and $\tilde{AG}(3)$ were classified up to conjugacy under $G(3)$ and $\tilde{G}(3)$, respectively, in Ref. 5. Further simplification of these subalgebras is being realized by $SL(2,R)$ automorphisms.

**Theorem 6.1:** Let $\alpha,\beta,\gamma,\lambda,\mu \in R$, and $\alpha > 0$, $\beta > 0$, $\gamma \neq 0$. The splitting subalgebras of the algebra $AG(3)$ are exhausted with respect to $Sch(3)$ conjugation by the splitting subalgebras of the algebra $AG(2)$ (see Ref. 2) and by the following algebras [the subalgebras preceded by the sign $\sim$ are subalgebras of $ASch(3)$]:

$\sim\langle G_1 + P_2, P_3\rangle$; $\langle G_1 + P_2, P_1 + \alpha P_3\rangle$; $\sim\langle G_1 + \gamma P_1 + P_3, G_2 + \alpha P_3\rangle$;
$\langle G_1 + \lambda P_1 + P_3, G_2 + \gamma P_1 + \alpha P_3\rangle$; $\langle G_1 + \lambda P_1 + P_3, G_2 + \alpha P_1\rangle$;
$\langle G_1 + P_2 + \alpha P_3, G_2 - P_1 + \beta P_2 + \lambda P_3\rangle$;
$\langle G_1 + P_2 + \alpha P_3, G_2 - P_1\rangle$; $\langle G_1 + P_2, G_2 - P_1 + \alpha P_2 + \beta P_3\rangle$;
$\sim\langle P_1, P_2, P_3\rangle$; $\sim\langle G_1, P_2, P_3\rangle$; $\langle G_1 + P_2, P_1, P_3\rangle$;
$\langle G_1, P_1 + \alpha P_2, P_3\rangle$; $\langle G_1 + P_3, G_2 + \alpha P_3, P_1\rangle$; $\langle G_1 + P_3, G_2, P_1\rangle$; $\langle G_1, G_2 + P_3, P_1\rangle$;
$\langle G_1 + \lambda P_1, G_2 + P_1, P_3\rangle$; $\sim\langle G_1 + \gamma P_1, G_2, P_3\rangle$; $\langle G_1 + \lambda P_1, G_2 + P_1, P_1 + \alpha P_3\rangle$;
$\langle G_1 + \lambda P_1, G_2, P_1 + \alpha P_3\rangle$; $\langle G_1 + P_2 + \alpha P_3, G_2 + \lambda P_3, P_1\rangle$; $\langle G_1 + P_2, G_2 + \alpha P_3, P_1\rangle$;
$\langle G_1 + P_2, G_2 - P_1, P_3\rangle$; $\langle G_1 + P_2, G_2 - P_1 + \alpha P_2, P_3\rangle$; $\langle G_1 + P_2 + \lambda P_3, G_2 + \mu P_3, P_1 + \alpha P_3\rangle$;
$\langle G_1 - P_2 + \alpha P_3, G_2 + P_1 + \beta P_2 + \lambda P_3, G_3 + \alpha P_1 + \lambda P_2 + \mu P_3\rangle$ $(\mu - \alpha^2\beta \neq 0)$;
$\langle G_1 - P_2, G_2 + P_1 + \beta P_2 + \alpha P_3, G_3 + \alpha P_2 + \gamma P_3\rangle$;
$\langle G_1 - P_2 + \alpha P_3, G_2 + P_1, G_3 + \alpha P_1 + \gamma P_3\rangle$; $\langle G_1, P_1, P_2, P_3\rangle$; $\langle G_1, G_2, P_1, P_3\rangle$;
$\langle G_1 + P_2, G_2, P_1, P_3\rangle$; $\langle G_1, G_2, + P_3, P_1 + \alpha P_3, P_2\rangle$; $\langle G_1, G_2, + P_3, P_1, P_2\rangle$;
$\langle G_1, G_2, P_1 + \alpha P_3, P_2\rangle$; $\langle G_1 + P_2, G_2 - P_1 + \alpha P_2, G_3 + \beta P_1 + \lambda P_2, P_3\rangle$;
$\langle G_1 + P_2, G_2 - P_1 + \alpha P_2, G_3 + \beta P_2, P_3\rangle$; $\langle G_1 + P_2, G_2 - P_1 + \alpha P_2, G_3, P_3\rangle$;
$\langle G_1, G_2 + P_2, G_3 + \alpha P_1 + \beta P_2, P_3\rangle$; $\langle G_1 + P_2, G_2 - P_1, G_3 + \alpha P_1, P_3\rangle$;
$\langle G_1 + P_2, G_2 - P_1, G_3, P_3\rangle$; $\langle G_1, G_2, P_1, P_2, P_3\rangle$;
$\langle G_1, G_2 + P_1, G_3, P_2, P_3\rangle$; $\langle G_1, G_2, G_3, P_1, P_2, P_3\rangle$;
$\langle T\rangle$: $\sim W[1,3]$, $\langle G_1 + P_2, P_1, P_3\rangle$, $\langle G_1, P_1, P_2, P_3\rangle$, $\langle G_1 + P_3, G_2, P_1, P_2\rangle$,
$\mathfrak{M}[1,2] + W[3]$, $\mathfrak{M}[1,3]$;
$\langle J_{12}\rangle$: $\sim W[3]$, $\mathfrak{M}[3]$, $\sim W[1,3]$, $\sim W[1,2] + V[3]$, $\mathfrak{L}_1 + W[3]$,
$V[3] + W[1,3]$, $\mathfrak{L}_1 + \mathfrak{M}[3]$, $\mathfrak{M}[1,2] + W[3]$, $\mathfrak{M}[1,3]$;
$\langle J_{12} + T\rangle$: $\sim W[3]$, $\mathfrak{M}[3]$, $\sim W[1,3]$, $V[3] + W[1,3]$, $\mathfrak{M}[1,2] + W[3]$, $\mathfrak{M}[1,3]$;
$\langle J_{12}, T\rangle$: $\sim W[3]$, $\mathfrak{M}[3]$, $\sim W[1,3]$, $V[3] + W[1,3]$, $\mathfrak{M}[1,2] + W[3]$, $\mathfrak{M}[1,3]$;
$AO(3)$: $\sim 0$, $\sim W[1,3]$, $\mathfrak{M}[1,3]$; $AO(3) \oplus \langle T\rangle$: $\sim 0$, $\sim W[1,3]$, $\mathfrak{M}[1,3]$.

**Theorem 6.2:** The nonsplitting subalgebras of the algebra $AG(3)$ are exhausted with respect to $Sch(3)$ conjugation by the nonsplitting subalgebras of the algebra $AG(2)^2$ and by the following algebras:

$\langle T + G_1\rangle$: $\sim W[2,3]$, $\langle P_1 + \alpha P_2, P_3\rangle$, $\langle G_2 + \alpha P_3, P_2\rangle$, $\langle G_2 + \alpha P_1 + \beta P_3, P_2\rangle$;
$W[1,3]$; $\langle G_2, P_2, P_3\rangle$, $\langle G_2 + \alpha P_1, P_2, P_3\rangle$, $\mathfrak{M}[2] + \langle P_1 + \alpha P_3\rangle$,
$\langle G_2 + \alpha P_3, P_1 + \beta P_3, P_2\rangle$, $\langle G_2 + \alpha P_3, P_1, P_2\rangle$, $V[2] + W[1,3]$, $\mathfrak{M}[2,3]$,
$\langle G_2 + \alpha P_1, G_3, P_2, P_3\rangle$, $\mathfrak{M}[2,3] + W[1]$ $(\alpha > 0, \beta > 0)$;
$\langle J_{12} + G_3\rangle$: $\sim 0$, $W[3]$, $\sim W[1,2]$, $\sim V[1,2]$, $W[1,3]$, $V[1,2] + W[3]$, $\mathfrak{M}[1,2]$, $\mathfrak{M}[1,2] + W[3]$;
$\langle J_{12} + T + \alpha G_3\rangle$: $\sim 0$, $W[3]$, $\sim W[1,2]$, $W[1,3]$, $\mathfrak{M}[1,2]$, $\mathfrak{M}[1,2] + W[3]$;
$\langle J_{12} + \alpha G_3\rangle$: $\mathfrak{L}_1$, $\mathfrak{L}_1 + W[3]$ $(\alpha > 0)$;
$\langle J_{12}, T + G_3\rangle$: $\sim 0$, $W[3]$, $\sim W[1,2]$, $W[1,3]$, $\mathfrak{M}[1,2]$, $\mathfrak{M}[1,2] + W[3]$;
$\langle J_{12} + \alpha G_3, T + G_3\rangle$: $W[3]$, $W[1,3]$, $\mathfrak{M}[1,2] + W[3]$ $(\alpha > 0)$;
$\langle J_{12} + G_3, T\rangle$: $W[3]$, $W[1,3]$, $\mathfrak{M}[1,2] + W[3]$;
$\langle J_{12} + P_3, T\rangle$: $\sim 0$, $\sim W[1,2]$, $\mathfrak{M}[1,2]$;
$\langle J_{12} + \alpha P_3, T + G_3\rangle$: $0$, $W[1,2]$, $\mathfrak{M}[1,2]$.

The written algebras are not mutually conjugated.

**Theorem 6.3:** The subalgebras of the algebra $A\tilde{G}(3)$ are exhausted with respect to $\widetilde{Sch}(3)$ conjugation by the subalgebras of the algebra $A\tilde{G}(2)$ (see Ref. 2), by the algebras preceded by the sign $\sim$ in Theorems 6.1 and 6.2, by algebras obtained from algebras written in Theorems 6.1 and 6.2 by adding the generator $M$, and by the following algebras:

$\langle T \pm M, P_1, P_2, P_3 \rangle$;

$\langle J_{12} + \alpha M \rangle$: $W[3]$, $W[1,3]$, $W[1,2] + V[3]$ $(\alpha > 0)$;

$\langle J_{12} + T \pm \alpha M \rangle$: $W[3]$, $W[1,3]$ $(\alpha > 0)$;

$\langle J_{12} + \alpha M, T \rangle$: $W[3]$, $W[1,3]$ $(\alpha > 0)$;

$\langle J_{12} + \alpha M, T \pm M \rangle$: $W[3]$, $W[1,3]$;

$\langle J_{12} + \alpha M, T + G_3 \rangle$: 0, $W[1,2]$ $(\alpha > 0)$;

$\langle J_{12} + \alpha P_3 + \beta M, T \pm M \rangle$: 0, $W[1,2]$ $(\alpha > 0, \beta > 0)$;

$\langle J_{12} + \alpha P_3, T \pm M \rangle$: 0, $W[1,2]$ $(\alpha > 0)$;

$\langle J_{12} + P_3 + \alpha M, T \rangle$: 0, $W[1,2]$ $(\alpha > 0)$;

$AO(3) \oplus \langle T \pm M \rangle$: 0, $W[1,3]$.

The written algebras are not mutually conjugated.

**Theorem 6.4:** Let $\alpha \in R$, $\alpha > 0$. The subalgebras of the algebra $ASch(3)$ which are nonconjugated to subalgebras of the algebras $AG(3)$ and $ASch(2)$ are exhausted with respect to $Sch(3)$ conjugation by the following algebras:

$\langle D \rangle$: $\sim W[1,3]$, $\sim \langle G_1, P_2, P_3 \rangle$, $\langle G_1, P_1 + \alpha P_2, P_3 \rangle$, $\langle G_1, G_2, P_1 + \alpha P_3, P_2 \rangle$,

$\langle G_1, G_2, P_1, P_3 \rangle$, $V[1] + W[1,3]$, $\overline{\mathfrak{M}}[1,2] + W[3]$, $\overline{\mathfrak{M}}[1,3]$;

$\langle S + T, G_1 - \lambda^{-1} P_2, G_2 + \lambda P_1, G_3, P_3 \rangle$ $(0 < \lambda \leqslant 1)$; $\langle S + T \rangle \oplus \overline{\mathfrak{M}}[1,3]$;

$\langle J_{12} + \alpha D \rangle$: $\sim W[3]$, $\overline{\mathfrak{M}}[3]$, $\sim W[1,3]$, $W[1,2] + V[3]$,

$W[1,2] + \overline{\mathfrak{M}}[3]$, $\overline{\mathfrak{M}}[1,2] + W[3]$, $\overline{\mathfrak{M}}[1,3]$;

$\langle S + T + \alpha J_{12} \rangle$: $\overline{\mathfrak{M}}[3]$, $\mathfrak{L}_1 + \overline{\mathfrak{M}}[3]$, $\mathfrak{N}_1 + \overline{\mathfrak{M}}[3]$, $\overline{\mathfrak{M}}[1,3]$;

$\langle S + T + 2J_{12}, G_1 + P_2 + \alpha P_3, G_2 - P_1 - \alpha G_3 \rangle$;

$\langle S + T + J_{12} \rangle$: $\langle G_1 + P_2 \rangle + \overline{\mathfrak{M}}[3]$, $\langle G_1 + P_2 \rangle + \mathfrak{N}_1 + \overline{\mathfrak{M}}[3]$;

$\langle D, T \rangle$: $\sim W[1,3]$, $V[1,j] + W[1,3]$ $(j = 1,2,3)$;

$\langle J_{12} + \alpha D, T \rangle$: $\sim W[3]$, $\overline{\mathfrak{M}}[3]$, $\sim W[1,3]$, $W[1,2] + \overline{\mathfrak{M}}[3]$, $\overline{\mathfrak{M}}[1,2] + W[3]$, $\overline{\mathfrak{M}}[1,3]$;

$\langle J_{12}, D \rangle$: $\sim W[3]$, $\overline{\mathfrak{M}}[3]$, $\sim W[1,3]$, $\sim W[1,2] + V[3]$, $W[1,2] + \overline{\mathfrak{M}}[3]$,

$\overline{\mathfrak{M}}[1,2] + W[3]$, $\overline{\mathfrak{M}}[1,3]$; $\langle J_{12}, S + T \rangle$: $\overline{\mathfrak{M}}[3]$, $\mathfrak{L}_1 + \overline{\mathfrak{M}}[3]$, $\overline{\mathfrak{M}}[1,3]$;

$\langle J_{12}, D, T \rangle$: $\sim W[3]$, $\overline{\mathfrak{M}}[3]$, $\sim W[1,3]$, $W[1,2] + \overline{\mathfrak{M}}[3]$,

$\overline{\mathfrak{M}}[1,2] + W[3]$, $\overline{\mathfrak{M}}[1,3]$; $ASL(2,R) \oplus \overline{\mathfrak{M}}[1,3]$;

$\langle J_{12} \rangle \oplus ASL(2,R)$: $\overline{\mathfrak{M}}[3]$, $\overline{\mathfrak{M}}[1,3]$; $AO(3) \oplus \langle D \rangle$: $\sim 0$, $\sim W[1,3]$, $\overline{\mathfrak{M}}[1,3]$;

$AO(3) \oplus \langle S + T \rangle$: $\sim 0$, $\overline{\mathfrak{M}}[1,3]$; $AO(3) \oplus \langle D, T \rangle$: $\sim 0$, $\sim W[1,3]$, $\overline{\mathfrak{M}}[1,3]$;

$AO(3) \oplus ASL(2,R)$: $\sim 0$, $\overline{\mathfrak{M}}[1,3]$;

$\langle S + T + J_{12} + \alpha(G_1 + P_2) \rangle$: $\overline{\mathfrak{M}}[3]$, $\langle G_2 - P_1 \rangle + \overline{\mathfrak{M}}[3]$, $\mathfrak{N}_1 + \overline{\mathfrak{M}}[3]$,

$\langle G_2 - P_1 \rangle + \mathfrak{N}_1 + \overline{\mathfrak{M}}[3]$.

The written algebras are not mutually conjugated.

**Theorem 6.5:** Let $\alpha, \beta, \gamma \in R$, and $\alpha > 0$, $\beta \neq 0$. The subalgebras of the algebra $\widetilde{ASch}(3)$ are exhausted with respect to $\widetilde{Sch}(3)$ conjugation by subalgebras of the algebra $A\tilde{G}(3)$, by subalgebras of the algebra $\widetilde{ASch}(2)$ (see Ref. 2), by algebras preceded by the sign $\sim$ in Theorem 6.4, by algebras obtained from algebras written in Theorem 6.4 by adding the generator $M$, and by the following algebras:

$\langle D + \beta M \rangle$: $W[1,3]$, $V[1] + W[2,3]$;

$\langle J_{12} + \alpha D + \beta M \rangle$: $W[3]$, $W[1,3]$,

$W[1,2] + V[3]$;

$\langle S + T + 2J_{12} + \beta M, G_1 + P_2 + \sqrt{2} P_3, G_2 - P_1 - \sqrt{2} G_3 \rangle$;

$\langle D + \beta M, T \rangle \oplus W[1,3]$;

$\langle J_{12} + \alpha M, D \rangle$: $W[3]$, $W[1,3]$, $W[1,2] + V[3]$;

$\langle J_{12} + \alpha M, D + \beta M \rangle$: $W[3]$, $W[1,3]$,

$W[1,2] + V[3]$;

$\langle J_{12}, D + \beta M \rangle$: $W[3]$, $W[1,3]$, $W[1,2] + V[3]$;

$\langle J_{12} + \alpha D + \beta M, T \rangle$: $W[3]$, $W[1,3]$;

$\langle J_{12} + \alpha M, D + \gamma M, T \rangle$: $W[3]$, $W[1,3]$;

$\langle J_{12}, D + \beta M, T \rangle$: $W[3]$, $W[1,3]$;

$AO(3) \oplus \langle D + \beta M \rangle$: 0, $W[1,3]$;

$AO(3) \oplus \langle S + T + \beta M \rangle$; $AO(3) \oplus \langle D + \beta M, T \rangle$: 0, $W[1,3]$.

The written algebras are not mutually conjugated.

## VII. CONCLUSIONS

The results of the present paper may be summarized in the following way.

(1) The completely reducible subalgebras of the algebra $AO(n) \oplus ASL(2,R)$ have been identified (Theorem 3.1).

(2) The subalgebras of $AO(n) \oplus ASL(2,R)$ which possess only splitting extensions in the algebra $ASch(n)$ have been described (Theorem 4.1).

(3) We have established that the description of the splitting subalgebras of the algebra $ASch(n)$ whose projections onto $ASL(2,R)$ are not equal to $\langle S + T \rangle$ is reduced to the description of the splitting subalgebras of $ASch(n)$ whose projections onto $AO(n)$ are equal to zero or to primary algebras (Theorem 4.2).

(4) The maximal Abelian subalgebras and the one-dimensional subalgebras of the algebras $ASch(n)$ and $\widetilde{ASch}(n)$ have been explicitly found (the corollaries to Theorems 5.1 and 5.2).

(5) The classification of the subalgebras of $ASch(3)$ and $\widetilde{ASch}(3)$ with respect to $Sch(3)$ conjugation and $\widetilde{Sch}(3)$ conjugation, respectively, has been carried out (Theorems 6.1–6.5). This classification gives the possibility to construct the wide classes of exact solutions of the nonlinear, Schrödinger-type equations in Refs. 15–18,

$$i\frac{\partial \Psi}{\partial t} - \Delta\Psi + \lambda\,|\Psi|^{4/3}\Psi = 0,$$

$$i\frac{\partial \Psi}{\partial t} - \Delta\Psi + \lambda\frac{\partial(\Psi^*\Psi)}{\partial X_a}\frac{\partial(\Psi^*\Psi)}{\partial X_a}\,(\Psi^*\Psi)^{-2}\cdot\Psi = 0,$$

which are invariant under Sch (3).


## ACKNOWLEDGMENT

We are grateful to the referee for his valuable remarks.



[1] W. I. Fushchich and R. M. Cherniha, J. Phys. A: Math. Gen. **18**, 3491 (1985).

[2] G. Burdet, J. Patera, M. Perrin, and P. Winternitz, Ann. Sci. Math. Québec **2**, 81 (1978).

[3] W. I. Fushchich and A. G. Nikitin, *Symmetry of Maxwell's Equations* (Reidel, Dordrecht, 1987).

[4] C. Boyer, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976).

[5] W. I. Fushchich, A. F. Barannik, and L. F. Barannik, "The continuous subgroups of the generalized Galilei group. I," Mathematical Institute preprint 85.19, Kiev, 1985.

[6] W. I. Fushchich, A. F. Barannik, L. F. Barannik, and V. M. Fedorchuk, J. Phys. A: Math. Gen. **18**, 2893 (1985).

[7] L. F. Barannik and A. F. Barannik, "Subalgebras of the generalized Galilei algebra," in *Group-Theoretical Studies of Equations of Mathematical Physics* (Mathematical Institute, Kiev, 1985).

[8] W. I. Fushchich, A. F. Barannik, and L. F. Barannik, Ukr. Math. J. **38**, 67 (1986).

[9] L. F. Barannik and W. I. Fushchich, J. Math. Phys. **28**, 1445 (1987).

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

[11] E. Goursat, Ann. Sci. École Norm. Sup. **6**, 9 (1889).

[12] N. Jacobson, *Lie Algebras* (Dover, New York, 1962).

[13] L. F. Barannik and W. I. Fushchich, "On continuous subgroups of the generalized Schrödinger groups," Mathematical Institute preprint 87.16, Kiev, 1987.

[14] S. Lang, *Algebra* (Addison–Wesley, Reading, MA, 1965).

[15] W. I. Fushchich, "Symmetry in the problems of mathematical physics," in *Algebraic-Theoretical Studies in Mathematical Physics*, Kiev, 1982.

[16] W. I. Fushchich and N. I. Serov, J. Phys. A: Math. Gen. **20**, L929 (1987).

[17] W. I. Fushchich and R. M. Cherniga "Exact solutions of multidimensional nonlinear Schrödinger-type equations," Mathematical Institute Preprint 86.85, Kiev, 1986.

[18] W. I. Fushchich and W. M. Shtelen,Theoret. Mat. Fiz. **56**, 387 (1983).

# Baker–Campbell–Hausdorff relations for the orthosymplectic group OSP(1/2)

B. W. Fatyga and V. Alan Kostelecký
*Physics Department, Indiana University, Bloomington, Indiana 47405*

D. Rodney Truax
*Department of Chemistry, University of Calgary, Calgary, Alberta, T2N 1N4 Canada*

(Received 23 June 1988; accepted for publication 21 September 1988)

Baker–Campbell–Hausdorff relations are presented for the connected supergroup associated with the orthosymplectic algebra osp(1/2).

## I. INTRODUCTION

In physical applications, the coordinate scheme in which Lie-group elements are most conveniently defined is often one in which practical calculations are awkward. Baker–Campbell–Hausdorff (BCH) formulas[1] are analytical expressions relating different coordinate schemes. These relations play an important role in physics, for example, in the theories of coherent or squeezed states: They have been extensively investigated and several techniques for their derivation are known.[2,3]

Supergroups are generalizations of Lie groups generated by elements of a superalgebra.[4] Scant attention has been given to the theory of BCH relations for supergroups, although their existence has been noted since the early days of supersymmetry.[5] In fact, the matrix methods frequently used[2] in the derivation of BCH relations for Lie groups are not as straightforwardly applied to the supergroup case. Despite the analogies between supergroups and Lie groups, novel features arise for supergroup BCH relations because the coordinates are Grassmann valued. For example, there are three natural choices of canonical coordinates for supergroups, of which only two have Lie-group counterparts.[6] Given the increasingly important role played by supersymmetry in physics, the development of explicit BCH formulas for supergroups is well worthwhile.

In the formulation discussed by Rogers,[7] supergroups are both abstract groups and superanalytic manifolds with composition mappings. Based on Ref. 7 and a theory of one-parameter subgroups of supergroups,[6] a framework has been established[8] for the analysis of BCH relations for supergroups. The key formulas result from the solution of systems of simultaneous Grassman-valued ordinary differential equations derived by methods analogous to those of Ref. 3. Using these techniques, BCH relations among both canonical and noncanonical coordinate schemes have been found for the connected supergroups associated with the three-generator quantum-mechanical superalgebra sqm(2),[8] the five-generator Inönü–Wigner contraction iosp(1/2) of the simple superalgebra osp(1/2),[8] and the 14-generator super-Poincaré algebra iosp(1/4).[9]

For the three supergroups mentioned above, the solution of the system of differential equations was relatively tractable. In part, this is because the simplified commutation relations of the contracted algebras result in a partial decoupling of the system of equations. In general, such simplifications do not occur. For example, for the connected supergroup COSP(1/2) associated with the simple superalgebra osp(1/2), the differential equations are more highly coupled and solving them is therefore considerably more subtle.

In this paper, we obtain BCH relations for COSP(1/2). The simplest relations result from solving a set of five coupled nonlinear Grassmann-valued differential equations, which is possible by fully exploiting the properties of Grassmann algebras. By a suitable transformation, these equations can be partially decoupled. Two of the resulting equations have a form corresponding to the key equations of Ref. 10; indeed, our solutions could be used to develop an explicit example for their approach.

As yet, no other BCH relations are known for simple superalgebras. The case of osp(1/2) is important because this superalgebra plays a role in many physical situations. For example, knowledge of its BCH relations would be useful in elucidating the nature of supercoherent states for the harmonic oscillator with spin-orbit coupling.[11] The techniques presented here should also be useful for other supergroups.

## II. THE BASIC BCH RELATION

The superalgebra osp(1/2) has three even generators $X_0$, $X_+$, and $X_-$ spanning su(1,1) and two odd generators $Q_+$ and $Q_-$. The graded commutation relations are

$$[X_+,X_-] = -2X_0, \quad [X_0,X_\pm] = \pm X_\pm, \tag{2.1a}$$

$$\{Q_\pm,Q_\pm\} = X_\pm, \quad \{Q_+,Q_-\} = X_0, \tag{2.1b}$$

$$[X_\pm,Q_\pm] = 0, \quad [X_\pm,Q_\mp] = \mp Q_\pm,$$
$$[X_0,Q_\pm] = \pm \tfrac{1}{2}Q_\pm. \tag{2.1c}$$

In this section, we present the BCH relation between canonical coordinates of the first and third kinds in normal sequence. These coordinate schemes have the form[6]

$$g_{\rm I} = \exp(aX_0 + bX_+ + cX_- + dQ_+ + eQ_-), \tag{2.2}$$

$$g_{\rm III} = \exp(\alpha X_0)\exp(\beta X_+)\exp(\gamma X_-) \times \exp(\delta Q_+)\exp(\epsilon Q_-). \tag{2.3}$$

In expressions (2.2) and (2.3), $\{a,b,c,\alpha,\beta,\gamma\} \in {}^0B_L$ and $\{d,e,\delta,\epsilon\} \in {}^1B_L$, where the even and odd parts of the Grassmann algebra $B_L$ with $L$ generators are denoted ${}^0B_L$ and ${}^1B_L$, respectively.

Following the method described in Ref. 8, we introduce a real parameter $t$ and write

 0022-2488/89/020291-04$02.50 © 1989 American Institute of Physics 

$$g_{\rm I}(t) := \exp[t(aX_0 + bX_+ + cX_- + dQ_+ + eQ_-)]$$
$$= g_{\rm III}(t), \tag{2.4}$$

where $g_{\rm III}$ is given by Eq. (2.3), but $\alpha, \beta, \gamma, \delta$, and $\epsilon$ are now functions of $t$. Next, we differentiate Eq. (2.4) with respect to $t$ and collect coefficients multiplying each generator. Independence of the generators then yields a system of five nonlinear, first-order, $B_L$-valued ordinary differential equations. Taking into account the properties of Grassmann variables, these may be rearranged to give

$$\dot\alpha = a + 2\beta c\exp(\alpha) + \delta e\exp(\tfrac12\alpha), \tag{2.5a}$$
$$\dot\beta = b\exp(-\alpha) - \beta^2 c\exp(\alpha) + \tfrac12\delta d\exp(-\tfrac12\alpha) - \tfrac12\beta\delta e\exp(\tfrac12\alpha), \tag{2.5b}$$
$$\dot\gamma = c\exp(\alpha) - \tfrac12\gamma(\gamma\delta+\epsilon)d\exp(-\tfrac12\alpha) + [\gamma\delta(1-\tfrac12\beta\gamma) + \tfrac12(1-\beta\gamma)\epsilon]e\exp(\tfrac12\alpha), \tag{2.5c}$$
$$\dot\delta = d\exp(-\tfrac12\alpha) + \beta e\exp(\tfrac12\alpha), \tag{2.5d}$$
$$\dot\epsilon = (1-\beta\gamma)(1-\tfrac12\delta\epsilon)e\exp(\tfrac12\alpha) - \gamma(1-\tfrac12\delta\epsilon)d\exp(-\tfrac12\alpha). \tag{2.5e}$$

The solution of this system, subject to the boundary conditions

$$\alpha(0) = \beta(0) = \gamma(0) = \delta(0) = \epsilon(0) = 0, \tag{2.5f}$$

yields the desired BCH relation when $t = 1$.

We remark here that by making the substitutions

$$\hat\alpha = \alpha,\quad \hat\beta = \beta e^{\alpha},\quad \hat\gamma = \gamma e^{-\alpha},\quad \hat\delta = \delta e^{(1/2)\alpha},\quad \hat\epsilon = \epsilon e^{-(1/2)\alpha}, \tag{2.6}$$

we can transform Eqs. (2.5) to a system in which the pair of equations for $\hat\beta$ and $\hat\delta$ are decoupled from the others. These equations have a similar form to Eqs. (11) of Ref. 10. Explicit examples for the approach of Ref. 10 may be found from the general solution that we present below.

The key to solving Eqs. (2.5) is to expand the dependent variables as polynomials in the odd variables $d$ and $e$. As $d^2 = e^2 = 0$, the polynomials contain only two terms:

$$A := \exp(\alpha) = A_0(a,b,c;t) + A_1(a,b,c;t)de = \exp[\alpha_0(a,b,c;t)](1 + \alpha_1(a,b,c;t)de), \tag{2.7a}$$
$$\beta = \beta_0(a,b,c;t) + \beta_1(a,b,c;t)de, \tag{2.7b}$$
$$\gamma = \gamma_0(a,b,c;t) + \gamma_1(a,b,c;t)de, \tag{2.7c}$$
$$\delta = \delta_0(a,b,c;t)d + \delta_1(a,b,c;t)e, \tag{2.7d}$$
$$\epsilon = \epsilon_0(a,b,c;t)d + \epsilon_1(a,b,c;t)e. \tag{2.7e}$$

We remark that this useful method is likely to have applications to other systems of Grassmann-valued equations.

We can find $A_0$ and $\beta_0$ in two steps. First, multiply Eq. (2.5a) by $A\beta de$ and Eq. (2.5b) by $Ade$ and add. Using Eqs. (2.7a) and (2.7b) and collecting coefficients of $de$ yields

$$\frac{d}{dt}(A_0\beta_0) = a(A_0\beta_0) + c(A_0\beta_0)^2 + b, \tag{2.8}$$

which is a Riccati equation.[12] Its solution is[3,9]

$$A_0\beta_0 = (b/K)S^{-1}\sinh(Kt), \tag{2.9}$$

where

$$K = \hat\sigma\Delta,\quad \Delta^2 = |\tfrac14 a^2 - bc|,\quad \sigma = \operatorname{sign}(\tfrac14 a^2 - bc),\quad \hat\sigma = \sqrt{\sigma}, \tag{2.10}$$

and

$$S = \cosh(Kt) - (a/2K)\sinh(Kt). \tag{2.11}$$

The second step is to multiply Eq. (2.5a) by $Ade$, yielding

$$\dot A_0 - [a + 2c(A_0\beta_0)]A_0 = 0 \tag{2.12}$$

subject to the boundary condition $A_0(0) = 1$. Using Eq. (2.9), the solution is

$$A_0(t) = S^{-2},\quad \alpha_0 = -2\ln S. \tag{2.13}$$

From Eq. (2.9) we find

$$\beta_0(t) = (b/K)S\sinh(Kt). \tag{2.14}$$

The next variables we consider are $\delta_0$ and $\delta_1$. Writing $\delta$ as in Eq. (2.7d) and collecting coefficients results in the differential equations

$$\dot\delta_0 = A_0^{-1/2},\quad \dot\delta_1 = \beta_0 A_0^{1/2}, \tag{2.15}$$

which have the solutions

$$\delta_0(t) = (1/K^2)(\dot S + a/2),\quad \delta_1(t) = (b/K^2)[\cosh(Kt) - 1] \tag{2.16}$$

where $\dot S = dS/dt$.

Next, we solve for $A_1$ and $\beta_1$. Substituting Eqs. (2.7a) and (2.7b) into Eqs. (2.5a) and (2.5b) and collecting coefficients of $de$ yields

$$\dot A_1 = (a + 4cA_0\beta_0)A_1 + (2cA_0^2)\beta_1 + \delta_0 A_0^{3/2}, \tag{2.17a}$$
$$\dot\beta_1 = -(2cA_0\beta_0)\beta_1 - (bA_0^{-2} + c\beta_0^2)A_1 - \tfrac12(\delta_0 A_0^{1/2}\beta_0 + \delta_1 A_0^{-1/2}). \tag{2.17b}$$

It follows that the quantity $X := (\beta_0 A_1 + A_0\beta_1)$ satisfies an inhomogeneous linear first-order ordinary differential equation:

$$\dot X = (a + 2cA_0\beta_0)X + \tfrac12(\delta_0 A_0^{3/2}\beta_0 - \delta_1 A_0^{1/2}), \tag{2.18}$$

with solution

$$\beta_0 A_1 + A_0\beta_1 = (b/2K^3)S^{-2}[\sinh(Kt) - Kt]. \tag{2.19}$$

Solving for $A_0\beta_1$ and substituting into Eq. (2.17a) yields an inhomogeneous first-order equation for $A_1$, with solution

$$A_1(t) = (1/K^2)S^{-2}\{S^{-1}[1 - (2bc/a)t\cosh(Kt)] - (2K^2/a)t - 1\}, \tag{2.20}$$
$$\alpha_1(t) = S^2 A_1(t).$$

The solution for $\beta_1$ follows from Eq. (2.19):

$$\beta_1(t) = -(b/2K^2)t + (b/K^3)\sinh(Kt)[S - \tfrac12 - t\dot S]. \tag{2.21}$$

It remains to find $\gamma$ and $\epsilon$. Substitution of Eq. (2.7c) into Eq. (2.5c) and collecting coefficients of unity in $B_L$ gives $\dot\gamma_0 = cA_0$, which has solution

$$\gamma_0(t) = (c/K)S^{-1}\sinh(Kt). \tag{2.22}$$

Equation (2.5e) then yields two differential equations, $\dot\epsilon_0 = -\gamma_0 A_0^{-1/2}$ and $\dot\epsilon_1 = (1 - \beta_0\gamma_0)A_0^{-1/2}$, which may be solved to give

$$\epsilon_0(t) = (c/K^2)[\cosh(Kt) - 1], \tag{2.23a}$$

$$\epsilon_1(t) = (1/K)\{\sinh(Kt) + (a/2K)[\cosh(Kt) - 1]\}. \tag{2.23b}$$

The one remaining unknown is $\gamma_1(t)$. The differential equation it satisfies is obtained as the *de* coefficient of Eq. (2.5c). After substitution for known dependent variables, a straightforward but somewhat involved integration in $t$ yields the desired result:

$$\begin{aligned}\gamma_1(t) = {} & \frac{mM^3}{2bK}(1-2M^2) - \frac{1}{2bm}t\,[m^2(1+M^4) - S^{-2}] \\ & - \frac{1}{2bK}S\,[2M\cosh(Kt) + (1+M^2)\sinh(Kt)] \\ & - \frac{M^2}{bK}\,[\sinh(Kt) + M\cosh(Kt)] \\ & - \frac{m}{2bK}\,[\sinh(Kt) - M\cosh(Kt)] \\ & + \frac{1}{2bk}S^{-1}[M - 2(1+M^2)\sinh(Kt)],\end{aligned} \tag{2.24}$$

where $M = a/2K$ and $m = (1-M^2)^{-1} = -K^2/bc$.

To summarize, the BCH relation between canonical coordinates of the first and third kinds in normal sequence is given by Eq. (2.4), with $t = 1$. The quantities $\alpha, \beta, \gamma, \delta$, and $\epsilon$ are expressed as functions of $a, b, c, d$, and $e$ through Eqs. (2.7) and the solutions (2.13), (2.14), (2.16), and (2.20)–(2.24).

## III. OTHER RELATIONS

From the basic BCH relation given in Sec. II, others may readily be found. Here, we sketch the derivation of several additional relations to illustrate the methods involved. Further examples of the methods may be found in Refs. 8 and 9.

First, consider a BCH formula involving canonical coordinates of the third kind in non-normal sequence:

$$\begin{aligned}&\exp(aX_0 + bX_+ + cX_- + dQ_+ + eQ_-) \\ &\quad = \exp(\beta' X_+)\exp(\delta' Q_+)\exp(\gamma' X_-) \\ &\qquad \times \exp(\epsilon' Q_-)\exp(\alpha' X_0).\end{aligned} \tag{3.1}$$

We believe expression (3.1) is important for the analysis of supercoherent states. In view of the results in Sec. II, it is sufficient to convert canonical coordinates of the third kind in normal sequence, Eq. (2.3), to the form of the right-hand side of Eq. (3.1). This is done by inserting the identity $I = \exp(\delta Q_+)\exp(-\delta Q_+)$ between the second and third exponentials of Eq. (2.3). Then, using the Campbell–Hausdorff theorems for superalgebras (Theorems 2 and 3 of Ref. 8) has the effect of interchanging the $X_-$ and $Q_+$ exponentials of Eq. (2.3) and modifying the coefficients of the $X_-$ and $Q_-$ exponentials in a well-defined manner. Repeating this procedure four times with insertions of $I = \exp(\alpha X_0)\exp(-\alpha X_0)$ in the appropriate places then effectively moves the $X_0$ exponential to right-hand side yielding the desired non-normal sequence. The coefficients $\alpha', \beta', \gamma', \delta'$, and $\epsilon'$ in Eq. (3.1) are thereby found to be expressed in terms of $\alpha, \beta, \gamma, \delta$, and $\epsilon$ as

$$\begin{aligned}&\alpha' = \alpha, \quad \beta' = \beta\exp(\alpha), \\ &\gamma' = \gamma(1 - \tfrac{1}{2}\delta\epsilon)\exp(-\alpha), \\ &\delta' = \delta\exp(\tfrac{1}{2}\alpha), \\ &\epsilon' = (\gamma\delta + \epsilon)\exp(-\tfrac{1}{2}\alpha).\end{aligned} \tag{3.2}$$

Together with the results of Sec. II, Eqs. (3.2) form the desired BCH relation (3.1).

Other non-normal sequences may be analyzed similarly. In particular, we have also found explicitly the BCH relation

$$\begin{aligned}&\exp(aX_0 + bX_+ + cX_- + dQ_+ + eQ_-) \\ &\quad = \exp(\delta' Q_+)\exp(\epsilon' Q_-)\exp(\alpha' X_0) \\ &\qquad \times \exp(\beta' X_+)\exp(\gamma' X_-).\end{aligned} \tag{3.3}$$

We find

$$\begin{aligned}&\alpha' = \alpha - \beta\gamma\delta\epsilon, \quad \beta' = \beta(1 + \tfrac{1}{2}\delta\epsilon), \\ &\gamma' = \gamma(1 - \tfrac{1}{2}\delta\epsilon), \\ &\delta' = [(1 - \beta\gamma)\delta - \beta\epsilon]\exp(\tfrac{1}{2}\alpha), \\ &\epsilon' = (\gamma\delta + \epsilon)\exp(-\tfrac{1}{2}\alpha),\end{aligned} \tag{3.4}$$

Again, combined with the results of Sec. II, Eqs. (3.4) define the BCH relation (3.3).

Consider next canonical coordinates of the second kind, given as[6]

$$\begin{aligned}g_{\rm II} = {} & \exp(a' X_0)\exp(b' X_+)\exp(c' X_-) \\ & \times \exp(b_+ X_+)\exp(d' Q_+)\exp(c_- X_-)\exp(e' Q_-),\end{aligned} \tag{3.5}$$

where $b_+$ and $c_-$ are defined in terms of the expansion coefficients of $d'$ and $e'$, respectively; see Eq. (7.26) of Ref. 6. Expression (3.5) may be converted to canonical coordinates of the third kind by appropriate insertions of identity. First, insert

$$\begin{aligned}I = {} & \exp(c_- d' Q_-)\exp(\tfrac{1}{2}c_- d' e' X_-) \\ & \times \exp(-\tfrac{1}{2}c_- d' e' X_-)\exp(-c_- d' Q_-)\end{aligned}$$

between the last two exponentials in Eq. (3.5). This effectuates the interchange of the $d' Q_+$ and $c_- X_-$ exponentials while changing the coefficients of the $X_-$ and $Q_-$ exponentials. Next, insert $\exp(-c' X_-)\exp(c' X_-)$ after $\exp(b_+ X_+)$. Using the BCH relations for su(1,1), which may be obtained from the basic relation in Sec. II by setting $d = e = 0$, the resulting expression can be converted into canonical coordinates of the third kind in normal sequence:

$$\begin{aligned}g_{\rm II} = {} & \exp[(a' + \alpha_2)X_0]\exp[(b'\exp(-\alpha_2) + \beta_2)X_+] \\ & \times \exp[(\gamma_2 + c' + c_- + \tfrac{1}{2}c_- d' e')X_-]\exp(d' Q_+) \\ & \times \exp[(e' - c_- d')Q_-],\end{aligned} \tag{3.6}$$

where $\alpha_2 = -2\ln(1 - c' b_+)$, $\beta_2 = b_+(1 - c' b_+)$, and $\gamma_2 = (c')^2 b_+/(1 - c' b_+)$.

Finally, as an example of noncanonical coordinates, consider

$$g_{\rm nc} = \exp(a'' X_0 + b'' X_+ + c'' X_-)\exp(d'' Q_+ + e'' Q_-). \tag{3.7}$$

Inserting the identity $\exp(\tfrac{1}{2}d'' e'' X_0)\exp(-\tfrac{1}{2}d'' e'' X_0)$ between the two exponentials and using a BCH relation for su(1,1) gives

$$g_{\rm nc} = \exp[(\alpha_0 + \tfrac{1}{2}d''e'')X_0]\exp[\beta_0(1 - \tfrac{1}{2}d''e'')X_+]$$
$$\times\exp[\gamma_0(1 + \tfrac{1}{2}d''e'')X_-]$$
$$\times\exp(d''Q_+)\exp(e''Q_-), \qquad (3.8)$$

where $\alpha_0$, $\beta_0$, and $\gamma_0$ have the functional form given in Eqs. (2.13), (2.14), and (2.22). This is an expression in terms of canonical coordinates of the third kind in normal sequence.

## ACKNOWLEDGMENTS

We thank Michael Martin Nieto for helpful discussions. Parts of this work were completed while we enjoyed the hospitality of the Los Alamos National Laboratory.

VAK was supported in part by the U.S. Department of Energy under Contract No. DE-AC02-84ER40125. DRT was supported in part by the National Science and Engineering Research Council of Canada.

[1]J. E. Campbell, Proc. London Math. Soc. **28**, (1) 381 (1897); H. F. Baker, *ibid.* **34**, (1) 347 (1902); **35**, 333 (1903); **2**, (2) 293 (1905); F. Hausdorff, Ber. Verh. Saechs. Akad. Wiss. Leipzig Math. Phys. Kl. **58**, 19 (1906).

[2]See, for example, R. M. Wilcox, J. Math. Phys. **8**, 962 (1967); B. Mielnik, Ann. Inst. H. Poincaré A **12**, 215 (1970); R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).

[3]D. R. Truax, Phys. Rev. D **31**, 1988 (1985).

[4]For an overview of superalgebras, see, for instance, M. Scheunert, "The theory of Lie superalgebras," *Lecture Notes in Mathematics*, Vol. 716 (Springer, New York, 1976).

[5]See, for example, F. A. Berezin and G. I. Kac, Math. USSR Sb. **11**, 311 (1970).

[6]D. R. Truax, V. A. Kostelecký, and M. M. Nieto, J. Math. Phys. **27**, 354 (1986).

[7]A. Rogers, J. Math. Phys. **22**, 939 (1981).

[8]V. A. Kostelecký, M. M. Nieto, and D. R. Truax, J. Math. Phys. **27**, 1419 (1986).

[9]V. A. Kostelecký and D. R. Truax, J. Math. Phys. **28**, 2480 (1987).

[10]J. Beckers, L. Gagnon, V. Hussin, and P. Winternitz, Lett. Math. Phys. **13**, 113 (1987).

[11]E. D'Hoker and L. Vinet, Commun. Math. Phys. **97**, 391 (1985); A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985); V. A. Kostelecký, M. M. Nieto, and D. R. Truax, Phys. Rev. D **32**, 2627 (1985).

[12]See, for example, E. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).

# Computation of elliptic functions

Charles L. Critchfield
*391 El Conejo, Los Alamos, New Mexico 87544*

(Received 11 March 1988; accepted for publication 12 October 1988)

The duplication formula for Weierstrass' elliptic function is the basis for a simple algorithm by which one can calculate values of that function. As in the case for sin $\theta$, this method is more efficient when the argument is reduced to a value near the origin by addition or subtraction of a suitable number of full periods of the function. When the invariants $g_2$ and $g_3$ are real numbers, $\wp(z;g_2,g_3)$ is periodic on both the real and imaginary axes of $z$, although these periods are not necessarily the basic ones. Convenient formulas for determining the relevant periods are given. They include an expression for $q(k^2)$ in terms of Gauss' arithmetico-geometrical sequence. The method is applied to Jacobian functions by finding the invariants that represent a given $k^2$. When this parameter is real the periods are well known and the Jacobian functions are calculated directly without transforming to those for which the parameter lies between zero and unity.

## I. THE RECURSION FORMULA

When numerical values for elliptic functions are needed, the functions usually belong to real, fixed values for the parameters and the variable is either real or simply imaginary. These conditions will be assumed to be met in the following presentation of a method for calculating $\wp(z;g_2,g_3)$. This method is analogous to finding $\sin\theta$ from $\theta/2^N$, viz., by choosing the integer $N$ large enough so that $\theta^2/4^N$ is an adequate approximation to $\sin^2(\theta/2^N)$ and applying $\sin^2 2x = 4\sin^2 x(1-\sin^2 x)$ $N$ times. Obviously it is desirable to reduce $\theta$ (assumed to be real) to a value between $-\pi$ and $+\pi$, for which the square of the sine is the same. Elliptic functions that have real parameters are periodic along both real and imaginary axes[1] so that the reduction can be made for both components of a complex argument. Usually, however, the relevant periods are not known beforehand and must be calculated from $g_2$ and $g_3$. Those formulas are given in Sec. II.

The Laurent series for $\wp(z;g_2,g_3)$ is

$$\wp z = z^{-2} + g_2 z^2/20 + g_3 z^4/28 + \cdots, \tag{1}$$

where the function is abbreviated to $\wp z$ and $g_2,g_3$ are the coefficients in the equation for $d(\wp z)/dz$,

$$(\wp' z)^2 = 4(\wp z)^3 - g_2\wp z - g_3. \tag{2}$$

The duplication formula is

$$\wp(2z) = \tfrac{1}{4}(\wp'' z/\wp' z)^2 - 2\wp z,$$

and a concise formulation of the algorithm is

$$w_0 \cong \wp z, \quad w_N = \wp(z/2^N),$$

$$w_{n-1} = \frac{(3w_n^2 - \frac{1}{4}g_2)^2}{4w_n^3 - g_2 w_n - g_3} - 2w_n, \tag{3}$$

$$n = N, N-1, \ldots, 2, 1.$$

Starting with $w_N = 4^N/z^2$, $N = 7$ is usually sufficient to obtain eight significant figures if $|z| < 3$. In a longhand calculation one can use all three terms in (1) for $w_N$ and then $N = 3$ should produce a comparable result. Note that zeros of the denominator in (3) that are nearest the origin occur where $\wp' z$ vanishes, i.e., at one-half period along either axis. Those values of $z$ on the real and imaginary axes will be denoted by $\omega_x$ and $\omega_y$, respectively. Thus, if an argument is reduced by $2n\omega_x$ or $2m\omega_y$ (where $n$ and $m$ are integers) to one between $\pm\omega_x$, $\pm\omega_y$, there will be no serious loss of significant figures in the recursion process.

## II. DETERMINATION OF PERIODS

The algorithm in (3) may be applied to empirical determinations of $\omega_x$ and $\omega_y$. One first plots rough values of $w_0$ near $z = 0$ in order to locate the nearest maximum or minimum. On the real axis $w_0$ will have a local minimum and in its vicinity $w_0$ is calculated, to the precision required for $\wp z$ in the final result, and plotted so as to determine the zero of $w_0' = (4w_0^3 - g_2 w_0 - g_3)^{1/2}$. That value of $z$ (on the positive axis) is $\omega_x$. Similarly, the least root of $w_0'$ on the positive $i$ axis (at a local maximum) yields $\omega_y$.

The numbers $\wp\omega_x$ and $\wp\omega_y$ are *critical values* of $\wp z$ and if $\wp\omega_x \neq \wp\omega_y$, in the calculation described above, $2\omega_x$ is the basic real period for $\wp(z;g_2,g_3)$ and $2\omega_y$ is its basic imaginary period. These will be denoted by $2\omega_1$ and $2\omega_3$, respectively. If, on the other hand, $\wp\omega_x$ turns out to be equal to $\wp\omega_y$, only one of the major axes is the direction of a basic period. When $g_3 \geqslant 0$, $2\omega_x$ remains as the basic real period but $2\omega_3$ is a complex number; when $g_3 < 0$, $2\omega_y$ is a basic period and *it* will be designated to be $2\omega_1$, whereas $2\omega_x$ is a composite period.

There are three finite critical values of $\wp z$ and they are the roots of $(\wp' z)^2$ in (2). In an arbitrary order they are denoted by $e_1$, $e_2$, $e_3$. Applying Cardan's rule they are

$$e_r = \tfrac{1}{2}\{g_3 + [g_3^2 - (g_2/3)^3]^{1/2}\}^{1/3} + \tfrac{1}{2}\{g_3 - [g_3^2 - (g_2/3)^3]^{1/2}\}^{1/3}, \tag{4}$$

where $r = 1,2,3$ refers to the three cube roots whose phases, for a given $r$, are equal in magnitude but of opposite sign in the two terms on the right-hand side of (4). Then (2) may be written

$$(\wp' z)^2 = 4(\wp z - e_1)(\wp z - e_2)(\wp z - e_3). \tag{5}$$

Comparing this with (2) yields the identities

$$e_1 + e_2 + e_3 = 0, \quad g_3 = 4e_1 e_2 e_3,$$

 0022-2488/89/020295-03$02.50 © 1989 American Institute of Physics 

$$g_2 = -4(e_1e_2 + e_1e_3 + e_2e_3), \tag{6}$$
$$\Delta = 16(e_1 - e_2)^2(e_1 - e_3)^3(e_2 - e_3)^2 = g_2^3 - 27g_3^2,$$

where $\Delta$ is the discriminant of (5). The square roots in (4) are then $(-\Delta/27)^{1/2}$.

The basic half-periods of $\wp z$ are defined by convention as the principal values of integrals of (5) for $z$, viz.,

$$\omega_r = \int_{e_r}^{\infty} \frac{dt}{2[(t-e_1)(t-e_2)(t-e_3)]^{1/2}}. \tag{7}$$

When all three $e_r$ are real their subscripts can be assigned so that

$$e_1 > e_2 > e_3 \tag{8}$$

[no two $e_r$ are equal, for if they were, the integral in (7) would be elementary rather than an elliptic integral]. The integral for $\omega_1$ in (7) is then real and, making the substitution $t = e_3 + (e_1 - e_3)\csc^2\psi$, it becomes

$$\omega_1 = (e_1 - e_3)^{-1/2}K(k), \tag{9}$$

where $K(k)$ is the complete elliptic integral of the first kind and

$$k^2 = (e_2 - e_3)/(e_1 - e_3). \tag{10}$$

From (8) one sees that $0 < k^2 < 1$; and from (10), that the exclusion of equal $e_r$ excludes $k^2 = 0, 1, \infty$. In (4) all three $e_r$ are real when $(g_2/3)^3 > g_3^2$, i.e., when $\Delta > 0$.

Also, because of (8), the integral in (7) from $e_2$ to $e_1$ is an imaginary number. Transforming from $t$ to $t'$ so that $(t' - e_2)(t - e_2) = (e_2 - e_1)(e_2 - e_3)$, that integral becomes one of the same form but with limits $e_3$ and $\infty$. Therefore $\omega_3$ is simply imaginary; setting $t = e_1 - (e_1 - e_2)\sin^2\psi$, the integral from $e_2$ to $e_1$ yields

$$\omega_3 = i(e_1 - e_3)^{-1/2}K'(k), \tag{11}$$

where $K'(k) = K(k')$ and $k' = +(1 - k^2)^{1/2}$. Thus when the $e_r$ are real, $2\omega_x = 2\omega_1$ and $2\omega_y = 2\omega_3$. The integral in (7) from $e_2$ to $\infty$ is $\omega_2$ and that is the sum of integrals from $e_2$ to $e_1$ to $\infty$. Therefore

$$\omega_2 = \omega_1 + \omega_3. \tag{12}$$

When $\Delta < 0$, i.e., $g_3^2 > (g_2/3)^3$, one root is real and the relations in (6) show that the other two are conjugate complex numbers. The real root will be denoted $e_1$ so that we can write

$$e_2 = -\tfrac{1}{2}e_1(1 - i\eta), \quad e_3 = -\tfrac{1}{2}e_1(1 + i\eta), \tag{13}$$

where $\eta$ is real and positive. With these definitions $g_3$ in (6) becomes $e_1^3(1 + \eta^2)$ and $e_1$ has the same sign as $g_3$; also, one sees from (10) that $k^2$, and therefore $k'^2$, are now complex numbers. The scale factor $(e_1 - e_3)^{-1/2}$ in (9) and (11) is complex as well and it turns out, when $g_3 > 0$, that $\omega_1$ is real and thus equal to $\omega_x$, but $\omega_3$ is complex. The imaginary half-period is then $\omega_y = \omega_1 + 2\omega_3$ and thus $\wp\omega_y = \wp(\omega_1 + 2\omega_3) = \wp\omega_1 = \wp\omega_x$ as anticipated above in the empirical calculation. Also, for this case,

$$(e_1 - e_3)^{1/2} = |e_1 - e_3|^{1/2}e^{(1/2)i\beta},$$
$$k'^2 = (e_1 - e_2)/(e_1 - e_3) = e^{-2i\beta}, \quad \tan\beta = \tfrac{1}{3}\eta, \tag{14}$$

and although tables are not available for complex moduli the elliptic integrals are easily computed when $|k'^2| = 1$.

When $e_1 < 0$ the scale factor $(e_1 - e_3)^{-1/2}$ acquires a factor $i$ and in that event $\omega_y = \omega_1$ and $\omega_x = \omega_1 + 2\omega_3$.

It is well known[2] that $K(k)$ is related to its asymptotic value $\frac{1}{2}\pi$ (as $k^2 \to 0$) by

$$K(k) = \frac{1}{2}\pi \prod_{n=1}^{\infty}(1 + k_n),$$
$$k_{n+1} = \frac{1 - k_n'}{1 + k_n'}, \quad k_0' = k'. \tag{15}$$

As $k^2 \to 1$ the asymptotic form is $-\ln(\frac{1}{4}k')$. Therefore as $k^2 \to 0$ the corresponding form for $K'(k)$ is $-\ln(\frac{1}{4}k)$. The $N$th convergent to $K(k)$ will be denoted by $K_N$ and is readily calculated by means of Gauss' arithmetico-geometrical sequences.[2] Carlson[3] has shown how to make the same calculations applicable to $K'(k)$. The pertinent definitions and identities are

$$a_0 = 1, \quad a_{n+1} = \tfrac{1}{2}(a_n + b_n), \quad k_n = c_n/a_n,$$
$$b_0 = k', \quad b_{n+1} = (a_nb_n)^{1/2}, \quad 1 + k_n = a_{n-1}/a_n,$$
$$c_0 = k, \quad c_{n+1} = \tfrac{1}{2}(a_n - b_n), \quad a_nc_n = \tfrac{1}{4}c_{n-1}^2,$$
$$k_0 = k, \quad k_0' = k', \quad \tfrac{1}{4}k_n = (1 + k_n)^2(\tfrac{1}{4}k_{n-1})^2. \tag{16}$$

The $N$th convergents to $K(k)$ and $K'(k)$ are

$$K_N = \pi/2a_N, \quad K_N' = -(2^{-N}/a_N)\ln(\tfrac{1}{4}k_N), \tag{17}$$

and that to the nome $q(k^2) = \exp[-\pi K'(k)/K(k)]$ is

$$q_N = (\tfrac{1}{4}k_N)^{2^{1-N}}. \tag{18}$$

Since $c_n$ approaches zero very rapidly as $n$ increases, the forms containing $k_N$ are of limited usefulness as they stand. However, the formula in (16) that relates $\frac{1}{4}k_n$ to $\frac{1}{4}k_{n-1}$ can be iterated so as to produce[4]

$$\frac{1}{4}k_N = \left(\frac{1}{4}k\right)^{2^N}\prod_{n=1}^{N}(1 + k_n)^{2^{N+1-n}}. \tag{19}$$

Using this in (18),

$$q_N = \frac{k^2}{16}\prod_{n=1}^{N}(1 + k_n)^{2^{2-n}} = q_{N-1}(a_{N-1}/a_N)^{2^{2-N}}. \tag{20}$$

Substituting $a_1 = \frac{1}{2}(1 + k')$, $a_2 = \frac{1}{4}(1 + \sqrt{k'})^2$ in $q^2$ one obtains the familiar approximation[5]

$$q_2 = \tfrac{1}{2}(1 - \sqrt{k'})/(1 + \sqrt{k'}). \tag{21}$$

When $0 < k^2 < 1$ the calculations deal only with real numbers. The case for $k' = e^{-i\beta}$ is almost as convenient because all $a_n$ and $b_n$ in (16), for $n > 0$, are real numbers times $e^{-(1/2)i\beta}$. Therefore $K_N$ in (17) is a real times $e^{(1/2)i\beta}$ and, when $e_1 > 0$, (14) shows that $(e_1 - e_3)^{-1/2}K_N$ will be real. Therefore the $N$th convergent to $\omega_1$ in (9) is real and equal to that for $\omega_x$,

$$\omega_1(N) = \pi/[2(e_1 - e_3)^{1/2}a_N] = \omega_x(N). \tag{22}$$

For the $N$th convergent to $\omega_3$ we write

$$\omega_3(N) = -i[\omega_1(N)/\pi]\ln q_N.$$

In (20) each $q_N$ is proportional to $k^2 = 1 - e^{-2i\beta} = 2ie^{-i\beta}\sin\beta$ and (for $n > 0$) inversely proportional to $a_n$ to the second power (and therefore to $e^{-i\beta}$). Thus the $q_N$ are purely imaginary and

$$\ln q_N = \tfrac{1}{2}i\pi + \ln|q_N|,$$

$$\omega_3(N) = \omega_1(N)(\tfrac{1}{2} - (i/\pi)\ln|q_N|). \tag{23}$$

Then $\omega_y(N) = 2\omega_3(N) - \omega_1(N)$. When $e_1 < 0$ the roles of $K_N$ and $K'_N$ are interchanged as described above.

## III. APPLICATION TO JACOBIAN FUNCTIONS

The definition of $\mathrm{sn}(u,k)$ in Weierstrass' theory is[6]

$$\mathrm{sn}(u,k) = [(e_1 - e_3)/(\wp z - e_3)]^{1/2}, \tag{24}$$

where $u$ is related to $z$ by the same factor that relates $K(k)$ to $\omega_1$,

$$u = (e_1 - e_3)^{1/2} z. \tag{25}$$

Thus, for example, when $z = \omega_1$, $\wp\omega_1 = e_1$ and the right-hand side of (24) is unity, when the positive sign applies, and $\mathrm{sn}(K,k) = 1$.

Actually the basic identities among Jacobian functions have their most symmetrical form for the three functions that have a pole at the origin. In modern notation these are

$$\mathrm{cs}\,u = \frac{\mathrm{cn}(u,k)}{\mathrm{sn}(u,k)}, \quad \mathrm{ds}\,u = \frac{\mathrm{dn}(u,k)}{\mathrm{sn}(u,k)}, \quad \mathrm{ns}\,u = \frac{1}{\mathrm{sn}(u,k)},$$

and for these the extension of (24) is

$$(e_1 - e_3)^{1/2}\,\mathrm{cs}\,u = (\wp z - e_1)^{1/2},$$
$$(e_1 - e_3)^{1/2}\,\mathrm{ds}\,u = (\wp z - e_2)^{1/2},$$
$$(e_1 - e_3)^{1/2}\,\mathrm{ns}\,u = (\wp z - e_3)^{1/2}.$$

Therefore, if $g_2$ and $g_3$ are calculated from the $e_r$ that satisfy

$$e_1 + e_2 + e_3 = 0, \quad e_1 - e_3 = 1, \quad e_2 - e_3 = k^2, \tag{26}$$

the $w_0$ resulting from (3), with $z = u$, determines these functions, up to a sign, as

$$\mathrm{cs}\,u = (w_0 - e_1)^{1/2}, \quad \mathrm{ds}\,u = (w_0 - e_2)^{1/2},$$
$$\mathrm{ns}\,u = (w_0 - e_3)^{1/2}. \tag{27}$$

The signs on the square roots are determined by the location of $u$ within a period for each of the functions. Solving (26) one finds

$$e_2 = \tfrac{1}{3}(2k^2 - 1), \quad e_1 = e_2 + 1 - k^2, \quad e_3 = e_2 - k^2, \tag{28}$$

and from the relations in (6) the special values of $g_2$ and $g_3$ are

$$g_2 = 1 + 3e_2^2, \quad g_3 = e_2(e_2^2 - 1). \tag{29}$$

Note that the relations in (27)–(29) place no restraint upon $k^2$. They merely guarantee that the invariants correspond to a given value and that $(e_1 - e_3)^{1/2} = 1$. Thus $k^2$ is not required to be between zero and unity so then $K(k)$ and $K'(k)$ may be complex numbers. Their contributions to $\omega_x$ and $\omega_y$, however, are well known.[7] They differ in the three regimes of $k^2$ and for $u$ and $iu$. Let $k_0^2$ be such that $0 < k_0^2 < 1$ and let its elliptic integrals be $K_0$ and $K'_0$. Then writing $k^2$ in terms of $k_0^2$ one has the following assignments:

| argument | $u$ | | $iu$ | |
|---|---|---|---|---|
| quarter-period | $\omega_x$ | $\omega_y$ | $\omega_x$ | $\omega_y$ |
| $k^2 = k_0^2$ | $K_0$ | $iK'_0$ | $K'_0$ | $iK_0$ |
| $k^2 = k_0^{-2}$ | $k_0K_0$ | $ik_0K'_0$ | $k_0K'_0$ | $ik_0K_0$ |
| $k^2 = -k_0^2/k_0'^2$ | $k'_0K_0$ | $ik'_0K'_0$ | $k'_0K'_0$ | $ik'_0K_0$. |

(30)

The problem that remains is to determine the correct signs on the square roots in (27). All three functions are odd functions of $u$. It is sufficient to describe the signs that pertain to positive $u$, on either axis.

On the real axis ns $u$ and ds $u$ resemble cosecants and thus are positive between $u = 0$ and $u = 2\omega_x$ with $\mathrm{ps}(2\omega_x - u) = \mathrm{ps}\,u$, where p = n or d, whereas cs $u$ is cotangentlike and positive for $0 < u < \omega_x$, negative for $\omega_x < u < 2\omega_x$. On the imaginary axis the three functions are purely imaginary but ns $u$ is the one that resembles a cotangent while cs $u$ and ds $u$ are cosecantlike. These rules suffice to determine the signs in (27) and so for all 12 pq $u = \mathrm{ps}\,u/\mathrm{qs}\,u$, where p and q are distinct choices of c, d, n, and s.

In addition to restricting attention to real $g_2$ and $g_3$, it has been assumed so far that $z$ is either real or simply imaginary. But, as with circular functions, values for complex arguments are calculated merely by finding those for the real and imaginary components separately and combining the results by means of the appropriate addition theorem. Thus for $z = x + iy$ one finds $\wp z$ from

$$w_0(x + iy) = \frac{1}{4}\left[\frac{w_0'(x) - w_0'(iy)}{w_0(x) - w_0(iy)}\right]^2 - w_0(x) - w_0(iy),$$
$$w_0' = (4w_0^3 - g_2w_0 - g_3)^{1/2}. \tag{31}$$

For the three ps $u$ (p = c,d,n) the addition formula is

$$\mathrm{ps}(a + b) = \frac{\mathrm{ps}\,a\,\mathrm{ps}'\,b - \mathrm{ps}'\,a\,\mathrm{ps}\,b}{\mathrm{ps}^2\,a - \mathrm{ps}^2\,b},$$
$$\mathrm{ps}'\,u = -\mathrm{qs}\,u\,\mathrm{rs}\,u, \tag{32}$$

where p, q, r is any permutation of c, d, n. The denominator $\mathrm{ps}^2\,a - \mathrm{ps}^2\,b$ is the same for all three choices of $p$. These formulas are useful also when a term in (27) is so close to vanishing that a more reliable value for ps $u$ would be obtained by expanding in a series about $\omega_x$ or $\omega_y$. Incidentally the form in (32) for $f(a + b)$, but not that for $f'(u)$, of course, holds as well for the circular functions that have a pole at the origin, i.e., $\cot(a + b)$ and $\csc(a + b)$.

[1] T. H. Southard, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, DC, 1970), p. 629.
[2] L. M. Milne-Thomson, in Ref. 1, p. 598.
[3] B. C. Carlson, J. Math. Phys. (Cambridge, Mass.) **44**, 36 (1965).
[4] See Ref. 3, p. 46.
[5] E. Jahnke, F. Emde, and F. Lösch, *Tables of Functions* (McGraw-Hill, New York, 1960), p. 86.
[6] See Ref. 5, p. 81.
[7] A. C. Dixon, *Elliptic Functions* (Macmillan, New York, 1894), p. 16.

# On the spectra of hypoelliptic operators in $R^2$

Terence Brenner
*Department of Mathematics, Hostos Community College of CUNY, 475 Grand Concourse, Bronx, New York 10451*

(Received 7 March 1988; accepted for publication 5 October 1988)

Sufficient conditions are given for the spectra and essential spectra of certain classes of operators in $R^2$ to be contained in an interval of the form $[d,\infty)$.

## I. INTRODUCTION

We investigate operators of the form

$$H=\sum_{k=1}^{e} f_k(x)(-\Delta)^k+q(x)$$

in $R^2$, where $\Delta$ is the Laplacian, $e$ is a positive integer greater than 1, $f_k(x)$ are real-valued functions in $R^2$ and $f_k\subset C^\infty$ for $k=1,2,\ldots,e$, and $q(x)$ is a real-valued function in $R^2$. This implies that $H$ has a self-adjoint extension $H$ that is bounded from below. For a self-adjoint operator the essential spectrum consists of those points in the spectrum that are not isolated eigenvalues of finite multiplicity.[1]

We are interested in finding (i) a lower bound for the spectrum $\sigma(H)$ of $H$, and (ii) a lower bound for the essential spectrum $\sigma_e(H)$ of $H$.

In the next section we present our main results. We shall prove these results, and the preliminary lemmas and theorems as well, in the last two sections.

Previous results can be found in Birman,[2] Brenner,[3,4] Glazman,[5] Muller-Pfeiffer,[6,7] Naimark,[8] Schechter,[1,9] and Yafaev.[10]

## II. STATEMENT OF RESULTS

We define

$$q_+(x)=\max(q(x),0),$$

$$q_-(x)=\max(-q(x),0),\quad q(x)=q_+(x)-q_-(x),$$

$$H_0=\sum_{k=1}^{e} f_k(x)(-\Delta)^k+q_+(x),$$

$$H=\sum_{k=1}^{e} f_k(x)(-\Delta)^k+q(x),$$

$\gamma\in D(H_0)$ if and only if $\gamma\in D(H)$, and $Q_{t,y}=I_1\times I_2$ is a square in $R^2$ centered at $y$ with edge length $t$.

We present our first theorem as follows.

**Theorem 1:** Letting $t>0$, if there exist constants $a_k$ such that $f_k(x)\geqslant a_k\geqslant 0$, for $k=1,2,\ldots,e$, and if there exists

$$0<g\leqslant t\sum_{k=1}^{e} a_k\pi^{2k}(2t)^{-2k},$$

such that for each $Q_{t,y}\in R^2$, $\int_{I_i} q_+(x)dx_i\geqslant g$, for $i=1,2$, then $H_0$ has a self-adjoint extension $H_0$ such that $\sigma(H_0)\subset[g/4t,\infty)$.

*Corollary 1:* If in addition $q_-(x)\leqslant g/4t-d$, where $g/4t\geqslant d>0$, then $H$ has a self-adjoint extension $H$ such that $\sigma(H)\subset[d,\infty)$.

**Theorem 2:** Under the hypothesis of Theorem 1, if

$$\liminf_{|y|\to\infty}\int_{I_i} q_+(x)dx_i\geqslant\lambda_i,$$

for $i=1,2$, $\lambda=\min(\lambda_1,\lambda_2)$, and $C\geqslant 1$ such that

$$g\leqslant\frac{\lambda}{C}\leqslant t\sum_{k=1}^{e} a_k\pi^{2k}(2t)^{-2k},$$

then $H_0$ has a self-adjoint extension $H_0$ such that $\sigma_e(H)\subset[\lambda/4Ct,\infty)$.

*Corollary 5:* If in addition $q_-(x)\leqslant\lambda/4Ct-d$, where $\lambda/4Ct\geqslant d>0$, then $H$ has a self-adjoint extension $H$ such that $\sigma_e(H)\subset[d,\infty)$.

For example,

$$H=(5x^8+y^4+3)(-\Delta)^r+(0.5x^8+6y^2+5)(-\Delta)^v-\Delta+x^2+y^2+1,$$

where $r>v>1$ are positive integers. If we use $t=1$, we get the following estimates:

$$\sigma(H)\subset[\tfrac{13}{48},\infty),$$

$$\sigma_e(H)\subset[3\pi^{2r}(4)^{-r-1}+5\pi^{2v}(4)^{-v-1}+\pi^2(4)^{-2},\infty).$$

## III. PRELIMINARY LEMMAS AND THEOREMS

*Lemma 1:* Let $I$ be an interval in $R$ of length $t$ and put $\|f\|_I=\int_I|f|^2\,dx$. If $f=0$ somewhere in $I$, then there exists $C_0$ such that $C_0\|f\|_I^2\leqslant\|f'\|_I^2$, where $C_0=\pi^2/4t^2$.

*Lemma 2:* For $a,b$ non-negative,

$$a\|u\|^2+b\|v\|^2\geqslant\tfrac{1}{2}\min(a,b)\|u+v\|^2.$$

*Lemma 3:* Let $Q$ be a square in $R^2$, and suppose there exists $R$ such that

$$\left(\sum_{k=1}^{e} f_k(x)(-\Delta)^k u,u\right)_Q+(q_+u,u)_Q\geqslant R\|u\|_Q^2.$$

If $q_-(x)\leqslant R-\epsilon$, where $R\geqslant\epsilon>0$, then

$$\left(\sum_{k=1}^{e} f_k(x)(-\Delta)^k u,u\right)_Q+(qu,u)_Q\geqslant\epsilon\|u\|^2.$$

*Lemma 4:* Let $P$ be a self-adjoint operator on $R^N$ with $C_0^\infty\subset D(P)$. If there exists $R>0$ such that $(P\varphi,\varphi)\geqslant\lambda\|\varphi\|^2$, $\forall\,\varphi\in C_0^\infty$ that vanish in $Q_{R,0}$, then $\sigma_e(P)\subset[\lambda,\infty)$.

Lemmas 1, 2, and 4 are proved in Ref. 3. The proof of Lemma 3 is similar to the proof of Lemma 3 in Ref. 3.

**Theorem 3:** Let $t>0,(y_1,y_2)\in R^2$. If there exist constants $a_k$ such that $f_k(x)\geqslant a_k\geqslant 0$, for $k=1,2,\ldots,e$, and if there exists

$$0<g\leqslant t\sum_{k=1}^{e} a_k\pi^{2k}(2t)^{-2k},$$

such that $\int_{I_i} q_+(x)dx_i\geqslant g$, $\forall y_i\in I_i$, for $i=1,2$, then

 0022-2488/89/020298-04$02.50 © 1989 American Institute of Physics 

(i) $\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q_+\right]u,u\right)_{Q_{t,y}} \geqslant \frac{g}{4t}\|u\|^2_{Q_{t,y}}$.

If in addition $q_-(x) \leqslant g/4t - d$, where $g/4t \geqslant d > 0$, then

(ii) $\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q\right]u,u\right)_{Q_{t,y}} \geqslant d\,\|u\|^2_{Q_{t,y}}$.

*Proof:*

Let

$$D_1^i u = \frac{\partial^i u}{\partial x_1^i}, \quad D_2^i u = \frac{\partial^i u}{\partial x_2^i},$$

for $i = 1,2,\ldots,e$, $Q = Q_{t,y}$, $Q = I_1 \times I_2$,

$$u_1^2(x_2) = \min_{x_1 \in I_1} u^2(x_1,x_2),$$

and

$$u_2^2(x_1) = \min_{x_2 \in I_2} u^2(x_1,x_2).$$

Then

$$\|u_i\|^2_{I_i} = \int_{I_i} u_i^2\,dx_i = t u_i^2 \Rightarrow u_i^2 = \frac{\|u_i\|^2_{I_i}}{t},$$

for $i = 1,2$. Then

$$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q_+(x)\right]u,u\right)_Q$$

$$\geqslant \left(\left[\sum_{k=1}^{e} a_k(-\Delta)^k + q_+(x)\right]u,u\right)_Q$$

$$\geqslant \int_{I_2}\int_{I_1} \sum_{k=1}^{e} a_k|D_1^k u|^2 + a_k|D_2^k u|^2 + q_+|u|^2\,dx_1\,dx_2.$$

First for $D_1^1$ and then $D_2^1$,

$$\geqslant \int_{I_2}\int_{I_1} \sum_{k=1}^{e} a_k|D_1^k(u \pm u_1)|^2 + q_+ u_1^2\,dx_1\,dx_2 + \int_{I_2}\int_{I_1} \sum_{k=1}^{e} a_k|D_2^k u|^2\,dx_1\,dx_2,$$

either $u + u_1$ or $u - u_1$ will vanish somewhere in $I_1$,

$$= \int_{I_2}\int_{I_1} \sum_{k=2}^{e} a_k|D_1^k(u \pm u_1)|^2 + a_1|D_1^1(u \pm u_1)|^2\,dx_1\,dx_2 + \text{''}$$

$$\geqslant \int_{I_2}\int_{I_1} \sum_{k=2}^{e} a_k C_0^{k-1}|D_1^1(u \pm u_1)|^2 + a_1|D_1^1(u \pm u_1)|^2 + q_+ u_1^2\,dx_1\,dx_2 + \text{''}$$ [repeated use of Lemma 1]

$$= \int_{I_2}\int_{I_1} \sum_{k=1}^{e} a_k C_0^{k-1}|D_1^1(u \pm u_1)|^2 + q_+ u_1^2\,dx_1\,dx_2 + \text{''}$$

$$\geqslant \int_{I_2} \sum_{k=1}^{e} a_k C_0^k \|u \pm u_1\|^2_{I_1} + u_1^2 \int_{I_1} q_+\,dx_1\,dx_2 + \text{''}$$ [Lemma 1]

$$\geqslant \int_{I_2} \sum_{k=1}^{e} a_k C_0^k \|u \pm u_1\|^2_{I_1} + u_1^2 g\,dx_2 + \text{''}$$

$$= \int_{I_2} \sum_{k=1}^{e} a_k C_0^k \|u \pm u_1\|^2_{I_1} + \frac{g}{t}\|u \pm u_1\|^2_{I_1}\,dx_2 + \text{''}$$

$$\geqslant \int_{I_2} \frac{g}{2t}\|u\|^2_{I_1}\,dx_2 + \text{''}$$ [Lemma 2]

$$= \int_{I_1}\int_{I_2} \frac{g}{2t}|u|^2 + \sum_{k=1}^{e} a_k|D_2^k u|^2\,dx_2\,dx_1$$

$$\geqslant \int_{I_1}\int_{I_2} \frac{g}{2t}|u|^2 + \sum_{k=1}^{e} a_k|D_2^k(u \pm u_2)|^2\,dx_2\,dx_1,$$

either $u + u_2$ or $u - u_2$ will vanish somewhere in $I_2$,

$$= \int_{I_1}\int_{I_2} \frac{g}{2t}u_2^2 + \sum_{k=2}^{e} a_k|D_2^k(u \pm u_2)|^2 + a_1|D_2^1(u \pm u_2)|^2\,dx_2\,dx_1$$

$$\geqslant \int_{I_1}\int_{I_2} \frac{g}{2t}u_2^2 + \sum_{k=2}^{e} a_k C_0^{k-1}|D_2^1(u \pm u_2)|^2 + a_1|D_2^1(u \pm u_2)|^2\,dx_2\,dx_1$$

[repeated use of Lemma 1]

$$= \int_{I_1}\int_{I_2} \frac{g}{2t}u_2^2 + \sum_{k=1}^{e} a_k C_0^{k-1}|D_2^1(u \pm u_2)|^2\,dx_2\,dx_1$$

$$\geqslant \int_{I_1} u_2^2 \int_{I_2} \frac{g}{2t}\,dx_2 + \sum_{k=1}^{e} a_k C_0^k \|u \pm u_2\|^2_{I_2}\,dx_1$$ [Lemma 1]

$$= \int_{I_1} \frac{g}{2t}\|u_2\|^2_{I_2} + \sum_{k=1}^{e} a_k C_0^k \|u \pm u_2\|^2_{I_2}\,dx_1$$

$$\geqslant \int_{I_1} \frac{g}{4t}\|u\|^2_{I_2}\,dx_1$$ [Lemma 2]

$$= \frac{g}{4t}\|u\|^2_Q.$$

Now for $D_2^1$ and then $D_1^1$,

$$\int_{I_1}\int_{I_2} \sum_{k=1}^{e} a_k|D_1^k u|^2 + a_k|D_2^k u|^2 + q_+|u|^2\,dx_2\,dx_1$$

$$\geqslant \int_{I_1}\int_{I_2} \sum_{k=1}^{e} a_k|D_2^k(u \pm u_2)|^2 + q_+ u_2^2\,dx_2 + \int_{I_1}\int_{I_2} \sum_{k=1}^{e} a_k|D_1^k u|^2\,dx_2\,dx_1,$$

either $u + u_2$ or $u - u_2$ will vanish somewhere in $I_2$,

$$= \int_{I_1}\int_{I_2} \sum_{k=2}^{e} a_k|D_2^k(u \pm u_2)|^2 + a_1|D_2^1(u \pm u_2)|^2 + q_+ u_2^2\,dx_2\,dx_1 + \text{''}$$

$$\geqslant \int_{I_1}\int_{I_2} \sum_{k=2}^{e} a_k C_0^{k-1}|D_2^1(u \pm u_2)|^2 + a_1|D_2^1(u \pm u_2)|^2 + q_+ u_2^2\,dx_2\,dx_1 + \text{''}$$

[repeated use of Lemma 1]

$$= \int_{I_1}\int_{I_2} \sum_{k=1}^{e} a_k C_0^{k-1}|D_2^1(u \pm u_2)|^2 + q_+ u_2^2\,dx_2\,dx_1 + \text{''}$$

$$\geqslant \int_{I_1} \sum_{k=1}^{e} a_k C_0^k \|u \pm u_2\|^2_{I_2} + u_2^2 \int_{I_2} q_+(x)\,dx_2\,dx_1 + \text{''}$$ [Lemma 1]

$$\geqslant \int_{I_1} \sum_{k=1}^{e} a_k C_0^k \|u \pm u_2\|_{I_2}^2 + u_2^2 g \, dx_1 + ''$$

$$= \int_{I_1} \sum_{k=1}^{e} a_k C_0^k \|u \pm u_2\|_{I_2}^2 + \frac{g}{t}\|u_2\|_{I_2}^2 \, dx_1 + ''$$

$$\geqslant \int_{I_1} \frac{g}{2t}\|u\|_{I_2}^2 \, dx_1 + '' \quad \text{[Lemma 2]}$$

$$= \int_{I_2}\int_{I_1} \frac{g}{2t}|u|^2 + \sum_{k=1}^{e} a_k |D_1^k u|^2 \, dx_1 \, dx_2$$

$$\geqslant \int_{I_2}\int_{I_1} \frac{g}{2t}u_1^2 + \sum_{k=1}^{e} a_k |D_1^k (u \pm u_1)|^2 \, dx_1 \, dx_2,$$

either $u + u_1$ or $u - u_1$ will vanish somewhere in $I_1$,

$$= \int_{I_2}\int_{I_1} \frac{g}{2t}u_1^2 + \sum_{k=2}^{e} a_k |D_1^k (u \pm u_1)|^2 + a_1 |D_1^1 (u \pm u_1)|^2 \, dx_1 \, dx_2$$

$$\geqslant \int_{I_2}\int_{I_1} \frac{g}{2t}u_1^2 + \sum_{k=2}^{e} a_k C_0^{k-1} |D_1^1 (u \pm u_1)|^2 + a_1 |D_1^1 (u \pm u_1)|^2 \, dx_1 \, dx_2$$

[repeated use of Lemma 1]

$$= \int_{I_2}\int_{I_1} \frac{g}{2t}u_1^2 + \sum_{k=1}^{e} a_k C_0^{k-1} |D_1^1 (u \pm u_1)|^2 \, dx_1 \, dx_2$$

$$\geqslant \int_{I_2} u_1^2 \int_{I_1} \frac{g}{2t} \, dx_1 + \sum_{k=1}^{e} a_k C_0^k \|u \pm u_1\|_{I_1}^2 \, dx_2 \quad \text{[Lemma 1]}$$

$$= \int_{I_2} \frac{g}{2t}\|u_1\|_{I_1}^2 + \sum_{k=1}^{e} a_k C_0^k \|u \pm u_1\|_{I_1}^2 \, dx_2$$

$$\geqslant \int_{I_2} \frac{g}{4t}\|u\|_{I_1}^2 \, dx_2 \quad \text{[Lemma 2]}$$

$$= \frac{g}{2t}\|u\|_Q^2.$$

Hence

$$\int_{I_1}\int_{I_2} \sum_{k=1}^{e} a_k |D_1^k u|^2 + a_k |D_2^k u|^2 + q_+ |u|^2 \, dx_2 \, dx_1 \geqslant \frac{g}{4t}\|u\|_Q^2.$$

Part 2 follows from Lemma 3.

**Theorem 4:** Let $t > 0$. If there exist constants $a_k$ such that $f_k(x) \geqslant a_k \geqslant 0$ for $k = 1,2,...,e$, and if there exists

$$0 < g \leqslant t \sum_{k=1}^{e} a_k \pi^{2k} (2t)^{-2k},$$

such that for each $Q_{t,y} \in R^2$, $\int_{I_1} q_+(x)dx_i \geqslant g$, $\forall y_i \in I_i$ for $i = 1,2$, then

(i) $$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q_+\right]u,u\right)_{Q_{t,y}} \geqslant \frac{g}{4t}\|u\|_{Q_{t,y}}^2.$$

If in addition $q_-(x) \leqslant g/4t - d$, where $g/4t \geqslant d > 0$, then

(ii) $$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q\right]u,u\right)_{Q_{t,y}} \geqslant d\,\|u\|_{Q_{t,y}}^2.$$

The proof follows from Theorem 3.

## IV. PROOFS OF MAIN THEOREMS

### A. Theorem 1

By Theorem 4, for each $Q_{t,y}$,

$$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q_+\right]u,u\right)_{Q_{t,y}} \geqslant \frac{g}{4t}\|u\|_{Q_{t,y}}^2, \quad \forall y \in R^2.$$

If we cover $E^2$ with such squares and sum, then

$$(g/4t)\|u\|^2 \leqslant (H_0 u,u) \Rightarrow \sigma(H_0) \subset [g/4t, \infty).$$

(See Ref. 9.)

### B. Corollary 1

By Theorem 4, part 2, for each $Q_{t,y}$,

$$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q\right]u,u\right)_{Q_{t,y}} \geqslant d\,\|u\|_{Q_{t,y}}^2, \quad \forall y \in R^2.$$

If we cover $E^2$ with such squares and sum, then

$$d\,\|u\|^2 \leqslant (Hu,u) \Rightarrow \sigma(H) \subset [d, \infty).$$

(See Ref. 9.)

### C. Theorem 2

Let $\epsilon > 0$ be given, and take $R$ so large that

$$\int_{I_i} q_+(x)dx_i \geqslant \frac{\lambda_i - \epsilon}{C},$$

for $i = 1,2$, and

$$\frac{\lambda - \epsilon}{C} = \min\left(\frac{\lambda_1 - \epsilon}{C}, \frac{\lambda_2 - \epsilon}{C}\right),$$

for $y$ outside the square $Q_{R,0}$. Then by Theorem 3,

$$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q_+\right]u,u\right)_{Q_{t,y}} \geqslant \frac{\lambda - \epsilon}{4Ct}\|u\|_{Q_{t,y}}^2 \quad \text{[for such } y\text{]}$$

$$\Rightarrow \sigma_e(H_0) \subset \left[\frac{\lambda - \epsilon}{4Ct}, \infty\right) \quad \text{[Lemma 4]}.$$

Since this is true for every $\epsilon \geqslant 0$,

$$\Rightarrow \sigma_e(H_0) \subset \left[\frac{\lambda}{4Ct}, \infty\right).$$

### D. Corollary 2

Let $\epsilon > 0$ be given, $q_-(x) \leqslant \lambda/4Ct - (d - \epsilon)$. Then by Theorem 3, part 2,

$$\left(\left[\sum_{k=1}^{e} f_k(x)(-\Delta)^k + q\right]u,u\right)_{Q_{t,y}} \geqslant (d - \epsilon)\|u\|_{Q_{t,y}}^2$$

[for $y$ outside $Q_{R,0}$] $\Rightarrow \sigma_e(H) \subset [d - \epsilon, \infty)$ [Lemma 4].

Since this is true for every $\epsilon \geqslant 0$,

$$\Rightarrow \sigma_e(H) \subset [d, \infty).$$


[1] M. Schechter, *Spectra of Partial Differential Operators* (North-Holland, Amsterdam, 1971).

[2] M. S. Birman, "The spectrum of singular boundary problems," Math. USSR-Sb 55, 125 (1961).

[3] T. Brenner, Ph.D. thesis, Yeshiva University, 1984.

[4] T. Brenner, "On the spectra of certain singular elliptic operators in $R^2$," Hostos Community College Mathematics Journal, Spring 1987.

[5]I. M. Glazman, *Direct Methods of Qualitative Spectral Analysis of Singular Differential Operators* (Israel Program of Scientific Translations, Jerusalem, 1965).
[6]E. Muller-Pfeiffer, "Spektraleigan-Schaften eindimensionaler Differential-operatoren hoherer Ordnung," Studia Math. **24**, 183 (1970).
[7]E. Muller-Pfeiffer, "Eine Bemerkung uber das Spektum des Schrodinger operators," Math. Nach. **58**, 299 (1973).
[8]M. A. Naimark, *Linear Differential Operators, Part II* (Ungar, New York, 1968).
[9]M. Schechter, "On the spectra of singular elliptic operators," Mathematika **23**, 107 (1976).
[10]D. R. Yafaev, "On the spectrum of the perturbed polyharmonic operator," *Topics in Mathematical Physics*, Vol. 5, edited by M. S. Birman (Plenum, New York, 1972), pp. 107–112.

# Solution to boundary value problems using the method of maximum entropy

James Baker-Jarvis
*Physics Department, North Dakota State University, Fargo, North Dakota 58105*

(Received 26 May 1988; accepted for publication 21 September 1988)

The method of maximum entropy is used to solve a class of linear boundary value problems. The method is based on using various moments of the differential equation as constraints when maximizing the entropy. Various examples are presented and compared to exact solutions for varying numbers of moments. It is found that the maximum entropy approximation is, in many cases, better than a Fourier series solution for a given number of expansion terms and moments. The method is very general and will find applications in many areas of physics. A comparison of the amount of work necessary for the maximum entropy solution versus finite difference techniques is presented and it is found that the maximum entropy technique shows promise as an alternative solution technique.

## I. INTRODUCTION

The concept of information entropy was presented by Shannon[1] to describe the information content in a signal. Following this pioneering work, Jaynes[2,3] developed the information theoretic approach to statistical physics. Since then, there have been many novel applications of the principle of maximum entropy in a number of areas.[1-9] For example, Burg[4,9] performed spectral analysis using maximum entropy techniques, Mead[5,6] has solved integral equations with this technique. Inguva and Baker-Jarvis[7] have studied the generalized inverse scattering method in the context of maximum entropy methods (MEM). The literature is extensive and a good bibliography is given in a paper by Smith and Inguva.[8]

Generally, when solving a linear boundary value problem with classical techniques, a Fourier series is utilized. However, Fourier series approaches can be problematic when the equation contains products of functions since the Fourier transform of products is not, in general, useful. Thus in obtaining solutions to general linear differential equations with nonconstant coefficients one generally must resort to finite difference or finite element techniques. These techniques require large-order matrices to be inverted and the amount of computation in the solution increases as $N^d$, where $N$ is the number of points and $d$ is the dimensionality. In this paper the method of maximum entropy will be used for the solution of certain classes of linear boundary value problems. The goal of this paper is to describe a general method by which maximum entropy can be utilized to solve many linear differential boundary value problems. The method should have applications in many areas of physics, for example, in solving the Fokker–Planck equation and problems where the boundary conditions or source functions are uncertain or noisy. The method is very general and can be thought of as a generalization of Fourier series. The method is based on previous work of the author in integral equations that has been generalized to the problem of a finite interval.[7] The advantage of this technique is its generality and efficiency. To the author's knowledge this technique has not been utilized previously for this problem. There have been a multitude of approaches used in applying the maximum entropy technique to physical problems. These approaches can, however, be classified into two main categories: (i) techniques that equate the probability function with the function to be determined and (ii) techniques where the expectation value of a function is determined using the probability distribution found from maximum entropy. For example, in Mead's[5] work the function itself is identified with the probability and thus only positive quantities can be utilized. He tries to overcome this characteristic by a change of variables to make all of the variables positive; however, it is a severe limitation. In the present theory the function can be either positive or negative, which is important since generally one does not know *a priori* whether the function takes on negative values. The method is presented in Secs. II–IV and various examples are presented in the next section. The maximum entropy solution technique is compared to the Fourier series solutions in Sec. V. The final section is a discussion of the technique's merits.

## II. DEVELOPMENT OF CONSTRAINT CONDITIONS

We consider a general linear, boundary value problem given in operator form as

$$L_1\langle V(r)\rangle = C(r)\,, \tag{1}$$

together with either Neumann or Dirichelet boundary conditions. Here, $\langle V\rangle$ denotes the expectation value. In Eq. (1) $L_1$ is a differential operator, $C$ is a source function, and $V$ is the function to be determined. In the solution technique we require an associated set of eigenfunctions that satisfy a differential equation of the form

$$L_2 f_n(r) = -\beta_n f_n(r)\,, \tag{2}$$

where $L_2$ is a differential operator and the $\beta_n$ are eigenvalues. We pick the set of eigenfunctions judiciously depending on the operator form of Eq. (1), the boundary conditions, and coordinate system utilized. The method of maximum entropy requires the specification of constraints. We form the constraints for the maximum entropy method by multiplying Eq. (1) by an eigenfunction satisfying Eq. (2) and then integrating over space:

$$\int f_n L_1\langle V(r)\rangle - C(r)d^3r = 0\,, \tag{3}$$

 0022-2488/89/020302-05$02.50 © 1989 American Institute of Physics 

for $n = 1,2,3,\ldots$ . Integration by parts and use of boundary conditions implies that

$$\int \alpha_n(r)\langle V(r)\rangle d^3r = A_n\,, \tag{4}$$

where $A_n$ and $\alpha_n$ contain boundary information and information on $C(r)$. The problem then reduces to maximizing the entropy subject to the constraints given in Eq. (4). The maximum of the entropy then gives the probability distribution subject to the available information. The problem of showing that the MEM solution converges to the solution of the differential equation is related to the classical moment problem. Mead[6] has studied the Hausdorff conditions necessary for convergence of the moment problem extensively in the context of MEM.

## III. THE ENTROPY FORMALISM

We consider the function $V(r)$ as represented by a discrete set of $N$ points $V(r_1) = V_1$, $V(r_2) = V_2,\ldots,V(r_N) = V_N$, so that in vector notation we can define

$$\mathbf{V} = (V_1,V_2,V_3,\ldots,V_N)^t\,. \tag{5}$$

Here $(t)$ denotes transpose. We define the information entropy as

$$S = -\int_V P(\mathbf{V})\ln(P(\mathbf{V}))d\mathbf{V}\,, \tag{6}$$

and $\int_V d\mathbf{V}$ denotes $\int dV_1 \int dV_2 \int dV_3\cdots dV_N$. Note that, in general, the entropy is a functional integral. We assume that the distribution is normalized so that

$$\int P(\mathbf{V})d\mathbf{V} = 1\,. \tag{7}$$

We define the expectation value of a function $V_j$ as

$$\langle V_j\rangle = \int P(\mathbf{V})V_j\,d\mathbf{V}\,. \tag{8}$$

We can write Eq. (4) as

$$\mathsf{B}\langle\mathbf{V}\rangle = \mathbf{A}\,, \tag{9}$$

where we discretized the integral

$$\mathsf{B} = [\alpha_{ki}]\,, \tag{10}$$

$$\mathbf{A} = [A_1,\ldots,A_M]^t\,. \tag{11}$$

In component form we can write Eq. (9) as

$$\sum_{i=1}^{N}\alpha_{ni}\langle V_i\rangle = A_n\,, \tag{12}$$

where $\alpha_{ni} = \alpha_n(r_i)\Delta r$, and $\Delta r$ is the grid size for a direct Riemann sum.

The entropy with $M$ constraints can be written as

$$\begin{aligned} S &= \int \{-P(\mathbf{V})\ln(P(\mathbf{V})) \\ &\quad -\lambda_0[P(\mathbf{V})-1] - [\lambda^t\mathbf{B}\mathbf{V}P(\mathbf{V}) - \mathbf{A}]\}d\mathbf{V} \\ &= \int\Big\{-P(\mathbf{V})\ln(P(\mathbf{V})) - \lambda_0[P(\mathbf{V})-1] \\ &\quad -\sum_{n=1}^{M}\lambda_n\Big[\sum_{j=1}^{N}[\alpha_{nj}V_jP(\mathbf{V})] - A_n\Big]\Big\}d\mathbf{V}\,, \end{aligned} \tag{13}$$

where $\lambda_i$ are Lagrange multipliers, $N$ is the number of points, and $M$ is the number of moments,

$$\boldsymbol{\lambda} = [\lambda_1,\ldots,\lambda_m]^t\,. \tag{14}$$

Performing a variation of $P$ yields

$$\delta S = \int\Big[-1-\ln(P(\mathbf{V})) - \lambda_0 - \sum_j \Gamma_jV_j\Big]\delta P\,d\mathbf{V}\,, \tag{15}$$

where we have defined

$$\Gamma_j = \sum_{n=1}^{M}\alpha_{nj}\lambda_n\,. \tag{16}$$

We then obtain

$$P(\mathbf{V}) = \exp[-(1+\lambda_0)]\exp[-\boldsymbol{\lambda}^t\mathbf{B}\mathbf{V}] \tag{17}$$

$$= \exp[-(1+\lambda_0)]\exp\Big[-\sum_j\Gamma_jV_j\Big]\,. \tag{18}$$

If we define the partition function by using Eq. (7)

$$Z = \exp[(1+\lambda_0)]\,, \tag{19}$$

we then obtain

$$P(\mathbf{V}) = \frac{\exp(-\boldsymbol{\lambda}^t\mathbf{B}\mathbf{V})}{Z} = \exp\Big[-\sum_j\Gamma_jV_j\Big]Z^{-1}\,. \tag{20}$$

Now by Eq. (8) we have

$$\langle V_i\rangle = \int_a^b dV_1\cdots dV_N\,V_i\exp\Big[-\sum\Gamma_jV_j\Big]Z^{-1}\,, \tag{21}$$

where $b$ and $a$ are the boundaries of the solution region. So, by integration we obtain the expectation value of $V_i$:

$$\langle V_i\rangle = \frac{1}{\Gamma_i} + \frac{b\exp(-\Gamma_ib) - a\exp(-\Gamma_ia)}{\exp(-\Gamma_ib) - \exp(-\Gamma_ia)}\,. \tag{22}$$

In the limit as $b\to\infty$, $a\to 0$, the second term in Eq. (22) drops out. Equation (22) can be expressed in terms of hyperbolic functions for various limiting cases. We can, for example, let $a = -b$ and obtain

$$\langle V_i\rangle = (1/\Gamma_i) - b\coth[b\Gamma_i]\,. \tag{23}$$

This form is particularly nice for numerical calculations when the function takes on both positive and negative values. For the case of positive functions ($z\in[0,b]$), Eq. (22) reduces to

$$\langle V_i\rangle = \frac{1}{\Gamma_i} - \frac{b}{2}\Big[\coth\Big(\frac{\Gamma_ib}{2}\Big) - 1\Big]\,. \tag{24}$$

For cases when $b\to\infty$ and $a\to -\delta$ ($\delta\geqslant 0$), we have

$$\langle V_i\rangle = (1/\Gamma_i) - \delta\,. \tag{25}$$

It is possible to convert the solution to a function of $r$ if we take the limit as $\Delta r\to 0$:

$$\begin{aligned} \langle V(r)\rangle &= \frac{1}{\Gamma(r)} \\ &\quad + \frac{b\exp[-\Gamma(r)b] - a\exp[-\Gamma(r)a]}{\exp[-\Gamma(r)b] - \exp[-\Gamma(r)a]}\,, \end{aligned} \tag{26}$$

where the Lagrange multipliers can be found from Eq. (4).

For the case of the integration interval $[a,b]$ $\to[-\delta,\infty]$ an exact solution for $\boldsymbol{\lambda}$ is possible. Let the vector $\mathbf{E}$ have components

$$E_i = 1/\{[(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}]_i + \delta\}, \quad (27)$$

where the $-1$ superscript indicates inverse. Therefore the solution is given by

$$\lambda = [\mathbf{B}\mathbf{B}']^{-1}\mathbf{B}\mathbf{E}. \quad (28)$$

For the other cases the Lagrange multipliers are found from Eq. (26) substituted into Eq. (12).

## IV. RELATION TO FOURIER SERIES

It is interesting to establish the relationship between the present method and the Fourier series. To this end assume that we can expand $V(z)$ in the Fourier series as

$$V(z) = \sum_m C_m f_m(z). \quad (29)$$

Now it is well known that a Fourier series expansion minimizes the $L_2$ norm

$$\int \left| V(z) - \sum_{m=1}^{M} C_m f_m(z) \right|^2 dz. \quad (30)$$

This is commonly called a least squares fit of the function. For each $M$ the $C_m$ are the best coefficients for the least squares fit. The MEM method finds the best function of the form of Eq. (22) that maximizes the entropy of the moments of the equation and interpolates the data given by the Fourier coefficients. It can be shown, when $[a,b] \to [-\infty,\infty]$, that the MEM solution with a constraint on the norm reduces exactly to Fourier series.

## V. SOLUTION OF DIFFERENTIAL EQUATIONS

Numerical solutions can be generated by solving Eq. (9) for Lagrange multipliers with $\langle V \rangle$ given by Eq. (22). The equations for Lagrange multipliers are a set of $M$ (number of moments) nonlinear equations. Once the Lagrange multipliers are found, Eq. (22) gives $\langle V \rangle$ [except for the case of Eq. (28) where $\lambda_i$ can be determined explicitly].

As a first example let us consider a very simple boundary value problem:

$$\frac{d^2V}{dz^2} = -3.2, \quad V(0) = V_0, \quad V(1) = V_1. \quad (31)$$

The constraints of Eq. (3) are in this case formed from the set of eigenfunctions on the interval [0,1],

$$f_n(z) = \sin(n\pi z), \quad n = 1,2,3,\ldots. \quad (32)$$

For this case we obtain

$$A_n = n\pi[V(1)\cos n\pi - V(0)] + (1/n\pi)[\cos n\pi - 1],$$

$$\alpha_{ni} = -(n\pi)^2 \sin n\pi z_i/N.$$

In Fig. 1 the exact solution to Eq. (31) is plotted against the maximum entropy solution. We see that even with only a few moments the MEM solution is very good. Furthermore, we see that the rate of convergence as the number of moments is increased is very good.

In the next example, a slightly more complicated differential equation is solved:

$$\frac{d^2V(z)}{dz^2} = -\beta V(z), \quad V(0) = V_0, \quad V(1) = V_1. \quad (33)$$

In this solution the eigenfunctions are given by Eq. (32). In Fig. 2 the exact solution and MEM solutions are plotted and compared with Fourier series solutions for two moments and for two Fourier expansion functions. We again see that the solution is quite good. We see that the solution with MEM is a better approximation for a given number of moments than the Fourier series for the same number of expansion functions as shown in Figs. 2 and 3.

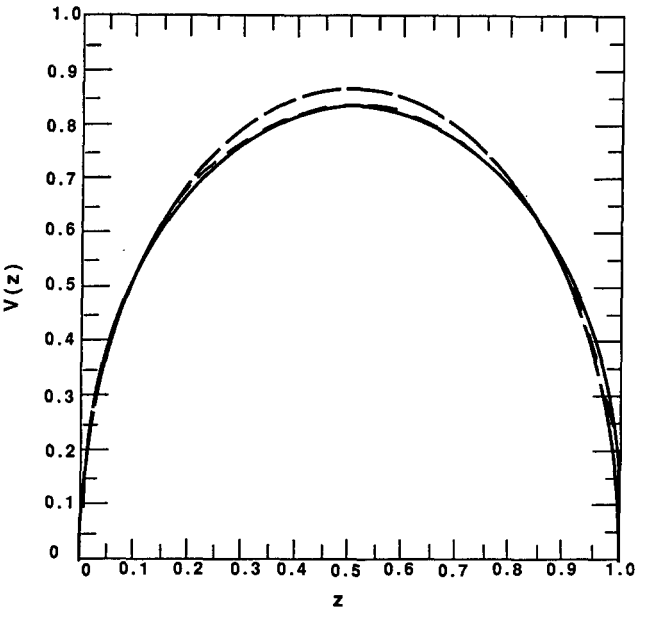



FIG. 1. Maximum entropy solution to Eq. (31) for two (– –) and ten (---) moments compared to the exact solution (—).

The next example is another step in complexity, here the functions are not eigenfunctions of the differential operator:

$$\frac{d^2V}{dz^2} + 2\frac{dV}{dz} + 2V = 0, \quad V(0) = V_0, \quad V(1) = V_1. \quad (34)$$

We use as moment functions those of Eq. (32). Again we see that the solution is reasonable in Fig. 4.

The technique can also be used for initial value problems and is shown in Fig. 5. In this case the function is

$$\frac{dV}{dz} = -V, \quad V(0) = V_0. \quad (35)$$

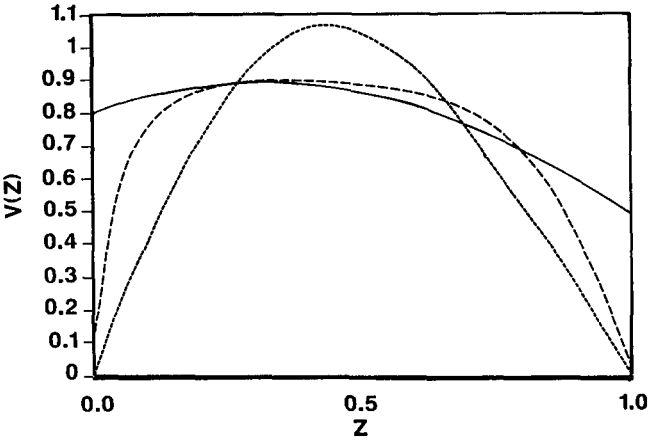



FIG. 2. Maximum entropy solution to Eq. (33) (– –) compared to the exact solution (—) and the Fourier series solution (----) for two moments and two expansion functions.

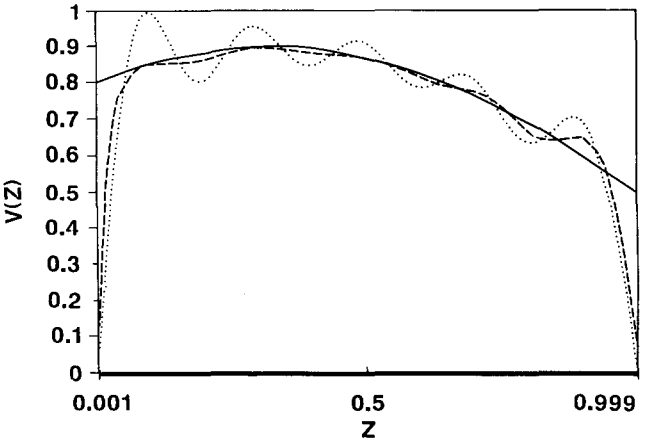



FIG. 3. Maximum entropy solution to Eq. (33) (---) compared to the exact solution (—) and the Fourier series solution (····) for ten moments and ten expansion functions.

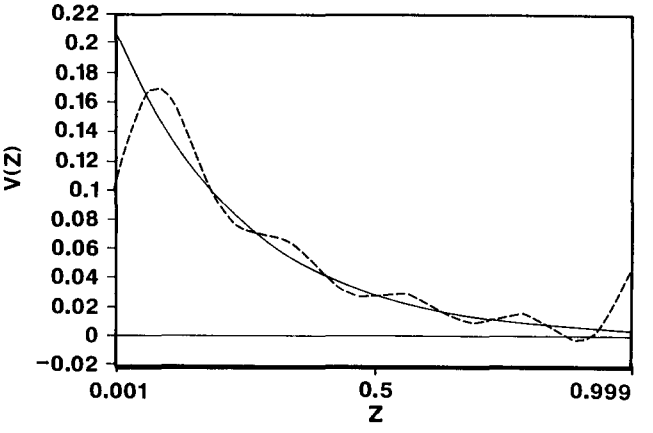



FIG. 5. Maximum entropy solution (----) for seven moments for Eq. (35) compared to the exact solution (—).

For general initial value problems in the interval $[0,\infty]$, the Laguerre polynomials form a complete set with weight function $\exp(-z)$. So in this case the moments of $V(z)$ are

$$\int_0^\infty \exp(-z)\mathscr{L}_n(z)[L_1V(z)-C(z)]dz=0\,. \quad (36)$$

Here $\mathscr{L}_n(z)$ are the Laguerre polynomials. For problems on the interval $[-\infty,\infty]$, the Hermite polynomials form a complete set with weight function $\exp(-z^2)$. As the next example, Mathieu's differential equation, which can be hard to solve using finite difference techniques, is solved using the MEM:

$$\frac{d^2V}{dz^2}+\eta\cos(\phi z)V=0\,, \quad (37)$$

where $\phi$ and $\eta$ are constants. In Fig. 6 the MEM solution is plotted together with the solution obtained by a finite difference solution. We see that the MEM solution does quite well. As a final example we consider the parabolic cylinder equation

$$\frac{d^2V}{dz^2}=-z^2V\,. \quad (38)$$

The exact solution by finite difference techniques is plotted in Fig. 7 versus the MEM solution.

The maximum entropy solution, of course, suffers the same pathologies as Fourier series solutions due to the nature of moment functions. For example, the moment functions $\sin(n\pi z)$ are zero at the end points, so that the solution is not good near the end points. However, the same techniques used to overcome these difficulties in Fourier series can be exploited in the MEM approach. A common technique is to write the solution as

$$V(z)=V'(z)+V''(z)\,, \quad (39)$$

where $V'(z)$ satisfies $d^2V'/dz^2=0$ and the boundary conditions. Then $V''(z)$ is solved for by the MEM method.

The values used for $b$ and $a$ in the probability integral of Eq. (21) influence the actual solution. The closer $[a,b]$ coincides with the actual limits of the function, the better MEM results (see Fig. 8). The MEM method truncates the probability distribution outside the limits $[a,b]$, thus by specifying

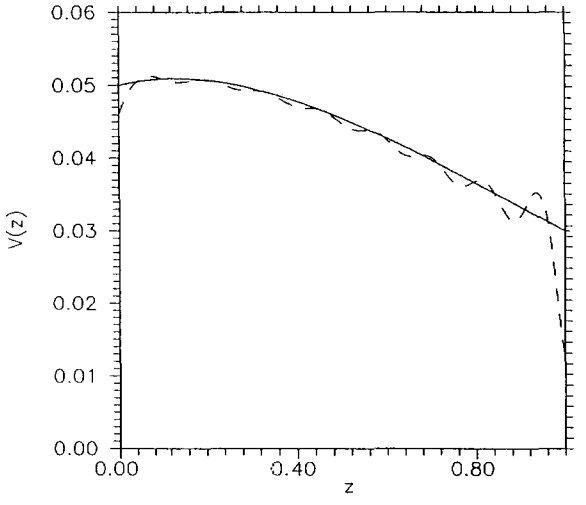



FIG. 4. Maximum entropy solution (----) for ten moments for Eq. (34) compared to the exact solution (—).

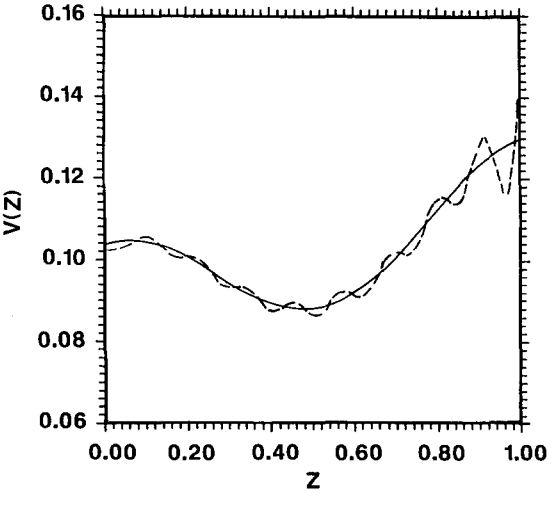



FIG. 6. Maximum entropy solution (----) for 12 moments for Eq. (37) compared to the exact solution (—).

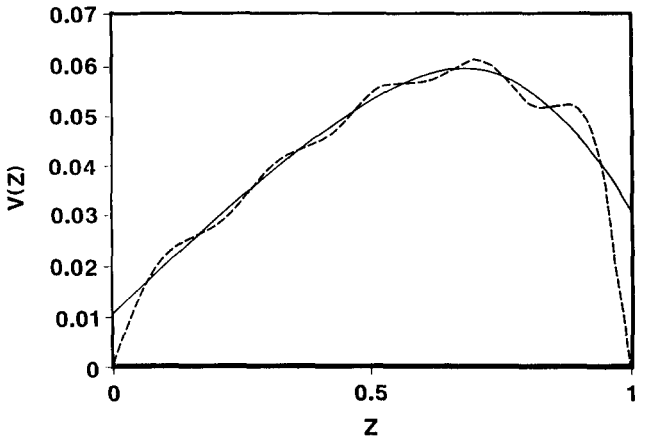



FIG. 7. Maximum entropy solution (----) for eight moments for Eq. (38) compared to the exact solution (—).

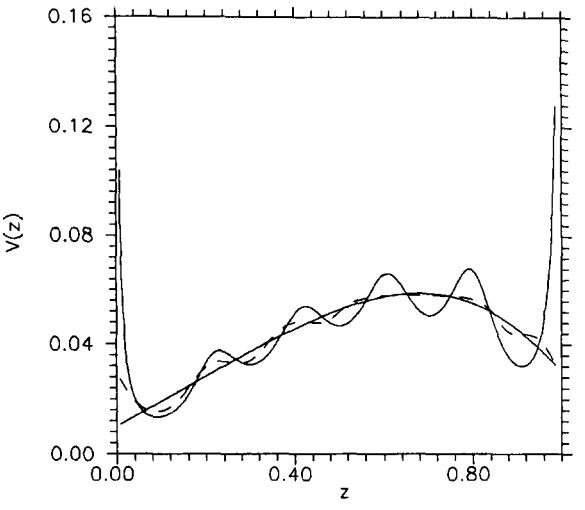



FIG. 8. Maximum entropy solution to Eq. (38) versus the finite difference solution for various values of $b$ in Eq. (26), $b = 1$ (—), $b = 0.1$(----).

$[a,b]$ accurately one is in effect giving more information toward the solution.

## VI. CONCLUSIONS

The MEM yields reasonable solutions to the equations tested. The solution generally appears to be better than Fourier series approximations for the cases tested and, furthermore, the method can be applied to problems where the Fourier series cannot be used easily. At this point it is of interest to compare strengths and weaknesses of the method. The technique is of interest in its own right from a formal aspect and the method should have many applications in various areas of theoretical physics. An advantage of the present technique over finite difference techniques is that the amount of work required in the solution is generally less. Whereas in the case of finite differences, doubling the number of grid points increases the computation by $n^d$ (where $d$ is the dimension), in the maximum entropy method the work only increases as $m^d$, where $m$ is the number of moments. So the work savings could be significant. The MEM method, like the Fourier series solution, does contain oscillations in the solutions that damp as the number of moments is increased; however, the MEM solution appears to be a better approximation in many cases for equal moments and expansion functions. Another advantage of the MEM solution is that additional information can be put into the solution using Lagrange multipliers. For example, if the function is everywhere positive then the probability can be integrated over only positive values, thus incorporating this information into the solution. As far as weaknesses go, the MEM method does require the solution of a system of $M$ (number of moments) nonlinear equations for some problems, which is more difficult than a system of linear equations. Also, how the method will extend to multiple dimensions is not, at present, known. It does appear that the method can be extended to certain classes on nonlinear differential equations. The goal of this paper was not to solve extremely complicated differential equations, but rather to develop the basis of the method. The application of the method to nonlinear problems is an area ripe for further research.

## ACKNOWLEDGMENTS

I would like to thank Professor Ramarao Inguva for many discussions. I would also like to thank Professor Mark Ordal for technical assistance with computer hardware.

Acknowledgment is given to the donors of Petroleum Research Fund administered by ACS for partial support of this research, Grant No. ACS-PRF #20304-G2.

[1]C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[2]E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[3]E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[4]J. P. Burg, Ph.D. dissertation, Stanford University, 1975.
[5]L. R. Mead, J. Math. Phys. **27**, 2903 (1986).
[6]L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
[7]R. Inguva and J. Baker-Jarvis, "Maximum entropy and inverse scattering," in *Maximum Entropy and Bayesian Methods in Applied Statistics* (Cambridge, London, 1986), pp. 300–315.
[8]R. Smith, R. Inguva, and R. L. Morgan, SIAM–AMS Proc. **14**, 127 (1984).
[9]A. van den Bos, IEEE Trans. Inf. Theory **IT-17**, 493 (1971).

# Eigenvalue curves and boundary conditions

Günther Otto Spies
*Max-Planck-Institut für Plasmaphysik, D-8046 Garching, West Germany*

(Received 26 April 1988; accepted for publication 28 September 1988)

Nonlinear eigenvalue problems are considered which are defined by linear homogeneous systems of ordinary differential equations, subject to linear homogeneous boundary conditions, both depending analytically on a complex eigenvalue parameter $\omega$, and on an additional small positive parameter $\epsilon$ such that all solutions have variation $O(\epsilon^{-1})$. There is, in general, a family of eigenvalues that, in the limit $\epsilon \to 0$, become densely spaced and form a definite curve in the complex $\omega$ plane (exceptions arise only for some special boundary conditions). This eigenvalue curve is constructed for arbitrary nonsingular systems without turning points. It depends on no details of the boundary conditions other than their type (i.e., the numbers of boundary conditions involving only the left, only the right, or both end points of the interval). In the special cases of either only two-point boundary conditions (such as periodicity) or only one-point boundary conditions with equally many conditions at each end point, and of differential equations involving only even derivatives whose coefficients are constant to leading order, the curve is simply given by the roots of the local dispersion relation. The distance of the eigenvalues from the curve is $O(\epsilon)$ in general, but $O(\epsilon^{1/l} \log \epsilon^{-1})$, with some positive integer $l$, in special cases.

## I. INTRODUCTION

In applications of eigenvalue problems one is frequently not interested in the precise locations of the eigenvalues, but rather only in certain global properties of the spectrum. In stability theory, for instance, where the eigenvalue parameter is the frequency $\omega$, one merely wants to know if the spectrum extends into the unstable lower half of the complex plane; if there is an eigenvalue curve, the question is whether this curve crosses the real axis.

To define eigenvalue curves, we recall that an "eigenvalue locus" is a curve (or a system of curves) in the complex plane on which the eigenvalues are located. If the eigenvalue problem depends on a real parameter $\epsilon$ such that an eigenvalue locus (or some part of it) becomes densely covered by eigenvalues in the limit $\epsilon \to 0$, and at the same time approaches a definite curve, then we call the latter an "eigenvalue curve." Equivalently, an eigenvalue curve is independent of $\epsilon$, and many eigenvalues are close to any given point of it for sufficiently small $\epsilon$. Unlike an eigenvalue locus, an eigenvalue curve is unique. It should be stressed that, in the limit $\epsilon \to 0$, eigenvalues approach one (or several), but by no means all points of the eigenvalue curve. Also, the curve need neither be uniformly approached by a locus, nor need the eigenvalue spacing become uniformly small.

As an example, consider the problem

$$\epsilon^2 \frac{d^2y}{dx^2} - \frac{\omega}{1+i\epsilon} y = 0, \quad y(0) = y(\pi) = 0, \tag{1}$$

with eigenvalues

$$\omega_n = (1+i\epsilon)\epsilon^2 n^2 \tag{2}$$

($n$ runs through all positive integers). Clearly, the straight line $\Im\omega = \epsilon\Re\omega$ is an eigenvalue locus. Although each individual eigenvalue (fixed $n$) approaches the origin in the limit $\epsilon \to 0$, the entire ray $\Im\omega = \epsilon\Re\omega > 0$ becomes densely covered with eigenvalues for small $\epsilon$, implying that the positive real axis is an eigenvalue curve. The distance of the eigenvalue $\omega_n$ from the curve is $O(\epsilon^3 n^2)$, thus being small or large depending on whether $n \ll \epsilon^{-3/2}$ or $n \gg \epsilon^{-3/2}$, and its distance to its nearest neighbor is $O(\epsilon^2 n)$, thus being small or large depending on whether $n \ll \epsilon^{-2}$ or $n \gg \epsilon^{-2}$. On the other hand, eigenvalues are at a distance $O(\omega_c \epsilon)$ from any given point $\omega_c > 0$ of the eigenvalue curve, and their spacing is $O(\omega_c^{1/2}\epsilon)$; both this distance and this spacing approach zero if $\epsilon$ approaches zero while $\omega_c$ is held fixed, but the approach is nonuniform in $\omega_c$.

Eigenvalue curves occur in many branches of mathematical physics. For instance, any system of ordinary differential equations, with adequate boundary conditions (as the above example), may yield an eigenvalue curve if some of its solutions become strongly varying functions of the independent variable when $\epsilon$ becomes small (viz., if derivatives carry small coefficients). In the trivial case of problems that are linear in the eigenvalue parameter, and self-adjoint, the real axis is an eigenvalue locus, implying that the eigenvalue curve is a subset of the latter. As to nonlinear and non-self-adjoint problems, eigenvalue curves were discovered both numerically[1] and by asymptotic analyses[2-9] in the theory of small nonideal effects upon normal modes in ideal magnetohydrodynamics. It appears that these curves are the same for different sets of boundary conditions. This property, although important because the boundary conditions are often not known explicitly,[7,8] has apparently been demonstrated only in one special case.[7]

In the present paper, we give a systematic discussion of eigenvalue curves and their dependence on boundary conditions for arbitrary non-self-adjoint systems of ordinary differential equations that, for simplicity, we assume to have no singular points (so that the spectrum is discrete), only strongly varying solutions (so that all solutions can be obtained using the eikonal expansion), and no turning points (so that the eikonal expansion is uniformly valid). We show

 0022-2488/89/020307-06$02.50 © 1989 American Institute of Physics 

that the spectrum contains two subsets with different properties: The first subset consists of "special" eigenvalues with spacing $O(1)$ that do depend on details of the boundary conditions. The second subset yields an eigenvalue curve that depends on the boundary condition type, but not on the coefficients in the boundary conditions.

Our analysis is heuristic in that we compute only the leading terms of the distance of the eigenvalues from the curve, assuming that the given data are entire analytic functions of the eigenvalue parameter. Thus we neither carry the expansion to higher orders, nor do we discuss its convergence or the question to what extent our analyticity assumption can be relaxed.

Our assumptions upon the differential equations are rather restrictive because they exclude many cases of interest. Nevertheless, we believe that our findings are generic. In other words, we conjecture that singularities, turning points, or some solutions with finite variation, do not invalidate our main conclusion that eigenvalue curves, if they exist at all, are independent of details of the boundary conditions.

## II. FORMULATION OF THE EIGENVALUE PROBLEM

We consider linear, homogeneous, $N$th-order ($N \geqslant 2$) systems of ordinary differential equations of the form

$$\epsilon \frac{dy_k}{dx} = \sum_{l=1}^{N} A_{kl}(x,\omega,\epsilon) y_l \quad (k = 1,...,N). \tag{3}$$

The real parameter $\epsilon$ is small and positive, and the coefficients $A_{kl}$ are series in non-negative powers of $\epsilon$. The coefficients of these series are assumed to be complex-valued analytic functions of the independent variable $x$ in a finite real interval $x_1 \leqslant x \leqslant x_2$ (this excludes singular points), and entire analytic functions of the complex eigenvalue parameter $\omega$.

The eikonal expansion[10] yields the formal fundamental solution matrix

$$(Y_{kl}) = \left(\eta_{kl} \exp\left[i\epsilon^{-1} \int dx\, \phi_l\right]\right), \tag{4}$$

where the amplitudes $\eta_{kl}(x,\omega,\epsilon)$ are series in non-negative powers of $\epsilon$, and the phase increments $\phi_l(x,\omega)$ are independent of $\epsilon$. The quantities $i\phi_l$ are the eigenvalues of the leading order $(A_{kl}^{(0)})$ of the matrix $(A_{kl})$, and the columns of the leading order $(\eta_{kl}^{(0)})$ of the amplitude matrix $(\eta_{kl})$ are eigenvectors,

$$\sum_{l=1}^{N} (A_{kl}^{(0)} - i\phi_m \delta_{kl}) \eta_{lm}^{(0)} = 0 \quad (m = 1,...,N). \tag{5}$$

The norm of these eigenvectors is determined from higher orders of the coefficients $A_{kl}$; it depends on $x$, and it has singularities at "turning points" where not all phase increments are distinct.

We assume that all solutions of the system (3) are strongly varying functions of $x$, with variation $O(\epsilon^{-1})$, and that the matrix (4) is a uniform asymptotic representation of an actual fundamental solution matrix (there are no turning points), so that its columns represent $N$ independent solutions. This amounts to requiring that none of the phase increments vanishes identically in both $x$ and $\omega$ [the determinant of the leading-order coefficient matrix $(A_{kl}^{(0)})$ does not vanish identically, and hence the order of the system (3) is not smaller than $N$], and that no two of them are equal at any isolated values of $x$, for any fixed $\omega$ (the discriminant of the characteristic polynomial has no isolated zeros). Thus, if two phase increments coalesce, they must do so identically in $x$; however, we assume that this happens only for isolated values of $\omega$. Trivial examples of systems meeting our assumptions are those whose leading-order coefficient matrices are independent of $x$, and have determinants and discriminants that do not vanish identically in $\omega$.

We impose $N$ boundary conditions of the form

$$\sum_{l=1}^{N} [\alpha_{kl}(\omega,\epsilon) y_l(x_1) + \beta_{kl}(\omega,\epsilon) y_l(x_2)] = 0 \quad (k = 1,...,N), \tag{6}$$

where the coefficients of $\alpha_{kl}$ and $\beta_{kl}$ are also series in non-negative powers of $\epsilon$ whose coefficients are entire analytic functions of $\omega$. In order that the boundary conditions be mutually independent, we require that the rectangular $(N \times 2N)$ matrix formed by all coefficients $\alpha_{kl}$ and $\beta_{kl}$ has rank $N$ throughout, and in particular for $\epsilon = 0$.

The "boundary condition type" is characterized by the ranks $R_\alpha$ and $R_\beta$ of the matrices $(\alpha_{kl})$ and $(\beta_{kl})$. We assume that these ranks are constant (i.e., independent of both $\omega$ and $\epsilon$). In order that both end points occur in the boundary conditions, we require that neither $R_\alpha$ nor $R_\beta$ vanish identically. Their sum cannot be smaller than $N$, and none of them can exceed $N$. Within these restrictions, the ranks are arbitrary. The number of boundary conditions involving only the left end point $x = x_1$ is $N_1 = N - R_\beta$, that involving only the right end point $x = x_2$ is $N_2 = N - R_\alpha$, and that involving both points is $N - N_1 - N_2 = R_\alpha + R_\beta - N$. Therefore, we may replace the original boundary conditions by a linear combination such that $\beta_{kl} = 0$ for $1 \leqslant k \leqslant N_1$, $\alpha_{kl} = 0$ for $N - N_2 + 1 \leqslant k \leqslant N$, but $|\alpha_{kl}|^2 + \cdots + |\alpha_{kN}|^2 \neq 0$ and $|\beta_{kl}|^2 + \cdots + |\beta_{kN}|^2 \neq 0$ for $N_1 + 1 \leqslant k \leqslant N - N_2$. The limitations upon the ranks $R_\alpha$ and $R_\beta$ imply that the numbers $N_1$ and $N_2$ are restricted by $0 \leqslant N_1 < N$, $0 \leqslant N_2 < N$, and $N_1 + N_2 \leqslant N$.

With the boundary conditions (6), the system (3) constitutes a nonlinear eigenvalue problem that yields, in general (there are exceptions to be discussed), a purely discrete spectrum consisting of a countable set of eigenvalues $\omega$.

## III. ASYMPTOTIC DISPERSION RELATION

We write the general solution of Eqs. (3) as a linear combination of the $N$ fundamental solutions (4),

$$y_k = \sum_{l=1}^{N} c_l \eta_{kl} \exp\left(i\epsilon^{-1} \int_{x_1}^{x} dx\, \phi_l\right), \tag{7}$$

with coefficients $c_l$ to be determined by the boundary conditions (6). The latter can now be written as

$$\sum_{l=1}^{N} \gamma_{kl} c_l = 0 \quad (k = 1,...,N), \tag{8}$$

where

$$\gamma_{kl} = \sum_{j=1}^{N} [\alpha_{kj} \eta_{jl}(x_1) + \beta_{kj} \eta_{jl}(x_2) \exp(i\epsilon^{-1} \Phi_l)], \tag{9}$$

and

$$\Phi_l(\omega) = \int_{x_1}^{x_2} dx\, \phi_l(x,\omega) \quad (l = 1,...,N) \tag{10}$$

are the phase integrals. Equation (8) is a system of homogeneous linear equations for the coefficients $c_l$ whose determinant $D = \det(\gamma_{kl})$ must vanish. Thus the eigenvalues $\omega$ are governed by the dispersion relation $D = 0$. When written out more explicitly, this is

$$\sum_{M=N_2}^{N-N_1} \sum_{l_1<\cdots<l_M} D_{l_1,...,l_M} \exp\left[i\epsilon^{-1}(\Phi_{l_1} + \cdots + \Phi_{l_M})\right] = 0, \tag{11}$$

where the coefficients $D_{l_1,...,l_M}$ are homogeneous polynomials of the amplitudes $\eta_{kl}$ at the two end points of the interval and of the coefficients $\alpha_{kl}$ and $\beta_{kl}$ in the boundary conditions; the exponent is defined to be zero if $M = 0$, and the corresponding coefficient is denoted by $D_0$.

The phase increments are the roots of the characteristic polynomial,

$$\det\left(A_{kl}^{(0)} - i\phi\delta_{kl}\right) = 0. \tag{12}$$

As functions of $\omega$ (i.e., for fixed $x$), they are therefore represented by the $N$ different sheets of the Riemann surface of one analytic function with no singularities other than algebraic branch points.[10] This is a consequence of the assumed analyticity of the coefficients of the differential equations. The branch points are those values of $\omega$ for which some phase increments coalesce. Therefore, any closed path about such points yields some permutation of the phase increments. The same is true for the phase integrals because two phase increments are assumed to coalesce identically in $x$ if they coalesce at all. This implies that the left-hand side of the dispersion relation is a power series in $\epsilon$ whose coefficients are single-valued analytic functions of $\omega$ even though the various terms of the sum are not.

The eigenvalues $\omega$ are (at least asymptotically) given by the formal roots of Eq. (11), with the possible exception of branch points [such a point is a root if the dimension of the eigenspace of the matrix $(A_{kl}^{(0)})$ is smaller than the multiplicity of the phase increment, but is an eigenvalue only for special boundary conditions].

It should be remarked that the boundary conditions can be such that all coefficients $D_{l_1,...,l_M}$ vanish identically in $\omega$, implying that the spectrum consists of the entire complex plane, or that one coefficient is nonzero throughout while all others vanish identically, implying that the spectrum is empty. To give an example of such highly degenerate situations, we consider the eigenmodes $\sim\exp(i\omega t)$ of the wave equation $\partial^2 y/\partial t^2 = \epsilon^2\,\partial^2 y/\partial x^2$ ($\epsilon$ is the propagation speed) with arbitrary one-point boundary conditions. The eigenfrequencies $\omega$ are governed by the system

$$\epsilon\frac{dy_1}{dx} = \omega y_2, \quad \epsilon\frac{dy_2}{dx} = -\omega y_1, \tag{13}$$

subject to the boundary conditions

$$y_1(0) + \alpha(\omega)y_2(0) = 0, \quad y_1(1) + \beta(\omega)y_2(1) = 0, \tag{14}$$

where $\alpha$ and $\beta$ are arbitrary analytic functions of $\omega$. The dispersion relation [now valid for all values of $\epsilon$ because Eqs. (13) are independent of $x$] turns out to be

$$(1 - i\alpha)(1 + i\beta)\exp(i\epsilon^{-1}\omega) - (1 + i\alpha)(1 - i\beta)\exp(-i\epsilon^{-1}\omega) = 0. \tag{15}$$

Hence the spectrum covers the entire complex plane if, and only if, either $\alpha = i$ and $\beta = i$, or $\alpha = -i$ and $\beta = -i$ (waves propagating from left to right, or from right to left). On the other hand, the spectrum is empty if, and only if, either $\alpha = i$ and $\beta = -i$, or $\alpha = -i$ and $\beta = i$ (ingoing waves or outgoing waves). Such situations thus do indeed occur in applications (for instance, in the theory of drift waves in a plasma[11–14]). However, since an arbitrarily small change of the boundary conditions removes the degeneracy, we ignore it in the remainder of this paper.

To relate to common terminology, we remark that the relation between the "local wave number" $k = \epsilon^{-1}\phi$ and the frequency $\omega$ [i.e., the characteristic equation (12)] is sometimes called "local dispersion relation",[15] as opposed to the "global" dispersion relation (11). Its roots $\omega(k,x)$, for real $k$ (the "local eigenfrequencies"), are certainly not roots of the global dispersion relation ("global eigenfrequencies"), at least unless they are independent of $x$. Accordingly, an unstable local eigenfrequency ($\Im\omega < 0$) which is unrelated to an unstable global eigenfrequency is said to give rise to a "convective" or "transient" instability,[16] as opposed to a so-called "absolute" instability that corresponds to an unstable global eigenvalue (the dependent variables grow exponentially with time at every point $x$). Since we consider global eigenfrequencies in the present paper, and not local ones, we are concerned with absolute instabilities, not with convective ones.

## IV. LEADING TERMS OF EIGENVALUES

To obtain the leading terms of the eigenvalues $\omega$, we need the coefficients $D_{l_1,...,l_M}$ in the dispersion relation (11) only for $\epsilon = 0$. In this approximation, the dispersion relation depends on $\epsilon$ only through the explicit factor $\epsilon^{-1}$ in the exponents. Hence that term dominates for small $\epsilon$ whose exponent has the largest real part (the phase integral sum in its exponent has the smallest imaginary part). To exploit this fact systematically, we divide the $\omega$ plane into regions such that each term dominates in exactly one region, thus having a one-to-one correspondence between the terms in Eq. (11) and the regions. This subdivision depends on the boundary condition type because the set of competing exponents depends on $N_1$ and $N_2$. At the borderline between two regions two of the exponents have equal real parts. Since the difference between two phase integral sums corresponding to adjacent regions equals either one of the phase integrals or the difference between two of them, an equation of the form

$$\Im(\Phi_k - \Phi_l) = 0 \quad \text{or} \quad \Im\Phi_k = 0 \tag{16}$$

holds at each borderline. In other words, each borderline is some part of one of the curves given by Eq. (16). These curves depend only on the leading orders of the coefficients of the differential equations (3), but what part of a curve is borderline depends on the boundary condition type. Different borderlines may be connected or disconnected, and one borderline may consist of several disconnected parts. There

may be no borderlines at all (this is the case, for instance, if the coefficients $A_{kl}^{(0)}$ are independent of $\omega$).

Since no general statement can be made about large eigenvalues because these depend on the behavior of the coefficients of the system (3) and (6) for large $\omega$, we consider only eigenvalues $O(1)$ or smaller. For small $\epsilon$, the dispersion relation has such roots near the zeros of the coefficient $D_{l_1,\ldots,l_M}$ of the dominating term, and near the borderlines, but nowhere else. Hence these roots form two subsets of the spectrum: The first subset consists of "special" eigenvalues whose leading terms are the zeros of the coefficients in their regions. These leading terms are independent of $\epsilon$, and therefore have spacing $O(1)$, and they depend on details of the boundary conditions. This subset is in general finite (because of our analyticity assumption it may contain infinitely many eigenvalues only in a region that extends to infinity).

To obtain the leading terms of the eigenvalues in the second subset, we consider one borderline, thus keeping only the two dominating terms of the dispersion relation (11), and writing it as

$$C_1 \exp(i\epsilon^{-1}\Sigma_1) + C_2 \exp(i\epsilon^{-1}\Sigma_2) = 0, \tag{17}$$

where $\Sigma_1$ and $\Sigma_2$ are the two phase integral sums with the smallest imaginary parts. Equation (17) is equivalent to

$$\Sigma_1 - \Sigma_2 = 2\pi n\epsilon + i\epsilon \log(-C_1/C_2), \tag{18}$$

with an arbitrary integer $n$. Since the product $\epsilon n$ is large for large $n$, no matter how small $\epsilon$ is, the first term at the right-hand side is not small. However, the second term is small unless the boundary conditions are such that $C_1/C_2$ has zeros or poles at the borderline of interest. Postponing discussion of this degenerate situation to Sec. VI (where we show that this term can always be dropped), we consider the approximation

$$\Sigma_1 - \Sigma_2 = 2\pi n\epsilon, \tag{19}$$

which is obviously independent of the details of the boundary conditions. Solving this for $\omega$ yields the leading terms of the eigenvalues $\omega_n$ in the second subset. For small $\epsilon$, these eigenvalues are closely spaced, and at a small distance from the borderline given by

$$\Im(\Sigma_1 - \Sigma_2) = 0. \tag{20}$$

As already mentioned, this is equivalent to Eq. (16). We conclude that the union of all borderlines is the eigenvalue curve.

Equation (19) shows that, in the limit $\epsilon \to 0$, each individual eigenvalue (fixed $n$) either approaches the point where $\Sigma_1 - \Sigma_2 = 0$ (if this point is on the given branch), or moves to some other branch to approach a similar point there, but that there are nevertheless eigenvalues near every point of the curve for arbitrarily small $\epsilon$. This is a consequence of the nonuniformity in $n$ of the limit $\epsilon \to 0$, or equivalently, of the noninterchangeability of the two limits $\epsilon \to 0$ and $n \to \pm\infty$: Letting first $\epsilon \to 0$, and then $n \to \pm\infty$, yields the above points, while the opposite procedure yields a point where $\Sigma_1 - \Sigma_2 = \infty$; arbitrary intermediate points are obtained by letting both $\epsilon \to 0$ and $n \to \pm\infty$, but keeping the product $\epsilon n$ fixed at arbitrary finite values.

## V. EIGENVALUE CURVES IN SPECIAL SYSTEMS

To construct the eigenvalue curve of a given system, one must solve two problems: First, one must determine, for the given boundary condition type, which of the equations (16) yield possible branches of the eigenvalue curve. Second, one must determine the phase increments as the roots of the characteristic polynomial (to compute from these the required phase integrals as functions of $\omega$).

Due to the complexity of the first problem, we have solved this only in some relatively simple cases. We label the phase integrals according to the magnitudes of their imaginary parts ($\Im\Phi_1 \leqslant \cdots \leqslant \Im\Phi_N$) in the following list.

(1) Systems of second order yield the following branches (for each possible boundary condition type, their equations are separated by commas):

$$N_1 = 0, \quad N_2 = 0: \quad \Im\Phi_1 = 0, \quad \Im\Phi_2 = 0, \tag{21}$$

$$N_1 = 0, \quad N_2 = 1: \quad \Im\Phi_2 = 0, \quad \Im\Phi_1 = \Im\Phi_2 > 0, \tag{22}$$

$$N_1 = 1, \quad N_2 = 0: \quad \Im\Phi_1 = 0, \quad \Im\Phi_1 = \Im\Phi_2 < 0, \tag{23}$$

$$N_1 = 1, \quad N_2 = 1: \quad \Im\Phi_1 = \Im\Phi_2. \tag{24}$$

(2) In systems of third order, the corresponding equations are

$$N_1 = 0, \quad N_2 = 0: \quad \Im\Phi_1 = 0, \quad \Im\Phi_2 = 0, \quad \Im\Phi_3 = 0, \tag{25}$$

$$N_1 = 0, \quad N_2 = 1: \quad \Im\Phi_2 = 0, \quad \Im\Phi_3 = 0, \quad \Im\Phi_1 = \Im\Phi_2 > 0, \tag{26}$$

$$N_1 = 1, \quad N_2 = 0: \quad \Im\Phi_1 = 0, \quad \Im\Phi_2 = 0, \quad \Im\Phi_2 = \Im\Phi_3 < 0, \tag{27}$$

$$N_1 = 2, \quad N_2 = 0: \quad \Im\Phi_1 = 0, \quad \Im\Phi_1 = \Im\Phi_2 < 0, \tag{28}$$

$$N_1 = 1, \quad N_2 = 1: \quad \Im\Phi_2 = 0, \quad \Im\Phi_2 = \Im\Phi_3 < 0, \quad \Im\Phi_1 = \Im\Phi_2 > 0, \tag{29}$$

$$N_1 = 0, \quad N_2 = 2: \quad \Im\Phi_3 = 0, \quad \Im\Phi_2 = \Im\Phi_3 > 0, \tag{30}$$

$$N_1 = 2, \quad N_2 = 1: \quad \Im\Phi_1 = \Im\Phi_2, \tag{31}$$

$$N_1 = 1, \quad N_2 = 2: \quad \Im\Phi_2 = \Im\Phi_3. \tag{32}$$

(3) In systems of arbitrary order, the boundary condition type $N_1 = N_2 = 0$ (only two-point, such as periodicity) yields the $N$ branches

$$\Im\Phi_l = 0 \quad (l = 1,\ldots,N) \tag{33}$$

(one of the $N$ phase integrals is real).

(4) In systems of arbitrary order, the boundary condition type $N_1 + N_2 = N$ (only one-point) yields the eigenvalue curve

$$\Im\Phi_{N_2} = \Im\Phi_{N_2+1}. \tag{34}$$

If $N$ is even, and if the characteristic polynomial contains only even powers of $\phi$ [this is so, for instance, in the frequently occurring case that our system of first-order equations is equivalent to one single $N$th-order equation that contains only even derivatives of some linear combination of the original dependent variables $y_k$ whose coefficients have variation $O(1)$], the phase integrals occur in pairs of oppositely equal quantities, and Eq. (34) requires that one of the phase integrals is real. Hence the eigenvalue curve is the same as in the case of two-pont boundary conditions.

The second problem, viz., that of finding the roots of the

characteristic polynomial (which can be formidable for $N>2$), can be sidestepped for the two most common boundary condition types $2N_1 = 2N_2 = N$ (standard boundary value problems with equally many boundary conditions at the two end points), and $N_1 = N_2 = 0$ (e.g., periodicity), if the coefficients $A_{kl}^{(0)}$ are independent of $x$ (e.g., equations with constant coefficients). In these cases the phase increments are also independent of $x$, implying that the phase integrals are proportional to them, and hence that reality of $\Phi_k$ is the same as reality of $\phi_k$. Instead of solving the characteristic equation for $\phi$, one may now solve it for $\omega$. The various branches of the eigenvalue curve are then traced by the roots $\omega(\phi^2)$ when $\phi^2$ is varied through all positive real numbers. In other words, the local dispersion relation yields the eigenvalue curve in these cases. Examples arise in the theory of small nonideal effects upon magnetohydrodynamic modes with small wavelengths.[7,8]

## VI. DISTANCE OF EIGENVALUES FROM EIGENVALUE CURVES

To estimate the distance of the eigenvalues from the eigenvalue curve, we now study eigenvalue loci for small $\epsilon$. We consider the dispersion relation in the form (17). Taking the magnitudes of its two terms we obtain

$$|C_1|\exp(-i\epsilon^{-1}\Im\Sigma_1) = |C_2|\exp(-i\epsilon^{-1}\Im\Sigma_2). \quad (35)$$

Since the roots of Eq. (17) satisfy Eq. (35), the latter describes one branch of an eigenvalue locus. This branch consists of small closed curves about zeros of the coefficients $C_1$ and $C_2$ (they have no singularities because of the assumed analyticity of the coefficients of the differential equations and of the boundary conditions), and of a curve that is near the eigenvalue curve. There is at most one closed curve about each zero, each of these contains exactly one "special eigenvalue."

To study the approach of the locus to the eigenvalue curve, we rewrite Eq. (35) as

$$\exp(\epsilon^{-1}\Im\Sigma) = |C_1/C_2|, \quad (36)$$

where $\Sigma = \Sigma_1 - \Sigma_2$ is either one of the phase integrals, or the difference between two of them. Looking at some vicinity of some point $\omega = \omega_0$ of the curve $\Im\Sigma = 0$, we introduce, instead of $\omega$, the new complex variable $\lambda = \Sigma(\omega)$. This relation can be inverted to yield $\omega$ as a function of $\lambda$ in some vicinity of $\lambda = \Sigma(\omega_0)$ if the derivative $d\Sigma/d\omega$ exists and is nonzero at $\omega = \omega_0$. Setting aside, for a moment, any critical points at which this is not the case, we rewrite Eq. (36) as

$$\exp(\epsilon^{-1}\Im\lambda) = |\psi(\lambda)|, \quad (37)$$

where $\psi$ is a given function that is determined by the boundary conditions. Regarding $\Re\lambda$ as given, we look for a solution $\Im\lambda$ that is $O(\epsilon)$, thus expanding it in powers of $\epsilon$. Then, to leading order,

$$\exp(\epsilon^{-1}\Im\lambda) = |\psi(\Re\lambda)|, \quad (38)$$

or, equivalently,

$$\Im\lambda = \epsilon\log|\psi(\Re\lambda)| + O(\epsilon^2). \quad (39)$$

Hence this part of the eigenvalue locus is indeed at a distance $O(\epsilon)$ from the eigenvalue curve as long as the function $\psi(\lambda)$ has neither zeros nor poles at the real axis, or equivalently, as long as neither of the functions $C_1(\omega)$ or $C_2(\omega)$ has zeros at the eigenvalue curve.

Thus the expansion in powers of $\epsilon$ fails near a special eigenvalue that happens to approach the eigenvalue curve in the limit $\epsilon\to 0$, and must be replaced by a different approximation. To treat such critical points (which occur only for special boundary conditions), it suffices, because of the symmetry of Eq. (35), to consider zeros of $C_1$. Thus assuming that $C_1$ has a zero of order $m$ for $\omega = \omega_0$, and that $\lambda_0 = \Sigma(\omega_0)$ is real, we set $\lambda = \lambda_0 + \tilde{\lambda}$ with small $\tilde{\lambda}$, and approximate Eq. (37) by

$$\exp(\epsilon^{-1}\Im\tilde{\lambda}) = c|\tilde{\lambda}|^m, \quad (40)$$

where $c$ is a given positive number. The two equations (38) and (40) coalesce if $|\Re\tilde{\lambda}| \ll |\Im\tilde{\lambda}| \ll 1$, indicating that their ranges of validity overlap: Indeed, for small values of $|\Re\tilde{\lambda}|$ Eq. (40) must be used, for some intermediate values both equations are valid, and for large values only Eq. (38) is valid. Solving Eq. (40) for $(\Re\tilde{\lambda})^2$, we see that this is a monotonically increasing function of $\Im\tilde{\lambda}$. It vanishes at some small negative value of $\Im\tilde{\lambda}$ which we denote by $-d$, and it attains finite values (thus invalidating the equation) long before $\Im\tilde{\lambda}$ reaches zero. Hence $d$ is the relevant maximum of $|\Im\lambda|$, and solves the equation

$$\exp(-\epsilon^{-1}d) = cd^m. \quad (41)$$

Since the left-hand side of this equation is a decreasing function of $d$ while its right-hand side is increasing, there is exactly one solution. As one verifies by substitution, this solution is

$$d = (m/2)\epsilon\log\epsilon^{-1} + O(\epsilon\log\log\epsilon^{-1}). \quad (42)$$

Since $d$ approaches zero slower than $\epsilon$, we conclude that the eigenvalue locus peaks away from the eigenvalue curve towards a special eigenvalue that approaches the curve in the limit $\epsilon\to 0$.

The foregoing analysis fails at an intersection of several branches of the eigenvalue curve because three or more terms in the dispersion relation (11) dominate, and the latter does not reduce to Eq. (17). It also fails where one branch of the eigenvalue curve intersects itself (the derivative $d\Sigma/d\omega$ is zero there), and at a branch point of the phase integrals where several phase integrals coalesce ($d\Sigma/d\omega$ diverges). At such points the function $\Sigma(\omega)$ cannot be inverted, and Eq. (37) is not valid. We state without proof that Eqs. (39) and (42) remain valid at intersections of several branches and at branch points of the phase integrals, but are replaced by

$$\Im\lambda = O(\epsilon^{1/l}), \quad (43)$$

and

$$\Im\lambda = O(\epsilon^{1/l}\log\epsilon^{-1}), \quad (44)$$

at a point of self-intersection of degree $l$. This shows that the approach to the eigenvalue curve can be arbitrarily slow in special cases.

## ACKNOWLEDGMENTS

Helpful discussions with K. Riedel, D. Pfirsch, D. Lortz, R. Meyer-Spasche, J. Li, and H. Stierstorfer are grate-

fully acknowledged.

This work was performed under the terms of the agreement of association between Max-Planck Institut für Plasmaphysik and Euratom.

[1]C. M. Ryu and R. C. Grimm, Princeton University Plasma Physics Laboratory Report PPL-2020, 1983.
[2]R. L. Dewar and B. Davies, Australian National University Report TPP 84/01, 1984.
[3]Y. Pao and W. Kerner, Phys. Fluids **28**, 287 (1985).
[4]R. Storer, Plasma Phys. **25**, 1279 (1983).
[5]D. Lortz and G. O. Spies, Phys. Lett. A **101**, 335 (1984).
[6]D. Lortz and G. O. Spies, *Plasma and Physics and Controlled Nuclear Fusion Research* (IAEA, Vienna, 1983), Vol. 2, p. 245.
[7]G. O. Spies and M. Faghihi, Phys. Fluids **30**, 1724 (1987).
[8]G. O. Spies, Plasma Phys. Controlled Fusion **30**, 1025 (1988).
[9]K. S. Riedel, Phys. Fluids **29**, 1093 (1986).
[10]W. Wasow, *Linear Turning Point Theory* (Springer, Berlin, 1985).
[11]L. D. Pearlstein and H. L. Berk, Phys. Rev. Lett. **23**, 220 (1969).
[12]D. W. Ross, Phys. Rev. Lett. **40**, 324 (1978).
[13]K. T. Tsang, J. C. Whitson, and J. Smith, Phys. Rev. Lett. **40**, 327 (1978).
[14]T. M. Antonsen, Phys. Rev. Lett. **41**, 33 (1978).
[15]H. L. Berk and D. Pfirsch, J. Math. Phys. **21**, 2054 (1980).
[16]A. I. Akhiezer, I. A. Akhiezer, R. V. Polovin, A. G. Sitenko, and K. N. Stepanov, *Collective Oscillations in a Plasma* (Pergamon, New York, 1967).

# Equations of motion and wave functions in spherical coordinates for an integrable Hamiltonian system

L. Gagnon[a)]
*Département de Physique et Centre de Recherches Mathématiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec H3C 3J7, Canada*

(Received 30 March 1988; accepted for publication 12 October 1988)

The classical equations of motion for the Hamiltonian $H = \Sigma_{\mu=0}^{n} (y_\mu^2/2 + u_\mu^2/x_\mu^2)$ (where $\Sigma_\mu x_\mu^2 = 1$, $y_\mu$ is the conjugate momentum to $x_\mu$, and $u_\mu$ is constant) are solved by separation of variables, in spherical coordinates, in the Hamilton–Jacobi equation. This flow is related to the one obtained from the projection of geodesic (free) flow on the sphere $S^{2n+1}$. Wave functions for the quantum case together with the level degeneracy for the generic case $u_\mu \neq 0$ are also given.

## I. INTRODUCTION

The purpose of this work is to analyze the classical and quantum equations of motion in spherical coordinates for the Hamiltonian

$$H = \sum_{\mu=0}^{n} \left( \frac{1}{2} y_\mu^2 + \frac{u_\mu^2}{x_\mu^2} \right) \tag{1.1}$$

defined on the phase space $T^*S^n$, where $S^n$ is the $n$-dimensional sphere, $y_\mu$ is the conjugate momentum to $x_\mu$, $u_\mu$ is constant, and

$$\sum_\mu x_\mu^2 = 1, \quad \sum_\mu x_\mu y_\mu = 0. \tag{1.2}$$

The above Hamiltonian is a particular case of the so-called Rosochatius Hamiltonian, having an additional "harmonic" term $\Sigma \alpha_\mu x_\mu^2$, which has been known since the 19th century.[1] It recently received more attention because of its complete integrability[2] and, in opposition to a wide variety of integrable systems such as the Toda lattice, Korteweg–de Vries equation, or sine–Gordon system, its *finite-dimensional* underlying Lie algebraic structure.[3] Those properties made the Hamiltonian (1.1) an interesting simple example to study the link between the theory of a completely integrable Hamiltonian system and that of Lie groups and Lie algebras.[4] In particular, it served, in a recent work,[5] to illustrate how it is possible to unify the geometrical and algebraic structure shared by many Hamiltonian systems.

In fact, the Hamiltonian (1.1) can be obtained from the free one

$$H = \sum_{\mu=0}^{n} \frac{1}{2} p_\mu \bar{p}_\mu \tag{1.3}$$

defined on the phase space $T^*S^{2n+1}$, where $p_\mu = y_\mu + iz_\mu$ are the conjugate momentum to the complex Cartesian coordinates $\{W_\mu = s_\mu + it_\mu\}$ on $\mathbb{C}^{n+1}$ and

$$\sum_\mu W_\mu^2 = 1, \quad \sum_\mu (W_\mu p_\mu + \bar{W}_\mu \bar{p}_\mu) = 0, \tag{1.4}$$

by applying the Weinstein–Marsden reduction method[6,7] of the phase space (see Ref. 8 for a reduction from $\mathbb{C}P^n$). It follows that the free flow of (1.3) on $S^{2n+1}$ given by

$$\begin{aligned} W_\mu &= w_\mu \cos|\xi|t + v_\mu \sin|\xi|t, \\ p_\mu &= -|\xi| w_\mu \sin|\xi|t + |\xi| \bar{v}_\mu \cos|\xi|t \end{aligned} \tag{1.5}$$

can be projected on $S^n$ to give[5]

$$\begin{aligned} x_\mu^2 = |W_\mu|^2 &= w_\mu^2 \cos^2|\xi|t + |v_\mu|^2 \sin^2|\xi|t \\ &\quad + 2w_\mu \operatorname{Re} \bar{v}_\mu \sin|\xi|t \cos|\xi|t, \end{aligned} \tag{1.6}$$

where $w_\mu \in \mathbb{R}$, $v_\mu \in \mathbb{C}$, and, in view of (1.4),

$$\sum_\mu w_\mu^2 = 1, \quad \sum_\mu |v_\mu|^2 = 1, \quad \operatorname{Re}\left( w_\mu \sum_\mu \bar{v}_\mu \right) = 0. \tag{1.7}$$

The constants $u_\mu$ in (1.1) appear to be the invariants associated with the geodesic flow on $S^{2n+1}$ and are related to the parameters $w_\mu$, $v_\mu$, and $|\xi|$ by[5]

$$u_\mu = (|\xi|/\sqrt{2}) w_\mu \operatorname{Im} \bar{v}_\mu. \tag{1.8}$$

On the other hand, one can obtain the flow in a more classical way by solving the Hamilton–Jacobi equation by separation of variables in, for example, elliptic coordinates.[5] This leads to hyperelliptic integrals and to a Jacobi inversion problem. The authors established the link between the geometric notion of reduction by symmetries and the algebraic linearization of the flow on a Jacobi variety.[5]

Here we propose (Sec. II) to make a similar study for spherical coordinates. The problem thus appears simpler since we can solve the integrals of the action-angle variables directly without all the mathematical tools of the Jacobi inversion problem. The link between the two methods will be made by relating the two sets of parameters involved. Furthermore, we will take advantage of the simple form of the flow (1.6) to investigate the conditions of "plane" trajectories on $S^2$.

In Sec. III, we present the solution of the quantum problem associated to (1.1) and give the degeneracies of the energy levels. Finally, some conclusions and future outlook are given in the last section.

## II. THE CLASSICAL PROBLEM

### A. Solution of the Hamilton–Jacobi equation

We introduce the spherical coordinates on the $n$-dimensional sphere $S^n$ as[9]

[a)] Present address: Laboratoire de Recherches en Optique et Laser, Département de Physique, Université Laval, Cité Universitaire, Québec G1K 7P4, Canada.

 0022-2488/89/020313-05$02.50 © 1989 American Institute of Physics 

$$
\begin{aligned}
x_0 &= \sin\phi_1 \cdots \sin\phi_{n-1} \sin\phi_n, \\
x_1 &= \sin\phi_1 \cdots \sin\phi_{n-1} \cos\phi_n, \\
x_2 &= \sin\phi_1 \cdots \cos\phi_{n-1}, \\
&\vdots \\
x_{n-1} &= \sin\phi_1 \cos\phi_2, \\
x_n &= \cos\phi_1, \quad 0 \leqslant \phi_n < 2\pi, \quad 0 \leqslant \phi_k < \pi \quad (k \neq n).
\end{aligned} \tag{2.1}
$$

This permits us to separate the Hamilton–Jacobi equation

$$
H\left(\phi_i, \frac{\partial S}{\partial \phi_i} = \dot{\phi}_i\right) = E \equiv P_1, \tag{2.2}
$$

where

$$
S = \sum_{i=1}^{n} S_i(\phi_i) \tag{2.3}
$$

is the generating function.

We obtain the $n$ equations

$$
\frac{1}{2}\left(\frac{\partial S_i}{\partial \phi_i}\right)^2 + V_i(\phi_i) = P_i, \quad i = 1,...,n, \tag{2.4}
$$

where

$$
V_i(\phi_i) = \frac{P_{i+1}}{\sin^2\phi_i} + \frac{u_{n-i+1}^2}{\cos^2\phi_i} \tag{2.5}
$$

is the potential associated to the coordinates $\phi_i$ and

$$
P_{n+1} \equiv u_0^2. \tag{2.6}
$$

The others $P_i$'s are separation constants.

Treating the $P_i$'s as the new momenta, we obtain, for the conjugate variables $Q_i$,

$$
\begin{aligned}
Q_i \equiv \frac{\partial S}{\partial P_i} &= \frac{1}{\sqrt{2}} \int [P_i - V_i(\phi_i)]^{-1/2} \, d\phi_i \\
&\quad - \frac{1}{\sqrt{2}} \int \frac{1}{\sin^2\phi_{i-1}} \\
&\quad \times [P_{i-1} - V_{i-1}(\phi_{i-1})]^{-1/2} \, d\phi_{i-1} \\
&= \delta_{i1} t, \quad i = 1,...,n,
\end{aligned} \tag{2.7}
$$

where $\delta_{i1}$ is the Kronecker symbol.

The integrals in (2.7) are easily performed by the substitution $x = \sin^2\phi_i$ and by observing from (2.4) and (2.5) that

$$
0 \leqslant V_{\min}(\phi_i) \leqslant P_i. \tag{2.8}
$$

We obtain, for the coordinate $\phi_1$,

$$
\begin{aligned}
\sin^2\phi_1 = (1/2P_1)[P_1 + P_2 - u_n^2 + [(P_1 + P_2 - u_n^2)^2 \\
- 4P_1P_2]^{1/2} \sin(2\sqrt{2P_1}t - \tau_1)]
\end{aligned} \tag{2.9a}
$$

and for $\phi_j, j \geqslant 2$,

$$
\begin{aligned}
\sin^2\phi_j = \frac{1}{2P_j}\Bigg[P_j + P_{j+1} - u_{n-j+1}^2 + [(P_j + P_{j+1} - u_{n-j+1}^2)^2 - 4P_jP_{j+1}]^{1/2} \\
\times \sin\left[\sin^{-1}\left\{\frac{(P_{j-1} + P_j - u_{n-j+2}^2)\sin^2\phi_{j-1} - 2P_j}{\sin^2\phi_{j-1}[(P_{j-1} + P_j - u_{n-j+2}^2 - 4P_{j-1}P_j]^{1/2}}\right\} + \tau_j\right]\Bigg],
\end{aligned} \tag{2.9b}
$$

where $\tau_i$, $i = 1,...,n$, are real integration constants.

Substituting (2.9) into (2.1) gives the solution of the classical equations of motion for the $x_\mu$'s in a less compact form than (1.6). We shall now relate those two equivalent expressions.

### B. Relation between the classical and projected flows

The task is now to relate the set of $3n + 1$ parameters

$$
\{P_i, u_0, u_i, \tau_i, \ i = 1,...,n\} \tag{2.10}
$$

involved in the Hamilton–Jacobi equation, to the set

$$
\{w_0, w_i, v_0, v_i, |\xi|, \ i = 1,...,n\}, \tag{2.11}
$$

constrained by (1.7), of the projection method.

A part of the work has already been done in view of relation (1.8); so only the $P_i$'s and the $\tau_i$'s will now retain our attention.

Substituting relations (1.6) and (2.9a) in Eq. (2.1) for $x_n^2$, we obtain by simply identifying coefficients, a relation for $P_1$, $P_2$, and $\tau_1$ (see Ref. 10):

$$
P_1 = \tfrac{1}{2}|\xi|^2, \tag{2.12}
$$

$$
P_2 = (|\xi|^2/2)(1 - w_n^2 - |v_n|^2 + w_n^2 \operatorname{Im}^2 \bar{v}_n), \tag{2.13}
$$

$$
\begin{aligned}
\sin\tau_1 = P_1(w_n^2 - |v_n|^2) \\
\times ([(P_1 + P_2 - u_n^2)^2 - 4P_1P_2]^{1/2})^{-1}.
\end{aligned} \tag{2.14}
$$

So we see that the parameter $|\xi|^2$ is, in fact, the energy of the system.

The approach used to find (2.12)–(2.14) becomes, however, more and more cumbersome for $x_{n-1},...,x_0$. We found the following one more efficient.

To obtain the separation constant $P_j$ ($j = 2,...,n - 1$), we evaluate (2.4) at $t = 0$. Remembering that

$$
\frac{\partial S_j}{\partial \phi_j} = \dot{\phi}_j(t), \tag{2.15}
$$

we get the $\dot{\phi}_j$'s by differentiating the last $(n - 1)$ relations in (2.1). Substituting into (2.4) and evaluating at $t = 0$ with

$$
x_\mu^2(0) = w_\mu^2, \quad x_\mu(0)\dot{x}_\mu(0) = |\xi| w_\mu \operatorname{Re} \bar{v}_\mu, \tag{2.16}
$$

$$
\sin^2\phi_j(0) = 1 - \frac{w_{n+1-j}^2}{(1 - w_n^2)\Pi_{i=2}^{j-1}\sin^2\phi_i(0)}, \tag{2.17}
$$

we obtain a recurrence relation giving $P_{j+1}$ in terms of $P_j$:

$$
\begin{aligned}
P_{j+1} &= (1 - w_{n-j+1}^2)P_j \\
&\quad - \frac{|\xi|^2}{2}\Bigg\{\left[w_{n-j+1}^2\left(1 - \sum_{i=0}^{j-2} w_{n-i}^2\right)^3\right]^{-1} \\
&\quad \times \Bigg[\left(1 - \sum_{i=0}^{j-2} w_{n-i}^2\right) w_{n-j+1} \operatorname{Re} \bar{v}_{n-j+1} \\
&\quad + w_{n-j+1}^2 \operatorname{Re}\left(\sum_{i=0}^{j-2} w_{n-i}\bar{v}_{n-i}\right)\Bigg]^2 \\
&\quad + \left(1 - \sum_{i=0}^{j-1} w_{n-i}^2\right)\operatorname{Im}^2 \bar{v}_{n-j+1}\Bigg\}, \quad j = 2,...,n-1.
\end{aligned} \tag{2.18}
$$

The last parameters to be related to those in (3.2) are the $\tau_j$'s, $j = 2,...,n$. This can clearly be done by an iterative procedure using the last $n-1$ relations in (2.1) together with (2.9), all evaluated at $t = 0$. The result involves the constant $P_i$ and $u_\mu$, already related to the parameters (2.11). The relations are long and will not be given here.

The fact that the two sets of parameters (2.10) and (2.11) can be related shows explicitly the equivalence between the flow (1.6), obtained by the projection method, and the classical equations of motion (2.1) and (2.9).

### C. Plane trajectories on $S^2$

We already see from (1.6) or (2.9) that all the trajectories are closed because of the trigonometric functions. Now an interesting physical question is to ask for which conditions a trajectory can live in a plane in $\mathbb{R}^{n+1}$. For simplicity, let us restrict to $n = 2$.

We calculate the tangent vector at each point of the parametric curve (1.6) and ask for what conditions will the vector product of any two of these tangent vectors always point in the same direction, though not always necessarily with the same norm (this is equivalent to imposing a vanishing torsion of the curve by using Frenet's formula[11]). The condition is that the ratio between the vector product's components must be constant. Thus fixing one of the tangent vectors at $t = 0$, we must have

$$\dot{x}_2(t)\,\mathrm{Re}\,\bar{v}_0 - \dot{x}_0(t)\,\mathrm{Re}\,\bar{v}_2 = C_1(\dot{x}_2(t)\,\mathrm{Re}\,\bar{v}_1 - \dot{x}_1(t)\,\mathrm{Re}\,\bar{v}_2), \tag{2.19a}$$

$$\dot{x}_1(t)\,\mathrm{Re}\,\bar{v}_0 - \dot{x}_0(t)\,\mathrm{Re}\,\bar{v}_1 = C_2(\dot{x}_2(t)\,\mathrm{Re}\,\bar{v}_1 - \dot{x}_1(t)\,\mathrm{Re}\,\bar{v}_2), \tag{2.19b}$$

where $C_1$ and $C_2$ are constants. Treating $\dot{x}_2$ as independent of $\dot{x}_0$ and $\dot{x}_1$ (i.e., trajectories are not restricted to a sphere), the less restrictive condition imposed by (2.19a) is

$$\dot{x}_1\,\mathrm{Re}\,\bar{v}_0 = \dot{x}_0\,\mathrm{Re}\,\bar{v}_1. \tag{2.20}$$

Thus the vector product has only two independent components and relation (2.19b) becomes irrelevant; in other words, it is satisfied for $C_2 = 0$.

Obviously, the two relations

$$\dot{x}_2\,\mathrm{Re}\,\bar{v}_0 = \dot{x}_0\,\mathrm{Re}\,\bar{v}_2, \tag{2.21}$$

$$\dot{x}_1\,\mathrm{Re}\,\bar{v}_2 = \dot{x}_2\,\mathrm{Re}\,\bar{v}_1 \tag{2.22}$$

also satisfy (2.19). Each of the last three relations is necessary to have a free plane trajectory in $\mathbb{R}^3$ but not sufficient if the particle moves along a closed path on a sphere. The complete conditions will be given by the explicit form of the flow (1.6) that contains all the information about the restriction to the sphere and the effect of the potential. In fact, integrating (2.20), we get

$$x_1\,\mathrm{Re}\,\bar{v}_0 = x_0\,\mathrm{Re}\,\bar{v}_1 + K, \tag{2.23}$$

where we show by substituting (1.6) and identifying independent coefficients that $K = 0$ and

$$w_1\,\mathrm{Re}\,\bar{v}_0 = w_0\,\mathrm{Re}\,\bar{v}_1, \tag{2.24a}$$

$$|v_0|\mathrm{Re}\,\bar{v}_1 = |v_1|\mathrm{Re}\,\bar{v}_0. \tag{2.24b}$$

Similar relations exist for (2.21) and (2.22) with an appropriate change in indices. Remembering that $w_\mu$ fixes the initial position, $\mathrm{Re}\,\bar{v}_\mu$ the initial speed, and $\mathrm{Im}\,\bar{v}_\mu$ (once $w_\mu$ is fixed) the magnitude of the potential [see (1.8)], we see that relations (2.24) impose conditions on these quantities in order to have a plane trajectory.

According to (2.23), the projection of the trajectories on the $(x_1,x_0)$ plane is a straight line passing through the origin. Thus the plane (oscillating) trajectories follow a segment of a great circle restricted to an octant if $u_\mu \neq 0$, $\mu = 0,1,2$, to a quadrant if one $u_\mu$ vanishes, and to a hemisphere if two vanish (the potential repels away from the coordinate planes). Obviously, for $V = 0$, the plane trajectories are great circles.

## III. THE QUANTUM PROBLEM

The separation of variables for the Schrödinger equation

$$\sum_{\mu=0}^{n}\left[-\frac{1}{2}\frac{\partial^2}{\partial x_\mu^2} + \frac{u_\mu^2}{x_\mu^2}\right]\psi = E\psi \equiv K_1\psi, \quad \psi = \prod_{i=1}^{n}\psi_i(\phi_i), \tag{3.1}$$

in spherical coordinates (2.1) yields the set of $n$ ordinary differential equations (ODE's)

$$\frac{1}{2}\frac{d^2\psi_n}{d\phi_n^2} + \left[K_n - \frac{u_0^2}{\sin^2\phi_n} - \frac{u_1^2}{\cos^2\phi_n}\right]\psi_n = 0, \tag{3.2a}$$

$$\frac{1}{2}\frac{d^2\psi_{n-k}}{d\phi_{n-k}^2} + \frac{k}{2\tan\phi_{n-k}}\frac{d\psi_{n-k}}{d\phi_{n-k}} + \left[K_{n-k} - \frac{K_{n-k+1}}{\sin^2\phi_{n-k}} - \frac{u_{k+1}^2}{\cos^2\phi_{n-k}}\right]\psi_{n-k} = 0, \quad k = 1,...,n-1, \tag{3.2b}$$

where $K_i$, $i = 2,...,n$, are positive separation constants since all separated potentials are positive wells.

We set

$$2u_\mu^2 = m_\mu(m_\mu + 1), \quad m_\mu \geqslant 0, \tag{3.3}$$

in both equations. Note that $m_\mu$ is not a quantum number since it is fixed from the "outside" by the values of the constants in the potential.

Relation (3.2a) is the Schrödinger equation for the well-known Pöschl–Teller potential,[12] which recently received particular attention in connection to the application of algebraic methods to scattering.[13,14] The two independent solutions for $u_0^2 > 0$ and $u_1^2 > 0$ can be given in terms of the hypergeometric function $F$ (see Ref. 15):

$$\psi_n \begin{cases} \sin^{m_0+1}\phi_n \cos^{m_1+1}\phi_n\, F(a,b,c,\sin^2\phi_n), & (3.4) \\ \sin^{-m_0}\phi_n \cos^{m_1+1}\phi_n\, F(1+a-c,1+b-c,2-c,\sin^2\phi_n), & (3.5) \end{cases}$$

where

$$a,b = \tfrac{1}{2}(m_0 + m_1 + 2 \mp \sqrt{2K_n}), \tag{3.6}$$

$$c = m_0 + \tfrac{3}{2}. \tag{3.7}$$

Only (3.4) is bounded at $\phi_n = 0, \pm\pi, \pm 2\pi, \ldots$. Furthermore, by using appropriate identities, one can see that the behavior of (3.4) for $\phi_n \to \pm\pi/2, \pm 3\pi/2, \ldots$ is

$$\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} \sin^{m_0+1}\phi_n \cos^{-m_1}\phi_n. \tag{3.8}$$

To avoid the divergence, we must truncate the hypergeometric series by imposing

$$a = -m, \quad m = 0,1,2,\ldots. \tag{3.9}$$

This leads to the relation

$$2K_n = (2m + m_0 + m_1 + 2)^2 \tag{3.10}$$

and permits us to give $\psi_n$ in terms of the Jacobi polynomials as

$$\psi_n(\phi_n) = A \sin^{m_0+1}\phi_n \cos^{m_1+1}\phi_n P_m^{(m_0+1/2, m_1+1/2)} \times (\cos 2\phi_n), \tag{3.11}$$

where $A$ is a normalization constant.

Finally, the periodicity $\psi_n(\phi_n) = \psi_n(\phi_n + 2\pi)$ imposes that

$$m_0 + m_1 = 1,2,3,\ldots. \tag{3.12}$$

The case $u_0 = 0$, $u_1^2 > 0$ may be transformed to the case $m_0 = m_1$ by the identity

$$\frac{1}{\cos^2\phi_n} = \frac{1}{4}\left[\frac{1}{\sin^2(\phi_n/2 - \pi/4)} + \frac{1}{\cos^2(\phi_n/2 - \pi/4)}\right]. \tag{3.13}$$

The solution is then given in terms of the Gegenbauer polynomials as

$$\psi_n(\phi_n) = A \cos^{m_1+1}\phi_n \, C_m^{m_1+1}(\sin\phi_n), \quad m_1 = 1,2,3,\ldots, \tag{3.14}$$

with

$$2K_n = (m + m_1 + 1)^2, \quad m = 0,1,2,\ldots. \tag{3.15}$$

Similarly, for the case $u_0^2 > 0$, $u_1 = 0$, we obtain

$$\psi_n(\phi_n) = A \sin^{m_0+1}\phi_n \, C_m^{m_0+1}(\cos\phi_n), \quad m_0 = 1,2,3,\ldots, \tag{3.16}$$

with

$$2K_n = (m + m_0 + 1)^2, \quad m = 0,1,2,\ldots. \tag{3.17}$$

Finally for $u_0 = u_1 = 0$, we have the two obvious independent solutions

$$\cos m\phi_n, \tag{3.18a}$$

$$\sin m\phi_n, \tag{3.18b}$$

with

$$2K_n = m^2, \quad m = 0,1,2,\ldots, \tag{3.19}$$

since this is the only case giving two independent bounded solutions.

To solve Eqs. (3.2b) we set, together with (3.3), the identities

$$2K_{n-k} = p_{n-k}(p_{n-k} + k), \quad p_{n-k} \geqslant 0, \tag{3.20}$$

$$2K_{n-k+1} = s_{n-k+1}(s_{n-k+1} + k - 1), \quad s_{n-k+1} = 0,1,2,\ldots. \tag{3.21}$$

Note that $s_{n-k+1}$ is fixed by the value of $K_{n-k+2}$ found from the $(k-1)$th ODE in the series (3.2) and is always a non-negative integer. The two independent solutions can still be given in terms of the hypergeometric function as

$$\psi_{n-k}(\phi_{n-k}) = \begin{cases} \tan^{s_{n-k+1}}\phi_{n-k} \cos^{p_{n-k}}\phi_{n-k} F(a,b,c, -\tan^2\phi_{n-k}), & (3.22) \\ \cot^{k+s_{n-k+1}-1}\phi_{n-k} \cos^{p_{n-k}}\phi_{n-k} F(1+a-c, 1+b, -c, 2-c, -\tan^2\phi_{n-k}), & (3.23) \end{cases}$$

where

$$a = \tfrac{1}{2}(s_{n-k+1} - p_{n-k} + m_{k+1} + 1), \tag{3.24}$$

$$b = \tfrac{1}{2}(s_{n-k+1} - p_{n-k} - m_{k+1}), \tag{3.25}$$

$$c = s_{n-k+1} + \tfrac{1}{2}(k+1). \tag{3.26}$$

For the case $m_{k+1} > 0$ and $s_{n-k+1} \geqslant 0$, only (3.22) is bounded at $\phi_{n-k} = 0, \pm\pi, \pm 2\pi, \ldots$ [note that (3.22) coincides with (3.23) for $s_{n-k+1} = 0$ and $k = 1$]. For $\phi_{n-k} = \pm\pi/2, \pm 3\pi/2, \ldots$, the hypergeometric function diverges while the factor $\tan^s\phi \cos^p\phi = \sin^s\phi \cos^{p-s}\phi$ converges for $p \geqslant s$. We thus have to truncate the hypergeometric series by imposing

$$s_{n-k+1} - p_{n-k} + m_{k+1} + 1 = -2l_{n-k}, \quad l_{n-k} = 0,1,2,\ldots. \tag{3.27}$$

This yields the relation

$$2K_{n-k} = (2l_{n-k} + s_{n-k+1} + m_{k+1} + 1)(2l_{n-k} + s_{n-k+1} + m_{k+1} + 1 + k). \tag{3.28}$$

Finally, the bounded solutions can be written in terms of the Jacobi polynomials as

$$\psi_{n-k}(\phi_{n-k}) = A \sin^{s_{n-k+1}}\phi_{n-k} \cos^{m_{k+1}+1}\phi_{n-k} \times P_{l_{n-k}}^{(s_{n-k+1}+(1/2)(k-1), m_{k+1}+1/2)}(\cos 2\phi_{n-k}). \tag{3.29}$$

Equation (3.29) satisfies $\psi(\phi) = \psi(\phi + 2\pi)$ if $s_{n-k+1} + m_{k+1} + 1$ is an integer; thus $m_{k+1} \in \mathbb{Z}^+$.

The last case to be solved is when $u_{k+1} = 0$. Let $m_{k+1} = -1$ in (3.22) with

$$s_{n-k+1} - p_{n-k} = -l_{n-k}, \quad l_{n-k} = 0,1,2,\ldots. \tag{3.30}$$

In fact, one of the two numbers $(s-p)/2$ or $(s-p+1)/2$ will be a nonpositive integer, so the hypergeometric series will be truncated. The eigenvalue $K_{n-k}$ is then given by

$$2K_{n-k} = (l_{n-k} + s_{n-k+1})(l_{n-k} + s_{n-k+1} + k) \tag{3.31}$$

and the bounded wave function by

$$\psi_{n-k}(\phi_{n-k}) = \tilde{A} \sin^{s_{n-k+1}} \phi_{n-k} \cos^{l_{n-k}} \phi_{n-k} \times F\left(\frac{-l_{n-k}}{2}, \frac{1-l_{n-k}}{2}, s_{n-k+1} + \frac{k+1}{2}, -\tan^2 \phi_{n-k}\right), \tag{3.32}$$

which can be reduced to the form[16]

$$\psi_{n-k}(\phi_{n-k}) = A \sin^{(1/2)(1-k)} \phi_{n-k} \times P^{s_{n-k+1}+(1/2)(k-1)}_{l_{n-k}+s_{n-k+1}+(1/2)(k-1)}(\cos \phi_{n-k}), \tag{3.33}$$

where $P$ is the associated Legendre function. For $k = 1$, we recognize the classical spherical harmonics.

The energy level can be calculated in terms of $l_{k-n}$ and $m$ by simply substituting the different iterative values for the $K_i$'s. For example, in the generic case $m_\mu \neq 0$, $\mu = 0,1,\ldots,n$, we have

$$\begin{aligned}
2K_n &= (2m + m_0 + m_1 + 2)^2 = s_n^2, \\
2K_{n-1} &= (2l_{n-1} + s_n + m_2 + 1) \times (2l_{n-1} + s_n + m_2 + 2) \\
&= s_{n-1}(s_{n-1} + 1), \\
2K_{n-2} &= (2l_{n-2} + s_{n-1} + m_3 + 1) \times (2l_{n-2} + s_{n-1} + m_3 + 3) \\
&= s_{n-2}(s_{n-2} + 2), \\
&\vdots \\
2K_2 &= (2l_2 + s_3 + m_{n-1} + 1)(2l_2 + s_3 + m_{n-1} + n - 1) \\
&= s_2(s_2 + n - 2), \\
2E \equiv 2K_1 &= (2l_1 + s_2 + m_n + 1)(2l_1 + s_2 + m_n + n),
\end{aligned} \tag{3.34}$$

which yields

$$2E = \left(2\sum_{k=1}^{n-1} l_k + 2m + \sum_{\mu=0}^{n} m_\mu + n + 1\right) \times \left(2\sum_{k=1}^{n-1} l_k + 2m + \sum_{\mu=0}^{n} m_\mu + 2n\right), \tag{3.35}$$

with $l_k, m = 0,1,2,\ldots$ .

Relation (3.35) shows that the degeneracy corresponds to that of the $n$-dimensional isotropic harmonic oscillator; i.e., the levels are degenerated according to the dimensions of the symmetric irreducible representations of SU($n$).

## IV. CONCLUSION

The Hamiltonian (1.1) has been solved by separation of variables in spherical coordinates for the classical and quantum cases. We related the two sets of parameters involved in the projected flow and the classical equation of motion to state the equivalence between the projection method and the Hamilton–Jacobi equation. We also found the conditions for degenerate classical trajectories, i.e., plane trajectories on the sphere $S^2$. We saw that the only possible cases are segments of great circles. Finally we gave the general wave function on $S^n$ for the quantum Hamiltonian in spherical coordinates. This last result may be the starting point of a future work that could consist of comparing the solution of the Schrödinger equation to the one obtained by the geometrical quantification method.[17] Finally we plan to make an analysis similar to that above for a different class of integrable systems already obtained from the Hermitian hyperbolic space HH($n$) (see Ref. 18) (rather than $S^{2n+1}$ or $\mathbb{C}P^n$) by quotienting the phase space by the maximal Abelian subgroups of SU($n$,1). All the systems can be integrable by the group projection method as well as other classical ones.


## ACKNOWLEDGMENTS

The author thanks P. Winternitz for helpful discussions and a critical reading of the manuscript.

He also thanks the Natural Sciences and Engineering Research Council of Canada and the Université de Montréal for fellowships.



[1] E. Rosochatius, dissertation, Göttingen, Gebr. Unger, Berlin, 1877.
[2] J. Moser, "Various aspects of integrable Hamiltonian systems," in *Dynamical Systems*, CIME Lectures, Bressanone, Italy, June 1978; *Progress in Mathematics*, Vol. 8 (Birkhauser, Boston, 1980).
[3] T. Ratiu, "The Lie algebraic interpretation of the complete integrability of the Rosochatius system," in *Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems*, AIP Conference Proceedings, Vol. 88 (AIP, New York, 1982).
[4] M. A. Olshanetsky and A. M. Perelomov, Phys. Rep. **71**, 313 (1981).
[5] L. Gagnon, J. Harnad, J. Hurtubise, and P. Winternitz, J. Math. Phys. **26**, 1605 (1985).
[6] J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[7] R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[8] C. P. Boyer, E. G. Kalnins, and P. Winternitz, SIAM J. Math. Anal. **16**, 93 (1985).
[9] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968).
[10] L. Gagnon, M.Sc. thesis, Montréal, 1984.
[11] See, for example, Carmo, Manfredo Perdigäo do, *Differential Geometry of Curves and Surfaces* (Prentice-Hall, New York, 1976).
[12] G. Pöschl and E. Teller, Z. Phys. **83**, 143 (1933).
[13] Y. Alhassid, J. Engel, and J. Wu, Phys. Rev. Lett. **53**, 17 (1984).
[14] A. Frank and K. B. Wolf, Phys. Rev. Lett. **52**, 1737 (1984).
[15] G. M. Murphy, *Ordinary Differential Equations and Their Solutions* (Van Nostrand, Princeton, NJ, 1960).
[16] W. Magnus, F. Oberhettinger, and F. G. Triconi, *Tables of Integral Transforms*, Bateman Manuscript Project, Vol. 2 (McGraw-Hill, New York, 1954).
[17] A. G. Reyman and M. A. Semenov-Tian-Shansky, Invent. Math. **54**, 81 (1979).
[18] C. P. Boyer, E. G. Kalnins, and P. Winternitz, J. Math. Phys. **24**, 2022 (1982).

# Classical relativistic spinning particles

Martín Rivas
*Departamento de Física Teórica, Facultad de Ciencias, Universidad del País Vasco-Euskal Herriko Unibertsitatea, Apdo. 644, 48080 Bilbao, Spain*

(Received 30 March 1988; accepted for publication 12 October 1988)

An elementary particle is defined as a mechanical system whose kinematical space is a homogeneous space of the Poincaré group. Lagrangians for describing these systems depend on higher-order derivatives and some of them are analyzed. For bradyons the Lagrangian depends on the acceleration and angular velocity of the particle and is characterized by two parameters $m$ and $s$, the rest mass and absolute value of spin, respectively. In general the spin is of kinematical nature, related to the rotation and internal orbital motion of the system. Two different kinds of bradyons appear according to the spin structure. One has a spin related to the generalized angular velocity while the other is a function of the generalized acceleration. Photons are massless particles with spin lying along the direction of motion and energy $h\nu$, where $\nu$ is the frequency of its rotational motion. Particles moving in circles with velocity $c$ in their center of mass frame are also predicted, showing a Dirac-type Hamiltonian. There also appear particles with tachyonic orbital motion whose center of mass has bradyonic motion. Transformation properties under space and time reversal are also analyzed.

## I. INTRODUCTION

Many attempts can be found in the literature to describe classical spinning particles, which in general endow the point particle with some additional degrees of freedom to give account of the spin structure.[1]

Recently,[2] the possible internal spaces that, in addition to the space-time position variables, describe spinning particles, have been completely classified and analyzed.

In the approach we present here we shall describe structured particles by adding some extra degrees of freedom and by allowing the Lagrangian to depend on higher-order derivatives. Lagrangian theory was generalized by Ostrogradsky[3] and since then several contributors have claimed to consider generalized Lagrangians for studying generalized electrodynamics[4] and the classical spinning particle.[5]

We are thus basically working in a generalized Lagrangian formalism, in which dependence on higher-order derivatives is assumed, and which is sketched in Sec. II. However, we assume that the dynamics is based upon the knowledge of the action function of the system, which is a real function defined on $X \times X$, where $X$ is the kinematical space of the system that will be conveniently defined and that is in general different from the configuration space and the phase space of the system. This means that a particular path followed by the system can be expressed in terms of endpoint conditions in $X$ space, as in Feynman's approach. Physical considerations lead us to define in Sec. II C an elementary particle as a system for which $X$ is a homogeneous space of the kinematic group $G$. This statement restricts the dependence on higher-order derivatives to the $G$ structure. In this work $G$ will be the Poincaré group $\mathscr{P}$, so that $X$ is at most ten dimensional, implying that the Lagrangian dependence on the derivatives is at most on the acceleration and angular velocity of the particle.

The remainder of this work shows that the proposed formalism is not empty, by explicitly constructing several Lagrangians. In order to work out a specific $X$ space we present in Sec. III a useful Poincaré group parametrization, where the parameters are the relative velocity and orientation and the space and time translation among inertial observers.

In Sec. IV we study the simplest case, that of a point particle, obtaining the habitual results, but preparing the ground for further applications. In Secs. V and VI we analyze two particular Poincaré homogeneous spaces: the most general bradyon and the kinematical space of particles that travel at the speed of light. In the first, two kinds of particles come out according to the kinematical structure of their spins. In the second group we have found the photon with its properties of having no transversal spin arriving at $H = h\nu$ for the expression of its energy, where $\nu$ is the frequency of its rotational motion along the spin direction.

However, we have also found particles that, although they travel at the speed of light, have a center of mass with a straight bradyonic motion with constant velocity below $c$. For these particles we have found a certain analogy between their Hamiltonian and Dirac's Hamiltonian, and a particular Lagrangian has been analyzed. In Sec. VII particles with internal orbital tachyonic motion are considered, having a center of mass that travels at velocity $u \leqslant c$. Section VIII is devoted to the analysis of the previous Lagrangians under the discrete symmetry operations of time and space inversion.

## II. GENERAL FORMALISM

### A. Generalized Lagrangian systems

Although the generalized Lagrangian formalism is well known, we shall sketch it briefly in order to enhance the role of the manifold $X$, the kinematical space of the system, and the action function on $X \times X$, which are defined later.

Let us consider those mechanical systems of $n$ degrees of freedom that can be described by means of a generalized Lagrangian function $\mathbf{L}(t, q_i^{(s)}(t))$, $i = 1,...,n$, $s = 0,1,...,k$, which depends on the time $t$ and on the $n$ generalized coordinates $q_i(t)$, and their derivatives up to order $k$. Here $q_i^{(s)}(t) = d^s q_i(t)/dt^s$.

 0022-2488/89/020318-12$02.50 © 1989 American Institute of Physics 

The action functional is defined as

$$A[q(t)] = \int_{t_1}^{t_2} L(t,q_i^{(s)}(t))dt, \tag{2.1}$$

where the condition that $A$ be extremal for the class of paths $q_i(t)$ with fixed end points [i.e., with fixed values $q_i^{(s)}(t_1)$, $q_i^{(s)}(t_2)$, $i=1,...,n$, $s=0,1,...,k-1$] implies that the functions $q_i(t)$ must necessarily satisfy the Euler–Lagrange dynamical equations.[6]

$$\sum_{s=0}^{k} (-1)^s \frac{d^s}{dt^s}\left(\frac{\partial L}{\partial q_i^{(s)}}\right) = 0, \quad i=1,...,n. \tag{2.2}$$

A generalization to systems for which the order $k$ is different for each generalized coordinate $q_i$ can be obtained easily, but in this work we shall consider for simplicity the same order $k$ in all variables.

Existence and uniqueness theorems imply that a particular solution of this $2k$ th-order system (2.2) is determined by giving the $2kn$ values $q_i^{(s)}(t)$ $i=1,...,n$, $s=0,1,...,2k-1$, at the initial time $t_1$. If we fix end-point conditions, i.e., the values $q_i^{(s)}(t_1)$ and $q_i^{(s)}(t_2)$, $i=1,...,n$, $s=0,1,...,k-1$, there will not exist, in general, a solution of (2.2), although the variational problem (2.1) leads to the system (2.2) for the class of paths with fixed end-point conditions. However, if there exist solutions, perhaps nonunique, with fixed end points, this means in some sense that the above initial conditions at time $t_1$ can be expressed, perhaps in a nonuniform way, in terms of the end-point conditions. Thus a particular solution is finally expressed as a function $q_i(t;q_j^{(s)}(t_1), q_j^{(s)}(t_2))$, $j=1,...,n$, $s=0,1,...,k-1$, of time and of $2kn$ independent parameters, related to end-point conditions, and we shall consider from now on those mechanical systems for which this holds. A generalized Lagrangian formalism and the existence of solutions with fixed end-point conditions are the basic assumptions of the formalism we propose.

By considering this particular solution, the action function is defined as the value of the functional (2.1) for this particular path. Thus the action function becomes a function of $2(kn+1)$ independent variables

$$A(t_1,q_i^{(s)}(t_1);t_2,q_i^{(s)}(t_2))$$

$$\equiv A(x_1,x_2), \quad i=1,...,n, \quad s=0,1,...,k-1,$$

with the property $A(x,x)=0$.

*Definition:* We shall call kinematical variables of the system to the time $t$ and the $n$ generalized coordinates and their derivatives up to order $k-1$ $q_i^{(s)}, s=0,1,...,k-1$, and they will be denoted by $x_j, j=0,1,...,kn$. The $(kn+1)$-dimensional manifold spanned by the kinematical variables is called the kinematical space of the system $X$.

If the trajectories are written in parametric form $\{t(\tau),q(\tau)\}$, in terms of some evolution parameter $\tau$, the Lagrangian can be expressed in terms of the kinematical variables and their first $\tau$ derivatives, and (2.1) appears:

$$A[(\tau),q(\tau)] = \int_{\tau_1}^{\tau_2} L\left(x(\tau), \frac{\dot{x}(\tau)}{\dot{t}(\tau)}\right)\dot{t}(\tau)d\tau$$

$$= \int_{\tau_1}^{\tau_2} \hat{L}(x(\tau),\dot{x}(\tau))d\tau, \tag{2.3}$$

where $\hat{L} = L\dot{t}(\tau)$.

Although (2.3) looks like a first-order system of $kn+1$ degrees of freedom, we see that there exist among the kinematical variables $(k-1)n$ nonholonomic differential constraints $q_i^{(s)}(\tau) = \dot{q}_i^{(s-1)}(\tau)/\dot{t}(\tau)$, $i=1,...,n$, $s=1,...,k-1$, where the overdot means a $\tau$ derivative.

We can see[7] that the function $\hat{L}$ is $\tau$ independent and homogeneous of first degree in terms of the derivatives of the kinematical variables giving rise to a further constraint

$$\hat{L} = \frac{\partial \hat{L}}{\partial \dot{x}^j}\dot{x}^j = F_j(x,\dot{x})\dot{x}^j, \tag{2.4}$$

which, together with the $(k-1)n$ differential constraints, reduces to $n$ the number of independent variables. The action functional in the form (2.3) is also independent with respect to parametric transformations, and the functions $F_j(x,\dot{x})$ are homogeneous functions of zero degree in the derivatives of the kinematical variables.

Conversely, if the system is described by the knowledge of the action function $A(x_1,x_2)$, which is assumed to be a continuous and differentiable function of the kinematical variables of the initial and final points, then the Lagrangian can be obtained by the limiting process:

$$\hat{L} = \lim_{y\to x} \frac{\partial A(x,y)}{\partial y^j}\dot{x}^j. \tag{2.5}$$

This can be seen from (2.3) by considering two close points and thus

$$A(x(\tau),x(\tau+d\tau)) = A(x(\tau),x(\tau)+\dot{x}(\tau)d\tau) = \hat{L}\,d\tau.$$

By making a Taylor expansion of $A$, taking into account the condition $A(x,x)=0$, we get (2.5).

The function of the kinematical variables and their derivatives (2.5) together with the homogeneity condition (2.4) and the differential constraints among the kinematical variables reduces the problem to that of a system of $n$ degrees of freedom, where its Lagrangian is a function of the derivatives of the generalized coordinates up to order $k$. From now on we shall delete the caret over the function $\hat{L}$, and we shall consider systems for which trajectories are written in parametric form.

What we want to emphasize is that the important dynamical object of the theory is the action function. Its knowledge determines by (2.5) the Lagrangian $L$, and thus the dynamical equations (2.2). Here $A(x_1,x_2)$ characterizes the dynamics globally. We have a similar situation in the quantum scattering theory, in which the dynamics is globally contained in the $S$ matrix. The Feynman path integral approach links both formalisms by relating the action function for a particular path with the phase of the corresponding probability amplitude. In quantum mechanics all paths can be followed, so that we have to add the corresponding probability amplitudes; while in classical mechanics the variational formalism singles out just one path, and thus the action function for that path contains the required dynamical information.

### B. The relativity principle

Let $G$ be the kinematical group[8] that acts transitively on the space-time $Y$ as a transformation group. The group $G$ defines the class of equivalent observers, called inertial ob-

servers, for which the laws of physics are the same, and we shall assume that a realization of $G$ on the kinematical space of the system $X$ is known.

The invariance of the dynamical equations for two inertial observers $O$ and $O'$ related by a transformation $g \in G$, implies[9] that the action function must transform according to

$$A(gx_1,gx_2) = A(x_1,x_2) + \alpha(g;x_2) - \alpha(g;x_1), \quad (2.6)$$

where $\alpha(g;x)$ is a function defined on $G \times X$, which verifies, for all $g,g' \in G$, and all $x \in X$,

$$\alpha(g';gx) + \alpha(g;x) - \alpha(g'g;x) = \xi(g',g), \quad (2.7)$$

where $\xi(g',g)$ is an exponent of $G$.[10]

This function $\alpha(g;x)$ is called a gauge function for the group $G$ and the kinematical space $X$. Different mechanical systems with the same kinematical space $X$ can be characterized by different gauge functions.

From (2.5), the Lagrangian transforms

$$L\left(gx(\tau),\frac{d(gx(\tau))}{d\tau}\right) = L(x(\tau),\dot{x}(\tau)) + \frac{d\alpha(g;x(\tau))}{d\tau}, \quad (2.8)$$

which, together with the homogeneity condition (2.4), will lead to certain transformation properties for the functions $F_j$ under the group $G$, giving us information about the structure of these functions. Expression (2.8) is the restriction imposed to the Lagrangian by the relativity principle.

Among the gauge functions there exists an equivalence relation.[9] Two gauge functions $\alpha_1$ and $\alpha_2$ are said to be equivalent if

$$\alpha_1(g;x) - \alpha_2(g;x) = \Phi(x) - \Phi(gx) + \sigma(g), \quad (2.9)$$

where $\Phi$ and $\sigma$ are some functions defined on $X$ and $G$, respectively. Thus with $G$ and $X$ fixed, to every $\alpha(g;x)$ solution of (2.7) up to an equivalence, the relativity principle in its form (2.8) will give us information about the Lagrangian mechanical systems whose dynamical laws are $g$ invariant.

In particular if $X$ is a homogeneous space of $G$ then (2.7) has the solution[9]

$$\alpha(g;x) = \xi(g,h_x), \quad (2.10)$$

where $h_x \in G$ is any element of the equivalence class $x \in X$.

In this paper $G$ will be the Poincaré group $\mathscr{P}$, and all its exponents are equivalent to zero[10] so that, for those mechanical systems for which $X$ is a homogeneous space of $\mathscr{P}$, the action function and the Lagrangian can be taken strictly invariant.

### C. Elementary systems

An elementary mechanical system will be defined as that system for which the evolution from the initial to the final point, if no interaction is present, is necessarily free.

Let us consider a system that is observed at instant $\tau$ by a certain inertial observer $O$. At instant $\tau + d\tau$ some physical observables will change their values as measured by $O$. However, if the system is elementary, then there will exist at instant $\tau + d\tau$ another inertial observer $O'$ for which the measurements of physical observables will give the same values as those obtained by $O$ at the earlier time $\tau$.

These two inertial observers will be related by some infinitesimal transformation $\delta g(\tau)$ of the kinematical group $G$. If the evolution of the elementary system is free, this means that the corresponding infinitesimal transformation $\delta g(\tau)$ must be independent of $\tau$. Otherwise we could distinguish one instant from another by looking at the different change in the physical observables, and thus concluding that this difference in the physical behavior of the system is produced by some interaction.

Thus if the evolution is free, the measurement of any observable by observer $O$ at instant $\tau + d\tau$ will be obtained from its measurement at instant $\tau$ by acting with $\delta g$ in the corresponding realization of the algebra of observables. Since $\delta g$ is constant, it generates a one-parameter subgroup of $G$, such that the evolution of any observable is the action of this one-parameter subgroup on its initial value. In this way, the free or inertial motions are identified with the one-parameter subgroups of $G$.

We have seen in Sec. II A that for Lagrangian systems the dynamical information is contained in the action function, which is a function of the kinematical variables at the initial and final points. If the evolution is free, the final point $x_2$ is obtained by acting on $x_1$ with the corresponding one-parameter subgroup generated by $\delta g$, and thus there exists a finite group element $g$ such that $x_2 = gx_1$.

Conversely, if we fix $x_1$ and $x_2$ and the evolution has to be free, then necessarily the kinematical space has to be a homogeneous space of $G$. Otherwise, if $X$ is not a homogeneous space of $G$, then in general there will not exist any group element and any one-parameter subgroup of $G$ that brings $x_1$ to $x_2$, and the evolution of the system will no longer be free.

*Definition:* An elementary classical particle is that mechanical system for which its kinematical space $X$ is a homogeneous space of the corresponding kinematical group $G$.

## III. THE POINCARE GROUP

The kinematical group for relativistic systems is the Poincaré group $\mathscr{P}$. The usual covariant parametrization of $\mathscr{P}$ is given by the four-vector $a^\mu$ of the space-time translations and the 16 elements $\Lambda^\mu{}_\nu$ of the Lorentz transformation, and we write it as $(a,\Lambda)$. The composition law is given by

$$(a',\Lambda')(a,\Lambda) = (\Lambda' a + a',\Lambda'\Lambda),$$

i.e.,

$$a''^\mu = \Lambda'^\mu{}_\nu a^\nu + a'^\mu, \quad \Lambda''^\mu{}_\nu = \Lambda'^\mu{}_\sigma \Lambda^\sigma{}_\nu.$$

However, the elements $\Lambda^\mu{}_\nu$ are not independent and they verify the ten relations $\Lambda^\mu{}_\sigma \eta^{\sigma\lambda} \Lambda^\nu_\lambda = \eta^{\mu\nu}$, where $\eta^{\mu\nu}$ is the Minkowski metric tensor. This parametrization is used in Refs. 1 and 11.

Instead of this covariant parametrization we give here an essential parametrization in terms of ten independent parameters.[12] Since every Lorentz transformation can always be written as a product $\Lambda = LR$ of a boost $L$ by a rotation $R$, we shall use the relative velocity vector $\mathbf{v}$ that characterizes $L$ and the three angular variables for the rotation $R$ as the six essential parameters without any further constraint. Now

the formulation is not manifestly covariant but the physical interpretation of these parameters as velocity and orientation will be shared later by the variables of the corresponding homogeneous $X$ spaces.

Then every element is parametrized in terms of the ten real parameters $g \equiv (b,\mathbf{a},\mathbf{v},\boldsymbol{\mu})$, where $b \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^3$ represent the time and space translation, $\mathbf{v} \in \mathbb{R}^3$, with $v < c$, is the relative velocity among observers, and $\boldsymbol{\mu} = \mathbf{e} \tan(\alpha/2)$ is the relative orientation of their spatial Cartesian coordinate frames. It is parametrized by the clockwise rotation around the direction given by the unit vector $\mathbf{e}$ and of value $\alpha \in [0,\pi]$ to get the $O'$ frame from that of $O$. With this parametrization, $\boldsymbol{\mu}$ takes values on a real three-dimensional compact manifold which we shall denote by $\mathbb{R}^3_c$.

The orthogonal rotation matrix $R(\boldsymbol{\mu})$ is given by

$$R(\boldsymbol{\mu})_{ij} = [1/(1+\mu^2)][(1-\mu^2)\delta_{ij} + 2\mu_i\mu_j - 2\epsilon_{ijk}\mu^k]. \tag{3.1}$$

The composition of two rotations $R(\boldsymbol{\mu}')R(\boldsymbol{\mu}) = R(\boldsymbol{\mu}'')$ is

$$\boldsymbol{\mu}'' = (\boldsymbol{\mu}' + \boldsymbol{\mu} + \boldsymbol{\mu}' \times \boldsymbol{\mu})/(1 - \boldsymbol{\mu}' \cdot \boldsymbol{\mu}). \tag{3.2}$$

The action of a group element $g$ on the Minkowski space-time $\mathscr{M}$ is

$$t' = \gamma t + \gamma(\mathbf{v} \cdot R(\boldsymbol{\mu})\mathbf{r})/c^2 + b, \tag{3.3}$$

$$\mathbf{r}' = R(\boldsymbol{\mu})\mathbf{r} + \gamma \mathbf{v} t + \gamma^2(\mathbf{v} \cdot R(\boldsymbol{\mu})\mathbf{r})\mathbf{v}/(1+\gamma)c^2 + \mathbf{a}, \tag{3.4}$$

where $(t,\mathbf{r})$ and $(t',\mathbf{r}')$ are the coordinates of the same space-time event for observers $O$ and $O'$, respectively, where $\gamma = (1 - v^2/c^2)^{-1/2}$.

The composition law of the Poincaré group in this parametrization $g'' = g'g$ can be expressed[12] as

$$b'' = \gamma' b + \gamma'(\mathbf{v}' \cdot R(\boldsymbol{\mu}')\mathbf{a})/c^2 + b', \tag{3.5}$$

$$\mathbf{a}'' = R(\boldsymbol{\mu}')\mathbf{a} + \gamma' \mathbf{v}' b + [\gamma'^2/(1+\gamma')c^2] \times (\mathbf{v}' \cdot R(\boldsymbol{\mu}')\mathbf{a})\mathbf{v}' + \mathbf{a}', \tag{3.6}$$

$$\mathbf{v}'' = \frac{R(\boldsymbol{\mu}')\mathbf{v} + \gamma'\mathbf{v}' + \gamma'^2 c^{-2}(\mathbf{v}' \cdot R(\boldsymbol{\mu}')\mathbf{v})\mathbf{v}'/(1+\gamma')}{\gamma'(1 + (\mathbf{v}' \cdot R(\boldsymbol{\mu}')\mathbf{v})/c^2)}, \tag{3.7}$$

$$\boldsymbol{\mu}' = \frac{\boldsymbol{\mu}' + \boldsymbol{\mu} + \boldsymbol{\mu}' \times \boldsymbol{\mu} + \mathbf{F}(\mathbf{v}',\boldsymbol{\mu}';\mathbf{v},\boldsymbol{\mu})}{1 - \boldsymbol{\mu}' \cdot \boldsymbol{\mu} + G(\mathbf{v}',\boldsymbol{\mu}';\mathbf{v},\boldsymbol{\mu})}, \tag{3.8}$$

where $\mathbf{F}(\mathbf{v}',\boldsymbol{\mu}';\mathbf{v},\boldsymbol{\mu})$ and $G(\mathbf{v}',\boldsymbol{\mu}';\mathbf{v},\boldsymbol{\mu})$ are the functions

$$\begin{aligned}\mathbf{F}(\mathbf{v}',\boldsymbol{\mu}';\mathbf{v},\boldsymbol{\mu}) = [\gamma'\gamma/(1+\gamma')(1+\gamma)c^2][&\mathbf{v} \times \mathbf{v}' + \mathbf{v}(\mathbf{v}' \cdot \boldsymbol{\mu}') \\ &+ \mathbf{v}'(\mathbf{v} \cdot \boldsymbol{\mu}) + \mathbf{v} \times (\mathbf{v}' \times \boldsymbol{\mu}') \\ &+ (\mathbf{v} \times \boldsymbol{\mu}) \times \mathbf{v}' + (\mathbf{v} \cdot \boldsymbol{\mu})(\mathbf{v}' \times \boldsymbol{\mu}') \\ &+ (\mathbf{v} \times \boldsymbol{\mu})(\mathbf{v}' \cdot \boldsymbol{\mu}') + (\mathbf{v} \times \boldsymbol{\mu}) \\ &\times (\mathbf{v}' \times \boldsymbol{\mu}')],\end{aligned} \tag{3.9}$$

$$\begin{aligned}G(\mathbf{v}',\boldsymbol{\mu}';\mathbf{v},\boldsymbol{\mu}) = [\gamma'\gamma/(1+\gamma')(1+\gamma)c^2] &\times [\mathbf{v} \cdot \mathbf{v}' + \mathbf{v} \cdot (\mathbf{v}' \times \boldsymbol{\mu}') + \mathbf{v}' \cdot (\mathbf{v} \times \boldsymbol{\mu}) \\ &- (\mathbf{v} \cdot \boldsymbol{\mu})(\mathbf{v}' \cdot \boldsymbol{\mu}') + (\mathbf{v} \times \boldsymbol{\mu}) \cdot (\mathbf{v}' \times \boldsymbol{\mu}')].\end{aligned} \tag{3.10}$$

The proper Lorentz group $\mathscr{L}$ is the set $\{(0,0,\mathbf{v},\boldsymbol{\mu}) \mid \mathbf{v} \in \mathbb{R}^3, v < c, \boldsymbol{\mu} \in \mathbb{R}^3_c\}$ so that every Lorentz transformation is parametrized in terms of the two three-vectors $\mathbf{v}$ and $\boldsymbol{\mu}$, $\Lambda(\mathbf{v},\boldsymbol{\mu})$ is a product of a pure Lorentz transformation $L(\mathbf{v})$ by the rotation $R(\boldsymbol{\mu})$, and $\Lambda(\mathbf{v},\boldsymbol{\mu}) = L(\mathbf{v})R(\boldsymbol{\mu})$. The expression (3.7) is the relativistic addition of the two velocities $\mathbf{v}'$ and $R(\boldsymbol{\mu}')\mathbf{v}$, because

$$\begin{aligned}L(\mathbf{v}'')R(\boldsymbol{\mu}'') &= L(\mathbf{v}')R(\boldsymbol{\mu}')L(\mathbf{v})R(\boldsymbol{\mu}) \\ &= L(\mathbf{v}')R(\boldsymbol{\mu}')L(\mathbf{v})R(-\boldsymbol{\mu}')R(\boldsymbol{\mu}')R(\boldsymbol{\mu}) \\ &= L(\mathbf{v}')L(R(\boldsymbol{\mu}')\mathbf{v})R(\boldsymbol{\mu}')R(\boldsymbol{\mu}),\end{aligned}$$

since $R(\boldsymbol{\mu}')L(\mathbf{v})R(-\boldsymbol{\mu}') = L(R(\boldsymbol{\mu}')\mathbf{v})$ and the composition of the two boosts $L(\mathbf{v}')L(R(\boldsymbol{\mu}')\mathbf{v}) = L(\mathbf{v}'')R(\mathbf{w})$, where $R(\mathbf{w})$ is the corresponding Wigner rotation. The expression (3.8) comes from $R(\boldsymbol{\mu}'') = R(\mathbf{w})R(\boldsymbol{\mu}')R(\boldsymbol{\mu})$.

The general continuous subgroups of $\mathscr{P}$ were classified by Patera *et al.*[13] and thus the homogeneous spaces of $\mathscr{P}$ can be obtained as the corresponding quotient structures. However, we are interested in those homogeneous spaces that describe particles with the maximum structure.

We devote the remaining sections of this work to analyze different homogeneous spaces; we begin with the Minkowski space-time to describe the simplest case, that of a point particle.

Later on we will be interested in those homogeneous spaces with higher dimenson, giving rise to systems with the highest number of degrees of freedom. We shall start with the Poincaré group itself, for describing general bradyons; the nine-dimensional manifold $X_c$, which describes particles that travel at the speed of light, defined as $X_c = \mathscr{P}/\mathscr{Y}$, where $\mathscr{Y}$ is the one-parameter subgroup of pure Lorentz transformations in a given direction; and finally the seven-dimensional manifold $X_T = \mathscr{P}/\mathrm{SO}(3)$ for particles with tachyonic internal orbital motion.

## IV. POINT PARTICLES

Let us consider first those mechanical systems for which the kinematical space $X = \mathscr{P}/\mathscr{L}$ is the Minkowski space-time $\mathscr{M}$. The purpose is to illustrate the method for obtaining their Lagrangians for further generalizations.

An element $x \in X$ is characterized by the four real numbers $(t(\tau),\mathbf{r}(\tau))$ that transform under $\mathscr{P}$ according to the formulas (3.3) and (3.4), and which are physically interpreted as the time and position of the system, respectively. There are no constraints among these four kinematical variables and only the homogeneity condition (2.4) will reduce the number of degrees of freedom to 3. The general Lagrangian for these systems can be written as

$$L = -H\dot{t} + \mathbf{p} \cdot \dot{\mathbf{r}}, \tag{4.1}$$

where $H$ and $\mathbf{p}$ are defined by $H = -\partial L/\partial \dot{t}$ and $p_i = \partial L/\partial \dot{r}^i$, are functions of $t,\mathbf{r}$, and are homogeneous of zero degree in terms of the derivatives $\dot{t}(\tau)$ and $\dot{\mathbf{r}}(\tau)$.

If we assume that the parameter $\tau$ is invariant under the group $\mathscr{P}$, the derivatives transform as

$$\dot{t}'(\tau) = \gamma \dot{t}(\tau) + \gamma(\mathbf{v} \cdot R(\boldsymbol{\mu})\dot{\mathbf{r}}(\tau))/c^2, \tag{4.2}$$

$$\dot{\mathbf{r}}'(\tau) = R(\boldsymbol{\mu})\dot{\mathbf{r}}(\tau) + \gamma \mathbf{v}\dot{t}(\tau) + [\gamma^2/(1+\gamma)c^2](\mathbf{v} \cdot R(\boldsymbol{\mu})\dot{\mathbf{r}}(\tau))\mathbf{v}, \tag{4.3}$$

and the invariance of $L$ under $\mathscr{P}$ leads for $\mathbf{p}$ and $H$ to the

transformation equations

$$H'(\tau) = \gamma H(\tau) + \gamma(\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{p}(\tau)), \tag{4.4}$$

$$\mathbf{p}'(\tau) = R(\boldsymbol{\mu})\mathbf{p}(\tau) + \gamma\mathbf{v}H(\tau)/c^2 + [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{p}(\tau))\mathbf{v}, \tag{4.5}$$

and thus since $H$ and $\mathbf{p}$ are invariant under translations they are functions independent of $t$ and $\mathbf{r}$ depending only on $\dot{t}$ and $\dot{\mathbf{r}}$.

From (4.2) and (4.3) we see that $c^2\dot{t}^2(\tau) - \dot{\mathbf{r}}^2(\tau)$ is group invariant and we shall consider those systems for which this invariant remains either greater than, equal to, or less than zero during the evolution.

If $c^2\dot{t}^2(\tau) - \dot{\mathbf{r}}^2(\tau) > 0$, for $\tau\in[\tau_1,\tau_2]$, then $\dot{t}(\tau)\neq 0$ for every inertial observer and then $t(\tau)$ can be inverted to obtain $\tau(t)$ and thus we can make a time evolution description $\mathbf{r}(t)$. The velocity is defined as $\mathbf{u}(\tau) = \dot{\mathbf{r}}(\tau)/\dot{t}(\tau)$, and $H$ and $\mathbf{p}$ are only functions of $\mathbf{u}$, with $u < c$. This particle is called a spinless bradyon.

If $c^2\dot{t}^2(\tau) - \dot{\mathbf{r}}^2(\tau) = 0$, for $\tau\in[\tau_1,\tau_2]$, then $\dot{t}(\tau)$ and $\dot{\mathbf{r}}(\tau)$ are different from zero for every observer. The velocity of this system $\mathbf{u} = \dot{\mathbf{r}}/\dot{t}$ is such that $u = c$, and this particle is called a spinless photon.

If $c^2\dot{t}^2(\tau) - \dot{\mathbf{r}}(\tau) < 0$, for $\tau\in[\tau_1,\tau_2]$, then there exist observers for whom $\dot{t}(\tau) = 0$ and it is not possible, in general, to make a time evolution description. However, $|\dot{\mathbf{r}}|\neq 0$ for every observer, so that the magnitude $\mathbf{l} = \ddot{\mathbf{r}}\dot{t}/\dot{\mathbf{r}}^2$ is homogeneous of zero degree and well defined. Here $H$ and $\mathbf{p}$ are only functions of $\mathbf{l}$, and the velocity of this system $\mathbf{u} = \dot{\mathbf{r}}/\dot{t}$ is always greater than $c$, and for some observers can become infinite. This system is a spinless tachyon.

Since the action function is invariant under $\mathscr{P}$, Noether's theorem defines the following constants of the motion:

under time translation, the energy

$$H = -\frac{\partial L}{\partial \dot{t}}; \tag{4.6}$$

under space translation, the linear momentum

$$p_i = \frac{\partial L}{\partial \dot{r}^i}; \tag{4.7}$$

under a Lorentz boost, the Poincaré momentum

$$\boldsymbol{\pi} = -H\mathbf{r}/c^2 + \mathbf{p}t; \tag{4.8}$$

under a rotation, the angular momentum

$$\mathbf{J} = \mathbf{r}\times\mathbf{p}. \tag{4.9}$$

From (4.4) and (4.5) $(H/c)^2 - p^2$ is group invariant and from (4.6) and (4.7) it is also a constant of the motion, and thus has to be independent of $\dot{\mathbf{r}}$ and $\dot{t}$. Consequently it defines a constant parameter that will be used to characterize the system.

Taking the $\tau$ derivative in (4.8), $\dot{\boldsymbol{\pi}} = 0 = -H\dot{\mathbf{r}}/c^2 + \mathbf{p}\dot{t}$ and then $\mathbf{p} = H\mathbf{u}/c^2$. In the bradyonic case, $u < c$ and thus $(H/c)^2 - p^2 = m^2c^2$ is positive and defines the constant parameter $m$, the rest mass of the system. Substituting the expression for $\mathbf{p}$ leads to $H = mc^2/(1 - u^2/c^2)^{1/2}$ and the Lagrangian (4.1) becomes

$$L = -mc^2(1 - c^2\dot{\mathbf{r}}^2/\dot{t}^2)^{1/2}\dot{t} = -mc(c^2 - u^2)^{1/2}\dot{t} \tag{4.10}$$

with the action function

$$A(x_1,x_2) = -mc\{c^2(t_2 - t_1)^2 - (\mathbf{r}_2 - \mathbf{r}_1)^2\}^{1/2}. \tag{4.11}$$

In the photonic case the parameter $m = 0$ and the Lagrangian and the action function are also identically zero. This formalism does not predict spinless photons.

For the pointlike tachyon we get $(H/c)^2 = p^2c^2l^2$, the invariant is negative and we write it as $-\alpha^2 = (H/c)^2 - p^2$. This $\alpha$ is the absolute value of the linear momentum carried by the particle for the observer for which $H = 0$, which corresponds to infinite speed. Thus the energy $H$ is given by

$$H = \mathbf{p}\cdot\mathbf{l}c^2, \tag{4.12}$$

the Lagrangian for the spinless tachyon is

$$L = \alpha(1 - c^2\dot{t}^2/\dot{\mathbf{r}}^2)^{1/2}|\dot{\mathbf{r}}|, \tag{4.13}$$

and the action function is

$$A(x_1,x_2) = \alpha\{(\mathbf{r}_2 - \mathbf{r}_1)^2 - c^2(t_2 - t_1)^2\}^{1/2}. \tag{4.14}$$

## V. GENERAL POINCARÉ BRADYONS

Let us consider the mechanical system for which $X = \mathscr{P}$. An element $x$ of $X$ will be given by the ten real numbers, $x\equiv(t(\tau),\mathbf{r}(\tau),\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))$ with domains $t\in\mathbb{R}$, $\mathbf{r}\in\mathbb{R}^3$, $\mathbf{u}\in\mathbb{R}^3$, $u < c$, and $\boldsymbol{\rho}\in\mathbb{R}^3_c$. Taking into account (3.5)–(3.8), they transform under $\mathscr{P}$ as follows:

$$t'(\tau) = \gamma t(\tau) + \gamma(\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{r}(\tau))/c^2 + b, \tag{5.1}$$

$$\mathbf{r}'(\tau) = R(\boldsymbol{\mu})\mathbf{r}(\tau) + \gamma\mathbf{v}t(\tau) + [\gamma^2/(1+\gamma)c^2]\times(\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{r}(\tau))\mathbf{v} + \mathbf{a}, \tag{5.2}$$

$$\mathbf{u}'(\tau) = \frac{R(\boldsymbol{\mu})\mathbf{u}(\tau) + \gamma\mathbf{v} + \gamma^2c^{-2}(\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{u}(\tau))\mathbf{v}/(1+\gamma)}{\gamma(1 + (\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{u}(\tau))/c^2)}, \tag{5.3}$$

$$\boldsymbol{\rho}'(\tau) = \frac{\boldsymbol{\mu} + \boldsymbol{\rho}(\tau) + \boldsymbol{\mu}\times\boldsymbol{\rho}(\tau) + \mathbf{F}(\mathbf{v},\boldsymbol{\mu};\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))}{1 - \boldsymbol{\mu}\cdot\boldsymbol{\rho}(\tau) + G(\mathbf{v},\boldsymbol{\mu};\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))}. \tag{5.4}$$

The way they transform allows us to interpret $t(\tau)$ as the time, $\mathbf{r}(\tau)$ as the position, $\mathbf{u}(\tau)$ as the velocity, and $\boldsymbol{\rho}(\tau)$ as the orientation of the system.

There are three additional constraints among the $x$ variables. The velocity $\mathbf{u}(\tau) = \dot{\mathbf{r}}(\tau)/\dot{t}(\tau)$, together with the homogeneity condition (2.4), reduces to six the number of degrees of freedom of the system. The six independent variables are $\mathbf{r}(t)$ and $\boldsymbol{\rho}(t)$ and the Lagrangian will depend up to the second derivative of $\mathbf{r}$ and only on the first derivative of $\boldsymbol{\rho}$. Since $u < c$ the system is called a bradyon.

Again assuming that the $\tau$ parameter is group invariant, taking the $\tau$ derivative in both sides of (5.1) (5.4) we get that $\dot{t}(\tau)$ and $\dot{\mathbf{r}}(\tau)$ transform like (4.2)–(4.3) and $\dot{\mathbf{u}}(\tau)$ and $\dot{\boldsymbol{\rho}}(\tau)$ in a very complicated way. However, instead of the derivatives $\dot{\mathbf{u}}(\tau)$ and $\dot{\boldsymbol{\rho}}(\tau)$, we shall define two other three-vectors

$\alpha$ and $\omega$ (see the Appendix)

$$\alpha(\tau) = \frac{\gamma(u)}{c} \times\left[\dot{\mathbf{u}}(\tau) + \frac{\gamma(u)^2}{(1+\gamma(u))}\frac{\mathbf{u}\cdot\dot{\mathbf{u}}}{c^2}\mathbf{u}(\tau) + \mathbf{u}(\tau)\times\boldsymbol{\omega}_0(\tau)\right], \tag{5.5}$$

$$\boldsymbol{\omega}(\tau) = \gamma(u)\boldsymbol{\omega}_0 - \frac{\gamma(u)^2}{(1+\gamma(u))}\frac{\mathbf{u}\cdot\boldsymbol{\omega}_0}{c^2}\mathbf{u} + \boldsymbol{\omega}_T \tag{5.6}$$

that are the strict components of an antisymmetric tensor $\Omega^{\mu\nu}$, $\alpha_i = \Omega^{i0}$, and $\omega_i = \frac{1}{2}\epsilon_{ijk}\Omega^{jk}$, and where the variables $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}_T$ are given by

$$\boldsymbol{\omega}_0 = 2(\dot{\boldsymbol{\rho}} + \boldsymbol{\rho}\times\dot{\boldsymbol{\rho}})/(1+\rho^2), \tag{5.7}$$

$$\boldsymbol{\omega}_T = -[\gamma(u)^2/(1+\gamma(u))](\mathbf{u}\times\dot{\mathbf{u}})/c^2. \tag{5.8}$$

If $\tau$ is the time, $\boldsymbol{\omega}_0$ is the instantaneous angular velocity and $\boldsymbol{\omega}_T$ the Thomas angular velocity. The new varibles $\alpha$ and $\omega$ transform under $\mathscr{P}$:

$$\alpha'(\tau) = \gamma R(\mu)\alpha(\tau) - [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\mu)\alpha(\tau))\mathbf{v} + (\gamma/c)(\mathbf{v}\times R(\mu)\omega(\tau)), \tag{5.9}$$

$$\omega'(\tau) = \gamma R(\mu)\omega(\tau) - [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\mu)\omega(\tau))\mathbf{v} - (\gamma/c)(\mathbf{v}\times R(\mu)\alpha(\tau)). \tag{5.10}$$

The homogeneity condition in terms of the variables $(t,\dot{\mathbf{r}},\dot{\mathbf{u}},\dot{\boldsymbol{\rho}})$ allows us to write $L$ in the form

$$L = -T\dot{t} + \mathbf{Q}\cdot\dot{\mathbf{r}} + \mathbf{U}\cdot\dot{\mathbf{u}} + \mathbf{V}\cdot\dot{\boldsymbol{\rho}}, \tag{5.11}$$

and in terms of the variables $(\dot{t},\dot{\mathbf{r}},\boldsymbol{\alpha},\boldsymbol{\omega})$,

$$L = -T\dot{t} + \mathbf{Q}\cdot\dot{\mathbf{r}} + \mathbf{D}\cdot\boldsymbol{\alpha} + \mathbf{S}\cdot\boldsymbol{\omega}, \tag{5.12}$$

where the functions are defined by

$$T = -\frac{\partial L}{\partial \dot{t}}, \quad Q_i = \frac{\partial L}{\partial \dot{r}^i}, \quad D_i = \frac{\partial L}{\partial \dot{\alpha}^i},$$

$$S_i = \frac{\partial L}{\partial \dot{\omega}^i}, \quad U_i = \frac{\partial L}{\partial \dot{u}^i}, \quad V_i = \frac{\partial L}{\partial \dot{\rho}^i},$$

and they are functions of the kinematical variables $(t,\mathbf{r},\mathbf{u},\boldsymbol{\rho})$ and homogeneous functions of zero degree in the variables $(\dot{t},\dot{\mathbf{r}},\boldsymbol{\alpha},\boldsymbol{\omega})$. The observable $T$ has the dimensions of energy, $\mathbf{Q}$ of linear momentum and $\mathbf{D}$ and $\mathbf{S}$ of angular momentum, and, being the Lagrangian invariant under $\mathscr{P}$, they transform

$$T'(\tau) = \gamma T(\tau) + \gamma(\mathbf{v}\cdot R(\mu)\mathbf{Q}(\tau)), \tag{5.13}$$

$$\mathbf{Q}'(\tau) = R(\mu)\mathbf{Q}(\tau) + \gamma\mathbf{v}T(\tau)/c^2 + [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\mu)\mathbf{Q}(\tau))\mathbf{v}, \tag{5.14}$$

$$\mathbf{D}'(\tau) = \gamma R(\mu)\mathbf{D}(\tau) - [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\mu)\mathbf{D}(\tau))\mathbf{v} - (\gamma/c)(\mathbf{v}\times R(\mu)\mathbf{S}(\tau)), \tag{5.15}$$

$$\mathbf{S}'(\tau) = \gamma R(\mu)\mathbf{S}(\tau) - [\gamma^2/(1+\gamma)c^2] \times(\mathbf{v}\cdot R(\mu)\mathbf{S}(\tau))\mathbf{v} + (\gamma/c)(\mathbf{v}\times R(\mu)\mathbf{D}(\tau)). \tag{5.16}$$

The observables $\mathbf{D}$ and $\mathbf{S}$ are the strict components of an antisymmetric tensor $S^{\mu\nu} = -S^{\nu\mu}$, $D_i = S^{0i}$, $S_i = \frac{1}{2}\epsilon_{ijk}S^{jk}$. We see that these functions are invariant under translations and thus independent of $t$ and $\mathbf{r}$, and since, for the bradyonic case, $\dot{t} \neq 0$ for every observer, they only have to be functions of $(\mathbf{u},\boldsymbol{\rho},\boldsymbol{\alpha}/\dot{t},\boldsymbol{\omega}/\dot{t})$. Since for the observer for which the system has zero velocity, $\mathbf{u} = 0$, $c\boldsymbol{\alpha}/\dot{t}$ reduces to $d\mathbf{u}/dt$ and $\boldsymbol{\omega}/\dot{t}$ to $\boldsymbol{\omega}_0/\dot{t}$, we say that $c\boldsymbol{\alpha}/\dot{t}$ and $\boldsymbol{\omega}/\dot{t}$ are, respectively, the generalized acceleration and generalized angular velocity of the system.

Noether's theorem defines the following constants of the motion:

energy

$$H = T - \frac{d\mathbf{U}}{dt}\cdot\mathbf{u}; \tag{5.17}$$

linear momentum,

$$\mathbf{p} = \mathbf{Q} - \frac{d\mathbf{U}}{dt}; \tag{5.18}$$

Poincaré momentum

$$\boldsymbol{\pi} = -H\mathbf{r}/c^2 + \mathbf{p}t + \mathbf{D}/c; \tag{5.19}$$

angular momentum,

$$\mathbf{J} = \mathbf{r}\times\mathbf{p} + \mathbf{S}; \tag{5.20}$$

where the function $\mathbf{U}(\tau)$ is given in terms of $\mathbf{D}$ and $\mathbf{S}$ by

$$\mathbf{U}(\tau) = \frac{\gamma(u)}{c}\left[\mathbf{D}(\tau) + \frac{\gamma(u)^2}{(1+\gamma(u))}\frac{\mathbf{u}\cdot\mathbf{D}}{c^2}\mathbf{u}(\tau) + \frac{\gamma(u)}{(1+\gamma(u))c}\mathbf{u}(\tau)\times\mathbf{S}(\tau)\right].$$

Energy and linear momentum transform as in (4.4) and (4.5). The center of mass observer is defined as that observer for which $\mathbf{p} = 0$ and $\boldsymbol{\pi} = 0$. These six conditions do not define uniquely an observer, but rather the class of observers for which the center of mass is at rest and placed at the origin of the coordinate frame. They are thus defined up to an arbitrary rotation and to an arbitrary time translation. We shall call to this class the center of mass observer as is usually done.

The observable $\mathbf{S}$ is called the spin of the system and is a constant of the motion only for the center of mass observer. Since we consider systems for which $H > 0$ we can define the observable $\mathbf{k} = c\mathbf{D}/H$ with dimensions of length such that taking the $\tau$ derivative in (5.19) we get

$$\dot{\boldsymbol{\pi}} = (\dot{\mathbf{r}} - \dot{\mathbf{k}})H/c^2 + \mathbf{p}\dot{t} = 0.$$

Thus $\mathbf{p} = (H/c^2)d(\mathbf{r}-\mathbf{k})/dt$ and defines for every observer the position of a point $\mathbf{q} = \mathbf{r} - \mathbf{k}$ that moves at constant velocity $d\mathbf{q}/dt$. We see that $\mathbf{q}$ is the position of the center of mass and thus $\mathbf{k}$ is the relative position of the system with respect to its center of mass. The absolute value of $\mathbf{k}$ gives information about a lower bound for the radius of the particle.

From these constants of the motion other constants can be defined:

$$w^0 = \mathbf{p}\cdot\mathbf{J} = \mathbf{p}\cdot\mathbf{S}, \tag{5.21}$$

$$\mathbf{W} = H\mathbf{J}/c - c\mathbf{p}\times\boldsymbol{\pi} = H(\mathbf{S} + \mathbf{k}\times\mathbf{p})/c. \tag{5.22}$$

These can be expressed in terms of the four-vector $p^\mu \equiv (H/c,\mathbf{p})$ and the antisymmetric tensor $S^{\mu\nu}$ in the form $w_\sigma = \frac{1}{2}\epsilon_{\sigma\mu\nu\lambda}S^{\mu\nu}p^\lambda$. Equations (5.21) and (5.22) are the

components of the Pauli–Lubanski four-vector. From (5.21), taking the time derivative we obtain $\mathbf{p}\cdot d\mathbf{S}/dt = 0$, and the spin variation is orthogonal to the center of mass motion, the helicity, or spin projection on that direction, remaining constant.

Thus $p_\mu p^\mu = m^2c^2$ and $-w^\mu w_\mu = m^2c^2s^2$ are two functionally independent invariants, constants of the motion, that define two constant parameters $m$ and $s$ called, respectively, the rest mass and the absolute value of the spin for the center of mass observer and we expect that the Lagrangian will be an explicit function of them.

Let $H_0$ and $\mathbf{p}_0$ be the energy and linear momentum for the observer for which the variables $\mathbf{u} = \boldsymbol{\rho} = 0$. For this observer they are only functions of $\boldsymbol{\alpha}/\dot{t}$ and $\boldsymbol{\omega}/\dot{t}$. For the general observer for which $\mathbf{u}$ and $\boldsymbol{\rho}$ are different from zero, $H$ and $\mathbf{p}$ are obtained from $H_0$ and $\mathbf{p}_0$ by the transformation equations (4.4) and (4.5) and thus

$$H(\mathbf{u},\boldsymbol{\rho},\boldsymbol{\alpha}/\dot{t},\boldsymbol{\omega}/\dot{t}) = \gamma(u)H_0 + \gamma(u)(\mathbf{u}\cdot R(\boldsymbol{\rho})\mathbf{p}_0), \tag{5.23}$$

$$\mathbf{p}(\mathbf{u},\boldsymbol{\rho},\boldsymbol{\alpha}/\dot{t},\boldsymbol{\omega}/\dot{t}) = R(\boldsymbol{\rho})\mathbf{p}_0 + \gamma(u)\mathbf{u}H_0/c^2 + [\gamma(u)^2/(1+\gamma(u))c^2](\mathbf{u}\cdot R(\boldsymbol{\rho})\mathbf{p}_0)\mathbf{u}. \tag{5.24}$$

Since the first half of the Lagrangian $-T\dot{t} + \mathbf{Q}\cdot\dot{\mathbf{r}} = -H\dot{t} + \mathbf{p}\cdot\dot{\mathbf{r}}$ is Poincaré invariant, substituting $H$ and $\mathbf{p}$ in terms of (5.23) and (5.24) reduces it to $-H_0\dot{t}/\gamma(u)$. Because $\dot{t}/\gamma(u) = (c^2\dot{t}^2 - \dot{\mathbf{r}}^2)^{1/2}/c$ is a Poincaré-invariant function, $H_0$ must also be an invariant function of their arguments.

Similarly, the second half of the Lagrangian $\mathbf{D}\cdot\boldsymbol{\alpha} + \mathbf{S}\cdot\boldsymbol{\omega} = \frac{1}{2} S_{\mu\nu}\Omega^{\mu\nu}$ is itself Poincaré invariant and we have to choose for $\mathbf{D}$ and $\mathbf{S}$ functions of $(\mathbf{u},\boldsymbol{\rho},\boldsymbol{\alpha}/\dot{t},\boldsymbol{\omega}/\dot{t})$ in order that this holds. From $\Omega^{\mu\nu}$ we can form the two invariants $\epsilon_{\mu\nu\sigma\lambda}\Omega^{\mu\nu}\Omega^{\sigma\lambda}$ and $\Omega_{\mu\nu}\Omega^{\mu\nu}$, which reduce, respectively, to $\boldsymbol{\alpha}\cdot\boldsymbol{\omega}$ and $\alpha^2 - \omega^2$. Expressed in terms of the kinematical variables and their derivatives these invariants are

$$\alpha^2 - \omega^2 = \frac{2\gamma(u)^2[\dot{\mathbf{u}}^2 + \dot{\mathbf{u}}\cdot(\mathbf{u}\times\boldsymbol{\omega}_0)]}{(1+\gamma(u))c^2} + \frac{(2+2\gamma(u)+\gamma(u)^2)\gamma(u)^4}{(1+\gamma(u))^2c^4} \times(\mathbf{u}\cdot\mathbf{u})^2 - \omega_0^2, \tag{5.25}$$

$$\boldsymbol{\alpha}\cdot\boldsymbol{\omega} = \frac{\gamma(u)}{c}\left[\dot{\mathbf{u}}\cdot\boldsymbol{\omega}_0 + \frac{\gamma(u)^2}{(1+\gamma(u))}\frac{(\mathbf{u}\cdot\dot{\mathbf{u}})(\mathbf{u}\cdot\boldsymbol{\omega}_0)}{c^2}\right]. \tag{5.26}$$

Thus two elementary choices for $\mathbf{D}$ and $\mathbf{S}$: First, to choose $\mathbf{D}$ and $\mathbf{S}$ proportional to $-\gamma(u)\boldsymbol{\alpha}/\dot{t}$ and $\gamma(u)\boldsymbol{\omega}/\dot{t}$, respectively, with the same proportionality coefficient, which has to be an invariant function. In this case the spin is proportional to the generalized angular velocity (suggesting an intrinsic angular momentum of a rotating nature) and the momentum $\mathbf{D}$ (and thus the relative position vector $\mathbf{k}$) has opposite direction to the generalized acceleration, suggesting for the *Zitterwebegung* a certain kind of generalized central motion.

On the other side we can choose $\mathbf{D}$ proportional to the function $\gamma(u)\boldsymbol{\omega}/\dot{t}$ and $\mathbf{S}$ to $\gamma(u)\boldsymbol{\alpha}/\dot{t}$ and we see in this case that $\mathbf{S}$ is by no means related to the angular velocity and the internal motion is no longer a generalized central motion.

These two possible elementary Lagrangians expressed in terms of the two invariants $A$ and $B$ will reduce to

$$L_B = -A(1 - c^{-2}\dot{\mathbf{r}}^2/\dot{t}^2)^{1/2}\dot{t} - B(\alpha^2 - \omega^2)(\dot{t}^2 - \dot{\mathbf{r}}^2/c^2)^{-1/2}, \tag{5.27}$$

$$L_F = -A(1 - c^{-2}\dot{\mathbf{r}}^2/\dot{t}^2)^{1/2}\dot{t} + B(\boldsymbol{\alpha}\cdot\boldsymbol{\omega})(\dot{t}^2 - \dot{\mathbf{r}}^2/c^2)^{-1/2}, \tag{5.28}$$

where $\alpha^2 - \omega^2$ and $\boldsymbol{\alpha}\cdot\boldsymbol{\omega}$ are given, respectively, in (5.25) and (5.26).

Lagrangians of the first type (5.27) can be found in the literature. Constantelos[14] quotes a Lagrangian in which $\boldsymbol{\omega}_0 = 0$ and thus the particle has internal orbital motion but no rotation. The Lagrangian depends on the velocity and acceleration of the particle but not on the angular variables. On the other hand, Hanson and Regge[11] when discussing the relativistic spherical top, assume $\dot{\mathbf{u}} = 0$ and thus the invariant $\alpha^2 - \omega^2$ reduces to $-\omega_0^2$. The particle has no internal orbital motion; its position coincides with its center of mass but it rotates with angular velocity $\boldsymbol{\Omega} = \boldsymbol{\omega}_0/\dot{t}$ since this rotation is responsible for the existence of spin. However, to the best of our knowledge, no Lagrangians of the form (5.28) have been studied before. These two Lagrangians lead to nonlinear dynamical equations.

The Lagrangian (5.27), for the center of mass observer, gives rise to the equations

$$\mathbf{r} = -\frac{2B^2}{mc^2}\gamma(u)^2\left[\frac{d\mathbf{u}}{dt} + \mathbf{u}\times\boldsymbol{\Omega} + \frac{\gamma(u)^2}{(1+\gamma(u))}\times\left(\frac{\mathbf{u}}{c}\cdot\frac{d\mathbf{u}}{dt}\right)\frac{\mathbf{u}}{c}\right], \tag{5.29}$$

$$\mathbf{S} = 2B^2\gamma(u)\left[\gamma(u)\boldsymbol{\Omega} - \frac{\gamma(u)^2}{(1+\gamma(u))c^2}\times\left((\mathbf{u}\cdot\boldsymbol{\Omega})\mathbf{u} + \mathbf{u}\times\frac{d\mathbf{u}}{dt}\right)\right], \tag{5.30}$$

where the spin $\mathbf{S}$ is constant. Solutions, with constant absolute value of velocity and angular velocity and $\boldsymbol{\Omega}\cdot\mathbf{u} = 0$, exist and show the motion displayed in Fig. 1.

Similarly, the Lagrangian (5.28) also describes motions with constant absolute value of velocity and angular velocity, where $\boldsymbol{\Omega}$ is orthogonal to $\mathbf{u}$, and in the center of mass

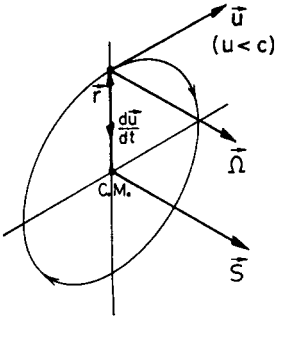



FIG. 1. Motion in the center of mass (C.M.) frame of particle (5.27).

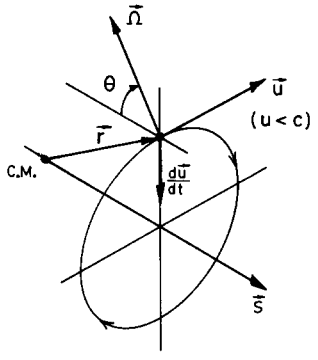



FIG. 2. Motion in the C.M. frame of particle (5.28).

frame they are given by

$$\mathbf{r} = \frac{B}{mc}\gamma(u)\left[\gamma(u)\boldsymbol{\Omega} - \frac{\gamma(u)^2}{(1+\gamma(u))c^2} \times\left((\mathbf{u}\cdot\boldsymbol{\Omega})\mathbf{u} + \mathbf{u}\times\frac{d\mathbf{u}}{dt}\right)\right], \tag{5.31}$$

$$\mathbf{S} = \frac{B}{c}\gamma(u)^2\left[\frac{d\mathbf{u}}{dt} + \mathbf{u}\times\boldsymbol{\Omega} + \frac{\gamma(u)^2}{(1+\gamma(u))} \times\left(\frac{\mathbf{u}}{c}\cdot\frac{d\mathbf{u}}{dt}\right)\frac{\mathbf{u}}{c}\right]. \tag{5.32}$$

One possible motion that verifies these is depicted in Fig. 2.

## VI. LUXONS

Now let us consider those mechanical systems whose kinematical space $X_C$ is the nine-dimensional manifold spanned by the variables $(t,(\tau),\mathbf{r}(\tau),\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))$ with domains $t\in\mathbb{R}$, $\mathbf{r}\in\mathbb{R}^3$, and $\boldsymbol{\rho}\in\mathbb{R}^3_c$ (as before) and $\mathbf{u}\in\mathbb{R}^3$, but with $u = c$. These particles are usually called luxons.

This manifold $X_c$ is a homogeneous space isomorphic to $\mathscr{P}/\mathscr{V}$, where $\mathscr{V}$ is the one-parameter subgroup of pure Lorentz transformations on a given direction. In fact, let $x \equiv (0,\mathbf{0},u,\mathbf{0})$, where $u = c$ is a point of this manifold $X_c$ The stabilizer group of this point is the subgroup of pure Lorentz transformations in the direction given by $\mathbf{u}$, $\mathscr{V}_u$. Thus $X_c \approx \mathscr{P}/\mathscr{V}_u$.

Again there are three constraints between the kinematical variables, $\mathbf{u} = \dot{\mathbf{r}}/\dot{t}$. The kinematical variables $t,\mathbf{r},\mathbf{u}$ transform like (5.1)–(5.3), while the transformation of $\boldsymbol{\rho}$ is obtained from (5.4) taking the limit $u = c$ on the right-hand side:

$$\boldsymbol{\rho}'(\tau) = \frac{\boldsymbol{\mu} + \boldsymbol{\rho}(\tau) + \boldsymbol{\mu}\times\boldsymbol{\rho}(\tau) + \mathbf{F}_c(\mathbf{v},\boldsymbol{\mu};\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))}{1 - \boldsymbol{\mu}\cdot\boldsymbol{\rho}(\tau) + G_c(\mathbf{v},\boldsymbol{\mu};\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))}, \tag{6.1}$$

where

$$\begin{aligned}\mathbf{F}_c(\mathbf{v},\boldsymbol{\mu};\mathbf{u},\boldsymbol{\rho}) &= [\gamma/(1+\gamma)c^2][\mathbf{u}\times\mathbf{v} + \mathbf{u}(\mathbf{v}\cdot\boldsymbol{\mu}) + \mathbf{v}(\mathbf{u}\cdot\boldsymbol{\rho}) \\ &\quad + \mathbf{u}\times(\mathbf{v}\times\boldsymbol{\mu}) + (\mathbf{u}\times\boldsymbol{\rho})\times\mathbf{v} + (\mathbf{u}\cdot\boldsymbol{\rho})(\mathbf{v}\times\boldsymbol{\mu}) \\ &\quad + (\mathbf{u}\times\boldsymbol{\rho})(\mathbf{v}\cdot\boldsymbol{\mu}) + (\mathbf{u}\times\boldsymbol{\rho})\times(\mathbf{v}\times\boldsymbol{\mu})],\end{aligned} \tag{6.2}$$

$$\begin{aligned}G_c(\mathbf{v},\boldsymbol{\mu};\mathbf{u},\boldsymbol{\rho}) &= [\gamma/(1+\gamma)c^2[\mathbf{u}\cdot\mathbf{v} + \mathbf{u}\cdot(\mathbf{v}\times\boldsymbol{\mu}) + \mathbf{v}\cdot(\mathbf{u}\times\boldsymbol{\rho}) \\ &\quad - (\mathbf{u}\cdot\boldsymbol{\rho})(\mathbf{v}\cdot\boldsymbol{\mu}) + (\mathbf{u}\times\boldsymbol{\rho})\cdot(\mathbf{v}\times\boldsymbol{\mu})].\end{aligned} \tag{6.3}$$

Since $u' = u = c$, Eq. (5.3), implies that $\mathbf{u}'$ is obtained from $\mathbf{u}$ by means of an orthogonal transformation:

$$\mathbf{u} = = R(\boldsymbol{\phi})\mathbf{u}, \tag{6.4}$$

where $\boldsymbol{\phi}$ is given by

$$\boldsymbol{\phi} = \frac{\boldsymbol{\mu} + \mathbf{F}_c(\mathbf{v},\boldsymbol{\mu};\mathbf{u}(\tau),\mathbf{0})}{1 + G_c(\mathbf{v},\boldsymbol{\mu};\mathbf{u}(\tau),\mathbf{0})}. \tag{6.5}$$

Equation (6.1) also corresponds to

$$R(\boldsymbol{\rho}') = R(\boldsymbol{\phi})R(\boldsymbol{\rho}), \tag{6.6}$$

with the same $\boldsymbol{\phi}$ as in (6.5).

Since $u(\tau) = c$, we have to distinguish two different kinds of systems. Taking the $\tau$ derivative of this identity we get $\mathbf{u}(\tau)\cdot\dot{\mathbf{u}}(\tau) = 0$, and we shall next consider systems for which $\dot{\mathbf{u}}(\tau) = 0$ and those for which $\dot{\mathbf{u}}(\tau) \neq 0$ and is orthogonal to $\mathbf{u}$.

### A. Massless particles

If $\dot{\mathbf{u}}(\tau) = 0$ then $\mathbf{u}(\tau)$ is constant, the system follows a straight trajectory with velocity $c$, and the kinematical space reduces to the seven-dimensional manifold $(t(\tau),\mathbf{r}(\tau),\boldsymbol{\rho}(\tau))$, with $t\in\mathbb{R}$, $\mathbf{r}\in\mathbb{R}^3$, $\boldsymbol{\rho}\in\mathbb{R}^3_c$.

The derivatives $\dot{t}$ and $\dot{\mathbf{r}}$ transform as in (4.2) and (4.3) and, instead of the variable $\dot{\boldsymbol{\rho}}$, we define the linear function of it $\boldsymbol{\omega}_0$ given by (5.7), which transforms under $\mathscr{P}$, as

$$\boldsymbol{\omega}_0'(\tau) = R(\boldsymbol{\phi})\boldsymbol{\omega}_0(\tau), \tag{6.7}$$

where $\boldsymbol{\phi}$ is given again by (6.5). The invariant $c^2\dot{t}^2 - \dot{\mathbf{r}}^2 = 0$, and $\dot{t} \neq 0$ and $\dot{\mathbf{r}} \neq 0$ for every observer.

For this system there are no differential constraints between the kinematical variables, the Lagrangian will only depend on the first derivatives of the variables $\mathbf{r}$ and $\boldsymbol{\rho}$, and the homogeneity condition (2.4) reduces to 6 the number of degrees of freedom of the system. This condition leads to a Lagrangian of the form

$$L = -H\dot{t} + \mathbf{p}\cdot\dot{\mathbf{r}} + \mathbf{S}\cdot\boldsymbol{\omega}_0, \tag{6.8}$$

where the functions

$$H = -\frac{\partial L}{\partial \dot{t}}, \quad p_i = \frac{\partial L}{\partial \dot{r}^i}, \quad S_i = \frac{\partial L}{\partial \omega_0^i}$$

will be functions of $(t,\mathbf{r},\boldsymbol{\rho})$, homogeneous of zero degree of $(\dot{t},\dot{\mathbf{r}},\boldsymbol{\omega}_0)$, and since $\dot{t} \neq 0$ they can be expressed as functions of $\mathbf{u} = \dot{\mathbf{r}}/\dot{t}$ and $\boldsymbol{\Omega} = \boldsymbol{\omega}_0/\dot{t}$, which are, respectively, the velocity and angular velocity of the system.

The invariance of (6.8) under $\mathscr{P}$ implies that these functions transform as

$$H'\tau = \gamma H(\tau) + \gamma(\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{p}(\tau)), \tag{6.9}$$

$$\begin{aligned}\mathbf{p}'(\tau) &= R(\boldsymbol{\mu})\mathbf{p}(\tau) + \gamma\mathbf{v}H(\tau)/c^2 \\ &\quad + [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{p}(\tau))\mathbf{v},\end{aligned} \tag{6.10}$$

$$\mathbf{S}'(\tau) = R(\boldsymbol{\phi})\mathbf{S}(\tau), \tag{6.11}$$

and being invariant under space and time translations they are only functions of $(\boldsymbol{\rho},\mathbf{u},\boldsymbol{\Omega})$ with the condition $u = c$.

Noether's theorem gives rise in this case to the following constants of the motion:

energy, $H$; (6.12)

linear momentum, $\mathbf{p}$; (6.13)

Poincaré momentum,

$$\boldsymbol{\pi} = -H\mathbf{r}/c^2 + \mathbf{p}t + \mathbf{S}\times\mathbf{u}/c^2; \tag{6.14}$$

angular momentum, $\mathbf{J} = \mathbf{r}\times\mathbf{p} + \mathbf{S}$. (6.15)

By analogy to the above case we say that $\mathbf{S}$ is the spin of the system. Taking in (6.15) the $\tau$ derivative we obtain $\dot{\mathbf{r}}\times\mathbf{p} + \dot{\mathbf{S}} = 0$ and thus $d\mathbf{S}/dt = \mathbf{p}\times\mathbf{u}$. Since $\mathbf{p}$ and $\mathbf{u}$ are constant vectors, the spin $\mathbf{S}$ has a constant time derivative. A continuously increasing spin system is far from being what we shall understand as an elementary particle, so that we shall only consider that system for which the constant $d\mathbf{S}/dt$ is zero. The spin is then a constant of the motion and due to(6.11) its absolute value is also a Poincaré-invariant parameter. In this case $\mathbf{p}$ and $\mathbf{u}$ are parallel vectors. In fact, by taking in (6.14) the $\tau$ derivative we get $\dot{\boldsymbol{\pi}} = 0 = -H\dot{\mathbf{r}}/c^2 + \mathbf{p}\dot{t}$ and thus $\mathbf{p} = H\mathbf{u}/c^2$.

Another group invariant and constant of the motion is $(H/c)^2 - p^2$, which vanishes identically; thus the mass of the particle is zero. Also the first two terms of the Lagrangian cancel out $-H\dot{t} + \mathbf{p}\cdot\dot{\mathbf{r}} = 0$, and $L$ reduces to the third term $L = \mathbf{S}\cdot\boldsymbol{\omega}_0$, where $\mathbf{S}$ is only a function of $\boldsymbol{\rho}$, $\mathbf{u}$, and $\boldsymbol{\Omega}$. We see that $\epsilon = \mathbf{S}\cdot\mathbf{u}$ is another group invariant and constant of the motion, and we expect that the Lagrangian will be dependent on these two parameters $S$ and $\epsilon$. If we take into account (6.4), (6.6), (6.7), and (6.11) the general solution for $\mathbf{S}$ is a vector function of $R(\boldsymbol{\rho})\mathbf{z}$, $\mathbf{u}$, and $\boldsymbol{\omega}_0/\omega_0$, where $\mathbf{z}$ is a constant unit vector.

A system with a nontransversal spin will be such that $\mathbf{S} = \epsilon S\mathbf{u}/c$, where $\epsilon = \pm 1$, and thus the Lagrangian becomes

$$L = \epsilon S(\dot{\mathbf{r}}\cdot\boldsymbol{\omega}_0)/c\dot{t}. \tag{6.16}$$

From this particular Lagrangian we get $H = -\partial L/\partial\dot{t} = \mathbf{S}\cdot\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the angular velocity of the particle. The linear momentum $p_i = \partial L/\partial\dot{r}^i = \epsilon S\Omega_i/c$ and the angular velocity lies in the direction of $\mathbf{u}$. Since $H$ has to be definite positive, $\boldsymbol{\Omega} = \epsilon\Omega\mathbf{u}/c$, leading to the expression $H = S\Omega$ for the energy. Experimentally $S = \hbar$ and $H = \hbar\Omega = h\nu$. We say that system (6.16) is a photon of spin $S$ and helicity $\epsilon$. Thus the frequency of a photon appears as the frequency of its rotational motion, causing the rotation axis to lie parallel to the velocity. We cannot define any size associated with the photon as we did before in connection with the general Poincaré bradyon. It must be remarked that, although the spin and the angular velocity are not related, they have the same direction.

### B. Massive particles

Now considering systems with $\dot{\mathbf{u}} \neq 0$ but orthogonal to $\mathbf{u}$, we have that the kinematical variables and the derivatives $\dot{t}$ and $\dot{\mathbf{r}}$ transform as before, and for $\dot{\mathbf{u}}$ and $\boldsymbol{\omega}_0$ we obtain

$$\dot{\mathbf{u}}' = R(\boldsymbol{\phi})\dot{\mathbf{u}} + \dot{R}(\boldsymbol{\phi})\mathbf{u}, \tag{6.17}$$

$$\boldsymbol{\omega}_0' = R(\boldsymbol{\phi})\boldsymbol{\omega}_0 + \boldsymbol{\omega}_\phi, \tag{6.18}$$

where $\boldsymbol{\phi}$ is again given by (6.5) and $\boldsymbol{\omega}_\phi$:

$$\boldsymbol{\omega}_\phi = \left(\gamma\frac{\mathbf{R}\mathbf{u}\times\mathbf{v}}{c^2} - (\gamma-1)\frac{R(\mathbf{u}\times\dot{\mathbf{u}})}{c^2} + \frac{2\gamma^2}{(1+\gamma)} \times\frac{[v\cdot R(\mathbf{u}\times\dot{\mathbf{u}})]}{c^3}\frac{\mathbf{v}}{c}\right)\left[\gamma\left(1+\frac{\mathbf{v}\cdot\mathbf{R}\mathbf{u}}{c^2}\right)\right]^{-1}. \tag{6.19}$$

Expression (6.17) can be rewritten in the form

$$\dot{\mathbf{u}}' = R(\boldsymbol{\phi})\dot{\mathbf{u}}/\gamma(1 + \mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{u}/c^2), \tag{6.20}$$

and the homogeneity condition leads to

$$L = -T\dot{t} + \mathbf{Q}\cdot\dot{\mathbf{r}} + \mathbf{U}\cdot\dot{\mathbf{u}} + \mathbf{Z}\cdot\boldsymbol{\omega}_0, \tag{6.21}$$

where

$$T = -\frac{\partial L}{\partial\dot{t}}, \quad Q_i = \frac{\partial L}{\partial\dot{r}^i}, \quad U_i = \frac{\partial L}{\partial\dot{u}^i}, \quad Z_i = \frac{\partial L}{\partial\omega_0^i}$$

and Noether's theorem again defines the constants of the motion:

energy, $H = T - \dfrac{d\mathbf{U}}{dt}\cdot\mathbf{u}$; (6.22)

linear momentum, $\mathbf{p} = \mathbf{Q} - \dfrac{d\mathbf{U}}{dt}$; (6.23)

Poincaré momentum,

$$\boldsymbol{\pi} = -H\mathbf{r}/c^2 + \mathbf{p}t + (\mathbf{S}\times\mathbf{u})/c^2; \tag{6.24}$$

angular momentum, $\mathbf{J} = \mathbf{r}\times\mathbf{p} + \mathbf{S}$; (6.25)

where the spin $\mathbf{S}$ is defined by

$$\mathbf{S} = \mathbf{u}\times\mathbf{U} + \mathbf{Z}. \tag{6.26}$$

The above definitions (6.22)–(6.25) lead for $H$ and $\mathbf{p}$ to the transformation equations of a four-momentum as in (4.4)–(4.5) and for the spin to

$$\mathbf{S}'(\tau) = \gamma R(\boldsymbol{\mu})\mathbf{S}(\tau) - [\gamma^2/(1+\gamma)c^2](\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{S}(\tau))\mathbf{v} + (\gamma/c^2)(\mathbf{v}\times R(\boldsymbol{\mu})(\mathbf{S}(\tau)\times\mathbf{u})), \tag{6.27}$$

which corresponds to the transformation equations of an antisymmetric tensor $S^{\mu\nu}$ with strict components $S^{0i} = (\mathbf{S}\times\mathbf{u})^i/c$, $S^{ij} = \epsilon^{ijk}S_k$.

From (6.24), by taking the $\tau$ derivative and doing the dot product with $\mathbf{u}$, we obtain

$$H = \mathbf{p}\cdot\mathbf{u} + \mathbf{S}\cdot\left(\frac{d\mathbf{u}}{dt}\times\mathbf{u}\right)c^{-2}. \tag{6.28}$$

In a certain sense this Hamiltonian looks like Dirac's Hamiltonian for a fermion $H = c\mathbf{p}\cdot\boldsymbol{\alpha} + \beta mc^2$, where $\boldsymbol{\alpha}$ and $\beta$ are Dirac's matrices, and, since $c\boldsymbol{\alpha}$ is identified with the local velocity $\mathbf{u}$, we have finally $H = \mathbf{p}\cdot\mathbf{u} + \beta mc^2$. In the identification the spin term gives rise to the mass term, suggesting a mass–spin relation. However, we shall not discuss any quantization procedure in this work and we delay these considerations to a subsequent paper.

From (6.25) we have that $d\mathbf{S}/dt = \mathbf{p}\times\mathbf{u}$ and we can again define the center of mass observer by $\mathbf{p} = \boldsymbol{\pi} = 0$. For this observer $\mathbf{S}$ is a constant of the motion and the energy does not necessarily vanish, defining the rest mass of the system. From (6.24), $\mathbf{r} = (\mathbf{S}\times\mathbf{u})/H$, and thus the internal motion lies on a plane orthogonal to $\mathbf{S}$. The velocity is then orthogonal to $\mathbf{S}$, and since $S$, $u = c$, and $H = mc^2$ are constants in this frame, the internal motion is a circle of radius $R_0 = S/mc$.

Equation (6.27) can also be written

$$\mathbf{S}' = \gamma(1 + \mathbf{v}\cdot R(\mu)\mathbf{u}/c^2)R(\phi)\mathbf{S}, \tag{6.29}$$

with the same $R(\phi)$ as in (6.6) and thus

$$\mathbf{S}'\cdot\mathbf{u}' = \gamma(1 + \mathbf{v}\cdot R(\mu)\mathbf{u}/c^2)\mathbf{S}\cdot\mathbf{u}$$

and $\mathbf{S}\cdot\dot{\mathbf{u}} = \mathbf{S}'\cdot\dot{\mathbf{u}}'$. Since, for the center of mass observer $\mathbf{S}\cdot\mathbf{u} = 0$ and $\mathbf{S}\cdot d\mathbf{u}/dt = 0$, the spin remains orthogonal to $\mathbf{u}$ and $d\mathbf{u}/dt$ for every observer.

The particle has mass and spin and moves in circles with velocity $c$ in a plane orthogonal to $\mathbf{S}$ for the center of mass observer. All these features are general and independent of the particular Lagrangian (6.21). The only thing that remains to be described is the angular motion. All the above considerations do not give information about the angular velocity of the system. Therefore the different kinds of Lagrangians of the form (6.21) will differ from each other by describing different angular motions, which will be related to other kinds of observables, such as, for instance, electromagnetic multipole momenta. However, since we are describing here free particles, we do not expect that such properties will appear in this setup.

Coming back to the general situation we see that the term $-T\dot{t} + \mathbf{Q}\cdot\dot{\mathbf{r}} = -H\dot{t} + \mathbf{p}\cdot\dot{\mathbf{r}}$ is a Poincaré-invariant term and then $\mathbf{U}\cdot\dot{\mathbf{u}} + \mathbf{Z}\cdot\boldsymbol{\omega}_0$ also has to be invariant. Thus the general Lagrangian seems to be the sum of two invariant terms depending on the two constant parameters $m$ and $s$ that are functions of the variables $(\mathbf{u},\boldsymbol{\rho})$ and homogeneous of first degree in the derivatives $(\dot{t},\dot{\mathbf{r}},\dot{\mathbf{u}},\dot{\boldsymbol{\rho}})$.

We find that the first degree term $\dot{\mathbf{u}} + \mathbf{u}\times\boldsymbol{\omega}_0 = (d\mathbf{u}/dt + \mathbf{u}\times\boldsymbol{\Omega})\dot{t} = \mathbf{y}$ transforms under $\mathscr{P}$ in the form $\mathbf{y}' = R(\phi)\mathbf{y}$, where $\phi$ is given in (6.5), so that $y^2 = (\dot{\mathbf{u}} + \mathbf{u}\times\boldsymbol{\omega}_0)^2$ is a second degree invariant term. Similarly $(\boldsymbol{\omega}_0\cdot\dot{\mathbf{u}})\dot{t}$ and $\dot{\mathbf{u}}^2\dot{t}^2$ are, respectively, third and fourth degree invariant terms. We can thus find several first degree invariant terms, and among others we quote

$$mc\frac{\dot{\mathbf{u}}^2\dot{t}}{\boldsymbol{\omega}_0\cdot\dot{\mathbf{u}}},\quad mc^2\dot{t}\left(\frac{\dot{\mathbf{u}}^2}{\mathbf{y}^2}\right)^{1/2},\quad \frac{mc^3(\boldsymbol{\omega}_0\cdot\dot{\mathbf{u}})\dot{t}}{y^2},$$

$$\frac{m^2c^5}{S^2}\frac{\dot{\mathbf{u}}^2\dot{t}^2}{y^3},\quad \frac{S\boldsymbol{\omega}_0\cdot\dot{\mathbf{u}}}{(\dot{\mathbf{u}}^2)^{1/2}},\quad S\frac{y}{c},$$

where the parameters $m$ and $S$ have been introduced by dimensionality considerations.

For instance, the first degree invariant Lagrangian

$$L = mc^3\frac{(\boldsymbol{\Omega}\cdot\dot{\mathbf{u}})}{(d\mathbf{u}/dt + \mathbf{u}\times\boldsymbol{\Omega})^2} - \frac{S}{c}\left[\left(\frac{d\mathbf{u}}{dt} + \mathbf{u}\times\boldsymbol{\Omega}\right)^2\right]^{1/2}\dot{t} \tag{6.30}$$

leads for the spin $\mathbf{S} = \mathbf{u}\times\mathbf{U} + \mathbf{Z}$ to

$$\mathbf{S} = mc^3\frac{(d\mathbf{u}/dt + \mathbf{u}\times\boldsymbol{\Omega})}{(d\mathbf{u}/dt + \mathbf{u}\times\boldsymbol{\Omega})^2}, \tag{6.31}$$

and we see that $\mathbf{S}\cdot\mathbf{u} = 0$. For the center of mass observer, $\mathbf{p} = 0$, where $\mathbf{S}$ is constant, and $\boldsymbol{\pi} = 0$ implies that (6.24) reads $mc^2\mathbf{r} = \mathbf{S}\times\mathbf{u}$. We can eliminate $\mathbf{u}$ from these two equations obtaining $\mathbf{u} = -mc^2S^{-2}\mathbf{S}\times\mathbf{r} = \boldsymbol{\Omega}_l\times\mathbf{r}$, where the orbital angular velocity $\boldsymbol{\Omega}_l$ has direction opposite to the spin and constant value $mc^2/S$.

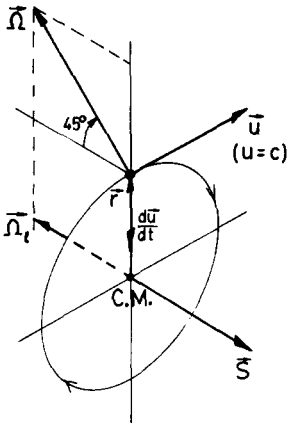



FIG. 3. Relative orientation of observables in the C.M. frame.

Since $\mathbf{S}\cdot d\mathbf{u}/dt = 0$ we get from (6.31) that

$$\left(\frac{d\mathbf{u}}{dt}\right)^2 + \left(\frac{d\mathbf{u}}{dt}\right)\cdot(\mathbf{u}\times\boldsymbol{\Omega}) = 0.$$

If we consider a body fixed coordinate frame that rotates with angular velocity $\boldsymbol{\Omega}_l$ with respect to the center of mass frame and we call $\Omega_s$, $\Omega_u$, and $\Omega_{\dot{u}}$ the components of $\boldsymbol{\Omega}$ along the three orthogonal directions of $\mathbf{S}$, $\mathbf{u}$, and the $d\mathbf{u}/dt$, respectively, we obtain $\Omega_l^2c^2 + \Omega_l\Omega_sc^2 = 0$, i.e., $\Omega_s = -\Omega_l$, since $\boldsymbol{\Omega}_l\times\mathbf{u} = d\mathbf{u}/dt$.

Taking the absolute value of (6.31) we get that $|(d\mathbf{u}/dt + \mathbf{u}\times\boldsymbol{\Omega})| = \Omega_lc$. Then its projection on the direction of $\mathbf{S}$ gives $S = mc^3(-c\Omega_{\dot{u}})/\Omega_l^2c^2$ and thus $\Omega_{\dot{u}} = -\Omega_l$.

Finally the condition $\mathbf{p} = \mathbf{Q} - d\mathbf{U}/dt = 0$ leads in this frame to

$$\mathbf{u}(\Omega^2) - \boldsymbol{\Omega}(\boldsymbol{\Omega}\cdot\mathbf{u}) - 2\frac{d\mathbf{u}}{dt}\times\boldsymbol{\Omega} - \frac{d^2\mathbf{u}}{dt^2} - \mathbf{u}\times\frac{d\boldsymbol{\Omega}}{dt} - c\frac{d\boldsymbol{\Omega}}{dt} = 0,$$

and, since $d^2\mathbf{u}/dt^2 = -\Omega_l^2\mathbf{u}$ and $d\boldsymbol{\Omega}/dt = (d\boldsymbol{\Omega}/dt)_b + \boldsymbol{\Omega}_l\times\boldsymbol{\Omega}$, where $(d\boldsymbol{\Omega}/dt)_b$ is the derivative of $\boldsymbol{\Omega}$ in the body fixed frame, we have

$$\mathbf{u}(3\Omega_l^2 + \Omega_u^2) - (\boldsymbol{\Omega} + \boldsymbol{\Omega}_l)c\Omega_u - 2\frac{d\mathbf{u}}{dt}\times\boldsymbol{\Omega} - \mathbf{u}\times\left(\frac{d\boldsymbol{\Omega}}{dt}\right)_b - c\left(\frac{d\boldsymbol{\Omega}}{dt}\right)_b - c(\boldsymbol{\Omega}_l\times\boldsymbol{\Omega}) = 0.$$

If we take the projection of this expression along the $\mathbf{S}$ direction, taking into account that $\Omega_s = -\Omega_l$ is constant, we obtain $\Omega_u = 0$, and thus the angular velocity $\boldsymbol{\Omega}$ is of constant value $\sqrt{2}\Omega_l$ and in the center of mass frame it rotates around the spin direction with angular velocity $\boldsymbol{\Omega}_l$. We see that for this system $\mathbf{S}$ and $\boldsymbol{\Omega}$ are not directly related (Fig. 3).

## VII. TACHYONS

If we consider the manifold spanned by the variables $(t,\mathbf{r},\mathbf{u},\boldsymbol{\rho})$ with domains $t\in\mathbf{R}$, $\mathbf{r}\in\mathbf{R}^3$, $\boldsymbol{\rho}\in\mathbf{R}_c^3$ as before and $\mathbf{u}\in\mathbf{R}^3$, $u > c$, we see that the transformation equations for $\boldsymbol{\rho}$ [(5.4)] do not give any real limit when $u > c$ because the $\gamma(u)$ terms

involved in the F and $G$ functions are imaginary. Thus for tachyons we are left with the seven-dimensional manifold $X_T$ spanned by the variables $(t(\tau),\mathbf{r}(\tau),\mathbf{u}(\tau))$, with $t\in\mathbb{R}$, $\mathbf{r}\in\mathbb{R}^3$, $\mathbf{u}\in\mathbb{R}^3$, and $u>c$, which is isomorphic to the homogeneous space $\mathscr{P}/\mathrm{SO}(3)$, and thus no angular variables can be defined. The kinematical variables transform as in (5.1)–(5.3). There exists the constraint $\mathbf{u}=\dot{\mathbf{r}}/\dot{t}$ and the homogeneity condition (2.4) allows us to write the Lagrangian

$$L=-T\dot{t}+\mathbf{Q}\cdot\dot{\mathbf{r}}+\mathbf{U}\cdot\dot{\mathbf{u}}, \tag{7.1}$$

where

$$T=-\frac{\partial L}{\partial \dot{t}},\quad Q_i=\frac{\partial L}{\partial \dot{r}^i},\quad U_i=\frac{\partial L}{\partial \dot{u}^i}.$$

Invariance of $L$ leads for $T$ and $\mathbf{Q}$ to the transformation equations (5.13) and (5.14) and for $\mathbf{U}$ to

$$\mathbf{U}'=\gamma\left(1+\frac{\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{u}}{c^2}\right)\left[R(\boldsymbol{\mu})\mathbf{U}+\frac{\gamma^2}{(1+\gamma)c^2}\times(\mathbf{v}\cdot R(\boldsymbol{\mu})\mathbf{U})\mathbf{v}+\frac{\gamma}{c^2}(\mathbf{u}\cdot\mathbf{U})\mathbf{v}\right] \tag{7.2}$$

and thus they are functions of $(\mathbf{u},\dot{t},\dot{\mathbf{r}},\dot{\mathbf{u}})$, being homogeneous of zero degree in terms of the derivatives.

Noether's theorem defines the constants of the motion to be

$$\text{energy},\quad H=T-\frac{d\mathbf{U}}{dt}\cdot\mathbf{u}; \tag{7.3}$$

$$\text{linear momentum},\quad \mathbf{p}=\mathbf{Q}-\frac{d\mathbf{U}}{dt}; \tag{7.4}$$

Poincaré momentum,

$$\boldsymbol{\pi}=-H\mathbf{r}/c^2+\mathbf{p}t+\mathbf{U}-[\mathbf{U}\cdot\mathbf{u}/c^2]\mathbf{u}; \tag{7.5}$$

$$\text{angular momentum},\quad \mathbf{J}=\mathbf{r}\times\mathbf{p}+\mathbf{u}\times\mathbf{U}. \tag{7.6}$$

The observable $\mathbf{U}-[(\mathbf{U}\cdot\mathbf{u})/c^2]\mathbf{u}$ is always different from zero since $u>c$, and if we define the relative position vector $\mathbf{k}$ as before by $H\mathbf{k}/c^2=\mathbf{U}-[(\mathbf{U}\cdot\mathbf{u})/c^2]\mathbf{u}$ then $H$ has also to be different from zero for every observer. This implies that the invariant and constant of the motion $(H/c)^2-p^2$ cannot be negative and thus the system has a tachyonic internal orbital motion while its center of mass has a velocity $\leqslant c$. The spin of the particle $\mathbf{S}=\mathbf{u}\times\mathbf{U}=H\mathbf{u}\times\mathbf{k}/c^2$ is a constant of the motion in the center of mass frame for nonzero mass particles, while for massless particles it precesses around the linear momentum, always being orthogonal to the velocity.

The invariant Lagrangian for $u>c$ particles,

$$L_T=A\dot{t}(u^2-c^2)^{1/2}-\frac{B\dot{t}}{(u^2-c^2)^{3/2}}\times\left[\left(\frac{d\mathbf{u}}{dt}\right)^2-\frac{(\mathbf{u}\cdot d\mathbf{u}/dt)^2}{u^2-c^2}\right], \tag{7.7}$$

leads in the center of mass frame to the dynamical equations

$$m\mathbf{r}=\frac{-2B}{(u^2-c^2)^{3/2}}\frac{d^2\mathbf{r}}{dt^2}. \tag{7.8}$$

The internal motion is a central motion, and, being the spin constant, this gives rise to a first integral $\mathbf{S}=-\mathbf{r}\times m\mathbf{u}$, the motion lying in a plane orthogonal to the constant vector $\mathbf{S}$.

In polar coordinates $(r,\theta)$ in this plane Eqs. (7.8) become

$$\frac{d^2r}{dt^2}-r\left(\frac{d\theta}{dt}\right)^2+\frac{m}{2B}\left[\left(\frac{dr}{dt}\right)^2+r^2\left(\frac{d\theta}{dt}\right)^2-c^2\right]^{3/2}r=0, \tag{7.9}$$

$$2\frac{dr}{dt}\frac{d\theta}{dt}+r\frac{d^2\theta}{dt^2}=0, \tag{7.10}$$

the first integral $d\theta/dt=S/mr^2$, and thus the radial equation (7.9) becomes

$$\frac{d^2r}{dt^2}-\frac{S^2}{m^2r^3}+\frac{m}{2B}\left[\left(\frac{dr}{dt}\right)^2+\frac{S^2}{m^2r^2}-c^2\right]^{3/2}r=0. \tag{7.11}$$

We see that this equation has solutions with constant $r\neq 0$ and the system in the center of mass frame follows a circle with constant velocity $u>c$. A general solution of (7.11) has not yet been obtained.

## VIII. INVERSIONS

Since space and time reversal are automorphisms of $\mathscr{P}$ given by

$$P\colon (b,\mathbf{a},\mathbf{v},\boldsymbol{\mu})\to(b,-\mathbf{a},-\mathbf{v},\boldsymbol{\mu}),$$

$$T\colon (b,\mathbf{a},\mathbf{v},\boldsymbol{\mu})\to(-b,\mathbf{a},-\mathbf{v},\boldsymbol{\mu}),$$

we extend this action on the general kinematical space $X=\mathscr{P}$ by

$$P\colon (t(\tau),\mathbf{r}(\tau),\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))\to(t(\tau),-\mathbf{r}(\tau),-\mathbf{u}(\tau),\boldsymbol{\rho}(\tau)),$$

$$T\colon (t(\tau),\mathbf{r}(\tau),\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))\to(-t(\tau),\mathbf{r}(\tau),-\mathbf{u}(\tau),\boldsymbol{\rho}(\tau)).$$

If we assume that $\tau$ parameter remains invariant under inversions, then we can define the $P$ and $T$ action on the derivatives as

$$P\colon (\dot{t}(\tau),\dot{\mathbf{r}}(\tau),\dot{\mathbf{u}}(\tau),\dot{\boldsymbol{\rho}}(\tau))\to(\dot{t}(\tau),-\dot{\mathbf{r}}(\tau),-\dot{\mathbf{u}}(\tau),\dot{\boldsymbol{\rho}}(\tau)),$$

$$T\colon (\dot{t}(\tau),\dot{\mathbf{r}}(\tau),\dot{\mathbf{u}}(\tau),\dot{\boldsymbol{\rho}}(\tau))\to(-\dot{t}(\tau),\dot{\mathbf{r}}(\tau),-\dot{\mathbf{u}}(\tau),\dot{\boldsymbol{\rho}}(\tau)),$$

and thus

$$P\colon (\boldsymbol{\alpha}(\tau),\boldsymbol{\omega}(\tau))\to(-\boldsymbol{\alpha}(\tau),\boldsymbol{\omega}(\tau)),$$

$$T\colon (\boldsymbol{\alpha}(\tau),\boldsymbol{\omega}(\tau))\to(-\boldsymbol{\alpha}(\tau),\boldsymbol{\omega}(\tau)).$$

Lagrangian (4.10) remains invariant under $P$ and changes its sign under $T$ so that dynamical equations are invariant under inversions. The Lagrangian (4.13) is itself invariant.

Similarly, Lagrangians (5.27) $L_B$ and (7.7) $L_T$ are invariant under $P$ and change sign under $T$, and the photonic Lagrangian (6.16) is affected by a minus sign under both inversions.

However, the Lagrangian (5.28) $L_F$ under parity operation has the first term invariant while the second one changes in sign. Under time reversal we have the opposite situation: the first term is affected by a minus sign but the second is left invariant, so that inversions alter this system.

Finally, the Lagrangian (6.30) has the same behavior as $L_F$ under inversions. Its two terms are separately affected by a minus sign by a different inversion, and we can see from Fig. 3 that the internal motion is reversed but the spin remains unchanged.

## APPENDIX: DEFINITION OF VARIOUS FUNCTIONS

If $\Lambda$ is a Lorentz transformation, then $\Lambda G \Lambda^T = G$ holds, where $G$ is the Minkowski metric tensor written in matrix form, i.e., diag( $-1,1,1,1$), and $\Lambda^T$ is the transpose of $\Lambda$.

If we form the matrix $\Lambda(\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))$ it also fulfills this condition. Taking the $\tau$ derivative,

$$\dot{\Lambda}(\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))G\Lambda^T(\mathbf{u}(\tau),\boldsymbol{\rho}(\tau)) + \Lambda(\mathbf{u}(\tau),\boldsymbol{\rho}(\tau))G\dot{\Lambda}^T(\mathbf{u}(\tau),\boldsymbol{\rho}(\tau)) = 0,$$

i.e., $\Omega + \Omega^T = 0$, and the antisymmetric matrix $\Omega$ is a linear function of the derivatives $\dot{\mathbf{u}}$ and $\dot{\boldsymbol{\rho}}$.

Under a Poincaré transformation, the variables $\mathbf{u}$ and $\boldsymbol{\rho}$ transform according to (5.3) and (5.4), which is equivalent to

$$\Lambda(\mathbf{u}',\boldsymbol{\rho}') = \Lambda(\mathbf{v},\boldsymbol{\mu})\Lambda(\mathbf{u},\boldsymbol{\rho}),$$

and thus

$$\dot{\Lambda}(\mathbf{u}',\boldsymbol{\rho}') = \Lambda(\mathbf{v},\boldsymbol{\mu})\dot{\Lambda}(\mathbf{u},\boldsymbol{\rho}),$$

so that

$$\begin{aligned}\Omega' &= \dot{\Lambda}(\mathbf{u}',\boldsymbol{\rho}')G\Lambda(\mathbf{u}',\boldsymbol{\rho}') \\ &= \Lambda(\mathbf{v},\boldsymbol{\mu})\dot{\Lambda}(\mathbf{u},\boldsymbol{\rho})G\Lambda(\mathbf{u},\boldsymbol{\rho})\Lambda^T(\mathbf{v},\boldsymbol{\mu}) \\ &= \Lambda(\mathbf{v},\boldsymbol{\mu})\Omega\Lambda^T(\mathbf{v},\boldsymbol{\mu}),\end{aligned}$$

which corresponds to the transformation properties of a tensor $\Omega^{\mu\nu}$.

Similarly, if $R$ is an orthogonal $3\times 3$ matrix, $RR^T = 1$. In particular, if we define the orthogonal matrix $R(\boldsymbol{\rho}(\tau))$, then taking the $\tau$ derivative we get $\dot{R}(\tau)R^T(\tau) + R(\tau)\dot{R}^T(\tau) = 0$, i.e., $\Omega + \Omega^T = 0$, and if we call $\omega_0$ the nondiagonal components of this $\Omega$ we get expression (5.7).


[1] A. P. Balachandran, G. Marmo, B. S. Skagerstam, and A. Stern, *Gauge Symmetries and Fibre Bundles, Lecture Notes in Physics*, Vol. 188 (Springer, Berlin, 1984); N. Mukunda, H. Van Dam, and L. C. Biedenharn, *Lecture Notes in Physics*, Vol. 165 (Springer, Berlin, 1982); J. M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970).

[2] M. V. Atre and N. Mukunda, J. Math. Phys. **27**, 2908 (1986); **28**, 792 (1987).

[3] M. Ostrogradsky, Mem. Acad. St. Petersburg **6(4)**, 385 (1850).

[4] F. Bopp, Ann. Phys. (NY) **38**, 345 (1940); B. Podolsky, Phys. Rev. **62**, 68 (1942); B. Podolsky and C. Kikuchi, *ibid.* **65**, 228 (1944); **67**, 184 (1945).

[5] F. Riewe, Lett. Nuovo Cimento **1**, 807 (1971).

[6] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., Cambridge, 1959), 4th ed.

[7] I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, NJ, 1963); R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. I (Interscience, New York, 1970), 8th ed.

[8] H. Bacry and J. M. Levy-Leblond, J. Math. Phys. **9**, 1605 (1968).

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

[10] V. Bargmann, Ann. Math. **59**, 1 (1954).

[11] A. J. Hanson and T. Regge, Ann. Phys. NY **87**, 498 (1974).

[12] M. Rivas, M. A. Valle, and J. M. Aguirregabiria, Eur. J. Phys. **7**, 1 (1986).

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

[14] G. C. Constantelos, Nuovo Cimento B **21**, 279 (1974).

# The trace identity, a powerful tool for constructing the Hamiltonian structure of integrable systems

Gui-zhang Tu[a)]
*Theoretical Physics Division, Nankai Institute for Mathematics, Tianjin, People's Republic of China*

(Received 5 June 1987; accepted for publication 21 September 1988)

A new approach to Hamiltonian structures of integrable systems is proposed by making use of a trace identity. For a variety of isospectral problems that can be unified to one model $\psi_x = U\psi$, it is shown that both the related hierarchy of evolution equations and the Hamiltonian structure can be obtained from the same solution of the equation $V_x = [U,V]$.

## I. INTRODUCTION

The theory of integrable Hamiltonian systems of infinite dimensions has undergone a rapid development since the late 1960's (see, e.g., Refs. 1–32 and references therein).

Nowadays it is well known that a great number of nonlinear evolution equations, arising from various branches of physics, are (formally) integrable, i.e., they admit the zero-curvature representation

$$U_t - V_x + [U,V] = 0 .$$

Given a properly chosen isospectral problem

$$\psi_x = U(\lambda)\psi , \tag{1}$$

with $\lambda$ being the spectral parameter for which $\lambda_t = 0$, one can relate it to a hierarchy of nonlinear evolution equations

$$U_t - V_x^{(n)} + [U,V^{(n)}] = 0 . \tag{2}$$

A central and very important subject in the theory of integrable systems is to search for a symplectic operator $J$ and a sequence of scalar functions $\{H_n\}$ such that Eqs. (2) can be cast in the Hamiltonian form

$$u_t = J\frac{\delta H_n}{\delta u} ,$$

where $u = (u_1,\ldots,u_N)$ is the "potential" contained in the matrix $U = U(\lambda) = U(u,\lambda)$, and $\delta/\delta u = (\delta/\delta u_i)$ stands for the variational derivatives

$$\frac{\delta}{\delta u_i} = \sum_{n\geq 0}(-\partial)^n \frac{\partial}{\partial u_i^{(n)}}\left(\partial = \frac{d}{dx},\ u_i^{(n)} = \partial^n u_i\right).$$

The Hamiltonians $\{H_n\}$ constitute, in fact, an infinite number of conserved densities of the hierarchy (2). Various techniques have been developed to calculate $\{H_n\}$. We mention in the following some of the typical techniques.

A standard method makes use of the transmission coefficients of the scattering date (see, e.g., Refs. 3–5). To apply this method we need first to do inverse scattering analysis; it is sometimes very difficult for a number of isospectral problems with higher singularities in $\lambda$.

Another method, proposed by Chen, Lee, and Liu,[10] is based on perturbation expansions. This method is of particular value when the corresponding Lax pairs are unknown. It is most suitable for a single concrete equation, not for a hierarchy of equations.

In their important monograph[5] Faddeev and his group formulated an ingenious method using a classical $r$ matrix, which originates from the famous Yang–Baxter equation. Presently, this method does not seem popular in the mathematical community. The search for appropriate $r$ matrices is by no means easy work. Some known $r$ matrices were constructed by an ansatz or guesswork. A systematic construction of $r$ matrices is given in Ref. 5, p. 528; however, it seems to the author that this construction works only when the corresponding symplectic operators are linearly dependent on the potentials.

Alberty, Koikawa, and Sasaki[14] have proposed an effective method for calculating the conserved densities of the hierarchy (2). First one solves the equation

$$V_x = [U,V] \tag{3}$$

to generate the hierarchy (2) and then one solves the ordinary differential equation

$$Y_x = UY - Y(UY)_D ,$$

where $H \equiv (UY)_D$ represents the diagonal part of $UY$, which is just the desired generating function of the conserved densities.

One goal of this paper is to point out that the information obtained from solving Eq. (3) is completely sufficient to construct $\{H_n\}$. Therefore searching for $Y$ is, in fact, unnecessary and thus the work can be considerably simplified.

In a beautiful paper,[7] Magri proposed the famous bi-Hamiltonian formalism. It always involves a recursion operator $L$. There are two coupling conditions that the symplectic operator $J$ and the recursion operator $L$ have to satisfy. To ensure that the scheme is workable, Magri imposes two other conditions which are equivalent to the properties of strong and hereditary symmetries[9,12] and one further initial condition on gradient reserving. The five conditions above, together with the condition that implies $J$ is a symplectic operator, require very heavy and tedious calculation even for the simple-looking operators $J$ and $L$.

The method presented in this paper requires only two conditions. The first one makes $J$ a symplectic operator, which is clearly necessary; the second one is the first coupling condition $JL = L^*J$, which is much simpler than the second coupling condition.

It is worthwhile to mention that the recursion operators exist not only in the case of $1+1$ dimensions, but also in the case of $1+2$ dimensions as shown by the recent important work of Fokas and Santini.[24] They successfully constructed

[a)] Mailing address: Computing Center, Academia Sinica, Beijing, China.

 0022-2488/89/020330-09$02.50 © 1989 American Institute of Physics 

a recursion operator for the Kadomtsev–Petviashvili (KP) hierarchy and proved that this well-known KP hierarchy is bi-Hamiltonian.

In a series of papers[15–21] we develop a simple approach to Hamiltonian structures of integrable systems. The essence of this approach is the use of the chain rule of variational derivatives. This method was developed further by Chowdhury and Swapna.[26] Based on this technique we proved an important fact—the matrix $V=(\delta H/\delta u)^T$ satisfies (3), where $H$ is calculated by the Alberty–Koikawa–Sasaki algorithm as mentioned above. This result suggested that we find in the general cases the relation between solutions of (3) and the conserved densities of the hierarchy. This is indeed the motivation of the present work. We then realized in Ref. 22 that the solution $V$ of (3) does carry much information on conserved densities. Bearing this clue in mind we finally established in 1985 a remarkable trace identity.[25] In the present work we are able to show that both the hierarchy (2) and the Hamiltonians $\{H_n\}$ can be derived from the single equation (3), and the trace identity provides us with a powerful tool for this goal. Furthermore, we have shown in a subsequent paper[32] that the Liouville integrability of zero-curvature equations and an explicit formula of Poisson brackets could be successfully established by making use of the trace identity.

The next two sections are of a preliminary nature; the main body of the paper is contained in Sec. IV, and the paper ends in Sec. V with concluding remarks and an open problem.

## II. BASIC NOTIONS

To fix the notation we recall briefly some basic notions from the theory of generalized Hamiltonian equations.

Let $S$ be the Schwartz space over $\mathbb{R}$, $S^M=S\otimes\cdots\otimes S$ ($M$ times). The operator $\partial=d/dx$ introduces an equivalence relation among elements of $S^M$,

$$f\sim g\Leftrightarrow\exists h \text{ such that } f-g=\partial h \quad (f,g,h\in S^M).$$

The equivalence class that contains $f$ is denoted by $\int f\,dx$,[8]

$$\int f\,dx=\{f+\partial h\,|h\in S^M\}.$$

The scalar product between $f$ and $g$ is defined by

$$(f,g)=\int f\cdot g\,dx=\int\sum_i f_i g_i\,dx.$$

Let $u=u(x,t)=(u_1,\cdots,u_M)$ be a smooth function that belongs to $S^M$ for any fixed $t$. A linear and skew-symmetric (with respect to the above scalar product) operator $J=J(u):S^M\to S^M$ is called symplectic[8,11,12] if it holds that

$$(J'(u)[Jf]g,h)+(J'(u)[Jg]h,f)+(J'(u)[Jh]f,g)=0,$$

for any triplet $f,g,h\in S^M$, where $J'(u)[f]$ represents the Fréchet derivative of $J$,

$$J'(u)[f]=\frac{d}{d\epsilon}J(u+\epsilon f)|_{\epsilon=0}.$$

An operator $N(u):S^M\to S^M$ is called hereditary[9] if

$$N'(u)[Nf]g-N'(u)[Ng]f=N(N'(u)[f]g-N'(u)[g]f).$$

It is known that if $J$ is symplectic then the bracket

$$\{f,g\}=\left(\frac{\delta f}{\delta u},J\frac{\delta g}{\delta u}\right) \tag{4}$$

is a well-defined Poisson bracket, and we call the equation $u_t=J\,\delta H/\delta u$ the generalized Hamiltonian equation.

It has been proved[12] that if the operators $J_1$, $J_2$, and $J_1+J_2$ are simultaneously symplectic, and $J_1$ is invertible, then $N=J_1^{-1}J_2$ is hereditary. This fact is closely related to the intriguing bi-Hamiltonian structures of integrable systems.[7]

## III. A MODEL ISOSPECTRAL PROBLEM

Let $G$ be a finite-dimensional Lie algebra over $\mathbb{C}$, and $\tilde{G}$ be the corresponding loop algebra[33]

$$\tilde{G}=G\otimes\mathbb{C}[\lambda,\lambda^{-1}],$$

where $\mathbb{C}[\lambda,\lambda^{-1}]$ is the set of Laurent polynomials in $\lambda$. Suppose that $\{E_1,\ldots,E_d\}$ is a basis of $G$. Then

$$\{E_1(n),\ldots,E_d(n)\,|n\in\mathbb{Z}\},$$

where $E_i(n)=E_i\otimes\lambda^n=E_i\lambda^n$ provides a basis for $\tilde{G}$. It can be transformed to other bases

$$\{\overline{E}_i(n),\ldots,\overline{E}_d(n)\},$$

with the property that

$$\overline{E}_i(n)=\overline{E}_i(0)\lambda^n.$$

We call an element $R\in\tilde{G}$ pseudoregular if, for

$$\ker\text{ad}\,R=\{x|x\in\tilde{G},\ [x,R]=0\},$$

$$\text{Im ad}\,R=\{x|\exists y\in\tilde{G}\text{ such that }x=[y,R]\},$$

it holds that

(i) $\tilde{G}=\ker\text{ad}\,R\oplus\text{Im ad}\,R$;

(ii) $\ker\text{ad}\,R$ is commutative.

From the Cartan decomposition it is easy to deduce the following proposition.

*Proposition 1:* If $A$ is a regular element of a semisimple Lie algebra $G$, then $R=A\otimes\lambda^n$ is pseudoregular in $\tilde{G}$.

There certainly exist other pseudoregular elements as shown in Sec. IV.

Different gradations of $\tilde{G}$ are available, one of them defined by

$$\deg(X\otimes\lambda^n)=n. \tag{5}$$

In the sequel of the section we shall fix one of the gradations. For $g\in\tilde{G}$, let $g=\Sigma_n g_n$, $\deg g_n=n$, be its gradation decomposition. We set

$$g_+=\sum_{n\geq\pi}g_n,\quad g_-=\sum_{n<\pi}g_n,$$

and call $g_+$ the positive part of $g$, where $\pi\in\mathbb{Z}$ is a properly chosen integer.

Since

$$\deg(x\lambda)+\deg(y)=\deg(x)+\deg(\lambda y)=\deg(\lambda[x,y]),$$

we have

$$\deg(x\lambda)-\deg(x)=\deg(y\lambda)-\deg(y),\quad x,y\in\tilde{G}. \tag{6}$$

We shall discuss in Sec. IV a variety of isospectral problems, $\psi_x=U\psi$; they can be unified in the following model:

$$U = R + u_1 e_1 + \cdots + u_p e_p ,$$

where $u_1,...,u_p \in S$ (the Schwartz space) and $R, e_1,...,e_p \in \tilde{G}$ meet the conditions

(i) $R, e_1,...,e_p$ are linear independent,

(ii) $R$ is pseudoregular,

(iii) $\alpha > 0, \quad \alpha > \epsilon_i \quad (i = 1,...,p)$,

where

$$\alpha = \deg(R), \quad \epsilon_i = \deg(e_i) . \tag{7}$$

In the one exceptional case that we shall deal with in Sec. IV, condition (iii) is weakened to

$$\alpha > 0, \quad \alpha \geqslant \epsilon_i \quad (i = 1,...,p) . \tag{8}$$

The aim of imposing the above conditions on $U$ is to ensure the solvability of (3). This idea can be traced back to Wilson[13] and Drinfeld and Sokolov.[28] A further analysis of this model isospectral problem is given in Ref. 29.

We need to introduce another concept, the rank for $\partial$, $u$, $\lambda$, and $x \in \tilde{G}$ such that if $ab$ makes sense for two entities $a$ and $b$, then

$$\operatorname{rank}(ab) = \operatorname{rank}(a) + \operatorname{rank}(b) .$$

We shall define the rank in such a way that the above element $U$ is of homogeneous rank, i.e., $\operatorname{rank}(R) = \operatorname{rank}(u_1 e_1) = \cdots = \operatorname{rank}(u_p e_p)$. To this end we take

$$\operatorname{rank}(x) = \deg(x), \quad x \in \tilde{G}, \tag{9a}$$

$$\operatorname{rank}(\lambda) = \deg(x\lambda) - \deg(x) , \tag{9b}$$

$$\operatorname{rank}(u_i) = \alpha - \epsilon_i \quad (i = 1,...,p) , \tag{9c}$$

$$\operatorname{rank}(\partial) = \alpha , \tag{9d}$$

$$\operatorname{rank}(\beta) = 0 \quad (\beta = \text{const}, \ \beta \neq 0) . \tag{9e}$$

Thus we have, for example,

$$\operatorname{rank}(\lambda u_i) = \operatorname{rank}(\lambda) + \operatorname{rank}(u_i),$$

$$\operatorname{rank}(\partial^2 u_3) = 2\alpha + \operatorname{rank}(u_3) .$$

Note that (9b) is well defined from (6).

In order to fix the integral constant arising from calculation we shall follow the *homogeneous rank convention*: both sides of an equation have the same rank. For example, the condition $\operatorname{rank}(a) = 1$ and $a_x = 0$ implies $a = 0$; here the integral constant is set to zero by convention. Originally in the literature (see, e.g., Refs. 6 and 13) the rank was defined only for $u_i$ and $\partial$ to ensure a unique choice of integral constants. Here we have extended the rank somewhat for convenience of a later formulation.

In the next section the scheme to construct the hierarchy (2) is as follows. First, we take a solution $V$ of (3); second, we search for a $\Delta_n \in \tilde{G}$ such that for

$$V^{(n)} = (\lambda^n V)_+ + \Delta_n$$

it holds that

$$-V_x^{(n)} + [U, V^{(n)}] \in \mathbb{C}e_1 + \cdots + \mathbb{C}e_p;$$

this requirement yields a hierarchy of evolution equations (2).

## IV. KILLING–CARTAN FORMS $\langle V, \partial U/\partial\lambda \rangle$ AS HAMILTONIAN

### A. The trace identity

Let $G$ be a matrix semisimple Lie algebra. It is well known that the Killing–Cartan form $\langle x,y \rangle$ is proportional to the trace $\operatorname{tr}(xy)$, $\langle x,y \rangle = \text{const} \operatorname{tr}(xy)$. For notational convenience, we thus take, in this section,

$$\langle x,y \rangle = \operatorname{tr}(xy) .$$

Let $U = U(\lambda,u)$ be an element of $\tilde{G}$ that depends on $\lambda$ and $u = (u_i)$. Suppose we have introduced in some manner the rank for $u$, $\partial$, $\lambda$, and $x \in \tilde{G}$. Under the supposition that the solution $V$ of (3), which is of given homogeneous rank, is unique up to a constant multiplier, we proved in Refs. 25 and 29 that for any solution $V$ of homogeneous rank, there exists a constant $\gamma$ such that for $\overline{V} = \lambda^\gamma V$, which is again a solution of (3), it holds that

$$\frac{\delta}{\delta u_i}\left\langle \overline{V}, \frac{\partial U}{\partial \lambda} \right\rangle = \frac{\partial}{\partial \lambda}\left\langle \overline{V}, \frac{\partial U}{\partial u_i} \right\rangle . \tag{10}$$

We have shown that a number of known formulas obtained in the literature by complicated calculations are special cases of the above remarkable trace identity.

By substituting $\overline{V} = \lambda^\gamma V$ in (10) we obtain

$$\frac{\delta}{\delta u_i}\left\langle V, \frac{\partial U}{\partial \lambda} \right\rangle = \left(\lambda^{-\gamma} \frac{\partial}{\partial \lambda} \lambda^\gamma\right)\left\langle V, \frac{\partial U}{\partial u_i} \right\rangle . \tag{11}$$

We shall always search for solutions $V$ of (3) of the form

$$V = \sum_{n=0}^{\infty} V_n(u) \lambda^{-n},$$

with

$$V_0(u) = \text{const} \neq 0 .$$

The following simple proposition will be frequently used in later discussion.

*Proposition 2:* Let $A = \Sigma A_n \lambda^{-n}$ and the operator $\Gamma_n$ be defined as $\Gamma_n A = A_n$. Then

$$\Gamma_n(\lambda^k A) = \Gamma_{n+k} A,$$

$$\Gamma_n\left(\lambda^{-\gamma} \frac{\partial}{\partial \lambda} \lambda^\gamma A\right) = (\gamma - n + 1)\Gamma_{n-1} A .$$

Magri[7] proposed a geometrical approach to Hamiltonian structures of integrable systems; however, the calculation involved in the verification of his two "coupling conditions" and others is tedious in most cases. For our present needs the following proposition is sufficient.

*Proposition 3:* Let $J,L$ be two linear operators mapping $S^M$ into itself. Suppose that (i) both $J$ and $JL$ are skew symmetric, i.e.,

$$J^* = -J, \quad (JL) = L^* J ; \tag{12}$$

(ii) there exists a series of scalar functions $\{H_n\}$ for which it holds that

$$L^n f(u) = \frac{\delta H_n}{\delta u}, \tag{13}$$

for some $f(u) \in S^M$. Then $\{H_n\}$ is a common series of conserved densities for the whole hierarchy of equations

$$u_t = JL^n f(u) \tag{14}$$

and we have

$$\{H_n,H_m\}=0. \tag{15}$$

*Proof:* By (4) and (12) we have

$$\{H_n,H_m\}=\left(\frac{\delta H_n}{\delta u},J\frac{\delta H_m}{\delta u}\right)=(L^n f(u),JL^m f(u))$$
$$=(L^n f(u),L^* JL^{m-1}f(u))$$
$$=(L^{n+1}f(u),JL^{m-1}f(u))=\{H_{n+1},H_{m-1}\}.$$

Repeating the above argument, we obtain $\{H_n,H_m\}=\{H_m,H_n\}$, while by (12) we should have $\{H_n,H_m\}=-\{H_m,H_n\}$. Thus Eq. (15) holds. The fact that $H_m$ is a conserved density for Eq. (14) is a consequence of (15) and the fact that $(H_m)_t\sim\{H_m,H_n\}$.

Now we turn to the search for the Hamiltonian structures for various integrable systems. In all the following cases Eqs. (12) hold, and, moreover, $J$ is symplectic and $L^*$ is hereditary.

### B. TC hierarchy[30]

First we discuss some details about a new hierarchy. The Lie algebra $G$ is taken to be $A_1$. Its basis is

$$\bar{h}=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{pmatrix},\quad e=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix},\quad f=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}.$$

Let $x(n)=x\otimes\lambda^n$ for $x=\bar{h},e,f$. The basis for $\tilde{G}=\tilde{A}_1$ is taken to be

$$\{\bar{h}(n),e_+(n),e_-(n)\,|\,n\in\mathbb{Z}\},$$

where $e_\pm(n)=(e(n-1)\pm f(n))/2$, for which it holds that

$$[e_-(m),e_+(n)]=\bar{h}(m+n-1),$$
$$[\bar{h}(m),e_\pm(n)]=e_\mp(m+n).$$

The isospectral problem is (1) with

$$U=\begin{pmatrix}0 & 1+(q+r)/2\lambda\\ \lambda+(q-r)/2 & 0\end{pmatrix}$$
$$=2e_+(1)+qe_+(0)+re_-(0).$$

By using the notation of Sec. III we set

$$R=2e_+(1),\quad e_1=e_+(0),$$
$$e_2=e_-(0),\quad u_1=q(x,t),\quad u_2=r(x,t).$$

It is easy to see that $R$ is pseudoregular. The gradation for $A_1$ is defined by

$$\deg e_\pm(n)=2n-1,\quad \deg\bar{h}(n)=2n.$$

Therefore we have, according to (7),

$$\alpha=1,\quad \epsilon_1=\epsilon_2=-1,$$

and, by (9a)–(9e), we have

$$\operatorname{rank}(\lambda)=2,\quad \operatorname{rank}(\partial)=1,\quad \operatorname{rank}(q)=\operatorname{rank}(r)=2.$$

For the elements $g\in\tilde{A}_1$,

$$g=\sum_{m\in\mathbb{Z}}(g_{1m}\bar{h}(m)+g_{2m}e_+(m)+g_{3m}e_-(m)),$$

we define

$$g_+=\sum_{m\geq 0}(g_{1m}\bar{h}(m)+g_{2m}e_+(m)+g_{3m}e_-(m)). \tag{16}$$

By using the notation of Sec. III we set $\pi=-1$. Let the solution of (3) be

$$V=\sum_{m\geq 0}(a_m\bar{h}(-m)+b_me_+(-m)+c_me_-(-m)). \tag{17}$$

By substituting (17) into (3) we obtain

$$a_{m+1}=(-qa_m-c_{mx})/2,$$
$$b_{m+1}=-\partial^{-1}ra_{m+1}, \tag{18}$$
$$c_{m+1}=(rb_m-qc_m-a_{m+1x})/2\quad (m\geq 0),$$

and the initial values

$$a_0=0,\quad b_0=\beta,\quad c_0=0.$$

From now on, $\beta$ will be used exclusively for nonzero constants. From the recurrence equation (18) and the convention on homogeneous rank we calculate successively that

$$a_1=0,\quad b_1=0,\quad c_1=\beta r/2,$$
$$a_2=-(\beta/4)r_x,\quad b_2=(\beta/8)r^2,$$
$$c_2=(\beta/8)(r_{xx}-2qr),$$

and so on. We note that

$$\operatorname{rank}(a_m)=2m-1,\quad \operatorname{rank}(b_m)=\operatorname{rank}(c_m)=2m$$

and

$$\operatorname{rank}(U)=1,\quad \operatorname{rank}(V)=-1.$$

Since (17) is a solution of (3), we have

$$-(\lambda^n V)_{+x}+[U,(\lambda^n V)_+]$$
$$=(\lambda^n V)_{-x}-[U,(\lambda^n V)_-], \tag{19}$$

where, by (16),

$$(\lambda^n V)_+=\sum_{m=0}^{n}(a_m\bar{h}(n-m)$$
$$+b_me_+(n-m)+c_me_-(n-m)).$$

Note that the terms in the left-hand side of (19) are of degree $\geq -2$, while the terms of the right-hand side are of degree $\leq -1$; therefore both sides of (19) are of degree $-1$ and $-2$. In other words, we have

$$-(\lambda^n V)_{+x}+[U,(\lambda^n V)_+]$$
$$\in\{\mathbb{C}e_+(0)+\mathbb{C}e_-(0)+\mathbb{C}\bar{h}(-1)\}.$$

To cancel the term from $\mathbb{C}\bar{h}(-1)$, we introduce

$$\Delta_n=((q/r)c_n-b_n)e_+(0).$$

Then it is easy to see that for $V^{(n)}=(\lambda^n V)_+ +\Delta_n$, we have

$$-V_x^{(n)}+[U,V^{(n)}]=-((q/r)c_n-b_n)_xe_+(0)$$
$$+(c_{nx}-(q/r)b_{nx})e_-(0),$$

from which we obtain the hierarchy

$$q_t=((q/r)c_n-b_n)_x,\quad r_t=(c_{nx}-(q/r)b_{nx}). \tag{20}$$

Taking $n=2$, $\beta=8$ in (20), we obtain the representative pair of equations in this hierarchy,

$$q_t=((q/r)r_{xx}-2q^2-r^2)_x,\quad r_t=r_{xxx}-2q_xr-4qr_x,$$

which reduces to the celebrated KdV equation when $q=\pm r$.

Now we proceed to search for the Hamiltonian structure of the hierarchy (20). To this end we write

$$V=a\bar{h}(0)+be_+(0)+ce_-(0)\,, \tag{21}$$

with

$$a=\sum_{m>0}a_m\lambda^{-m},\quad b=\sum_{m>0}b_m\lambda^{-m},$$
$$c=\sum_{m>0}c_m\lambda^{-m}. \tag{22}$$

It is easy to see that

$$\frac{\partial U}{\partial\lambda}=f(0)-\frac{(q+r)e(-2)}{2}.$$

Since $\langle\bar{h},\bar{h}\rangle=\frac{1}{2}$, $\langle e,f\rangle=1$, we have

$$\left\langle V,\frac{\partial U}{\partial\lambda}\right\rangle=\left\langle a\bar{h}+\frac{b+c}{2\lambda}e+\frac{(b-c)f}{2},f-\frac{q+r}{2\lambda^2}\right\rangle$$
$$=(b+c)/(2\lambda)-(b-c)(q+r)/(4\lambda^2)\,,$$

and

$$\left\langle V,\frac{\partial U}{\partial q}\right\rangle=\frac{b}{2\lambda},\quad\left\langle V,\frac{\partial U}{\partial r}\right\rangle=\frac{-c}{2\lambda}.$$

Therefore we obtain, by using the trace identity (11), that

$$\begin{pmatrix}\delta/\delta q\\ \delta/\delta r\end{pmatrix}\left(\frac{b+c}{2\lambda}-\frac{(b-c)(q+r)}{4\lambda^2}\right)$$
$$=(\lambda^{-\gamma}\frac{\partial}{\partial\lambda}\lambda^{\gamma})\begin{pmatrix}b/(2\lambda)\\ -c/(2\lambda)\end{pmatrix}.$$

By applying the operator $\Gamma_{n+2}$ given in Proposition 2 to both sides of the above equation we deduce that

$$\begin{pmatrix}\delta/\delta q\\ \delta/\delta r\end{pmatrix}\left(\frac{b_{n+1}+c_{n+1}}{2}-\frac{(q+r)(b_n-c_n)}{4}\right)$$
$$=(\gamma-n-1)\begin{pmatrix}b_n/2\\ -c_n/2\end{pmatrix}.$$

To fix the constant $\gamma$, we simply set $n=0$ in the above equation; then

$$\frac{(\gamma-1)\beta}{2}=\frac{(\gamma-1)b_0}{2}=\frac{\delta}{\delta q}\left(\frac{\beta r}{4}-\frac{(q+r)\beta}{4}\right)=\frac{-\beta}{4},$$

from which we obtain $\gamma=\frac{1}{2}$. Therefore we establish the following equation:

$$\begin{pmatrix}-b_n\\ c_n\end{pmatrix}=\begin{pmatrix}\delta/\delta q\\ \delta/\delta r\end{pmatrix}H_n$$

with

$$H_n=(2/(2n+1))[(q+r)(c_n-b_n)/2$$
$$+(b_{n+1}+c_{n+1})]\,, \tag{23}$$

and consequently the hierarchy (20) takes its Hamiltonian form

$$\begin{pmatrix}q_t\\ r_t\end{pmatrix}=J\begin{pmatrix}-b_n\\ c_n\end{pmatrix}=J\begin{pmatrix}\delta/\delta q\\ \delta/\delta r\end{pmatrix}H_n\,, \tag{24}$$

where

$$J=\begin{pmatrix}\partial & \partial(q/r)\\ (q/r)\partial & \partial\end{pmatrix}.$$

It can be verified that

$$\begin{pmatrix}-b_{n+1}\\ c_{n+1}\end{pmatrix}=L\begin{pmatrix}-b_n\\ c_n\end{pmatrix}$$

with

$$L=-\frac{1}{4}\begin{pmatrix}2\partial^{-1}q\partial & 2\partial^{-1}r\partial\\ 2r-\partial(q/r)\partial & 2q-\partial^2\end{pmatrix}.$$

It is easy to see that Eqs. (12) hold for the present pair $J$ and $L$; therefore by Proposition 3 we conclude that the Hamiltonians (23) are conserved densities for the whole hierarchy of equations (24) and they are in involution in pairs.

The above argument, which leads to (24) by using the trace identity, is much simpler than those in Ref. 30 where we directly used the constrained variational calculus (CVC) technique to draw the same conclusion.

### C. TA hierarchy[15]

Let $G$, $\{E_i\}$, $\{\bar{E}_i(n)\}$, deg, and $\pi$ be the same as in the TC hierarchy. We set

$$U=\begin{pmatrix}0 & 1\\ q+r\lambda^{-1}+\lambda & 0\end{pmatrix}$$
$$=2e_+(1)+q(e_+(0)-e_-(0))$$
$$+r(e_+(-1)-e_-(-1))\,,$$

or, by the notation of Sec. III,

$$R=2e_+(1),\quad e_1=e_+(0)-e_-(0),$$
$$e_2=e_+(-1)-e_-(-1)\,,$$
$$u_1=q(x,t),\quad u_2=r(x,t)\,.$$

The ranks are as follows:

$$\operatorname{rank}(q)=2,\quad\operatorname{rank}(r)=4,$$
$$\operatorname{rank}(\lambda)=2,\quad\operatorname{rank}(\partial)=1\,.$$

Let the solution of (3) be represented by (17) and (21). We have

$$a_{mx}=-qb_{m-1}-rb_{m-2}-2c_m-qc_{m-1}-rc_{m-2}\,,$$
$$b_{mx}=qa_m+ra_{m-1}\,,$$
$$c_{mx}=-qa_m-2a_{m+1}-ra_{m-1}\quad(m\geqslant 2)\,,$$

and

$$a_0=c_0=0,\quad b_0=2\beta,\quad a_1=b_1=0,\quad c_1=-\beta q\,.$$

By setting $\bar{b}_n=(b_n+c_n)/2$, we obtain from the above equations that

$$\bar{b}_{n+1}=(\tfrac{1}{4}\partial^2-q+\tfrac{1}{2}\partial^{-1}q_x)\bar{b}_n$$
$$+(-r+\tfrac{1}{2}\partial^{-1}r_x)\bar{b}_{n-1}\,,$$

with $\bar{b}_0=\beta$, $\bar{b}_1=-(\beta/2)q$. Note that we have

$$\operatorname{rank}(a_m)=2m-1,$$
$$\operatorname{rank}(b_m)=\operatorname{rank}(c_m)=\operatorname{rank}(\bar{b}_m)=2m\,,$$

and thus $\operatorname{rank}(V)=-1$.

The hierarchy (2) reads

$$\begin{pmatrix}q\\ r\end{pmatrix}_t=\begin{pmatrix}-2\bar{b}_{nx}\\ 2r\bar{b}_{n-1,x}+r_x\bar{b}_{n-1}\end{pmatrix}, \tag{25}$$

which can be derived from (2) by taking

$$V^{(n)} = (\lambda^n V)_+ + \Delta_n ,$$

with

$$\Delta_n = -2\bar{b}_n e_+(0) + r\bar{b}_{n-1}(e_+(-1) - e_-(-1)) .$$

To write (25) in its Hamiltonian form we use the trace identity and obtain

$$\begin{pmatrix}\delta/\delta q\\ \delta/\delta r\end{pmatrix}\frac{(b+c)(-r\lambda^{-3}+\lambda^{-1})}{2}$$
$$= \begin{pmatrix}\lambda^{-\gamma}\dfrac{\partial}{\partial\lambda}\lambda^{\gamma}\end{pmatrix}\frac{b+c}{2}\begin{pmatrix}\lambda^{-1}\\ \lambda^{-2}\end{pmatrix}.$$

By applying the operator $\Gamma_n$ to both sides of the above equation we have

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)(-r\bar{b}_{n-3}+\bar{b}_{n-1}) = (\gamma-n+1)(\bar{b}_{n-2},\bar{b}_{n-3}) .$$

By setting $n=2$ in the equation we find that $\gamma=\frac{1}{2}$, and consequently we establish the following equation:

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)H_n = (\bar{b}_n,\bar{b}_{n-1}) , \tag{26}$$

with

$$H_n = (2/(2n+1))(r\bar{b}_{n-1} - \bar{b}_{n+1}) .$$

Equation (26) suggests that we find the operators $J$ and $L$ such that

$$\begin{pmatrix}-2\bar{b}_{nx}\\ 2r\bar{b}_{n-1,x} + r_x\bar{b}_{n-1}\end{pmatrix} = J\begin{pmatrix}\bar{b}_n\\ \bar{b}_{n-1}\end{pmatrix},$$
$$L\begin{pmatrix}\bar{b}_n\\ \bar{b}_{n-1}\end{pmatrix} = \begin{pmatrix}\bar{b}_{n+1}\\ \bar{b}_n\end{pmatrix}.$$

The above equations together with the conditions (12) lead us to the following expressions:

$$J = \begin{pmatrix}-2\partial & 0\\ 0 & r_x + 2r\partial\end{pmatrix},$$
$$L = \begin{pmatrix}\frac{1}{4}\partial^2 - q + \frac{1}{2}\partial^{-1}q_x & -r + \frac{1}{2}\partial^{-1}r_x\\ 1 & 0\end{pmatrix}.$$

The hierarchy, therefore, takes the form

$$u_t = JL^n\begin{pmatrix}\beta\\ 0\end{pmatrix} = J\begin{pmatrix}\bar{b}_n\\ \bar{b}_{n-1}\end{pmatrix} = J\frac{\delta H_n}{\delta u},$$

where $u=(q,r)^T$ and $(\beta,0)=(\bar{b}_0,0)$. A representative member of the hierarchy is

$$q_t = (q_{xx} - 3q^2 + 4r)_x, \quad r_t = -2(r_x q + 2rq_x) ,$$

which corresponds to the case where $n=2, \beta=4$.

The TA hierarchy reduces to the KdV hierarchy when $r=0$, and there are several generalizations; see, e.g., Ref. 18.

### D. Ablowitz–Kaup–Newell–Segur (AKNS) hierarchy[3,4]

This is a well-known hierarchy. The underlying Lie algebra is $G=A_1$ with the basis $h=2\bar{h}$, $e$, and $f$. The basis for $\tilde{G}$ is $\{h(n),e(n),f(n)\,|\,n\in\mathbb{Z}\}$, where $x(n)=x\otimes\lambda^n$. The gradation is $\deg x(n)=n$. The isospectral problem is (1) with

$$U = \begin{pmatrix}-\lambda & q\\ r & \lambda\end{pmatrix} = -h(1) + qe + rf .$$

In other words we set

$$R = -h(1), \quad e_1 = e, \quad e_2 = f, \quad u_1 = q, \quad u_2 = r .$$

The ranks are

$$\operatorname{rank}(q) = \operatorname{rank}(r) = \operatorname{rank}(\lambda) = \operatorname{rank}(\partial) = 1 .$$

The solution of (3) is

$$V = \sum_{m\geqslant 0} V_m\lambda^{-m}$$
$$= \sum_{m\geqslant 0}(a_m h(-m) + b_m e(-m) + c_m f(-m))$$
$$= ah + be + cf, \tag{27}$$

for which it holds that

$$a_{mx} = qc_m - rb_m ,$$
$$b_{mx} = -2b_{m+1} - 2a_m q ,$$
$$c_{mx} = 2c_{m+1} + 2a_m r, \quad m\geqslant 0 ,$$

with

$$a_0 = -\beta, \quad b_0 = c_0 = 0 .$$

We note that

$$\operatorname{rank}(a_m) = \operatorname{rank}(b_m) = \operatorname{rank}(c_m) = m .$$

The hierarchy reads

$$\begin{pmatrix}q\\ r\end{pmatrix}_t = \begin{pmatrix}-2b_{n+1}\\ 2c_{n+1}\end{pmatrix}, \tag{28}$$

which can be derived from (2) by taking $\Delta_n = 0$ and thus $V^{(n)} = (\lambda^n V)_+$, where the index for the positive part is $\pi = 0$.

The trace identity yields, in this case, the following equation:

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)H_n = (c_{n+1},b_{n+1}) ,$$

Applying the operator $\Gamma_{n+2}$ to both sides we obtain

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)H_n = (c_{n+1},b_{n+1}) ,$$

with

$$H_n = (2/(n+1))a_{n+2} .$$

Therefore, the hierarchy (28) takes the form

$$u_t = \begin{pmatrix}q\\ r\end{pmatrix}_t = JL^n\begin{pmatrix}\beta r\\ \beta q\end{pmatrix} = J\begin{pmatrix}c_{n+1}\\ b_{n+1}\end{pmatrix} = J\frac{\delta H_n}{\delta u},$$

where $(\beta r,\beta q)=(c_1,b_1)$, and the pair of operators $J$ and $L$ are defined by

$$J = \begin{pmatrix}0 & -2\\ 2 & 0\end{pmatrix}, \qquad L = \begin{pmatrix}\frac{1}{2}\partial - r\partial^{-1}q & r\partial^{-1}r\\ -q\partial^{-1}q & -\frac{1}{2}\partial + q\partial^{-1}r\end{pmatrix}.$$

### E. N-AKNS hierarchy (see, e.g., Ref. 20)

This is a straightforward generalization of the AKNS hierarchy. We set $G=A_{N-1}$ and use the gradation (5) for $\tilde{G}$. For the isospectral problem (1) we set

$$U = \begin{pmatrix}\alpha_1\lambda & q_{12} & \cdots & q_{1N}\\ q_{21} & \alpha_2\lambda & \cdots & q_{2N}\\ \cdots & \cdots & \cdots & \cdots\\ q_{N1} & q_{N2} & \cdots & \alpha_N\lambda\end{pmatrix} = \lambda A + Q ,$$

where

$$A=\sum_i \alpha_i E_{ii},\quad Q=\sum_{i\neq j} q_{ij}E_{ij},\quad E_{ij}=(\delta_{ki}\delta_{lj})\ .$$

In this case we have $\alpha=\deg R=\deg(\lambda A)=1$ and rank $(q_{ij})=\operatorname{rank}(\partial)=\operatorname{rank}(\lambda)=1$. The solution $V$ of (3) is

$$V=\sum_{m\geq 0} V_m\lambda^{-m},\quad V_m=\sum_{ij} V_m^{(ij)}E_{ij}\ .$$

For matrices $K=\Sigma K_{ij}E_{ij}\in A_{N-1}$ we set

$$K_D=\sum_i K_{ii}E_{ii}\in \ker \operatorname{ad} R,\quad K_F=\sum_{i\neq j} K_{ij}E_{ij}\in \operatorname{Im}\operatorname{ad} R\ .$$

The hierarchy is

$$Q_t=[A,V_{n+1\,F}]\ ,$$

which is derived from (2) in the same way as in the AKNS case by setting $\Delta_n=0$ and $\pi=0$.

The trace identity yields

$$\frac{\delta}{\delta Q}\left(\sum_i V^{ii}\alpha_i\right)=\frac{\partial}{\partial\lambda}V_F^T,\quad \frac{\delta}{\delta Q}=\frac{\delta}{\delta q_{ij}}\ ,$$

from which we obtain

$$\frac{\delta H_n}{\delta Q}=V_{n+1\,F}^T\ ,$$

$$H_N=-\frac{1}{n+1}\sum_i V_{n+2}^{ii}\alpha_i\ .$$

Therefore the hierarchy is

$$Q_t=JL^n(\beta Q^T)=JV_{n+1\,F}^T=J\frac{\delta H_n}{\delta Q}\ ,$$

where $Q^T=V_{1F}^T$ and

$$JQ=[A,Q^T]\ ,$$

$$LQ=-(\operatorname{ad}A)^{-1}(Q_x+[P^T,Q]_F$$
$$-[P^T,I[P^T,Q]_D])\ .$$

Note that the conditions (12) hold again in this case. The conjugation $*$ is defined by taking the inner product

$$(A,B)=\int \operatorname{tr}(A^TB)dx=\int\sum_{ij}A_{ij}B_{ij}\,dx\ .$$

### F. Kaup–Newell (KN) hierarchy[34]

Let $G$, $\{E_i\}$, $\{\overline{E}_i(n)\}$, and deg be the same as in the AKNS hierarchy and set $\pi=1$. The isospectral problem is (1) with

$$U=\begin{pmatrix}-i\lambda^2 & \lambda q\\ \lambda r & i\lambda^2\end{pmatrix}$$
$$=-ih(2)+qe(1)+rf(1)=\lambda^2A+\lambda Q,$$

where $A=-ih$, $Q=qe+rf$. Thus we have

$$R=-ih(2),\quad e_1=e(1),\quad e_2=f(1),\quad u_1=q,\quad u_2=r,$$

and the ranks are

$$\operatorname{rank}(\lambda)=\operatorname{rank}(q)=\operatorname{rank}(r)=1,\quad \operatorname{rank}(\partial)=2\ .$$

Let the solution of (3) be defined by (27). The hierarchy is

$$Q_t=V_{2m+1\,Fx}$$

which can be derived from (2) by taking $n=2m+1$, $\Delta_{2m+2}=0$, and thus $V^{(n)}=(\lambda^{2m+2}V)_+$.

By the trace identity (11) we have

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)(-4ia\lambda+rb+qc)=\frac{\partial}{\partial\lambda}(c\lambda,b\lambda)\ .$$

Applying $\Gamma_{2m+1}$ to both sides we obtain

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)H_m=(c_{2m+1},b_{2m+1})$$

with

$$H_m=(4ia_{2m+2}-rb_{2m+1}-qc_{2m+1})/2m\ .$$

Therefore the hierarchy is

$$Q_t=JV_{2m+1\,F}^Y=JL^m(\beta Q_T)=J\frac{\delta H_m}{\delta Q}\ ,$$

where $\beta Q=V_{1F}$, and (see, e.g., Ref. 35)

$$JP=\partial P^T\ ,$$

$$LP=(-[A,P_x]+\{Q^T,\partial^{-1}\{Q^T,P_x\}\})/4$$
$$(\{A,B\}=AB+BA)\ .$$

### G. Wadati–Konno–Ichikawa (WKI) hierarchy[36,37]

Let $G$, $\{E_i\}$, $\{\overline{E}_i(n)\}$, deg, and $\pi$ be the same as in the KN hierarchy. The element $U$ is

$$U=\begin{pmatrix}\lambda & \lambda q\\ \lambda r & -\lambda\end{pmatrix}=h(1)+qe(1)+rf(1)=\lambda(A+Q)\ .$$

This choice amounts to taking

$$R=h(1),\quad e_1=e(1),\quad e_2=f(1),\quad u_1=q,\quad u_2=r\ .$$

The ranks are

$$\operatorname{rank}(\lambda)=\operatorname{rank}(\partial)=\operatorname{rank}(q)=\operatorname{rank}(r)=0\ .$$

Note that this is the exceptional case we mentioned in Sec. III, where condition (iii) is weakened to (8).

Let the solution $V$ of (3) be defined again by (27). The trace identity (11) leads to

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)(2a+qc+rb)=\frac{\partial}{\partial\lambda}(\lambda c,\lambda b)\ ,$$

or, equivalently,

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)H_n=(c_n,b_n)\ ,$$

with

$$H_n=-(2a_n+qc_n+rb_n)/(n-1)\ .$$

Note that the above Hamiltonian is not defined when $n=1$. To find $H_1$ we calculate as follows. We have

$$a_0=\frac{1}{(2\sqrt{1+qr})},\quad b_0=\bar{q}=\frac{q}{(2\sqrt{1+qr})},$$

$$c_0=\bar{r}=\frac{r}{(2\sqrt{1+qr})}$$

and

$$a_1=\frac{qr_x-rq_x}{8(\sqrt{1+qr})^3},\quad b_1=\frac{q_x}{4(\sqrt{1+qr})^3},$$

$$c_1=\frac{-r_x}{4(\sqrt{1+qr})^3}\ .$$

By the standard procedure (see, e.g., Ref. 38) for solving the inverse problem of variations, we find

$$H_1 = ((rq_x - qr_x)/(4qr))(1 - 1/(\sqrt{1+qr}))\,.$$

Substituting into (2)

$$V^{(n)} = (\lambda^{n+1}V)_+ + \Delta_n, \quad \Delta_n = \lambda(\bar{V}_n - V_n)\,,$$

where

$$\bar{V}_{nD} = 0, \quad \bar{V}_{nF} = (\mathrm{ad}\,A)^{-1}V_{n-1\,Fx}\,,$$

we obtain the hierarchy

$$u_t = \begin{pmatrix} q \\ r \end{pmatrix}_t = \frac{1}{2}\begin{pmatrix} b_{nxx} \\ -c_{nxx} \end{pmatrix} = JL^n\begin{pmatrix} \beta\bar{r} \\ \beta\bar{q} \end{pmatrix}$$

$$= J\begin{pmatrix} c_n \\ b_n \end{pmatrix} = J\frac{\delta H_n}{\delta u}\,,$$

where (see Refs. 21 and 37)

$$J = \frac{1}{2}\partial^2\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$L = \begin{pmatrix} -\frac{1}{2}\partial + \bar{r}\,\partial^{-1}\bar{q}\,\partial^2 & -\bar{r}\,\partial^{-1}r\,\partial^2 \\ \bar{q}\,\partial^{-1}\bar{q}\,\partial^2 & \frac{1}{2}\partial - \bar{q}\,\partial^{-1}\bar{r}\,\partial^2 \end{pmatrix}.$$

### H. Boite–Pempinelli–Tu (BPT) hierarchy[21]

Let $G$ be $A_1$. The base is taken to be

$$\bar{h}, \quad e^+ = (e+f)/2, \quad e^- = (e-f)/2\,.$$

The base for $\tilde{G}$ is

$$\{\bar{h}(n), e^+(n), e^-(n) \mid n \in \mathbf{Z}\}$$

with $x(n) = x \otimes \lambda^n$. The gradation is again defined by (5) and $\pi = 0$. For the isospectral problem we set

$$U = \begin{pmatrix} \lambda + s/(2\lambda) & (q + r\lambda^{-1})/2 \\ (q - r\lambda^{-1})/2 & -\lambda - s/(2\lambda) \end{pmatrix}$$

$$= 2\bar{h}(1) + qe^+(0) + re^-(-1) + s\bar{h}(-1)\,.$$

The ranks are

$$\mathrm{rank}(\lambda) = \mathrm{rank}(\partial) = \mathrm{rank}(q) = 1,$$

$$\mathrm{rank}(r) = \mathrm{rank}(s) = 2\,.$$

The solution of (3) is represented as

$$V = a\bar{h} + be^+ + ce^- = \sum_{m>0}(a_m\bar{h}(-m) + b_m e^+(-m) + c_m e^-(-m))\,.$$

We have

$$a_{nx} = -qc_n + rb_{n-1}\,,$$

$$b_{nx} = 2c_{n+1} + sc_{n-1} - ra_{n-1}\,,$$

$$c_{nx} = 2b_{n+1} + sb_{n-1} - qa_n, \quad n \geqslant 1\,,$$

and

$$a_0 = 2\beta, \quad b_0 = c_0 = 0, \quad a_{2k+1} = b_{2k} = c_{2k+1} = 0\,.$$

Then

$$\mathrm{rank}(a_m) = \mathrm{rank}(b_m) = \mathrm{rank}(c_m) = m, \quad \mathrm{rank}(V) = 0\,.$$

The hierarchy is

$$\begin{pmatrix} q \\ r \\ s \end{pmatrix}_t = \begin{pmatrix} 2c_{2m} \\ -sb_{2m-1} \\ -rb_{2m-1} \end{pmatrix}, \tag{29}$$

which corresponds to the choice

$$n = 2m+1, \quad V^{(n)} = (\lambda^n V)_+, \quad \Delta_n = 0\,. \tag{30}$$

By the trace identity we find

$$\left(\frac{\delta}{\delta q}, \frac{\delta}{\delta r}, \frac{\delta}{\delta s}\right)((1 - (s/2)\lambda^{-2})a - (r/2)\lambda^{-2}c)$$

$$= \frac{\partial}{\partial\lambda}\left(\frac{b}{2}, -\frac{c}{2\lambda}, \frac{a}{2\lambda}\right) \quad (\gamma = 0)\,,$$

from which we obtain

$$\left(\frac{\delta}{\delta q}, \frac{\delta}{\delta r}, \frac{\delta}{\delta s}\right)H_m = (b_{2m+1}, -c_{2m}, a_{2m})\,,$$

where $H_m = (sa_{2m} - rc_{2m} - 2a_{2m+2})/(2m+1)$. The hierarchy (29), therefore, takes the following Hamiltonian form:

$$u_t = \begin{pmatrix} q \\ r \\ s \end{pmatrix}_t = JL^m\begin{pmatrix} \beta q \\ 0 \\ 2\beta \end{pmatrix} = J\begin{pmatrix} b_{2m+1} \\ -c_{2m} \\ a_{2m} \end{pmatrix} = J\frac{\delta H_m}{\delta u}\,,$$

where

$$J = \begin{pmatrix} 0 & -2 & 0 \\ 2 & \partial & -q \\ 0 & q & -\partial \end{pmatrix},$$

$$L = \frac{1}{4}\begin{pmatrix} \partial^2 - 2s - q^2 + 2q\,\partial^{-1}r + q\,\partial^{-1}q_x & \partial s - q\,\partial^{-1}(qs) & \partial r - q\,\partial^{-1}(qr) \\ -2\partial & -2s & -2r \\ -2q + \partial^{-1}(2q_x + 4r) & -2\,\partial^{-1}(qs) & -2\,\partial^{-1}(qr) \end{pmatrix}.$$

### I. TB hierarchy[19]

Let $G$, $\{E_i\}$, $\{\bar{E}_i(n)\}$, deg, and $\pi$ be the same as in the BPT hierarchy. We take

$$U = \begin{pmatrix} \lambda + \epsilon r/2\lambda & (q + r\lambda^{-1})/2 \\ (q - r\lambda^{-1})/2 & -\lambda - \epsilon r/2\lambda \end{pmatrix} = 2\bar{h}(1) + qe^+(0) + r(e^-(-1) + \epsilon\bar{h}(-1)), \quad \epsilon = \pm 1\,.$$

The ranks for $q, r, \lambda$, and $\partial$ and the form of $V$ are the same as in the BPT hierarchy. Again we have (30) with respect to the derivation of the hierarchy.

The trace identity now reads

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)\left(\left(1-\frac{1}{2}\epsilon r\lambda^{-2}\right)a+\frac{r\lambda^{-1}c}{2}\right)=\frac{\partial}{\partial\lambda}\left(\frac{b}{2},-\frac{1}{2}c\lambda^{-1}+\frac{1}{2}\epsilon a\lambda^{-1}\right)\quad(\gamma=0)\,,$$

or $H_m=(\epsilon r a_{2m}-r c_{2m}-2a_{2m+2})/(2m+1)$ and

$$\left(\frac{\delta}{\delta q},\frac{\delta}{\delta r}\right)H_m=(b_{2m+1},\epsilon a_{2m}-c_{2m})\,.$$

The operator $L$ mapping $\delta H_m/\delta u$ to $\delta H_{m+1}/\delta u$ is

$$L=\frac{1}{4}\begin{pmatrix}-2\epsilon r-q^2+q\,\partial^{-1}(2r+q_x)+\partial^2 & \epsilon\,\partial r-\epsilon q\,\partial^{-1}qr\\ \epsilon\,\partial^{-1}(4r+2q_x)-2\epsilon q-2\,\partial & -2\,\partial^{-1}qr-2\epsilon r\end{pmatrix}$$

and the operator $J$ is

$$J=\begin{pmatrix}\partial & \epsilon r\\ -\epsilon r & 0\end{pmatrix}.$$

The hierarchy is therefore the following:

$$u_t=\begin{pmatrix}q\\ r\end{pmatrix}_t=\begin{pmatrix}2c_{2m+2}\\ -\epsilon r b_{2m+1}\end{pmatrix}=JL^m\begin{pmatrix}\beta q\\ 2\epsilon\beta\end{pmatrix}=J\begin{pmatrix}b_{2m+1}\\ \epsilon a_{2m}-c_{2m}\end{pmatrix}=J\frac{\delta H_m}{\delta u}\,.$$

Note that the TB hierarchy is a reduction of the BPT hierarchy; however, the Hamiltonian structure undergoes nontrivial change.

## V. CONCLUDING REMARKS

We have shown that the trace identity does provide us with a powerful tool to transform the hierarchy of integrable systems to its Hamiltonian form. Furthermore, the operator $L$ maps $\delta H_n/\delta u$ to $\delta H_{n+1}/\delta u$, while $\delta H_n/\delta u$ can be determined by means of the trace identity; therefore the present approach suggests the possibility of analyzing the algebraic structure of hereditary and symplectic operators and to construct more new ones starting from the model isospectral problem presented in Sec. III. Progress has been made toward this goal[27]; we leave the further study in this direction to later papers.

However, one open problem remains. We need to fix the constant $\gamma$ each time. Can one obtain an explicit formula for it? It seems that the constant $\gamma$ is closely related to the asymptotic behavior of the eigenfunctions of the corresponding spectral problem.


## ACKNOWLEDGMENT

This paper was supported by the National Science Foundation.



[1]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, London, 1980).
[2]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).
[3]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, PA, 1981).
[4]A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, PA, 1985).
[5]L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, Berlin, 1987).
[6]M. D. Kruskal, R. M. Miura, and C. S. Gardner, J. Math. Phys. **11**, 952 (1970).
[7]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[8]I. M. Gel'fand and I. Y. Dorfman, Funk. Anal. Appl. **13**, 13 (1979) (in Russian).
[9]B. Fuchssteiner, Nonlinear Anal. Th. Method. Appl. **3**, 849 (1979).
[10]H. H. Chen, Y. C. Lee, and C. S. Liu, Phys. Scr. **20**, 490 (1979).
[11]P. J. Olver, Math. Proc. Cambridge Philos. Soc. **88**, 71 (1980).
[12]B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[13]G. Wilson, Ergod. Th. Dynam. Sys. **1**, 361 (1981).
[14]T. M. Alberty, T. Koikawa, and R. Sasaki, Physica D **5**, 43 (1982).
[15]G. Z. Tu, Nuovo Cimento B **73**, 15 (1983).
[16]G. Z. Tu, Sci. Exploration **2**, 85 (1982).
[17]M. Boiti and G. Z. Tu, Nuovo Cimento B **75**, 145 (1983).
[18]M. Boiti, C. Laddomada, F. Pempinelli, and G. Z. Tu, J. Math. Phys. **24** 2035 (1983).
[19]G. Z. Tu, Phys. Lett. A **94**, 340 (1983).
[20]G. Z. Tu, Arab. J. Sci. Eng. **9**, 187 (1984).
[21]M. Boiti, F. Pempinelli, and G. Z. Tu, Nuovo Cimento B **79**, 231 (1984).
[22]G. Z. Tu, J. Eng. Math. **1**, 7 (1984) (in Chinese).
[23]G. Z. Tu, in *Dynamical Problems in Soliton Systems*, edited by S. Takeno (Springer, Berlin, 1985), Vol. 23.
[24]A. S. Fokas and P. M. Santini, Stud. Appl. Math. **75**, 179 (1986).
[25]G. Z. Tu, Sci. Sin. A **29**, 138 (1986).
[26]A. Roy Chowdhury and S. Roy, J. Math. Phys. **27**, 2464 (1986).
[27]G. Z. Tu, Commun. Appl. Math. Comp. **1**, 80 (1987).
[28]V. G. Drinfeld and V. V. Sokolov, "Lie algebras and equations of KdV type," Moscow University preprint (in Russian).
[29]G. Z. Tu, Adv. Sci. China Ser. Math. **2**, 45 (1989).
[30]G. Z. Tu, Sci. Sin. A **31**, 28 (1988).
[31]G. Z. Tu and D. Z. Men, "The trace identity, a powerful tool for constructing the Hamiltonian structure of integrable systems. II," Acta Math. Appl. Sin. (to appear).
[32]G. Z. Tu, "On Liouville integrability of zero curvature equations and the Yang hierarchy," J. Phys. A (to appear).
[33]V. G. Kac, *Infinite Dimensional Lie Algebras* (Birkhauser, Boston, 1983).
[34]D. J. Kaup, and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
[35]M. Boiti, C. Laddomada, F. Pempinelli, and G. Z. Tu, Physica D **9**, 425 (1983).
[36]M. Wadati, K. Konno, and Y. H. Ichikawa, J. Phys. Soc. Jpn. **47**, 1698 (1979).
[37]M. Boiti, F. Pempinelli, and G. Z. Tu, Prog. Theor. Phys. **69**, 48 (1983).
[38]G. Z. Tu, J. Math. Anal. Appl. **94**, 348 (1983).

# Displacement theorems for spherical solutions of the linear Navier–Stokes equations

B. U. Felderhof
*Institut für Theoretische Physik A, R.W.T.H. Aachen, Templergraben 55, 5100 Aachen, West Germany*

R. B. Jones
*Department of Physics, University of London, Queen Mary College, Mile End Road, London E1 4NS, England*

(Received 30 August 1988; accepted for publication 12 October 1988)

Displacement theorems are presented for the solutions in spherical coordinates of the linear Navier–Stokes equations for time-independent flow in an incompressible viscous fluid. The theorems allow one to expand spherical solutions about a chosen center into spherical solutions centered elsewhere.

## I. INTRODUCTION

The solutions in spherical coordinates of the linear Navier–Stokes equations for steady flow in an incompressible viscous fluid were found long ago by Lamb.[1] The solutions are described in detail by Happel and Brenner.[2] In problems of hydrodynamic interactions between spherical particles, it is necessary to expand spherical solutions centered about one particle into solutions centered about a different particle. For a special choice of coordinates, the desired transformation was found by Schmitz and Felderhof.[3] Here we present the general transformation for an arbitrary direction of the vector connecting the two centers. Regarded as a function of this vector, most of the transformation coefficients satisfy Laplace's equation. Our derivation is based on addition theorems for spherical wave solutions of the vector Helmholtz equation, as presented by Felderhof and Jones.[4] The derivation is straightforward but lengthy and we present only the final results.

## II. SOLUTIONS IN SPHERICAL COORDINATES

In order to establish notation, we recall here the solutions in spherical coordinates of the linear Navier–Stokes equations. The equations read

$$\eta\nabla^2\mathbf{v} - \nabla p = 0, \quad \nabla\cdot\mathbf{v} = 0, \tag{2.1}$$

where $\eta$ is the shear viscosity, $\mathbf{v}(\mathbf{r})$ is the flow velocity, and $p(\mathbf{r})$ is the pressure, which is determined by the condition of incompressibility. A complete set of solutions of (2.1) in spherical coordinates was first presented by Lamb.[1] Here we employ the notation of Cichocki *et al.*[5] The set of solutions appropriate to spherical symmetry is given by

$$\begin{aligned}
&\mathbf{v}^+_{lm0}(\mathbf{r}) = r^{l-1}\hat{\mathbf{A}}_{lm}(\hat{\mathbf{r}}), \quad p^+_{lm0}(\mathbf{r}) = 0,\\
&\mathbf{v}^+_{lm1}(\mathbf{r}) = ir^l\hat{\mathbf{C}}_{lm}(\hat{\mathbf{r}}), \quad p^+_{lm1}(\mathbf{r}) = 0,\\
&\mathbf{v}^+_{lm2}(\mathbf{r}) = r^{l+1}[\,[(l+1)(2l+3)/2l\,]\hat{\mathbf{A}}_{lm}(\hat{\mathbf{r}}) + \hat{\mathbf{B}}_{lm}(\hat{\mathbf{r}})\,],\\
&p^+_{lm2}(\mathbf{r}) = \eta[(l+1)(2l+1)(2l+3)/l\,]r^l\hat{Y}_{lm}(\hat{\mathbf{r}}).
\end{aligned} \tag{2.2}$$

Here $\hat{Y}_{lm}(\mathbf{r})$ is an unnormalized spherical harmonic related to the usual $Y_{lm}(\hat{\mathbf{r}})$, in the notation of Edmonds,[6] by

$$\hat{Y}_{lm}(\hat{\mathbf{r}}) = n_{lm}Y_{lm}(\hat{\mathbf{r}}) = (-1)^m P_l^m(\cos\theta)e^{im\varphi}, \tag{2.3}$$

with the normalization coefficient

$$n_{lm} = [[4\pi/(2l+1)]\,[(l+m)!/(l-m)!]]^{1/2}. \tag{2.4}$$

The vector spherical harmonics appearing in (2.2) are defined by

$$\begin{aligned}
\hat{\mathbf{A}}_{lm} &= l\hat{Y}_{lm}\mathbf{e}_r + \frac{\partial\hat{Y}_{lm}}{\partial\theta}\mathbf{e}_\theta + \frac{1}{\sin\theta}\frac{\partial\hat{Y}_{lm}}{\partial\varphi}\mathbf{e}_\varphi,\\
\hat{\mathbf{B}}_{lm} &= -(l+1)\hat{Y}_{lm}\mathbf{e}_r + \frac{\partial\hat{Y}_{lm}}{\partial\theta}\mathbf{e}_\theta + \frac{1}{\sin\theta}\frac{\partial\hat{Y}_{lm}}{\partial\varphi}\mathbf{e}_\varphi,\\
\hat{\mathbf{C}}_{lm} &= \frac{1}{\sin\theta}\frac{\partial\hat{Y}_{lm}}{\partial\varphi}\mathbf{e}_\theta - \frac{\partial\hat{Y}_{lm}}{\partial\theta}\mathbf{e}_\theta,
\end{aligned} \tag{2.5}$$

where $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_\varphi$ are unit vectors in spherical coordinates. The above vector spherical harmonics are related to the $\mathbf{Y}_{JLM}$ defined by Edmonds[6] by

$$\begin{aligned}
\hat{\mathbf{A}}_{lm} &= n_{lm}\sqrt{l(2l+1)}\,\mathbf{Y}_{l\,l-1\,m},\\
\hat{\mathbf{B}}_{lm} &= n_{lm}\sqrt{(l+1)(2l+1)}\,\mathbf{Y}_{l\,l+1\,m},\\
\hat{\mathbf{C}}_{lm} &= -in_{lm}\sqrt{l(l+1)}\,\mathbf{Y}_{l\,lm}.
\end{aligned} \tag{2.6}$$

We shall denote the set of solutions (2.2) by $\{\mathbf{v}^+_{lm\sigma}(\mathbf{r})\}$, where the angular quantum number $l$ takes the values $l = 1,2,\ldots$, the azimuthal quantum number $m$ takes the values $-l,\ldots,l$, and the subscript $\sigma$ takes the values 0,1,2.

It is convenient to define an adjoint set of functions $\{\mathbf{w}^+_{lm\sigma}(r)\}$. We require this set of functions to be orthonormal to the set $\{\mathbf{v}^+_{lm\sigma}\}$ on a sphere of arbitrary radius $b$. It is easily seen that such a set of functions is given by

$$\begin{aligned}
\mathbf{w}^+_{lm0}(\mathbf{r}) &= [\,l(2l+1)n_{lm}^2\,]^{-1}r^{-l}\\
&\quad\times[\hat{\mathbf{A}}_{lm} - \tfrac{1}{2}(2l+3)\hat{\mathbf{B}}_{lm}\,],\\
\mathbf{w}^+_{lm1}(\mathbf{r}) &= i[\,l(l+1)n_{lm}^2\,]^{-1}r^{-l-1}\hat{\mathbf{C}}_{lm},\\
\mathbf{w}^+_{lm2}(\mathbf{r}) &= [\,(l+1)(2l+1)n_{lm}^2\,]^{-1}r^{-l-2}\hat{\mathbf{B}}_{lm}.
\end{aligned} \tag{2.7}$$

These functions satisfy the orthonormality relations

$$\langle\mathbf{w}^+_{lm\sigma}\delta_b|\mathbf{v}^+_{l'm'\sigma'}\rangle = \delta_{ll'}\delta_{mm'}\delta_{\sigma\sigma'}, \tag{2.8}$$

where we have introduced the scalar product

$$\langle\mathbf{f}|\mathbf{g}\rangle = \int\mathbf{f}^*(\mathbf{r})\cdot\mathbf{g}(\mathbf{r})\,d\mathbf{r} \tag{2.9}$$

and the abbreviation

$$\delta_b = (1/b)\delta(r-b). \tag{2.10}$$

 0022-2488/89/020339-04$02.50 © 1989 American Institute of Physics 

We also consider a set of solutions $\{\mathbf{v}^-_{lm\sigma}(\mathbf{r})\}$ that tend to zero at infinity and satisfy the flow equations (2.1) everywhere, except at the origin. This set of solutions is given by

$$\mathbf{v}^-_{lm0}(\mathbf{r}) = \frac{1}{(2l+1)^2} r^{-l}\left[\frac{l+1}{l(2l-1)}\hat{\mathbf{A}}_{lm} - \frac{1}{2}\hat{\mathbf{B}}_{lm}\right],$$

$$p^-_{lm0}(\mathbf{r}) = \eta \frac{1}{2l+1} r^{-l-1}\hat{Y}_{lm},$$

$$\mathbf{v}^-_{lm1}(\mathbf{r}) = \frac{i}{l(l+1)(2l+1)} r^{-l-1}\hat{\mathbf{C}}_{lm}, \quad p^-_{lm1}(\mathbf{r}) = 0,$$

$$\mathbf{v}^-_{lm2}(\mathbf{r}) = \frac{l}{(l+1)(2l+1)^2(2l+3)} r^{-l-2}\hat{\mathbf{B}}_{lm},$$

$$p^-_{lm2}(\mathbf{r}) = 0. \tag{2.11}$$

We define again a set of adjoint functions $\{\mathbf{w}^-_{lm\sigma}(\mathbf{r})\}$ that satisfy the orthonormality relations

$$\langle \mathbf{w}^-_{lm\sigma}\delta_b | \mathbf{v}^-_{l'm'\sigma'}\rangle = \delta_{ll'}\delta_{mm'}\delta_{\sigma\sigma'}, \tag{2.12}$$

for any radius $b$. It is easily seen that such a set of functions is given by

$$\mathbf{w}^-_{lm0}(\mathbf{r}) = \frac{(2l-1)(2l+1)}{(l+1)n^2_{lm}} r^{l-1}\hat{\mathbf{A}}_{lm},$$

$$\mathbf{w}^-_{lm1}(\mathbf{r}) = i\frac{2l+1}{n^2_{lm}} r^{l}\hat{\mathbf{C}}_{lm}, \tag{2.13}$$

$$\mathbf{w}^-_{lm2}(\mathbf{r}) = \frac{(2l+1)(2l+3)}{ln^2_{lm}} r^{l+1}\left[\frac{2l-1}{2}\hat{\mathbf{A}}_{lm} + \hat{\mathbf{B}}_{lm}\right].$$

The solutions presented in this section are useful in the problem of flow about a single spherical particle. In treating the problem of hydrodynamic interactions between spherical particles, one wishes to expand the solutions defined relative to a chosen center in terms of solutions centered elsewhere. We address this problem in the next section.

## III. DISPLACEMENT THEOREMS

In this section we present the desired displacement theorems. We consider three vectors $\mathbf{r}$, $\mathbf{r}_>$, and $\mathbf{r}_<$ related by

$$\mathbf{r} = \mathbf{r}_> + \mathbf{r}_<, \tag{3.1}$$

where $|\mathbf{r}_>|$ is larger than $|\mathbf{r}_<|$. Our displacement theorems for the regular solutions take the form

$$\mathbf{v}^+_{lm\sigma}(\mathbf{r}) = \sum_{l'm'\sigma'} S^{++}(\mathbf{r}_>;l'm'\sigma',lm\sigma)\mathbf{v}^+_{l'm'\sigma'}(\mathbf{r}_<) \tag{3.2a}$$

$$= \sum_{l'm'\sigma'} S^{++}(\mathbf{r}_<;l'm'\sigma',lm\sigma)\mathbf{v}^+_{l'm'\sigma'}(\mathbf{r}_>). \tag{3.2b}$$

The displacement theorems for the singular solutions read

$$\mathbf{v}^-_{lm\sigma}(\mathbf{r}) = \sum_{l'm'\sigma'} S^{+-}(\mathbf{r}_>;l'm'\sigma',lm\sigma)\mathbf{v}^+_{l'm'\sigma'}(\mathbf{r}_<) \tag{3.3a}$$

$$= \sum_{l'm'\sigma'} S^{--}(\mathbf{r}_<;l'm'\sigma',lm\sigma)\mathbf{v}^-_{l'm'\sigma'}(\mathbf{r}_>). \tag{3.3b}$$

Using the orthonormality relations (2.8) and (2.12), we may write the superposition coefficients as matrix elements of spherical functions centered about two different centers, $\mathbf{R}_1$ and $\mathbf{R}_2$. We identify $\mathbf{R}_1$ with the origin and take the solutions on the right-hand side of Eqs. (3.2) and (3.3) to be centered about $\mathbf{R}_2$. We then find from (3.2), in obvious notation,

$$S^{++}(\mathbf{R};l'm'\sigma',lm\sigma) = \langle \mathbf{w}^+_{l'm'\sigma'}(2)\delta_b(2)|\mathbf{v}^+_{lm\sigma}(1)\rangle, \tag{3.4}$$

where $\mathbf{R} = \mathbf{R}_2 - \mathbf{R}_1$. Similarly, we find from (3.3a)

$$S^{+-}(\mathbf{R};l'm'\sigma',lm\sigma) = \langle \mathbf{w}^+_{l'm'\sigma'}(2)\delta_b(2)|\mathbf{v}^-_{lm\sigma}(1)\rangle, \tag{3.5}$$

with the condition $b < R$. From (3.3b) we find

$$S^{--}(\mathbf{R};l'm'\sigma',lm\sigma) = \langle \mathbf{w}^-_{l'm'\sigma'}(2)\delta_b(2)|\mathbf{v}^-_{lm\sigma}(1)\rangle, \tag{3.6}$$

with the condition $b > R$.

We present the values for the coefficients without derivation. We comment briefly on the derivation in the next section. The coefficients (3.4) may be written in the form

$$\begin{aligned}
S^{++}(\mathbf{R};l'm'0,lm0) &= s^{++}(l'm'0,lm0)R^{l-l'}\hat{Y}_{l-l',\mu}(\hat{\mathbf{R}}),\\
S^{++}(\mathbf{R};l'm'0,lm1) &= s^{++}(l'm'0,lm1)R^{l-l'+1}\\
&\quad\times\hat{Y}_{l-l'+1,\mu}(\hat{\mathbf{R}}),\\
S^{++}(\mathbf{R};l'm'0,lm2)&\\
&= R^{l-l'+2}[s_0^{++}(l'm'0,lm2)\hat{Y}_{l-l',\mu}(\hat{\mathbf{R}})\\
&\quad + s_2^{++}(l'm'0,lm2)\hat{Y}_{l-l'+2,\mu}(\hat{\mathbf{R}})],\\
S^{++}(\mathbf{R};l'm'1,lm1) &= s^{++}(l'm'1,lm1)R^{l-l'}\hat{Y}_{l-l',\mu}(\hat{\mathbf{R}}),\\
S^{++}(\mathbf{R};l'm'1,lm2) &= s^{++}(l'm'1,lm2)R^{l-l'+1}\\
&\quad\times\hat{Y}_{l-l'+1,\mu}(\hat{\mathbf{R}}),\\
S^{++}(\mathbf{R};l'm'2,lm2) &= s^{++}(l'm'2,lm2)R^{l-l'}\hat{Y}_{l-l',\mu}(\hat{\mathbf{R}}),
\end{aligned} \tag{3.7}$$

where $\mu = m - m'$ and the scalar coefficients are given by

$$\begin{aligned}
s^{++}(l'm'0,lm0) &= \frac{l'+1}{l+1}s^{++}(l'm'1,lm1),\\
s^{++}(l'm'0,lm1) &= \frac{(l'+1)(ml'-m'l-m')}{l'(l+1)(l-l'+\mu+1)}s^{++}(l'm'1,lm1),\\
s_0^{++}(l'm'0,lm2) &= \frac{(2l+1)(2l+3)(l'+1)}{2l(2l-2l'+3)}s^{++}(l'm'1,lm1),\\
s_2^{++}(l'm'0,lm2) &= \frac{(2l+1)(l'+1)(ll'-2l+3l'-3)}{ll'(l+1)(2l'-1)(2l-2l'+3)(l-l'+\mu+1)(l-l'+\mu+2)}\\
&\quad\times[l'(l+1)(l-l'+2)-(l-l'+2)(2l-2l'+3)m'^2+2m'\mu(2l'-1)(l-l'+2)\\
&\quad-2\mu^2l'^2+l'\mu(\mu-2m')+\mu m']s^{++}(l'm'1,lm1),
\end{aligned} \tag{3.8}$$

$$s^{++}(l'm'1,lm2)=\frac{(2l+1)(2l+3)(ml'-m'l-m')}{ll'(l+1)(l-l'+\mu+1)}s^{++}(l'm'1,lm1),$$

$$s^{++}(l'm'2,lm2)=\frac{l'(2l+1)(2l+3)}{l(2l'+1)(2l'+3)}s^{++}(l'm'1,lm1),$$

with the single coefficient $s^{++}(l'm'1,lm1)$ given by

$$s^{++}(l'm'1,lm1)=\delta_{m,\mu+m'}\frac{l+1}{l'+1}\frac{(l+m)!}{(l'+m')!(l-l'+\mu)!}. \tag{3.9}$$

All coefficients not listed explicitly in (3.7) vanish. For given $l$ the number $l'$ can take only the values $1,\ldots,l$.

The coefficients (3.5) may be written in the form

$$\begin{aligned}
&S^{+-}(\mathbf{R};l'm'0,lm0)\\
&\quad=R^{-l-l'+1}[s_0^{+-}(l'm'0,lm0)\hat{Y}_{l+l',\mu}(\hat{\mathbf{R}})\\
&\qquad+s_2^{+-}(l'm'0,lm0)\hat{Y}_{l+l'-2,\mu}(\hat{\mathbf{R}})],\\
&S^{+-}(\mathbf{R};l'm'0,lm1)\\
&\quad=s^{+-}(l'm'0,lm1)R^{-l-l'}\hat{Y}_{l+l'-1,\mu}(\hat{\mathbf{R}}),\\
&S^{+-}(\mathbf{R};l'm'0,lm2)\\
&\quad=s^{+-}(l'm'0,lm2)R^{-l-l'-1}\hat{Y}_{l+l',\mu}(\hat{\mathbf{R}}),\\
&S^{+-}(\mathbf{R};l'm'1,lm0)\\
&\quad=s^{+-}(l'm'1,lm0)R^{-l-l'}\hat{Y}_{l+l'-1,\mu}(\hat{\mathbf{R}}),\\
&S^{+-}(\mathbf{R};l'm'1,lm1)\\
&\quad=s^{+-}(l'm'1,lm1)R^{-l-l'-1}\hat{Y}_{l+l',\mu}(\hat{\mathbf{R}}),\\
&S^{+-}(\mathbf{R};l'm'2,lm0)\\
&\quad=s^{+-}(l'm'2,lm0)R^{-l-l'-1}\hat{Y}_{l+l',\mu}(\hat{\mathbf{R}}),
\end{aligned} \tag{3.10}$$

where $\mu=m-m'$ and the scalar coefficients are given by

$$s_0^{+-}(l'm'0,lm0)=\frac{1}{2}\frac{(l+1)(l'+1)}{2l+2l'-1}s^{+-}(l'm'1,lm1),$$

$$\begin{aligned}
&s_2^{+-}(l'm'0,lm0)\\
&\quad=\frac{(l+1)(l'+1)(ll'-2l-2l'+1)}{ll'(2l-1)(2l'-1)(2l+2l'-1)}\\
&\qquad\times\frac{1}{(l+l'-m+m')(l+l'-m+m'-1)}\\
&\qquad\times[-ll'(l+l')+2m'^2l^2+2m^2l'^2\\
&\qquad+(4mm'+1)ll'-m'(2m+m')l\\
&\qquad-m(2m'+m)l'+mm']s^{+-}(l'm'1,lm1),
\end{aligned}$$

$$s^{+-}(l'm'0,lm1)=\frac{(m'l+ml')(l'+1)}{ll'(l+l'-m+m')}s^{+-}(l'm'1,lm1),$$

$$s^{+-}(l'm'0,lm2)=-\frac{l(l'+1)}{(2l+1)(2l+3)}s^{+-}(l'm'1,lm1),$$

$$s^{+-}(l'm'1,lm0)=-\frac{(m'l+ml')(l+1)}{ll'(l+l'-m+m')}s^{+-}(l'm'1,lm1),$$

$$s^{+-}(l'm'2,lm0)=-\frac{(l+1)l'}{(2l'+1)(2l'+3)}s^{+-}(l'm'1,lm1),$$

with the single coefficient $s^{+-}(l'm'1,lm1)$ given by

$$\begin{aligned}
&s^{+-}(l'm'1,lm1)\\
&\quad=(-1)^{l'+m'+1}\frac{1}{(l+1)(2l+1)(l'+1)}\\
&\qquad\times\frac{(l+l'-m+m')!}{(l-m)!(l'+m')!}.
\end{aligned} \tag{3.11}$$

All coefficients not listed explicitly in (3.10) vanish. The numbers $l$ and $l'$ independently take all positive integer values. For the special case where $\mathbf{R}$ is in the $z$ direction, the coefficients (3.10) reduce to those given by Schmitz and Felderhof,[3] see also Cichocki *et al.*[5]

The coefficients (3.6) are related simply to the coefficients (3.4). We find

$$\begin{aligned}
&S^{--}(l'm'\sigma',lm\sigma)\\
&\quad=(-1)^{l+l'+m+m'}[(2l'+1)/(2l+1)]\\
&\qquad\times S^{++}(l-m\sigma,l'-m'\sigma').
\end{aligned} \tag{3.13}$$

For given $l'$ the number $l$ can take only the values $1,\ldots,l'$.

## IV. DERIVATION

We comment briefly on the derivation of the above results. We start from the addition theorems for spherical wave solutions of the vector Helmholtz equation, as presented in Ref. 4. These may be used to find addition theorems for the spherical wave solutions of the time-dependent linear Navier–Stokes equations, as given in Ref. 7. These solutions are stationary in time, so that one obtains relations for each frequency. The spherical solutions of Sec. II are a linear combination of the solutions of Ref. 7 taken at zero frequency. The relation is found by a succession of two linear transformations that have been given explicitly in Sec. 8 of Ref. 7 and in Appendix B of Ref. 5. The results presented in Sec. III are obtained by applying these transformations to the addition theorems for the spherical wave solutions and by simplifying the result by use of standard properties of the Clebsch–Gordan coefficients.

[1] H. Lamb, *Hydrodynamics* (Dover, New York, 1945), pp. 595 and 632.
[2] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics* (Noordhoff, Leyden, 1973), p. 62.

[3]R. Schmitz and B. U. Felderhof, "Friction matrix for two spherical particles with hydrodynamic interaction," Physica A **113**, 103 (1982).
[4]B. U. Felderhof and R. B. Jones, "Addition theorems for spherical wave solutions of the vector Helmholtz equation," J. Math. Phys. **28**, 836 (1987).
[5]B. Cichocki, B. U. Felderhof, and R. Schmitz, "Hydrodynamic interactions between two spherical particles," PhysicoChem. Hyd. **10**, 383 (1988).
[6]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).
[7]B. U. Felderhof and R. B. Jones, "Hydrodynamic scattering theory of flow about a sphere," Physica A **136**, 77 (1986).

# Analytic solutions of a fourth-order differential equation with a second-order turning point

Hans O. Åkerstedt
*Department of Technology, Uppsala University, Box 534, S-751 21 Uppsala, Sweden*

(Received 3 May 1988; accepted for publication 5 October 1988)

Analytic solutions of a certain fourth-order differential equation with a second-order turning point are given. The solutions can be expressed in terms of products of the well-known parabolic-cylinder functions, and the equation might therefore be useful as a comparison equation in the theory of asymptotic expansions of ordinary differential equations.

## I. INTRODUCTION

The theory of asymptotic solutions of linear differential equations is a powerful tool for investigating many physical phenomena in various branches of theoretical physics. The most well-known such method is perhaps the WKB method, which is frequently used, for instance, in quantum mechanics when solving the second-order Schrödinger equation. This method, however, breaks down near certain points called turning points.

One method for obtaining asymptotic solutions valid even at the turning points has been given by Langer.[1] His idea is to transform the differential equation into one equation that is close to a simpler equation, called a comparison equation, which has the same qualitative features as the original equation. This method works when the comparison equation can be solved in terms of already well-known transcendental functions. For second-order differential equations this method is well established for equations containing turning points of any order.[2]

For singularly perturbed differential equations of fourth order with turning points, the general theory is, however, more difficult. Fourth-order differential equations are important in connection with studying the stability of nonturbulent viscous flows in hydrodynamics. The stability is then described by the so-called Orr–Sommerfeld equation.[3] For this equation, uniformly valid asymptotic solutions have been obtained for the case of one single turning point.[3]

In a recent study on the effects of gyroviscosity on the stability properties of the Z-pinch,[4] a fourth-order differential equation similar to the Orr–Sommerfeld equation appears. For this case, however, the turning point is of second order. The comparison equation for this case turns out to be a fourth-order differential equation, which can be completely solved in terms of products of the well-known parabolic-cylinder functions. Since we believe that this comparison equation could probably be applied to other problems as well, we will present the solutions of this equation here.

## II. ANALYTIC SOLUTION

We will here consider the solutions $\chi(\eta)$ of the equation

$$\chi'''' + a\eta^2\chi'' + 3a\eta\chi' + b\chi = 0\,, \tag{1}$$

where primes denote differentiation with respect to $\eta$, and $a$ and $b$ are constants. This equation can be solved by using the integral transform

$$\chi(\eta) = \int_\Gamma \exp(\eta t)v(t)dt\,, \tag{2}$$

where $\Gamma$ is a curve in the complex $t$ plane, and must be chosen such that

$$[v(t)]_{\partial\Gamma} = 0 \quad \text{and} \quad [v'(t)]_{\partial\Gamma} = 0\,. \tag{3}$$

Transforming Eq. (1) gives an equation for $v(t)$ of the form

$$t^2(t^2v)'' - 3t(t^2v)' + (b/a + 3 + t^4/a)t^2v = 0\,,$$

with Bessel function solutions

$$v(t) = J_\mu(\tfrac{1}{2}s^2)\,, \tag{4}$$

where

$$\mu = \tfrac{1}{2}\sqrt{(a-b)/a} \quad \text{and} \quad s = a^{-1/4}t\,.$$

One of the solutions of Eq. (1) can then be written

$$\chi(\eta) = \int_\Gamma J_\mu\left(\frac{1}{2}s^2\right)\exp(\zeta s)ds\,, \tag{5}$$

where $\zeta = a^{1/4}\eta$. The curve $\Gamma$ can be chosen as in Fig. 1(a).

If $\mathrm{Re}(\mu) > -\frac{1}{2}$, the path $\Gamma$ can be deformed into ABCD as in Fig. 1(b), and then

$$\chi(\eta) = \int_{\mathrm{ABCD}} J_\mu\left(\frac{1}{2}s^2\right)\exp(\zeta s)ds$$
$$= 2i\sin(\mu\pi)\int_0^\infty \exp(-\zeta x)J_\mu\left(\frac{1}{2}x^2\right)dx\,. \tag{6}$$

This integral can be expressed in terms of the well-known parabolic-cylinder functions[5]

$$\chi(\eta) = 2i\sin(\mu\pi)\left[\Gamma(\mu+\tfrac{1}{2})/\sqrt{\pi}\right]U(\mu,\zeta e^{i\pi/4})$$
$$\times\, U(\mu,\zeta e^{-i\pi/4})\,. \tag{7}$$

The parabolic-cylinder functions $U(\mu,x)$ are solutions of the second-order differential equation

$$Y'' - (\tfrac{1}{4}x^2 + \mu)Y = 0\,. \tag{8}$$

The second solution to (8), linearly independent to $U(\mu,x)$, is denoted by $V(\mu,x)$.

Since $U(\mu,x)$ is analytic for all $\mu$ and $x$, the solution $\chi(\eta)$ given by (7) is analytic for all $\mu$ and $x$ by analytic continuation to $\mathrm{Re}(\mu) \leqslant -\frac{1}{2}$. We have thus found one solution to (1), i.e.,

 0022-2488/89/020343-02$02.50 © 1989 American Institute of Physics 

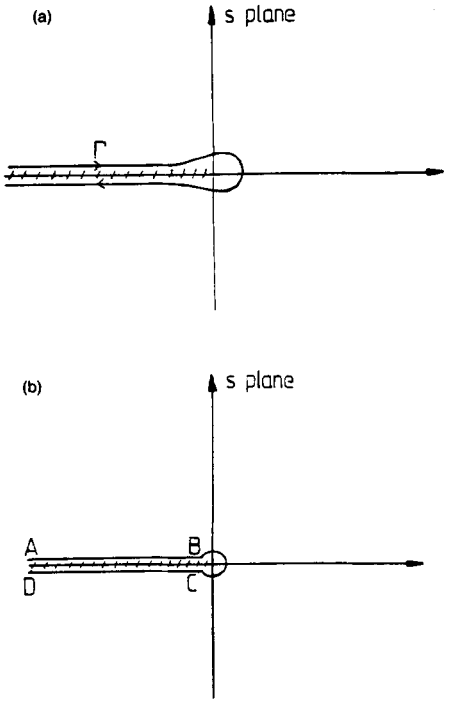



FIG. 1. (a) Integration contour Γ. (b) Deformed integration contour ABCD, where the small circle radius approaches zero.

$$\chi_1 = U(\mu,\zeta e^{i\pi/4})\,U(\mu,\zeta e^{-i\pi/4})\ . \tag{9}$$

The other three linearly independent solutions can easily be verified to be

$$\chi_2(\eta) = U(\mu,\zeta e^{i\pi/4})\,V(\mu,\zeta e^{-i\pi/4})\ ,$$

$$\chi_3(\eta) = U(\mu,\zeta e^{-i\pi/4})\,V(\mu,\zeta e^{i\pi/4})\ ,$$

$$\chi_4(\eta) = V(\mu,\zeta e^{-i\pi/4})\,V(\mu,\zeta e^{i\pi/4})\ .$$

We have thus found four linearly independent solutions to (1), which can be expressed in terms of products of the parabolic-cylinder functions. Since the parabolic-cylinder functions have well-known asymptotic properties, they might be useful in the theory of asymptotic expansions of differential equations, for cases when Eq. (1) can be used as a comparison equation.

Finally, we mention that the asymptotic properties of the solutions of a similar but slightly more general equation have been considered by Paris and Wood.[6] This equation is of the form

$$u^{\mathrm{iv}} + (\alpha z^2 + \beta)u'' + \gamma z u' + \delta u = 0\ .$$

The solutions to this equation can also be written in terms of an integral transform of the type (2) above. However, this integral can be performed only for special choices of the constants $\alpha,\beta,\gamma,\delta$. One such choice is $\beta = 0$, $\gamma = 3\alpha$, for which the solutions are identical to the ones given here in this paper.

[1]R. E. Langer, Trans. Am. Math. Soc. **33**, 23 (1931).
[2]A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).
[3]P. G. Drazin and W. H. Reid, *Hydrodynamic Stability* (Cambridge U.P., New York, 1981).
[4]H. O. Åkerstedt, Phys. Scr. **37**, 117 (1988).
[5]I. S. Gradsheteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[6]R. B. Paris and A. D. Wood, Philos. Trans. R. Soc. London Ser. A **293**, 511 (1980).

# Poincaré invariance—classical electrodynamics

Abraham Lozada and P. L. Torres
*Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, A. P. 20513, Caracas 1020-A, Venezuela*

(Received 1 March 1988; accepted for publication 12 October 1988)

In a previous paper by the authors [J. Math. Phys. **29**, 1361 (1988)] it was shown that, following the standard definition, the angular momentum tensor corresponding to the Coulomb field does not exist (due to the long range of the field). In the present paper, for a general classical field theory, nonconventional definitions for the four-momentum and angular momentum tensor are given. These definitions are applied to the particular case of a charged classical point particle in an external electromagnetic field, showing that the total four-momentum $P^\mu$ and the total angular momentum tensor $M^{\lambda\mu}$ of the system exist and are conserved quantities. In this framework, no asymptotic conditions on the four-acceleration of the particle are needed. The problem of extracting an equation of motion for the particle from these conservation laws is discussed. Also, the different parts (such as the radiated one and the bound one) of $P^\mu$ and $M^{\lambda\mu}$, and their corresponding energy momentum tensors and angular momentum density tensors, are considered.

## I. INTRODUCTION

In classical field theories, given an energy momentum tensor $\Theta_A^{\mu\nu}$, an angular momentum density $M_A^{\lambda\mu\nu}$, and any auxiliary timelike world line (TWL) with parametric equation $x^\mu = z^\mu(\tau)$, where $\tau$ is the proper time given along it, it has been customary[1-3] to define the corresponding four-momentum $p_A^\mu$ and angular momentum tensor $m_A^{\lambda\mu}$ as

$$p_A^\mu(\tau) \equiv \int_\sigma \Theta_A^{\mu\nu}\, d\sigma_\nu\,, \tag{1.1}$$

$$m_A^{\lambda\mu}(\tau) \equiv \int_\sigma M_A^{\lambda\mu\nu}\, d\sigma_\nu\,, \tag{1.2}$$

where $\sigma$ is any spatial hypersurface that cuts the TWL at $z^\mu(\tau)$. Definitions (1.1) and (1.2) are usually given without the use of a TWL because this is not really necessary, but we do so in order to apply these definitions directly to the case of classical electrodynamics.

For the case of a charged classical point particle in an external electromagnetic field the TWL corresponds to the particle world line (PWL), and $\sigma$ is restricted to a spatial hypersurface that cuts the PWL orthogonally at $z(\tau)$.[4-9] In a previous paper[10] we showed that in this case the integral (1.2) defining the total angular momentum tensor *does not exist*, and hence Lorentz invariance of the theory cannot be treated by definition (1.2). It is important to notice that the nonexistence of $m^{\lambda\mu}$ we are talking about is *not* due to the divergence of the theory at the PWL (which can be dealt with by the standard renormalization procedure[5,8-12]), but is due to the *asymptotic behavior* of the electromagnetic field at spatial infinity.[10] Precisely because of this, it is expected that the integral (1.2) does not exist for *any* extended body with nonzero net charge[10] (for instance, for Schwinger's charged spherical shell,[13] $m^{\lambda\mu}$ does *not* exist[10]).

In order to avoid the above-mentioned problem, we set in Sec. II A, for a general classical field theory, nonconventional definitions for the four-momentum $P_A^\mu$ and angular momentum tensor $M_A^{\lambda\mu}$ corresponding to $\Theta_A^{\mu\nu}$ and $M_A^{\lambda\mu\nu}$. In Sec. II B, we show that the total four-momentum $P^\mu$ *and* the total angular momentum tensor $M^{\lambda\mu}$ of a charged classical point particle in an external electromagnetic field *exist* and are conserved quantities. In Sec. II C general properties of the new definitions are discussed.

Also, in light of the new definitions we undertake the study of different items for a charged classical particle, previously considered in the literature in light of definitions (1.1) and (1.2). In Sec. III, we discuss the radiated part of the four-momentum and angular momentum tensor together with their splitting in a spin and orbital part. The problem of the corresponding energy momentum and angular momentum density tensors is also discussed. In Sec. IV, the bound part of the four-momentum and angular momentum is considered. We use two different methods to reach the PWL, which lead to two different results. The problem of the corresponding bound energy momentum and angular momentum density tensors is also addressed. In Sec. V, from the conservation law of $P^\mu$ and $M^{\lambda\mu}$, we consider the problem of extracting an equation of motion for the particle.

Throughout this paper we emphasize obtaining results in which *both* the four-momentum and the angular momentum tensor are inextricably linked.

We shall follow Rowe's[14] notation, i.e., the metric tensor $g$ has signature $+2$ and the speed of the light is taken as 1. When convenient, indices on vectors and tensors will be omitted, and scalar products will be indicated by a dot. A parenthesis $(\cdot,\cdot)$ or a bracket $[\cdot,\cdot]$ will denote symmetrization or antisymmetrization, respectively, of the enclosed variables (without a factor $\frac{1}{2}$). The PWL is $z(\tau)$, where $\tau$ is the proper time; $v(\tau) \equiv v(v^2 = -1)$ and $a(\tau) \equiv a$ $(v \cdot a = 0)$ are its four-velocity and four-acceleration, respectively. The components of the total electromagnetic energy tensor $\Theta_{\rm elm}$ for a charged point particle and an external electromagnetic field are

$$\Theta_{\rm elm}^{\mu\nu} = (1/4\pi)(F^{\mu\alpha}F^{\nu}{}_{\alpha} - \tfrac{1}{4} g^{\mu\nu} F^{\alpha\beta}F_{\alpha\beta})\,, \tag{1.3}$$

where

 0022-2488/89/020345-09$02.50 © 1989 American Institute of Physics 

$$F^{\mu\nu} = F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{ext}} . \tag{1.4}$$

The electromagnetic field tensor $F^{\mu\nu}_{\text{ret}}$ is the one corresponding to the *retarded* Lienard–Wiechert potential. The nonsingular external electromagnetic field $F^{\mu\nu}_{\text{ext}}$ satisfies Maxwell's equations for a vacuum and vanishes adequately asymptotically (see Sec. II B). Corresponding to the superposition showed in (1.4) we obtain (in obvious notation[14]), using (1.3),

$$\Theta_{\text{elm}} = \Theta_{\text{ret}} + \Theta_{\text{mix}} + \Theta_{\text{ext}} . \tag{1.5}$$

Retarded coordinates will be used here (see, e.g., Refs. 9 and 14). Then for any space-time point $x$, we define $R \equiv x - z(\tau)$, $R^2 = 0$ $(R^0 > 0)$, $\rho \equiv -v \cdot R$, and $u \equiv R/\rho - v$. In these coordinates,

$$\Theta_{\text{ret}} = \frac{e^2}{4\pi}\left(\frac{1}{2}g + vv - uu\right)\frac{1}{\rho^4} + \frac{e^2}{4\pi}\left(a - ua\cdot u, \frac{R}{\rho}\right)\frac{1}{\rho^3} + \frac{e^2}{4\pi}[a^2 - (a\cdot u)^2]\frac{RR}{\rho^4} . \tag{1.6}$$

The tensor $\Theta_{\text{ret}}$ has been split in several ways.[5,15,16] Here we shall consider the splitting proposed in Refs. 15 and 16, that is,

$$\Theta_{\text{ret}} = \Theta_B + \Theta_S + \Theta_L , \tag{1.7}$$

where

$$\Theta_S \equiv \frac{e^2}{4\pi}\left[\frac{(a,R)}{\rho} - \frac{4a\cdot RRR}{\rho^3} + \frac{a\cdot R(v,R)}{\rho^2}\right]\frac{1}{\rho^3} \tag{1.8}$$

and

$$\Theta_L \equiv \frac{e^2}{4\pi}[a^2 - (a\cdot u)^2]\frac{RR}{\rho^4} . \tag{1.9}$$

In Rowe's notation, $\Theta_B \equiv \Theta_1$, $\Theta_S \equiv \Theta_2$, $\Theta_L \equiv \Theta_3$.

Given an arbitrary energy momentum tensor $\Theta^{\mu\nu}_A$, we define its corresponding angular momentum tensor density $M^{\lambda\mu\nu}_A$ as

$$M^{\lambda\mu\nu}_A \equiv x^\lambda \Theta^{\mu\nu}_A - x^\mu \Theta^{\lambda\nu}_A . \tag{1.10}$$

The material properties of the particle will be characterized by the standard[14,16] bare energy momentum tensor $\Theta^{\mu\nu}_0$ given by

$$\Theta^{\mu\nu}_0 = \int m_0(\tau) v^\mu(\tau) v^\nu(\tau) \delta[x - z(\tau)]\, d\tau , \tag{1.11}$$

where $m_0(\tau)$ is an arbitrary scalar, and by a bare angular momentum tensor density, given by (see, e.g., Ref. 6)

$$\mathcal{M}^{\lambda\mu\nu}_0 = M^{\lambda\mu\nu}_0 + \mathcal{S}^{\lambda\mu\nu}_0 , \tag{1.12}$$

where $\mathcal{S}^{\lambda\mu\nu}_0$ is an intrinsic angular momentum density, antisymmetric in the indices $\lambda$ and $\mu$. We shall choose $\mathcal{S}^{\lambda\mu\nu}_0$ as

$$\mathcal{S}^{\lambda\mu\nu}_0 = \int S^{\lambda\mu}(\tau) v^\nu(\tau) \delta[x - z(\tau)]\, d\tau , \tag{1.13}$$

where $S^{\lambda\mu}(\tau)$ is an arbitrary antisymmetric tensor.

Then, the total energy momentum tensor $\Theta^{\mu\nu}$ for a charged classical point particle in an external field will be given by

$$\Theta^{\mu\nu} = \Theta^{\mu\nu}_0 + \Theta^{\mu\nu}_{\text{elm}} , \tag{1.14}$$

and the total angular momentum tensor density $M^{\lambda\mu\nu}$ will be given by

$$M^{\lambda\mu\nu} = \mathcal{M}^{\lambda\mu\nu}_0 + M^{\lambda\mu\nu}_{\text{elm}} , \tag{1.15}$$

where $M^{\lambda\mu\nu}_{\text{elm}}$ is the tensor corresponding to $\Theta^{\mu\nu}_{\text{elm}}$.

Given an arbitrary TWL with parametric equation $z^\mu(\tau)$, we shall define several hypersurfaces related to it that will be used throughout the paper. We shall designate the future light cone with apex at $z^\mu(\tau)$ by $C(\tau)$. We shall designate a connected band on $C(\tau)$ by $\sigma_c$ (or $\sigma'_c$), $\mathscr{E}$ being defined by $\mathscr{E} \equiv \inf_{\sigma_c} \rho$. A special band on $C(\tau)$, denoted by $C(\tau, E_1, E_2)$, is determined by the intersection of $C(\tau)$ and two Bhabha tubes[9,17] of radius $E_1$ and $E_2$ with $E_1 < E_2$ (here $\mathscr{E} = E_1$). Another special band on $C(\tau')$ denoted by $\mathscr{C}(\tau', \rho_0, E)$ is determined by the intersection of $C(\tau')$ with the Bhabha tube of radius $E$, and the intersection of $C(\tau')$ with the hyperplane with normal $n^\mu(\tau)$, where $\tau' < \tau$, $E > \sup \rho_0$ ($\rho_0$ being the set of values of the $\rho$ coordinate on the intersection with the hyperplane; $\mathscr{E} = \inf \rho_0$). We shall designate an arbitrary connected surface laying between $C(\tau_1)$ and $C(\tau_2)$ and intersecting them ($\tau_1 < \tau_2$, $E \equiv \inf_\Sigma \rho$) by $\Sigma(E, \tau_1, \tau_2)$. For the special case when $\Sigma(E, \tau_1, \tau_2)$ is a segment of the Bhabha tube of radius $E$, we shall use the notation $\Sigma \equiv B(E, \tau_1, \tau_2)$. We know[9,14] that the hypersurface elements of a Bhabha tube $dB_\nu$ and of a light cone $dC_\nu$ are $dB_\nu = E^2\, d\tau\, d\Omega (u_\nu + a\cdot u R_\nu)$ and $dC_\nu = -\rho\, d\rho\, d\Omega\, R_\nu$.

## II. CONSERVATION LAWS

### A. General theory

Given an auxiliary TWL, let $\tau_0$ be an arbitrary proper time referred to the TWL. We shall consider all the physical phenomena on the space-time region $\Omega_0$ contained in the absolute future light cone $C(\tau_0)$. Let us note that the limit $\tau_0 \to -\infty$ is not excluded here.

Let us consider energy momentum tensors $\Theta^{\mu\nu}_A$ and angular momentum density tensors $M^{\lambda\mu\nu}_A$ in a general classical field theory, for which the following three conditions hold. Given a TWL, the following statements hold.

(a) There exists in $\Omega_0$ a world volume $W$ whose intersection with every lightlike surface is bounded, and such that $\partial_\nu \Theta^{\mu\nu}_A = 0$, $\partial_\nu M^{\lambda\mu\nu}_A = 0$ on $\Omega'_0 \equiv \Omega_0 - W$ (this means that the field sources are contained in $W$).

(b) The integrals

$$\int_{C(\tau)} \Theta^{\mu\nu}_A\, dC_\nu \tag{2.1}$$

and

$$\int_{C(\tau)} M^{\lambda\mu\nu}_A\, dC_\nu \tag{2.2}$$

exist for all $\tau \in [\tau_0, \infty)$.

(c) There exists a surface $\Sigma(E, \tau_0, \tau)$ such that the integrals

$$\lim_{E\to\infty} \int_{\Sigma(E,\tau_0,\tau)} \Theta^{\mu\nu}_A\, d\Sigma_\nu \tag{2.3}$$

and

$$\lim_{E\to\infty} \int_{\Sigma(E,\tau_0,\tau)} M^{\lambda\mu\nu}_A\, d\Sigma_\nu \tag{2.4}$$

exist for all $\tau \in [\tau_0, \infty)$.

The existence of the integrals (2.1) and (2.2) imply[18,19] that, for $\tau \in [\tau_0, \infty)$,

$$\lim_{\mathscr{C}\to\infty}\int_{\sigma_c}\Theta_A^{\mu\nu}\,dC_\nu=0\,, \tag{2.5}$$

and that

$$\lim_{\mathscr{C}\to\infty}\int_{\sigma_c}M_A^{\lambda\mu\nu}\,dC_\nu=0\,, \tag{2.6}$$

for any band $\sigma_c$ on $C(\tau)$.

Let us now show that the imposed conditions (a)–(c) imply that the integrals (2.3) and (2.4) are independent of the surface $\Sigma(E,\tau_0,\tau)$. In fact, let $\Sigma(E,\tau_0,\tau)$ and $B(E',\tau_0,\tau)$ on $\Omega_0'$ $(E'<E)$ define the bands $\sigma_c'$ and $\sigma_c$ on $C(\tau_0)$ and $C(\tau)$, respectively (see Fig. 1). Then, by Gauss' theorem we obtain that

$$\int_{\Sigma(E,\tau_0,\tau)}\Theta_A^{\mu\nu}\,d\Sigma_\nu-\int_{B(E',\tau_0,\tau)}\Theta_A^{\mu\nu}\,dB_\nu$$
$$=\int_{\sigma_c'}\Theta_A^{\mu\nu}\,dC_\nu-\int_{\sigma_c}\Theta_A^{\mu\nu}\,dC_\nu\,. \tag{2.7}$$

Taking the limits $E\to\infty$ and $E'\to\infty$, the proof for $\Theta_A$ is now complete using (2.5). Obviously, the same procedure holds for $M_A^{\lambda\mu\nu}$.

For arbitrary tensors $\Theta_A^{\mu\nu}$ and $M_A^{\lambda\mu\nu}$ for which conditions (a)–(c) hold, we define for any $\tau\in[\tau_0,\infty)$ the corresponding four-momentum $P_A^\mu(\tau)$ and the corresponding angular momentum tensor $M_A^{\lambda\mu}(\tau)$ from $\tau_0$ up to the proper time $\tau$ as (see Fig. 1)

$$P_A^\mu(\tau)=\int_{C(\tau)}\Theta_A^{\mu\nu}\,dC_\nu+\lim_{E\to\infty}\int_{\Sigma(E,\tau_0,\tau)}\Theta_A^{\mu\nu}\,d\Sigma_\nu\,, \tag{2.8}$$

$$M_A^{\lambda\mu}(\tau)=\int_{C(\tau)}M_A^{\lambda\mu\nu}\,dC_\nu+\lim_{E\to\infty}\int_{\Sigma(E,\tau_0,\tau)}M_A^{\lambda\mu\nu}\,d\Sigma_\nu\,, \tag{2.9}$$

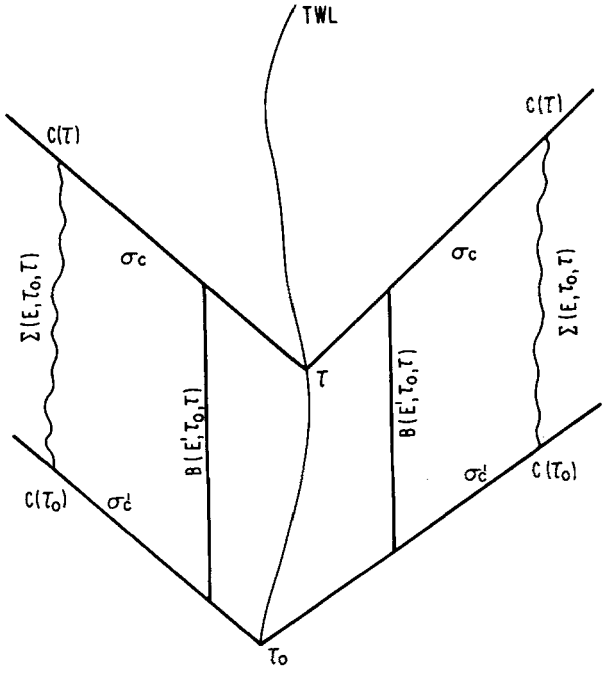



FIG. 1. Hypersurfaces used to define $P_A^\mu$ and $M_A^{\lambda\mu}$ and to evaluate $P_{\text{mix}}^\mu$ and $M_{\text{mix}}^{\lambda\mu}$.

where the surface $\Sigma$ that connects $C(\tau_0)$ and $C(\tau)$ is arbitrary. Since we showed that the last integral appearing in (2.8) or (2.9) does not depend on $\Sigma$, $P_A^\mu$ and $M_A^{\lambda\mu}$ are unambiguously defined. Hence in the following we shall use the special surface $B(E,\tau_0,\tau)$ instead of a generic $\Sigma(E,\tau_0,\tau)$ whenever necessary.

Now, we shall show that if $\Theta_A^{\mu\nu}$ and $M_A^{\lambda\mu\nu}$ obey a conservation law in differential form, that is,

$$\partial_\nu\Theta_A^{\mu\nu}=0\,,\quad \partial_\nu M_A^{\lambda\mu\nu}=0\,, \tag{2.10}$$

everywhere in $\Omega_0$, then $P_A^\mu$ and $M_A^{\lambda\mu}$ are conserved quantities for all $\tau\in[\tau_0,\infty)$; that is, we obtain conserved quantities in integral form.

For the proof, we use Gauss' theorem and (2.10) in the region shown in Fig. 2. We obtain

$$\left(\int_{C(\tau_2,0,E_2)}\Theta_A^{\mu\nu}\,dC_\nu+\int_{B(E_2,\tau_0,\tau_2)}\Theta_A^{\mu\nu}\,dB_\nu\right)$$
$$-\left(\int_{C(\tau_1,0,E_1)}\Theta_A^{\mu\nu}\,dC_\nu+\int_{B(E_1,\tau_0,\tau_1)}\Theta_A^{\mu\nu}\,dB_\nu\right)$$
$$=\int_{C(\tau_0,E_1,E_2)}\Theta_A^{\mu\nu}\,dC_\nu\,. \tag{2.11}$$

Then taking the limits $E_2\to\infty$ and $E_1\to\infty$ in (2.11) and using (2.8) and (2.5), it follows that

$$P_A^\mu(\tau_1)=P_A^\mu(\tau_2)\,. \tag{2.12}$$

Following the same steps, we obtain that

$$M_A^{\lambda\mu}(\tau_1)=M_A^{\lambda\mu}(\tau_2)\,. \tag{2.13}$$

Let us obtain $P_A^\mu(\tau)-p_A^\mu(\tau)$ and $M_A^{\lambda\mu}(\tau)-m_A^{\lambda\mu}(\tau)$ for a $\Theta_A^{\mu\nu}$ and $M_A^{\lambda\mu\nu}$ for which conditions (2.10), (b), and (c) hold, and where the integrals (1.1) and (1.2) defining $p_A^\mu$ and $m_A^{\lambda\mu}$ exist [to perform the comparison we take the limit $\tau_0\to-\infty$ in (2.8) and (2.9)]. Using Gauss' theorem and (2.10) in the region $\Omega$ shown in Fig. 3, it is easily obtained, if the limits $E\to\infty$, $\tau_0\to-\infty$ are taken, that

$$P_A^\mu(\tau)-p_A^\mu(\tau)=\lim_{\tau_0\to-\infty}\lim_{E\to\infty}\int_{\mathscr{C}(\tau_0,\rho_0,E)}\Theta_A^{\mu\nu}\,dC_\nu\,, \tag{2.14}$$

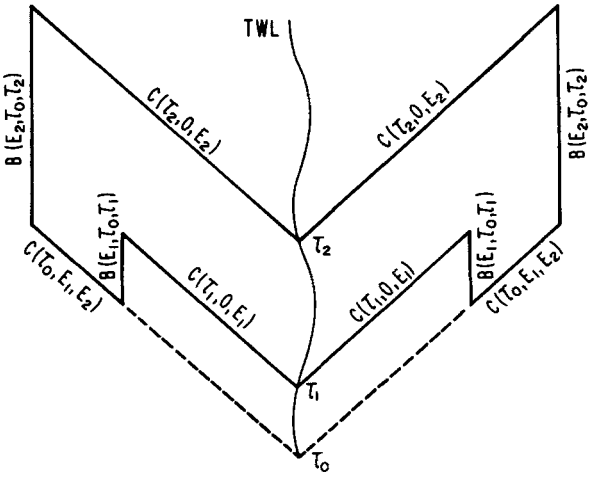



FIG. 2. Space-time region used to prove the conservation laws in integral form.

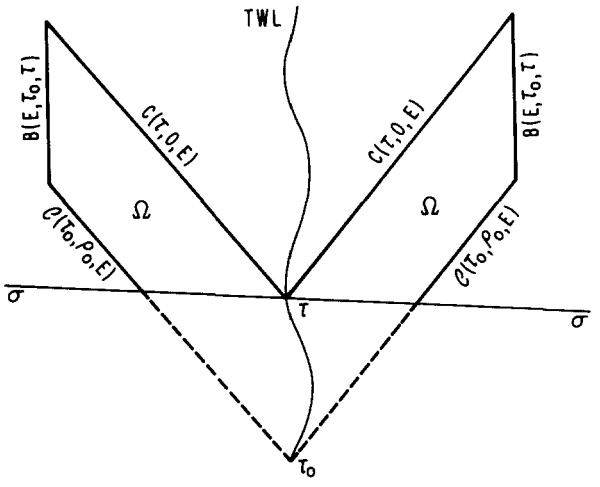



FIG. 3. Space-time region used to relate $P_A^\mu$ and $p_A^\mu$ or $M_A^{\lambda\mu}$ and $m_A^{\lambda\mu}$.

$$M_A^{\lambda\mu}(\tau) - m_A^{\lambda\mu}(\tau) = \lim_{\tau_0 \to -\infty} \lim_{E \to \infty} \int_{\mathscr{C}(\tau_0,\rho_0,E)} M_A^{\lambda\mu\nu}\, dC_\nu \,. \tag{2.15}$$

### B. Electrodynamics

We shall apply the general setting of Sec. II A to a charged classical point particle in an external field. We need to verify if conditions (a)–(c) hold for each of the parts of the total $\Theta^{\mu\nu}$ and $M^{\lambda\mu\nu}$ defined by (1.5), (1.6)–(1.9), and (1.10)–(1.13). We shall impose that $F_{\text{ext}}^{\mu\nu} = 0$ in a region $\Omega_0'$ [see the definition of $\Omega_0'$ in Sec. II A (a)].

The singularities of this theory at the PWL lead us to restrict the TWL to the PWL in order to carry on the renormalization procedure. We know (see Refs. 12, 15, and 16, for $\Theta_B$, $\Theta_S$, $\Theta_L$, and $\Theta_{\text{mix}}$) that

$$\partial_\nu \Theta_B^{\mu\nu} = \partial_\nu \Theta_S^{\mu\nu} = \partial_\nu \Theta_L^{\mu\nu} = \partial_\nu \Theta_{\text{mix}}^{\mu\nu} = \partial_\nu \Theta_{\text{ext}}^{\mu\nu} = 0 \,,$$

outside the PWL (and the same for each of the corresponding parts of $M^{\lambda\mu\nu}$, since the theory is well defined and symmetric there). Hence condition (a) is satisfied for each of these parts of $\Theta^{\mu\nu}$ and $M^{\lambda\mu\nu}$. For condition (b) it is obvious that the improper integrals of $\Theta_L$, $\Theta_S$, $\Theta_{\text{mix}}$, $\Theta_{\text{ext}}$, $M_L$, $M_S$, $M_{\text{mix}}$, and $M_{\text{ext}}$ over $C(\tau)$ are absolutely convergent (and hence they exist), since $\Theta_L^{\mu\nu}\, dC_\nu = \Theta_S^{\mu\nu}\, dC_\nu = 0$ and $M_L^{\lambda\mu\nu}\, dC_\nu = M_S^{\lambda\mu\nu}\, dC_\nu = 0$, and at infinity on $C(\tau)$ we have that $\Theta_{\text{ext}}^{\mu\nu} = \Theta_{\text{mix}}^{\mu\nu} = 0$ and $M_{\text{ext}}^{\lambda\mu\nu} = M_{\text{mix}}^{\lambda\mu\nu} = 0$, because of the assumed properties of $F_{\text{ext}}^{\mu\nu}$. The existence of the improper integral of $\Theta_B^{\mu\nu}$ and $M_B^{\lambda\mu\nu}$ on $C(\tau,E,\infty)$, where $E>0$, follows directly[18,19] from the fact that the integrals

$$\lim_{E' \to \infty} \int_{C(\tau,E,E')} |\Theta_B^{\mu\nu} R_\nu|\, \rho\, d\rho\, d\Omega \,, \tag{2.16}$$

$$\lim_{E' \to \infty} \int_{C(\tau,E,E')} |M_B^{\lambda\mu\nu} R_\nu|\, \rho\, d\rho\, d\Omega \tag{2.17}$$

exist, as can be verified easily (notice that since $\Theta_0^{\mu\nu}$ and $M_0^{\lambda\mu\nu}$ vanish outside the PWL, the above discussion is unnecessary for them). To treat the singularities of $\Theta_B^{\mu\nu}$ and $M_B^{\lambda\mu\nu}$ at the PWL [i.e., the respective integrals over $C(\tau,0,E)$], we shall see in Sec. IV that the renormalization procedure with $\Theta_0^{\mu\nu}$ and $M_0^{\lambda\mu\nu}$ can be carried on.

Condition (c) is automatically satisfied by $\Theta_0$, $\Theta_{\text{ext}}$, $\Theta_{\text{mix}}$, $M_0$, $M_{\text{ext}}$, and $M_{\text{mix}}$, since these quantities vanish on the surface $B(E,\tau_0,\tau)$, for $E$ big enough [and hence the respective integrals (2.3) and (2.4) vanish]. The existence of the integrals (2.3) and (2.4) for $\Theta_B$, $\Theta_S$, $\Theta_L$, $M_B$, $M_S$, and $M_L$ on the surface $B(E,\tau_0,\tau)$ are well known results.

From the discussion carried on above, it follows that the parts of the $P^\mu$ and $M^{\lambda\mu}$ corresponding to each of the terms $\Theta_S$, $\Theta_L$, $\Theta_{\text{mix}}$, $\Theta_{\text{ext}}$, $M_S$, $M_L$, $M_{\text{mix}}$, and $M_{\text{ext}}$ *exist*. Moreover, if a renormalization procedure is taken at the PWL, the same statement holds for $\Theta_0 + \Theta_B$ and $M_0 + M_B$. Hence we have proved that the total four-momentum $P^\mu$ and the total angular momentum tensor $M^{\lambda\mu}$ *exist* for each $\tau \in [\tau_0, \infty)$. The evaluation of $P^\mu$ and $M^{\lambda\mu}$ will be performed throughout Secs. III–V.

Now, let us consider the conservation laws. If (2.10) is assumed for $\Theta^{\mu\nu}$ and $M^{\lambda\mu\nu}$, it follows from the existence of $P^\mu$ and $M^{\lambda\mu}$ that they are *conserved quantities* (see Sec. II A). Moreover, since $\partial_\nu \Theta_{\text{ext}}^{\mu\nu} = 0$ everywhere in $\Omega_0$ is assumed, it also follows that $P_{\text{ext}}^\mu$ and $M_{\text{ext}}^{\lambda\mu}$ obtained from $\Theta_{\text{ext}}^{\mu\nu}$ and $M_{\text{ext}}^{\lambda\mu\nu}$, by Eqs. (2.8) and (2.9), are independently conserved quantities.

### C. Discussion

Let us elaborate on the physical relevance of the new definitions, comparing them with the standard ones. For this purpose, the following numbering with a (i) will refer to the general theory, and with a (ii) will refer to classical electrodynamics.

(i.1) A physical advantage of definitions (2.8) and (2.9) is that the tensors $\Theta_A^{\mu\nu}$ and $M_A^{\lambda\mu\nu}$ may be specified only on $\Omega_0$, in contraposition with definitions (1.1) and (1.2), where this specification is not enough. The choice of the complete Minkowski space-time instead of an $\Omega_0$ can be physically too demanding, since information on the system outside an $\Omega_0$ can be unnecessary or even physically inaccessible for its description.

(i.2) With definitions (2.8) and (2.9), if $P_A^\mu$ and $M_A^{\lambda\mu}$ exist and (2.10) holds, then *automatically* $P_A^\mu$ and $M_A^{\lambda\mu}$ are conserved quantities. This is not generally the case for $p_A^\mu$ and $m_A^{\lambda\mu}$.

(i.3) If, besides the conditions imposed in order to obtain (2.14) and (2.15), it happens that $p_A^\mu$ and $m_A^{\lambda\mu}$ are conserved quantities, then the right-hand sides of (2.14) and (2.15) are *constant tensors* (independents of $\tau$). Hence, in this case, $p_A^\mu$ and $P_A^\mu$ (or $m_A^{\lambda\mu}$ and $M_A^{\lambda\mu}$) will carry the same physical information.

Moreover, if the fields vanish outside a volume $W$ of the type specified in (a) of Sec. II A, as is often supposed (see, e.g., Ref. 1), the right-hand sides of Eqs. (2.14) and (2.15) vanish automatically [also, in this case, (1) conditions (a), (b), and (c) hold trivially; and (2) the only integrals that survive in definitions (2.18) and (2.19) are the ones over $C(\tau)$].

(i.4) Definitions (2.8) and (2.9) depend on the refer-

ence time $\tau_0$, so for different reference times the values of $P^\mu_A$ and $M^{\lambda\mu}_A$ are different. However, as is clear from the definitions, the differences between these values are *constant* tensors, independent of $\tau$ (moreover, for an asymptotic behavior of the fields as in Ref. 1, these constant tensors are the zero tensors). Hence the same physical consequences will be obtained.

(i.5) It is important to observe that, as was mentioned for definitions (1.1) and (1.2), definitions (2.8) and (2.9) can similarly be given without the use of an auxiliary TWL. This is so because $z^\mu(\tau_0)$ can be chosen as an arbitrary point in space-time, $z^\mu(\tau)$ being another arbitrary point on $\Omega_0$; $\Sigma(E,\tau_0,\tau)$ is then an arbitrary surface connecting $C(\tau)$ and $C(\tau_0)$.

(ii.1) For a charged point particle the difference between the existence found for $M^{\lambda\mu}$ and $m^{\lambda\mu}$ comes basically from the way the infinity is reached. For $m^{\lambda\mu}$ the infinity is reached following a spatial hypersurface, and for $M^{\lambda\mu}$ the infinity is reached following a future light cone. We can say that the Coulomb field has (as far as the angular momentum is concerned) a "bad" asymptotic behavior on spatial hypersurfaces, but a "good" asymptotic behavior for lightlike surface.

Because of that we expect that definitions (2.8) and (2.9) will be useful for extended charged bodies [but not definitions (1.1) and (1.2)[10]], since asymptotically the behavior of the field of a charged body is like the field of a point particle.[20]

Moreover, for Schwinger's[13] charged spherical model [see his relation (30)], it is obtained (the calculations are straightforward) that the total angular momentum $M^{\lambda\mu}$ *exists* and that $M^{\lambda\mu}(\tau) = [z^\lambda(\tau), P^\mu(\tau)]$, where $z^\lambda(\tau)$ characterizes the four-position of the charge center, and $P^\mu$ is the total four-momentum.

(ii.2) For the case of a charged point particle we saw that for the existence of $P^\mu$ and $M^{\lambda\mu}$ *no asymptotic conditions* need to be imposed on the four-acceleration of the particle. Treatments involving dropping this condition have been unsuccessful (see, e.g., Ref. 21).

## III. RADIATION PART OF THE LINEAR AND ANGULAR MOMENTUM

We define $P^\mu_{\text{rad}}$, the radiation part of the four-momentum, and $M^{\lambda\mu}_{\text{rad}}$, the radiation part of the angular momentum tensor (between $\tau_0$ and $\tau$), as

$$P^\mu_{\text{rad}}(\tau) \equiv \lim_{E\to\infty} \int_{\Sigma(E,\tau_0,\tau)} \Theta^{\mu\nu}_{\text{ret}}\, d\Sigma_\nu\,, \tag{3.1}$$

$$M^{\lambda\mu}_{\text{rad}}(\tau) \equiv \lim_{E\to\infty} \int_{\Sigma(E,\tau_0,\tau)} M^{\lambda\mu\nu}_{\text{ret}}\, d\Sigma_\nu\,, \tag{3.2}$$

where $\Sigma(E,\tau_0,\tau)$ is arbitrary. These definitions are the usual ones, and that they are unambiguous (i.e., independent on $\Sigma$) is shown in Refs. 4 and 22–24 (see, also, our Sec. II B).

The evaluation of $P^\mu_{\text{rad}}$ and $M^{\lambda\mu}_{\text{rad}}$ is straightforward and well known (see, e.g., Refs. 4, 7, 8, and 22). That is,

$$P^\mu_{\text{rad}}(\tau) = \frac{2}{3}e^2 \int_{\tau_0}^{\tau} a^2(\tau') v^\mu(\tau')\, d\tau'\,, \tag{3.3}$$

$$M^{\lambda\mu}_{\text{rad}}(\tau) = \frac{2e^2}{3}\int_{\tau_0}^{\tau} [v^\lambda(\tau'), a^\mu(\tau')]\, d\tau' + \int_{\tau_0}^{\tau} [z^\lambda(\tau'), \dot{P}^\mu_{\text{rad}}(\tau')]\, d\tau'\,. \tag{3.4}$$

Let us observe that the same results (3.3) and (3.4) are obtained if $\Theta^{\mu\nu}$ and $M^{\lambda\mu\nu}$ are used in Eqs. (3.1) and (3.2) instead of $\Theta^{\mu\nu}_{\text{ret}}$ and $M^{\lambda\mu\nu}_{\text{ret}}$ (this is so because of the properties assumed for $F^{\mu\nu}_{\text{ret}}$).

Several splittings of $\Theta^{\mu\nu}_{\text{ret}}$ and $M^{\lambda\mu\nu}_{\text{ret}}$ have been exposed in the literature (see Refs. 5, 15, and 16). Given a splitting

$$\Theta^{\mu\nu}_{\text{ret}} \equiv \sum_{i=1}^{n} \Theta^{\mu\nu}_i\,,$$

and

$$M^{\lambda\mu\nu}_{\text{ret}} \equiv \sum_{i=1}^{n} \mathcal{M}^{\lambda\mu\nu}_i\,,$$

if to each term $\Theta^{\mu\nu}_i$ an autonomous physical meaning is adscript, each part $\mathcal{M}^{\lambda\mu\nu}_i$ should be constructed in terms of $\Theta^{\mu\nu}_i$ (only), i.e., $\mathcal{M}^{\lambda\mu\nu}_i \equiv \mathcal{M}^{\lambda\mu\nu}_i(\Theta_i)$. Although this condition is rather physical, some authors dispense of it (see, e.g., Refs. 7 and 9).

The different splittings of $\Theta^{\mu\nu}_{\text{ret}}$ and $M^{\lambda\mu\nu}_{\text{ret}}$ have been considered in order to obtain (among other things) a part that can be interpreted as a radiation energy momentum tensor and an angular momentum density tensor, which leads to a local criteria of radiation. Let us state the properties we shall impose for a symmetric radiation energy momentum tensor $\Theta^{\mu\nu}_R$ and the associate angular momentum density tensor $M^{\lambda\mu\nu}_R$.

(a) The radiation angular momentum density $M^{\lambda\mu\nu}_R$ is given by (1.10).

(b) Off the PWL, we must have

$$\partial_\nu \Theta^{\mu\nu}_R = 0\,. \tag{3.5}$$

(c) Off the PWL, we must have

$$R_\nu \Theta^{\mu\nu}_R = 0\,. \tag{3.6}$$

(d) Finally,

$$P^\mu_{\text{rad}}(\tau) = P^\mu_R(\tau)\,, \tag{3.7}$$

$$M^{\lambda\mu}_{\text{rad}}(\tau) = M^{\lambda\mu}_R(\tau)\,, \tag{3.8}$$

where $P^\mu_R$ and $M^{\lambda\mu}_R$ are given by Eqs. (2.8) and (2.9), respectively.

Condition (a) has been used (see, e.g., Refs. 15 and 25), although some authors dispense of it (see, e.g., Refs. 7 and 9).

From the symmetry of $\Theta^{\mu\nu}_R$ and requirements (a) and (b), it follows that, off the PWL,

$$\partial_\nu M^{\lambda\mu\nu}_R = 0\,. \tag{3.9}$$

From requirements (a) and (c), it follows that, off the PWL,

$$R_\nu M^{\lambda\mu\nu}_R = 0\,. \tag{3.10}$$

Requirements (b) and (c) are widely used, and have

been discussed in the literature (see, e.g., Refs. 5, 7, 9, and 26).

Condition (3.7) is always used in the literature,[5,7,9,15,16] but condition (3.8) is not imposed [together with requirement (a)]. However, *both* conditions are pretty physical, since they are crucial for the interpretation of $\Theta^{\mu\nu}_R$ (and its corresponding $M^{\lambda\mu\nu}_R$) as a *radiation* part of the energy momentum tensor, i.e., with these tensors we must obtain the radiated four-momentum $P^\mu_{\rm rad}$ *and* radiated angular momentum tensor $M^{\lambda\mu}_{\rm rad}$ given by relations (3.3) and (3.4).

Requirement (d) imposes severe restrictions on the admissible $\Theta^{\mu\nu}_R$ tensors.

It is very important to observe that for a *general* $\Theta^{\mu\nu}_R$ and $M^{\lambda\mu\nu}_R$ for which conditions (a)–(d) hold, instead of evaluating $P^\mu_R$ and $M^{\lambda\mu}_R$ with Eqs. (2.8) and (2.9), we can use the following ones:

$$P^\mu_R(\tau)=\int_{\Sigma(E,\tau_0,\tau)}\Theta^{\mu\nu}_R\,d\Sigma_\nu\,, \tag{3.11}$$

$$M^{\lambda\mu}_R(\tau)=\int_{\Sigma(E,\tau_0,\tau)}M^{\lambda\mu\nu}_R\,d\Sigma_\nu\,. \tag{3.12}$$

This statement can be made because, from Eqs. (3.5) and (3.6) and Gauss' theorem, it follows directly that the evaluation of integrals in (3.11) and (3.12) does not depend on the surface chosen between the two light cones (see, e.g., Refs. 9 and 26). Then, since from Eqs. (3.6) and (3.10) it follows that $\Theta^{\mu\nu}_R\,dC_\nu=0$ and $M^{\lambda\mu\nu}_R\,dC_\nu=0$, we have that the limit $E\to\infty$ is not necessary in Eqs. (2.8) and (2.9).

Since the limit $E\to\infty$ is not involved in Eqs. (3.11) and (3.12), we have through them a *local* radiation criteria (for discussions on local radiation criteria, see, e.g., Refs. 4, 5, and 26).

For a radiation energy momentum tensor consistent with requirements (a)–(d), we propose

$$\Theta^{\mu\nu}_R\equiv\Theta^{\mu\nu}_S+\Theta^{\mu\nu}_L\,. \tag{3.13}$$

For this particular choice, Eqs. (3.5) and (3.6) have been proved in Refs. 15 and 16, and Eqs. (3.7) and (3.8) follow by a direct evaluation of the integrals.

In the literature (see, e.g., Refs. 5, 7, 9, 15, 16, and 25) the tensor $\Theta_L$ has been the standard choice for the radiation part of the energy momentum tensor, which satisfies[5] Eqs. (3.5)–(3.7). However, this tensor is not consistent with (a) and (3.8). For example, in Refs. 15, 16, and 25, requirement (a) is accepted, but (3.8) does not hold; in Refs. 7 and 9, (3.8) is satisfied, but their angular momentum density leading to $M^{\lambda\mu}_R$ is not constructed in terms of the radiated part of the energy momentum tensor ($\Theta_L$) alone.

Equation (3.4) suggests[7,9] that the radiation angular momentum $M^{\lambda\mu}_{\rm rad}$ is composed of two parts: one that is translationally invariant, which we shall call the *spin* part $M^{\lambda\mu}_{\rm spin}$, and the other one that depends on the origin of the reference system, which we shall call the *orbital* part $M^{\lambda\mu}_{\rm orb}$. That is,

$$M^{\lambda\mu}_{\rm rad}=M^{\lambda\mu}_{\rm spin}+M^{\lambda\mu}_{\rm orb}\,, \tag{3.14}$$

where

$$M^{\lambda\mu}_{\rm spin}(\tau)\equiv\frac{2e^2}{3}\int_{\tau_0}^{\tau}[v^\lambda(\tau'),a^\mu(\tau')]\,d\tau'\,. \tag{3.15}$$

The term $M^{\lambda\mu}_{\rm spin}$ is a kind of classical "analog" of the total spin (or helicity) of the photons in a quantized theory.

The tensor $\Theta^{\mu\nu}_R$ chosen in (3.13) can be split in two parts $\Theta^{\mu\nu}_S$ and $\Theta^{\mu\nu}_L$ (already defined in Sec. I), which can be interpreted as a spin energy momentum tensor and an orbital energy momentum tensor, respectively, since we have that requirements (b) and (c) hold for them, and that, if their associate $M^{\lambda\mu\nu}_S$ and $M^{\lambda\mu\nu}_L$ are obtained through requirement (a), it is easy to show that

$$M^{\lambda\mu}_{\rm spin}(\tau)=M^{\lambda\mu}_S(\tau)\,, \tag{3.16}$$

$$M^{\lambda\mu}_{\rm orb}(\tau)=M^{\lambda\mu}_L(\tau)\,, \tag{3.17}$$

where $M^{\lambda\mu}_S$ and $M^{\lambda\mu}_L$ are given by (2.9). Furthermore, $P^\mu_S$ given by (2.8) is easily evaluated and gives (cf. Ref. 15)

$$P^\mu_S(\tau)=0\,. \tag{3.18}$$

Hence it follows that

$$P^\mu_L(\tau)=P^\mu_R(\tau)\,. \tag{3.19}$$

Equation (3.18) is crucial in order to interpret $\Theta^{\mu\nu}_S$ as a *spin* energy momentum tensor.

As for the tensor $\Theta^{\mu\nu}_R$, we can show that

$$P^\mu_S(\tau)=\int_{\Sigma(E,\tau_0,\tau)}\Theta^{\mu\nu}_S\,d\Sigma_\nu\,, \tag{3.20}$$

$$M^{\lambda\mu}_S(\tau)=\int_{\Sigma(E,\tau_0,\tau)}M^{\lambda\mu\nu}_S\,d\Sigma_\nu\,, \tag{3.21}$$

$$P^\mu_L(\tau)=\int_{\Sigma(E,\tau_0,\tau)}\Theta^{\mu\nu}_L\,d\Sigma_\nu\,, \tag{3.22}$$

$$M^{\lambda\mu}_L(\tau)=\int_{\Sigma(E,\tau_0,\tau)}M^{\lambda\mu\nu}_L\,d\Sigma_\nu\,, \tag{3.23}$$

the integrals being independent of the surface chosen between the two light cones, and hence providing *local* criteria.

## IV. BOUND PART OF THE LINEAR AND ANGULAR MOMENTUM

From the bare energy momentum tensor $\Theta^{\mu\nu}_0$ its corresponding $M^{\lambda\mu\nu}_0$, and Eqs. (2.8) and (2.9), the bare four-momentum $P^\mu_0$ and the angular momentum tensor $M^{\lambda\mu}_0$ are obtained. They are

$$P^\mu_0(\tau)=m_0(\tau)v^\mu(\tau)\,, \tag{4.1}$$

$$M^{\lambda\mu}_0(\tau)=[z^\lambda(\tau),P^\mu_0(\tau)]\,, \tag{4.2}$$

where $m_0(\tau)$ is the particle bare mass.

Using definition (2.9) for $\mathcal{S}^{\lambda\mu\nu}_0$ defined in (1.13), it is found that

$$\mathcal{S}^{\lambda\mu}_0(\tau)=S^{\lambda\mu}(\tau)\,. \tag{4.3}$$

Then, the bare angular momentum tensor is given by

$$\mathcal{M}^{\lambda\mu}_0(\tau)=M^{\lambda\mu}_0(\tau)+\mathcal{S}^{\lambda\mu}_0(\tau)\,. \tag{4.4}$$

Let us notice that

$$P^\mu_0=\int_{C(\tau)}\Theta^{\mu\nu}_0\,dC_\nu\,, \tag{4.5}$$

$$\mathcal{M}^{\lambda\mu}_0=\int_{C(\tau)}\mathcal{M}^{\lambda\mu\nu}_0\,dC_\nu\,. \tag{4.6}$$

It is now natural to define $P^\mu_{\rm bound}$, the bound part of the

electromagnetic four-momentum of the particle, and $M^{\lambda\mu}_{\text{bound}}$, the bound part of the electromagnetic angular momentum tensor of the particle, as

$$P^{\mu}_{\text{bound}} \equiv P^{\mu}_{\text{ret}} - P^{\mu}_{\text{rad}} , \tag{4.7}$$

$$M^{\lambda\mu}_{\text{bound}} \equiv M^{\lambda\mu}_{\text{ret}} - M^{\lambda\mu}_{\text{rad}} , \tag{4.8}$$

where $P^{\lambda\mu}_{\text{ret}}$ and $M^{\lambda\mu}_{\text{ret}}$ are evaluated from $\Theta^{\mu\nu}_{\text{ret}}$ and its corresponding $M^{\lambda\mu\nu}_{\text{ret}}$, through Eqs. (2.8) and (2.9), and where $P^{\mu}_{\text{rad}}$ and $M^{\lambda\mu}_{\text{rad}}$ are defined by Eqs. (3.1) and (3.2). Hence it follows that

$$P^{\mu}_{\text{bound}}(\tau) = \int_{C(\tau)} \Theta^{\mu\nu}_{\text{ret}}\, dC_{\nu} , \tag{4.9}$$

$$M^{\lambda\mu}_{\text{bound}}(\tau) = \int_{C(\tau)} M^{\lambda\mu\nu}_{\text{ret}}\, dC_{\nu} . \tag{4.10}$$

For $\Theta^{\mu\nu}_{\text{ret}}$ and its corresponding $M^{\lambda\mu\nu}_{\text{ret}}$ we have seen that definitions (2.8) and (2.9) are particularly suited to discuss the bound and radiated parts in integral form. The integral on $C(\tau)$ of definitions (2.8) and (2.9) are explicitly independent on the history of the particle as it should be for the consistency of the interpretation as *bound* parts; cf. (4.9) and (4.10). The remaining integrals in definitions (2.8) and (2.9) correspond to the standard identification as *radiation* parts [cf. (3.1) and (3.2)].

We know that the evaluation of Eqs. (4.9) and (4.10) gives an infinite $P^{\mu}_{\text{bound}}$ and $M^{\lambda\mu}_{\text{bound}}$, because of the singularity at the PWL. In order to deal with these infinities[8,27] the integrals in Eqs. (4.9) and (4.10) are evaluated on a surface that avoids the PWL and that depends on a scalar $\varepsilon$ ($\varepsilon > 0$), such that under the limit $\varepsilon \to 0$, it goes to $C(\tau)$. Then, once the integrals are evaluated on the chosen surface (with $\varepsilon > 0$), the standard renormalization[5,8,11,27] procedure is performed with $P^{\mu}_{0}$ and $\mathcal{M}^{\lambda\mu}_{0}$ (where the limit $\varepsilon \to 0$ is involved).

We know[8,12,27,28] that the results obtained with the above-mentioned procedure depend on the chosen surface (i.e., on the chosen way to reach the singularity). We shall consider two different approaches here.

(a) A spacelike approach (cf. Ref. 8) consists of evaluating

$$\lim_{E\to\infty} \int_{\mathscr{C}(\tau-\varepsilon,\rho_0,E)} \Theta^{\mu\nu}_{\text{ret}}\, dC_{\nu} , \tag{4.11}$$

$$\lim_{E\to\infty} \int_{\mathscr{C}(\tau-\varepsilon,\rho_0,E)} M^{\lambda\mu\nu}_{\text{ret}}\, dC_{\nu} , \tag{4.12}$$

where $\mathscr{C}(\tau-\varepsilon,\rho_0,E)$ has been defined in Sec. I. Notice that

$$\lim_{\varepsilon\to 0} \mathscr{C}(\tau-\varepsilon,\rho_0,E) = C(\tau,0,E) .$$

The evaluation of (4.11) gives[8]

$$\frac{e^2}{6} \frac{[4v(\tau)\cdot v(\tau-\varepsilon)v^{\mu}(\tau-\varepsilon) + v^{\mu}(\tau)]}{v(\tau)\cdot[z(\tau)-z(\tau-\varepsilon)]} \equiv f^{\mu}_{\tau}(\varepsilon) . \tag{4.13}$$

The evaluation of (4.12) gives[8]

$$[z^{\lambda}(\tau-\varepsilon), f^{\mu}_{\tau}(\varepsilon)] . \tag{4.14}$$

Then, through the mass renormalization procedure, it is obtained[8] that

$$P^{\mu}_{0}(\tau) + P^{\mu}_{\text{bound}}(\tau) = \lim_{\varepsilon\to 0} [P^{\mu}_{0} + f^{\mu}_{\tau}(\varepsilon)] = m(\tau)v^{\mu}(\tau) - \tfrac{2}{3}e^2 a^{\mu}(\tau) , \tag{4.15}$$

and that

$$\mathcal{M}^{\lambda\mu}_{0}(\tau) + M^{\lambda\mu}_{\text{bound}}(\tau) = \lim_{\varepsilon\to 0} \{\mathcal{M}^{\lambda\mu}_{0} + [z^{\lambda}(\tau-\varepsilon), f^{\mu}_{\tau}(\varepsilon)]\} = [z^{\lambda}(\tau), m(\tau)v^{\mu}(\tau) - \tfrac{2}{3}e^2 a^{\mu}(\tau)] + \mathcal{S}^{\lambda\mu}_{0}(\tau) , \tag{4.16}$$

where $m(\tau)$ is the renormalized mass.

(b) A lightlike approach (cf. Ref. 27) consists of evaluating

$$\lim_{E\to\infty} \int_{C(\tau,\varepsilon,E)} \Theta^{\mu\nu}_{\text{ret}}\, dC_{\nu} , \tag{4.17}$$

$$\lim_{E\to\infty} \int_{C(\tau,\varepsilon,E)} M^{\lambda\mu\nu}_{\text{ret}}\, dC_{\nu} . \tag{4.18}$$

The evaluation of (4.17) gives[27]

$$(e^2/2\varepsilon)\, v^{\mu}(\tau) . \tag{4.19}$$

The evaluation of (4.18) is straightforward, and gives

$$[z^{\lambda}(\tau), (e^2/2\varepsilon)v^{\mu}(\tau)] . \tag{4.20}$$

Then, through the mass renormalization procedure, it is obtained[27] that

$$P^{\mu}_{0}(\tau) + P^{\mu}_{\text{bound}}(\tau) = \lim_{\varepsilon\to 0} [P^{\mu}_{0} + (e^2/2\varepsilon)v^{\mu}] = m(\tau)v^{\mu}(\tau) . \tag{4.21}$$

Also, it follows that

$$\mathcal{M}^{\lambda\mu}_{0}(\tau) + M^{\lambda\mu}_{\text{bound}}(\tau) = \lim_{\varepsilon\to 0} \{\mathcal{M}^{\lambda\mu}_{0} + [z^{\lambda}, (e^2/2\varepsilon)v^{\mu}]\} = [z^{\lambda}(\tau), m(\tau)v^{\mu}(\tau)] + \mathcal{S}^{\lambda\mu}_{0}(\tau) . \tag{4.22}$$

We want to close this section discussing the possibility of having an energy momentum tensor $\Theta^{\mu\nu}_{b}$ that can be interpreted as a *bound* part of $\Theta^{\mu\nu}_{\text{ret}}$. That is, a $\Theta^{\mu\nu}_{b}$ such that

$$P^{\mu}_{b}(\tau) = P^{\mu}_{\text{bound}}(\tau) \tag{4.23}$$

and

$$M^{\lambda\mu}_{b}(\tau) = M^{\lambda\mu}_{\text{bound}}(\tau) , \tag{4.24}$$

where $P^{\mu}_{b}$ and $M^{\lambda\mu}_{b}$ are evaluated with $\Theta^{\mu\nu}_{b}$ and its associated $M^{\lambda\mu\nu}_{b}$, through Eqs. (2.8) and (2.9). Also, off the PWL, we must have

$$\partial_{\nu}\Theta^{\mu\nu}_{b} = 0 . \tag{4.25}$$

Given an arbitrary radiation energy momentum tensor $\Theta^{\mu\nu}_{R}$ satisfying requirements (a)–(d) set in Sec. III, it follows trivially that

$$\Theta^{\mu\nu}_{b} \equiv \Theta^{\mu\nu}_{\text{ret}} - \Theta^{\mu\nu}_{R} \tag{4.26}$$

is a tensor that can be interpreted as a bound part of $\Theta^{\mu\nu}_{\text{ret}}$. Furthermore, it also follows [Eq. (4.26)] that

$$P^{\mu}_{b}(\tau) = \int_{C(\tau)} \Theta^{\mu\nu}_{b}\, dC_{\nu} , \tag{4.27}$$

$$M_b^{\lambda\mu}(\tau) = \int_{C(\tau)} M_b^{\lambda\mu\nu}\, dC_\nu\,, \tag{4.28}$$

as it is evident from Eqs. (3.6)–(3.8) and (3.10). With our chosen radiation energy momentum tensor $\Theta_R^{\mu\nu}$ given by (3.13) it follows that the bound energy momentum tensor is given by $\Theta_B^{\mu\nu}$ [see Eq. (1.7)], i.e., $\Theta_B \equiv \Theta_b$.

## V. EQUATION OF MOTION

With $\Theta_{\text{mix}}^{\mu\nu}$ and its corresponding $M_{\text{mix}}^{\lambda\mu\nu}$, we shall evaluate $P_{\text{mix}}^{\mu}$ and $M_{\text{mix}}^{\lambda\mu}$ defined by Eqs. (2.8) and (2.9). Since off the PWL $\partial_\nu \Theta_{\text{mix}}^{\mu\nu} = 0$, using Gauss' theorem on Fig. 1, we obtain

$$\int_{\sigma_c} \Theta_{\text{mix}}^{\mu\nu}\, dC_\nu + \int_{\Sigma(E,\tau_0,\tau)} \Theta_{\text{mix}}^{\mu\nu}\, d\Sigma_\nu = \int_{\sigma_c'} \Theta_{\text{mix}}^{\mu\nu}\, dC_\nu + \int_{B(E',\tau_0,\tau)} \Theta_{\text{mix}}^{\mu\nu}\, d\Sigma_\nu\,. \tag{5.1}$$

Let us call

$$\mathscr{P}^\mu(\tau_0) \equiv \int_{C(\tau_0)} \Theta_{\text{mix}}^{\mu\nu}\, dC_\nu\,. \tag{5.2}$$

Performing the limits $E' \to 0$ and $E \to \infty$ in (5.1), we obtain

$$P_{\text{mix}}^{\mu}(\tau) = -e\int_{\tau_0}^{\tau} F_{\text{ext}}^{\mu\nu}(\tau')v_\nu(\tau')d\tau' + \mathscr{P}^\mu(\tau_0)\,, \tag{5.3}$$

where

$$F_{\text{ext}}^{\mu\nu}(\tau) \equiv F_{\text{ext}}^{\mu\nu}[z(\tau)] \equiv \lim_{E\to 0} \frac{1}{4\pi}\int d\Omega\, F_{\text{ext}}^{\mu\nu}(E,\tau,\Omega)\,. \tag{5.4}$$

In the same way, we obtain that

$$M_{\text{mix}}^{\lambda\mu}(\tau) = -e\int_{\tau_0}^{\tau} [z^\lambda(\tau'), F_{\text{ext}}^{\mu\alpha}(\tau')] \times v_\alpha(\tau')d\tau' + \mathscr{N}^{\lambda\mu}(\tau_0)\,, \tag{5.5}$$

where

$$\mathscr{N}^{\lambda\mu}(\tau_0) \equiv \int_{C(\tau_0)} M_{\text{mix}}^{\lambda\mu\nu}\, dC_\nu\,. \tag{5.6}$$

Let us observe that we have

$$\lim_{\tau_0 \to -\infty} \mathscr{P}^\mu(\tau_0) = 0 \quad \text{and} \quad \lim_{\tau_0 \to -\infty} \mathscr{N}^{\lambda\mu}(\tau_0) = 0\,,$$

because of the properties usually[12] assumed for $F_{\text{ext}}^{\mu\nu}$. Furthermore, if it is the case that $F_{\text{ext}}^{\mu\nu} = 0$ on $C(\tau_0)$, we shall have that $\mathscr{P}^\mu(\tau_0) = 0$ and $\mathscr{N}^{\lambda\mu}(\tau_0) = 0$ automatically.

The total four-momentum $P^\mu$ and the total angular momentum tensor $M^{\lambda\mu}$, given by

$$P^\mu(\tau) = P_0^\mu(\tau) + P_B^\mu(\tau) + P_R^\mu(\tau) + P_{\text{mix}}^\mu(\tau) + P_{\text{ext}}^\mu(\tau)\,, \tag{5.7}$$

$$M^{\lambda\mu}(\tau) = \mathscr{M}_0^{\lambda\mu}(\tau) + M_B^{\lambda\mu}(\tau) + M_R^{\lambda\mu}(\tau) + M_{\text{mix}}^{\lambda\mu}(\tau) + M_{\text{ext}}^{\lambda\mu}(\tau) \tag{5.8}$$

are now completely specified. From the conservation law proved for them in Sec. II B, it follows that

$$\dot{P}^\mu(\tau) = 0\,, \quad \dot{M}^{\lambda\mu}(\tau) = 0\,. \tag{5.9}$$

As we shall see, the consistency of the equations in (5.9) will imply some restrictions on the phenomenology of the particle (together with its equation of motion), and hence there is no need to impose from the beginning strong assumptions about the phenomenology.

From the discussion in Sec. IV, it is clear that different evaluations of $P_0^\mu + P_B^\mu$ and $\mathscr{M}_0^{\lambda\mu} + M_B^{\lambda\mu}$ (which depend on the way the PWL is reached) will lead to different expressions for the Eqs. (5.7) and (5.8). We shall discuss the two approaches taken in Sec. IV.

(a) With the spacelike approach, it follows from Eqs. (3.3), (3.4), (4.15), (4.16), (5.3), (5.5), (5.7), and (5.8) and the fact that $P_{\text{ext}}^\mu$ and $M_{\text{ext}}^{\lambda\mu}$ are independently conserved quantities (see, Sec. II B) that

$$\dot{P}^\mu(\tau) = \frac{d}{d\tau}[m(\tau)v^\mu(\tau)] - \frac{2}{3}e^2\dot{a}^\mu(\tau) + \frac{2}{3}e^2a^2(\tau)v^\mu(\tau) - eF_{\text{ext}}^{\mu\nu}(\tau)v_\nu(\tau)\,, \tag{5.10}$$

$$\dot{M}^{\lambda\mu}(\tau) = [z^\lambda(\tau), \dot{P}^\mu(\tau)] + \dot{\mathscr{S}}_0^{\lambda\mu}(\tau)\,. \tag{5.11}$$

The conservation law for $P^\mu$ applied to the expression (5.10) leads to a *conserved* renormalized mass $m(\tau)$ and to the Lorentz–Dirac equation of motion. Furthermore, (5.11) shows that the conservations laws for $P^\mu$ and $M^{\lambda\mu}$ are compatible only for a particle with an independently *conserved* intrinsic angular momentum $\mathscr{S}_0^{\lambda\mu}(\tau)$.

(b) With the lightlike approach, it follows from Eqs. (3.3), (3.4), (4.21), (4.22), (5.3), (5.5), (5.7), and (5.8) and the fact that $P_{\text{ext}}^\mu$ and $M_{\text{ext}}^{\lambda\mu}$ are independently conserved quantities that

$$\dot{P}^\mu(\tau) = \frac{d}{d\tau}[m(\tau)v^\mu(\tau)] + \frac{2}{3}e^2a^2(\tau)v^\mu(\tau) - eF_{\text{ext}}^{\mu\nu}(\tau)v_\nu(\tau)\,, \tag{5.12}$$

$$\dot{M}^{\lambda\mu}(\tau) = [z^\lambda(\tau), \dot{P}^\mu(\tau)] + \tfrac{2}{3}e^2[v^\lambda(\tau), a^\mu(\tau)] + \dot{\mathscr{S}}_0^{\lambda\mu}(\tau)\,. \tag{5.13}$$

The conservation law for $P^\mu$ applied to the expression (5.12) leads to the Bonnor equation of motion[29] and hence to a *variable* renormalized mass $m(\tau)$. For physical discussions on this equation see, e.g., Refs. 14, 27, and 29.

Equation (5.13) shows that the conservation laws for $P^\mu$ and $M^{\lambda\mu}$ are compatible only for a particle with a *variable* intrinsic angular momentum $\mathscr{S}_0^{\lambda\mu}(\tau)$, such that

$$\dot{\mathscr{S}}_0^{\lambda\mu}(\tau) = -\tfrac{2}{3}e^2[v^\lambda(\tau), a^\mu(\tau)]\,. \tag{5.14}$$

In this case, the change in the spin part of the radiated field [see Eq. (3.15)] is supplied by the change in the intrinsic *mechanical* angular momentum of the particle [see Eq. (5.14)], just as the energy rate of radiation $\mathscr{R}$ emitted by the particle (i.e., $\mathscr{R} \equiv -v_\mu \dot{P}_{\text{rad}}^\mu$) is supplied by the change in the renormalized mass.[29]

It has been said[30] that Bonnor's equation does not conserve angular momentum. However, as we have seen, if an intrinsic (mechanical) angular momentum for the particle is assumed such that Eq. (5.14) holds for it, Bonnor's equation is consistent with both the conservation of four-momentum and angular momentum. The fact that Bonnor's equation is not in conflict with the conservation laws, provided that the particle has an intrinsic angular momentum, has been pointed out before; see the second footnote on p. 47 of Ref. 9.

[1]C. Møller, *The Theory of Relativity* (Oxford U.P., New York, 1969).
[2]A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).
[3]N. N. Bogoliubov and D. V. Shirkov, *Introduction to Theory of Quantized Fields* (Wiley–Interscience, New York, 1980).
[4]F. Rohrlich, *Classical Charged Particles* (Addison–Wesley, Reading, MA, 1965).
[5]C. Teitelboim, Phys. Rev. D **1**, 1572 (1970).
[6]Ch. G. Van Weert, Physica **76**, 345 (1974).
[7]C. A. López and D. Villarroel, Phys. Rev. D **11**, 2724 (1975).
[8]R. Tabensky and D. Villarroel, J. Math. Phys. **16**, 1380 (1975).
[9]C. Teitelboim, D. Villarroel, and Ch. G. Van Weert, Riv. Nuovo Cimento **3**, 1 (1980).
[10]A. Lozada and P. L. Torres, J. Math. Phys. **29**, 1361 (1988).
[11]P. A. M. Dirac, Proc. R. Soc. London Ser. A **167**, 148 (1938).
[12]R. Tabensky, Phys. Rev. D **13**, 267 (1976).
[13]J. Schwinger, Found. Phys. **13**, 373 (1983).
[14]E. G. P. Rowe, Phys. Rev. D **12**, 1576 (1975).
[15]C. A. López, Phys. Rev. D **17**, 2004 (1978).
[16]E. G. P. Rowe, Phys. Rev. D **18**, 3639 (1978).
[17]H. J. Bhabha, Proc. R. Soc. London Ser. A **172**, 384 (1939).
[18]W. H. Fleming, *Functions of Several Variables* (Springer, New York, 1977).
[19]V. A. Ilyin and E. G. Poznyak, *Fundamentals of Mathematical Analysis* (Mir, Moscow, 1982).
[20]J. N. Goldberg and R. P. Kerr, J. Math. Phys. **5**, 172 (1964).
[21]P. A. Hogan, Nuovo Cimento B **15**, 136 (1973); **19**, 122 (1964).
[22]A. Schild, J. Math. Anal. Appl. **1**, 127 (1960).
[23]J. Cohn, J. Math. Phys. **10**, 799 (1969).
[24]Ch. G. Van Weert, Physica **66**, 79 (1973).
[25]E. G. P. Rowe, Phys. Rev. D **20**, 580 (1979).
[26]D. Villarroel, Ann. Phys. (NY) **90**, 113 (1975).
[27]E. Marx, Int. J. Theor. Phys. **14**, 55 (1975).
[28]A. Lozada and P. L. Torres, Phys. Rev. D **35**, 2064 (1987).
[29]W. B. Bonnor, Proc. R. Soc. London Ser. A **337**, 591 (1974).
[30]E. Honig and G. Szamosi, Phys. Lett. A **93**, 319, 1983.

# Lagrangian variational principle in stochastic mechanics: Gauge structure and stability

Maria I. Loffredo[a)]
*Department of Mathematics, Princeton University, Princeton, New Jersey 08544*

Laura M. Morato
*Dipartimento di Fisica "G. Galilei," Universitá di Padova, I-35131 Padova, Italy*

(Received 6 June 1988; accepted for publication 5 October 1988)

The Lagrangian variational principle with the classical action leads, in stochastic mechanics, to Madelung's fluid equations, if only irrotational velocity fields are allowed, while new dynamical equations arise if rotational velocity fields are also taken into account. The new equations are shown to be equivalent to the (gauge invariant) system of a Schrödinger equation involving a four-vector potential $(\mathbf{A},\Phi)$ and the coupled evolution equation (of magnetohydrodynamical type) for the vector field $\mathbf{A}$. A general energy theorem can be proved and the stability properties of irrotational and rotational solutions investigated.

## I. INTRODUCTION

As shown in a series of recent papers,[1–4] stochastic variational principles with classical action appear to be a natural way of introducing quantum dynamics.

This must be understood as a completion of the work initiated by Nelson in 1966, who gave the first (Newtonian) dynamical formulation of stochastic mechanics.[5]

After the theorem by Carlen[6] on the existence of nonsmooth extremal diffusions for the above-mentioned variational problems (see also Refs. 7–9), one can claim that stochastic mechanics represents nowadays a mathematically well-founded reformulation, *with real time*, of Feynman's path integral.

It turns out that, like in hydrodynamics, two different types of variational principles can be stated: the Eulerian principle, where the action functional is written in terms of a velocity field, and the Lagrangian one, where the action is seen as a functional on the paths.

In hydrodynamics, even in the simplest case of an ideal fluid, the two principles are not trivially equivalent, since in the Eulerian principle the solutions are restricted to the set of irrotational flows, unless, as discovered by Lin in 1963, the identity of the fluid particles is taken as an explicit constraint.[10]

The problem of considering also rotational motions does not seem to be of any interest in stochastic mechanics, since the fluid-dynamical version of the Schrödinger equation, that is Madelung's fluid equations, describes the evolution of an irrotational flow.

Indeed Madelung's equations are recovered in the Eulerian picture without any additional constraint,[1] while for the Lagrangian problem irrotationality must be explicitly assumed.[3]

Moreover, the whole class of Lagrangian solutions can be explicitly calculated, leading to new (quantum) equations of motion.[4]

In the classical limit such equations do not reduce themselves to the Hamilton–Jacobi fluid equations, but to the more general Euler equations, where rotational velocity fields are also allowed.[4]

In this work we study some mathematical aspects of the unrestricted Lagrangian picture.

In particular, we shall show that the new equations can be transformed into a Schrödinger equation for a quantum particle in a vector potential, whose time evolution is coupled with that of the wave function. This also leads to a gauge invariant form of the original dynamical equations.

Finally, by exploiting the first result and variational techniques we prove a general energy theorem.

## II. THE LAGRANGIAN VARIATIONAL PRINCIPLE IN STOCHASTIC MECHANICS

For the sake of simplicity we shall limit the discussion to the simple Lagrangian of a particle of mass $m$ in a scalar potential $\Phi^{\text{ext}}$.

The classical action functional, for the time interval $[t_a,t_b]$, is

$$A^{\text{cl}}_{[t_a,t_b]}(\mathbf{q}(\cdot)) = \int_{t_a}^{t_b} \left[\frac{1}{2} m\dot{\mathbf{q}}(t)^2 - \Phi^{\text{ext}}(\mathbf{q}(t),t)\right] dt\,, \qquad (2.1)$$

where $\mathbf{q}(t)$ is the position of the particle at time $t$.

In stochastic mechanics the classical kinematical position defining the velocity $\mathbf{v}(t)$,

$$\dot{\mathbf{q}}(t) = \mathbf{v}(t)\,, \qquad (2.2)$$

is generalized through the stochastic differential equation

$$d\mathbf{q}(t) = \mathbf{v}_+(\mathbf{q}(t),t)dt + (\hbar/m)^{1/2}\, d\mathbf{w}(t)\,, \qquad (2.3)$$

where $\mathbf{w}(t)$ is a standard Wiener process and the drift $\mathbf{v}_+$ is a velocity field, to be determined by the initial conditions and dynamical constraints.

Denoting by $\{t_i\}_{i=1,\ldots,N}$ an equipartition of the interval $[t_a,t_b]$, and by $\Delta$ the time difference $(t_b - t_a)/N$, the mean discretized action is defined by

[a)] On leave from Dipartimento di Matematica, Universitá di Siena, I-53100 Siena, Italy.

 0022-2488/89/020354-07$02.50 © 1989 American Institute of Physics 

$$A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot)) = \mathscr{E}\left\{\sum_{i=1}^{N}\left(\frac{1}{2}m\frac{(\mathbf{q}(t_{i+1})-\mathbf{q}(t_i))^2}{\Delta^2} - \Phi^{\text{ext}}(\mathbf{q}(t_i),t_i)\right)\Delta\right\}, \tag{2.4}$$

where $\mathscr{E}$ denotes the integration with respect to the Wiener measure.

As shown by Nelson,[2] (2.3) can be iterated to estimate $\mathbf{q}(t_{i+1})-\mathbf{q}(t_i)$ to the order $\Delta^{3/2}$, so that the following decomposition holds:

$$\mathscr{E}_{\mathbf{q}(t)}\left(\frac{1}{2}m\frac{(\mathbf{q}(t_{i+1})-\mathbf{q}(t_i))^2}{\Delta^2}\right) = \mathscr{E}_{\mathbf{q}(t)}\left(\frac{1}{2}m\mathbf{v}_+^2 + \frac{\hbar}{2}\nabla\cdot\mathbf{v}_+ + \frac{3}{2}\frac{\hbar}{\Delta} + o(\Delta)\right), \tag{2.5}$$

where $\mathscr{E}_{\mathbf{q}(t)}$ denotes the conditional expectation given the $\sigma$ algebra generated by $\mathbf{q}(t)$.

Let now $\mathbf{q}(t)$ be a smooth diffusion in the sense of Nelson[2] and Carlen,[6] so that it has a positive definite probability density. Then the diffusion has a backward representation

$$d\mathbf{q}(t) = \mathbf{v}_-(\mathbf{q}(t),t)dt + (\hbar/m)^{1/2}\,d\mathbf{w}^*(t), \tag{2.6}$$

where $\mathbf{w}^*(t)$ is a "reversed" Wiener process and $\mathbf{v}_-$ the backward drift, that, if we denote by $\rho$ the probability density of $\mathbf{q}(t)$, is defined by the equality

$$\mathbf{v}_- = \mathbf{v}_+ - (\hbar/m)\nabla\ln\rho. \tag{2.7}$$

Then, by means of straightforward calculations one can get from (2.5) the equality

$$\mathscr{E}\left(\frac{1}{2}m\frac{(\Delta^+\mathbf{q}(t_i))^2}{\Delta^2}\right) = \mathscr{E}\left(\frac{1}{2}m\mathbf{v}_+\cdot\mathbf{v}_-\right) + \frac{3}{2}\frac{\hbar}{\Delta} + o(\Delta), \tag{2.8}$$

that, by the Markov property of $\mathbf{q}(t)$, is equivalent to

$$\mathscr{E}\left(\frac{1}{2}m\frac{(\Delta^+\mathbf{q}(t_i))^2}{\Delta^2}\right) = \mathscr{E}\left(\frac{1}{2}m\frac{\Delta^+\mathbf{q}(t_i)\cdot\Delta^-\mathbf{q}(t_i)}{\Delta^2}\right) + \frac{3}{2}\frac{\hbar}{\Delta} + o(\Delta), \tag{2.9}$$

where we have defined $\Delta^+\mathbf{q}(t)\equiv\mathbf{q}(t+\Delta)-\mathbf{q}(t)$ and $\Delta^-\mathbf{q}(t)\equiv\mathbf{q}(t)-\mathbf{q}(t-\Delta)$. Equalities (2.5) and (2.8) are typical of the Eulerian picture, while (2.9) leads to the Lagrangian one.

In particular, (2.5) allows one to consider the regularized action in the continuum limit as a functional of the drift field $\mathbf{v}_+$.

Thus the variational principle of stochastic mechanics in the Eulerian form can be stated by asking the stationarity of the mean classical action with respect to variations of the velocity field $\mathbf{v}_+$ (that, of course, cancels the divergent term). This has been done in Ref. 1, by exploiting a type of stochastic control technique. One gets the following sufficient and necessary conditions for the action to be stationary:

$$\frac{\mathbf{v}_+ + \mathbf{v}_-}{2} = \frac{\nabla S}{m},$$
$$\partial_t\rho = -\nabla\cdot\left(\frac{\nabla S}{m}\rho\right), \tag{2.10}$$
$$\partial_t S + \frac{(\nabla S)^2}{2m} - \frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}} + \Phi^{\text{ext}} = 0.$$

The more general case of a four-vector potential $(\mathbf{A}^{\text{ext}},\Phi^{\text{ext}})$ does not give rise to difficulties, and yields[2]

$$\frac{\mathbf{v}_+ + \mathbf{v}_-}{2} + \frac{\mathbf{A}^{\text{ext}}}{m} = \frac{\nabla S}{m},$$
$$\partial_t\rho = -\nabla\cdot\left(\frac{\nabla S - \mathbf{A}^{\text{ext}}}{m}\rho\right), \tag{2.10'}$$
$$\partial_t S + \frac{(\nabla S - \mathbf{A}^{\text{ext}})^2}{2m} - \frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}} + \Phi^{\text{ext}} = 0.$$

It is not surprising that the irrotationality condition comes from the principle.

Extending Lin's constraint to this case should be possible in principle, but we do not do this in the present work.

By going to the Lagrangian picture, we are forced to perform the calculus of variations with the discretized time: in fact, $\mathbf{q}(t)$ has no differentiable sample paths, so that the most strict analog of $\dot{\mathbf{q}}(t)$ is the ratio $(\mathbf{q}(t_{i+1})-\mathbf{q}(t_i))/\Delta$.

Consequently we consider the discretized action (2.4) and use the decomposition of the kinetic term (2.9).

The procedure is the following: for every Wiener path the corresponding one for the diffusion $\mathbf{q}(t)$ is varied by means of an arbitrary change of the drift. Then, by estimating the stochastic increments to the order $\Delta^{3/2}$, a discrete "integration by parts" is performed. Finally one integrates over all paths with respect to the Wiener measure and goes to the continuum limit.

The variations are stochastic processes, which can be characterized as follows: by setting

$$\delta\mathbf{q}(t) = \epsilon\mathbf{h}(t), \tag{2.11}$$

so that

$$\delta\mathbf{v}_+(\mathbf{q}(t),t) = \epsilon\mathbf{f}(\mathbf{q}(t),t) + o(\epsilon) \tag{2.12}$$

for a certain vector field $\mathbf{f}$, one gets that $\mathbf{h}(t)$ satisfies the linear differential equation

$$\dot{\mathbf{h}}(t) = \sum_{j=1}^{3}(\partial_j\mathbf{v}_+(\mathbf{q}(t),t))h_j(t) + \mathbf{f}(\mathbf{q}(t),t), \quad \partial_j\equiv\frac{\partial}{\partial x_j}. \tag{2.13}$$

As a consequence $\mathbf{h}(t)$ is not a Markov process and has differentiable sample paths.

Since Eq. (2.13) is of first order, we can only impose either $\mathbf{h}(t_a)=0$ *or* $\mathbf{h}(t_b)=0$. Nevertheless we can fix, for example, like the classical case, the initial position and the final momentum.

In this case, therefore, with the proper definition of the action functional

$$A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b}) = \mathscr{E}\left\{\left(\sum_{i=1}^{N}\frac{1}{2}m\frac{(\Delta^+\mathbf{q}(t_i))^2}{\Delta^2} - \Phi^{\text{ext}}(\mathbf{q}(t_i),t_i)\right)\Delta + \mathbf{p}_{t_b}\cdot\mathbf{q}(t_b)\right\}, \tag{2.14}$$

the Lagrangian variational problem is formulated as follows: for any $[t_a,t_b]$ find a (smooth) diffusion $\mathbf{q}(\cdot,\cdot)$ such that

$$\lim_{N\uparrow\infty} \delta A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b}) = o(\epsilon)\,, \tag{2.15}$$

for all admissible variations such that $\delta\mathbf{q}(t_a)=0$.

Since, like in the Eulerian case, the variations leave the diffusion coefficient unchanged, the variation is finite in the continuum limit.

One could say that, up to a proper final condition, and some attention to the measurability properties of $\mathbf{h}(t)$, the procedure (with discretized time) to get a sufficient condition for the stationarity of the action is conceptually the classical one.

On the contrary, some troubles arise in proving the necessity of the same condition; since $\mathbf{h}(t)$ is nonmeasurable with respect to the $\sigma$ algebra generated by $\mathbf{q}(t)$, the standard argument fails.

We summarize the sufficient condition in the following theorem.

**Theorem 1:** Let $\mathbf{q}(t)$ be a smooth diffusion with stochastic differential equation (2.3). Also let $\rho$ denote the probability density of $\mathbf{q}(t)$, $\mathbf{v}_-$ the backward drift, and set

$$(\mathbf{v}_+ + \mathbf{v}_-)/2 = \mathbf{v}, \quad (\hbar/2m)\nabla \ln\rho = \mathbf{u}\,. \tag{2.16}$$

Finally assume $A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b})$ to be defined by (2.14), where $\mathbf{p}_{t_b}$ is a measurable random variable. Then, if $\rho(\mathbf{q}(t),t)$ and $\mathbf{v}(\mathbf{q}(t),t)$ satisfy the system of equations

$$\partial_t\rho = -\nabla\cdot(\rho\mathbf{v})\,, \tag{2.17a}$$

$$\partial_t\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} - \left[\frac{\hbar}{2m}\nabla^2\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}\right] - \left(\mathbf{u} + \frac{\hbar}{2m}\nabla\right)\wedge\nabla\wedge\mathbf{v} = -\frac{\nabla\Phi^{\mathrm{ext}}}{m}\,, \tag{2.17b}$$

and the final condition $\mathbf{v}(\mathbf{q}(t_b),t_b)=\mathbf{p}_{t_b}/m$, one has

$$\lim_{N\uparrow\infty} \delta A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b}) = o(\epsilon), \tag{2.18}$$

whenever the variation is defined through (2.11)–(2.13) and preserves the initial position.

The proof we give here for completeness is rephrased from Refs. 3 and 4.

*Proof:* First we decompose the variation of the action as follows:

$$\delta A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b}) = \epsilon\sum_{i=1}^{N}\frac{m}{2}\mathscr{E}\left(\frac{\Delta^+\mathbf{q}(t_i)\cdot\Delta^-\mathbf{h}(t_i)}{\Delta^2} + \frac{\Delta^-\mathbf{q}(t_i)\cdot\Delta^+\mathbf{h}(t_i)}{\Delta^2} + o(\Delta)\right)\Delta$$
$$-\epsilon\sum_{i=1}^{N}\mathscr{E}(\nabla\Phi^{\mathrm{ext}}(\mathbf{q}(t_i),t_i)\cdot\mathbf{h}(t_i))\Delta$$
$$-\epsilon\mathscr{E}(\mathbf{p}_{t_b}\cdot\mathbf{h}(t_b)) + o(\epsilon)\,.$$

The divergent term $3\hbar/2\Delta$ does not give a contribution to the variation and, since $\Delta^{\pm}\mathbf{h}(t)$ are of first order in $\Delta$, we can *a priori* neglect terms of order $\Delta^{3/2}$ in the estimate of $\Delta^{\pm}\mathbf{q}(t)$.

The analysis of the kinetic terms gives

$$\mathscr{E}\left(\frac{\Delta^+\mathbf{q}(t)\cdot\Delta^-\mathbf{h}(t)}{\Delta^2}\right) = \frac{1}{\Delta}\mathscr{E}(\mathbf{v}_+(\mathbf{q}(t),t)\cdot\Delta^-\mathbf{h}(t) + o(\Delta))$$

and

$$\mathscr{E}\left(\frac{\Delta^-\mathbf{q}(t)\cdot\Delta^+\mathbf{h}(t)}{\Delta^2}\right) = \frac{1}{\Delta}\mathscr{E}\left(\mathbf{v}_-(\mathbf{q}(t),t)\cdot\Delta^+\mathbf{h}(t) + \left(\frac{\hbar}{m}\right)^{1/2}\Delta^-\mathbf{w}_*(t)\cdot\dot{\mathbf{h}}(t) + o(\Delta)\right),$$

where we have exploited the measurability properties of $\mathbf{h}(t)$ and the fact that $\Delta^+\mathbf{w}(t)$ is independent of the past of $\mathbf{q}(t)$, while $\Delta^-\mathbf{w}_*(t)$ is independent of the future and not of the past of $\mathbf{q}(t)$. Moreover we have taken into account that the term of order $\Delta^{3/2}$ in the expansion of $\Delta^+\mathbf{h}(t)$ gives, when multiplied by $\Delta^-\mathbf{w}_*(t)$, a contribution with zero mean [see, for details, Eq. (15′) in Ref. 3].

A discrete integration by parts then gives

$$\mathbf{v}_-(\mathbf{q}(t),t)\cdot\Delta^+\mathbf{h}(t) = \Delta^+[\mathbf{v}_-(\mathbf{q}(t),t)\cdot\mathbf{h}(t)] - \Delta^+\mathbf{v}_-(\mathbf{q}(t),t)\cdot\mathbf{h}(t) + o(\Delta)$$

and

$$\mathbf{v}_+(\mathbf{q}(t),t)\cdot\Delta^-\mathbf{h}(t) = \Delta^-[\mathbf{v}_+(\mathbf{q}(t),t)\cdot\mathbf{h}(t)] - \Delta^-\mathbf{v}_+(\mathbf{q}(t),t)\cdot\mathbf{h}(t) + o(\Delta)\,.$$

We now exploit Ito's rule to get

$$\mathscr{E}(\Delta^+\mathbf{v}_-(\mathbf{q}(t),t)\cdot\mathbf{h}(t)) = \mathscr{E}(D_+\mathbf{v}_-(\mathbf{q}(t),t)\cdot\mathbf{h}(t)\Delta + o(\Delta))$$

and

$$\mathscr{E}(\Delta^-\mathbf{v}_+(\mathbf{q}(t),t)\cdot\mathbf{h}(t)) = \mathscr{E}\left(D_-\mathbf{v}_+(\mathbf{q}(t),t)\cdot\mathbf{h}(t)\Delta + \left(\frac{\hbar}{m}\right)^{1/2}\sum_{j=1}^{3}\Delta^- w^j_*\partial_j\mathbf{v}_+(\mathbf{q}(t),t)\cdot\mathbf{h}(t) + o(\Delta)\right),$$

where $D_+$ and $D_-$ denote the mean forward and backward conditioned derivatives.

Therefore, by collecting all terms, one finds

$$\lim_{N\uparrow\infty}\delta A^{N}_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b})$$
$$= \epsilon\mathscr{E}\left\{\int_{t_a}^{t_b}\left[-\frac{1}{2}m(D_+\mathbf{v}_- + D_-\mathbf{v}_+)(\mathbf{q}(t),t) - \nabla\Phi^{\mathrm{ext}}(\mathbf{q}(t),t)\right]\cdot\mathbf{h}(t)dt\right\}$$
$$+\epsilon\mathscr{E}\left\{\left[\frac{1}{2}m(\mathbf{v}_+ + \mathbf{v}_-)(\mathbf{q}(t_b),t_b) - p_{t_b}\right]\cdot\mathbf{h}(t_b)\right\}$$
$$+ o(\epsilon) + \lim_{N\uparrow\infty}\epsilon\mathscr{E}\left\{\sum_{i=1}^{N}\frac{1}{2}m\left(\frac{\hbar}{m}\right)^{1/2}\sum_{j=1}^{3}\left[\dot{h}_j(t_i) - \sum_{k=1}^{3}h_k(t_i)\partial_j v^k_+(\mathbf{q}(t_i),t_i)\right]\Delta^- w^j_*(t_i)\right\}.$$

We now observe that, from (2.3) and (2.6), we have the following representation of $\Delta^-\mathbf{w}_*(t)$:

$$\Delta^-\mathbf{w}_*(t) = 2(m/\hbar)^{1/2}\mathbf{u}(\mathbf{q}(t),t)\Delta + \Delta^+\mathbf{w}(t-\Delta) + o(\Delta)\,,$$

so that, exploiting (2.13) and Ito's rule, we have

$$\mathscr{E}\left\{\sum_{j=1}^{3}\left[\dot{h}_j(t_i)-\sum_{k=1}^{3}h_k(t_i)\partial_j v^k_+(\mathbf{q}(t_i),t_i)\right]\Delta^- w^j_*(t_i)\right\}$$
$$=\mathscr{E}\left\{\sum_{k=1}^{3}\sum_{j=1}^{3}2\left(\frac{m}{\hbar}\right)^{1/2}[(\partial_k v^j_+ - \partial_j v^k_+)(\mathbf{q}(t),t)]u^j(\mathbf{q}(t),t)h_k(t)\Delta\right\}$$
$$+\mathscr{E}\left\{\sum_{k=1}^{3}\left[\sum_{j=1}^{3}\left(\frac{\hbar}{m}\right)^{1/2}\partial_j[(\partial_k v^j_+ - \partial_j v^k_+)(\mathbf{q}(t),t)]\Delta+o(\Delta)\right]h_k(t)\right\}.$$

Recalling that, for any smooth real-valued function $f(\mathbf{q}(t),t)$, one has[2]

$$D_\pm f(\mathbf{q}(t),t)=(\partial_t+\mathbf{v}_\pm\cdot\nabla\pm(\hbar^2/2m)\nabla^2)f(\mathbf{q}(t),t),$$

we finally get

$$\lim_{N\uparrow\infty}\delta A^N_{[t_a,t_b]}(\mathbf{q}(\cdot,\cdot);\mathbf{p}_{t_b})$$
$$=\epsilon\mathscr{E}\left\{\int_{t_a}^{t_b}\left[-\partial_t\mathbf{v}-(\mathbf{v}\cdot\nabla)\mathbf{v}+\frac{\hbar}{2m}\nabla^2\mathbf{u}+(\mathbf{u}\cdot\nabla)\mathbf{u}\right.\right.$$
$$\left.\left.+\left(\mathbf{u}+\frac{\hbar}{2m}\nabla\right)\wedge\nabla\wedge\mathbf{v}-\frac{\nabla\Phi^{\rm ext}}{m}\right]\cdot\mathbf{h}(t)dt\right\}$$
$$+\epsilon\mathscr{E}\left\{\tfrac{1}{2}m\mathbf{v}(\mathbf{q}(t_b),t_b)-\mathbf{p}_{t_b}\right\}+o(\epsilon).$$

Since for any diffusion satisfying (2.3) and (2.6) the continuity equation (2.17a) also holds,[5] the theorem is proved. ●

A necessary condition, which works in the same way in the irrotational and rotational cases, is proved in Ref. 3, p. 1986.

In the irrotational case, by setting $\mathbf{v}=\nabla S/m$, (2.17) immediately gives Madelung's equations.

In the general case the dynamical equations are different, due to the presence of a new quantum corrective term (of first order in $\hbar$), that depends on the vorticity of the current velocity $\mathbf{v}$.

Notice that, by the definition of $\mathbf{u}$ and (2.7) one has

$$\mathbf{u}=(\mathbf{v}_+-\mathbf{v}_-)/2, \tag{2.19}$$

so that, while $\mathbf{v}_+$, $\mathbf{v}_-$, and $\mathbf{v}$ change sign by time inversion, $\mathbf{u}$ does not.

One can see that, as a consequence, Madelung's fluid equations are time-reversal invariant, while in the new equations such an invariance is broken. The reason lies in the nature of the set of variations we have chosen: In spite of the fact that they are the most natural ones (we have simply asked that the variations change diffusions into diffusions with the same coefficient), one can see from (2.13) that they are *not* time-reversal invariant, since they are functionals of the past of $\mathbf{q}(t)$. Thus they are measurable with respect to the $\sigma$ algebra generated by the past of $\mathbf{q}(t)$ but not with respect to that generated by the future. This asymmetry is responsible for the rotational terms in (2.17b).

The time-reversed version of (2.17) is

$$\partial_t\rho=-\nabla\cdot(\rho\mathbf{v}), \tag{2.17a'}$$

$$\partial_t\mathbf{v}+(\mathbf{v}\cdot\nabla)\mathbf{v}-\left[\frac{\hbar}{2m}\nabla^2\mathbf{u}+(\mathbf{u}\cdot\nabla)\mathbf{u}\right]+\left(\mathbf{u}+\frac{\hbar}{2m}\nabla\right)\wedge\nabla\wedge\mathbf{v}=-\frac{\nabla\Phi^{\rm ext}}{m}, \tag{2.17b'}$$

which describes a flow with initial current velocity $\mathbf{v}=\mathbf{p}_{t_b}/m$.

## III. GAUGE STRUCTURE

The previous quantum dynamical equations (2.17) and (2.17') can be put into a more useful form, partially recovering the Schrödinger equation structure. Let $(\rho,\mathbf{v})$ satisfy (2.17) or (2.17') and let $S$ be an arbitrary smooth scalar field. Then, by defining the vector field $\mathbf{A}$ by

$$\mathbf{A}=\nabla S-m\mathbf{v}, \tag{3.1}$$

and exploiting the equality

$$\nabla(\mathbf{v}^2/2)=(\mathbf{v}\cdot\nabla)\mathbf{v}+\mathbf{v}\wedge(\nabla\wedge\mathbf{v}), \tag{3.2}$$

Eqs. (2.17b) and (2.17b') can be written, respectively, in the form

$$(1/m)\partial_t\mathbf{A}+(\mathbf{v}\pm\mathbf{u})\wedge(\nabla\wedge\mathbf{v})\pm(\hbar/2m)\nabla\wedge(\nabla\wedge\mathbf{v})$$
$$=\frac{1}{m}\nabla\left(\partial_t S+\frac{1}{2}m\mathbf{v}^2-\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}}+\Phi^{\rm ext}\right). \tag{3.3}$$

By the arbitrariness of $S$, we can fix $S=S_0$, where $S_0$ satisfies the equality

$$\partial_t S_0+\frac{1}{2}m\mathbf{v}^2-\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}}+\Phi^{\rm ext}=0. \tag{3.4}$$

But it is immediately seen that (3.4) and the continuity equations (2.17a) or (2.17a') formally give Madelung's fluid equations for a quantum particle with the four-vector potential $(\mathbf{A}_0,\Phi^{\rm ext})$, where $\mathbf{A}_0$ is defined by (3.1) for $S\equiv S_0$. Hence, by defining

$$\Psi_0=\rho^{1/2}\exp(iS_0/\hbar), \tag{3.5}$$

and recalling that, by (3.1), $m\nabla\wedge\mathbf{v}=-\nabla\wedge\mathbf{A}_0$, Eqs. (2.17) and (2.17') can be rewritten as the system of a Schrödinger equation for a quantum particle in the vector potential $\mathbf{A}_0$ and the coupled evolution equation for $\mathbf{A}_0$, that is, respectively,

$$i\hbar\,\partial_t\Psi_0=(1/2m)(i\hbar\nabla+\mathbf{A}_0)^2\Psi_0+\Phi^{\rm ext}\Psi_0, \tag{3.6a}$$

$$\partial_t\mathbf{A}_0=\mathbf{v}_\pm\wedge(\nabla\wedge\mathbf{A}_0)\pm(\hbar/2m)\nabla\wedge(\nabla\wedge\mathbf{A}_0), \tag{3.6b}$$

where $\mathbf{v}_\pm$ are defined by

$$m\mathbf{v}_\pm=\nabla S_0\pm(\hbar/2)\nabla\ln\rho-\mathbf{A}_0, \tag{3.7}$$

and then must be seen as a functional of $\Psi_0$ and $\mathbf{A}_0$.

It is worth noticing that (3.6b) has a strict analog in magnetohydrodynamics.[11] Indeed, the evolution equation for the "magnetic field" $\mathbf{B}\equiv\nabla\wedge\mathbf{A}_0$, obtained by taking the curl of both sides of (3.6b), is formally the same as in magnetohydrodynamics.

It is also possible to rewrite (3.6) in a gauge invariant form. In fact, for any smooth scalar field $\phi$, let us define

$$S=S_0+\phi,$$
$$\mathbf{A}=\mathbf{A}_0+\nabla\phi, \tag{3.8}$$
$$\Phi=\Phi^{\rm ext}-\partial_t\phi.$$

As a consequence (3.6) transform to

$$i\hbar\,\partial_t\Psi = (1/2m)(i\hbar\nabla + \mathbf{A})^2\Psi + \Phi\Psi\,, \tag{3.9a}$$

$$\partial_t\mathbf{A} = \mathbf{v}_\pm \wedge (\nabla\wedge\mathbf{A}) \pm (\hbar/2m)\nabla\wedge(\nabla\wedge\mathbf{A}) + \nabla(\Phi^{\rm ext} - \Phi)\,, \tag{3.9b}$$

where $(\mathbf{A},\Phi)$ is now defined up to a gauge transformation. Notice that $\Phi$ is not coupled with $\Psi$ and is simply fixed by the gauge.

## IV. ENERGY THEOREM AND STABILITY OF IRROTATIONAL MOTIONS

By substituting into (3.9) $\nabla\wedge\mathbf{A} = 0$, one can prove, by straightforward calculations, that the system is reduced to the usual linear Schrödinger equation

$$i\hbar\,\partial_t\Psi = (-(\hbar^2/2m)\nabla^2 + \Phi^{\rm ext})\Psi, \tag{4.1}$$

so that, defining the energy function $E$ by

$$E = \tfrac{1}{2}m\mathbf{v}^2 + \tfrac{1}{2}m\mathbf{u}^2 + \Phi^{\rm ext}, \tag{4.2}$$

and the Hamiltonian operator $H$ by

$$H = -(\hbar^2/2m)\nabla^2 + \Phi^{\rm ext}\,, \tag{4.3}$$

one has, if $\partial_t\Phi^{\rm ext} = 0$, the well-known equality[5]

$$\frac{d}{dt}\mathscr{E}\{E(\mathbf{q}(t),t)\} = \frac{d}{dt}\langle\Psi,H\Psi\rangle, \tag{4.4}$$

where $\langle\ ,\ \rangle$ denotes the inner product in $L^2(R^3,d^3x)$.

We now want to study the time evolution of $\mathscr{E}\{E(\mathbf{q}(t),t)\}$, where $E$ is still being defined by (4.2), in the general case of rotational velocity fields.

To do this let $(\bar\rho,\bar S,\bar{\mathbf{A}})$ denote a solution of (3.9) in the gauge $\Phi = \Phi^{\rm ext}$, and let us consider the linear Schrödinger equation

$$i\hbar\,\partial_t\Psi = (1/2m)(i\hbar\nabla + \bar{\mathbf{A}})^2\Psi + \Phi^{\rm ext}\Psi, \tag{4.5}$$

whose real and imaginary parts are

$$\partial_t\rho = -\nabla\cdot((\nabla S - \bar{\mathbf{A}})\rho/m)\,, \tag{4.6a}$$

$$\partial_t S + \frac{(\nabla S - \bar{\mathbf{A}})^2}{2m} - \frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}} + \Phi^{\rm ext} = 0\,. \tag{4.6b}$$

The energy conservation law reads still as in (4.4), where $E$ is defined by (4.2), with the only change that now we have $m\mathbf{v} = \nabla S - \bar{\mathbf{A}}$.[2]

We define the Hamiltonian density by

$$\mathscr{H} = E\rho\,, \tag{4.7}$$

that is, by explicitly displaying the dependence on $\rho$ and $S$,

$$\mathscr{H} = \mathscr{H}(\rho,S,\bar{\mathbf{A}}) = \left[\frac{(\nabla S - \bar{\mathbf{A}})^2}{2m} + \frac{1}{2m}\left(\frac{\hbar}{2}\nabla\ln\rho\right)^2 + \Phi^{\rm ext}\right]\rho\,. \tag{4.8}$$

We shall show that (4.6) can be derived from a variational principle with Lagrangian density

$$\mathscr{L} = S\dot\rho - \mathscr{H}\,. \tag{4.9}$$

Let $S$ and $\rho$ denote smooth scalar fields and consider, for any time interval $[t_a,t_b]$, the functional

$$I_{[t_a,t_b]}(\rho,S) = \int_{t_a}^{t_b} dt \int_{\rm Vol} d^3x\,[S\dot\rho - \mathscr{H}(S,\rho,\bar{\mathbf{A}})](\mathbf{x},t)\,. \tag{4.10}$$

We seek the sufficient and necessary conditions to ensure the stationarity conditions of such a functional with respect to independent variations of $S$ and $\rho$ that vanish at the boundary.

Denoting by $\delta_{(S)}$ and $\delta_{(\rho)}$ the variations with respect to $S$ and $\rho$, respectively, we have

$$\delta_{(S)}I_{[t_a,t_b]}(\rho,S) = \int_{t_a}^{t_b} dt \int_{\rm Vol} d^3x\,[\dot\rho\delta S - \delta_{(S)}\mathscr{H}](\mathbf{x},t)\,, \tag{4.11a}$$

$$\delta_{(\rho)}I_{[t_a,t_b]}(\rho,S) = \int_{t_a}^{t_b} dt \int_{\rm Vol} d^3x\,[-\dot S\delta\rho - \delta_{(\rho)}\mathscr{H}](\mathbf{x},t)\,, \tag{4.11b}$$

and also

$$\delta_{(S)}\mathscr{H} = \frac{1}{m}\{\nabla\cdot[(\nabla S - \bar{\mathbf{A}})\rho\delta S] - \nabla\cdot[(\nabla S - \bar{\mathbf{A}})\rho]\delta S\}\,, \tag{4.12a}$$

$$\delta_{(\rho)}\mathscr{H} = \left[\frac{1}{2m}(\nabla S - \bar{\mathbf{A}})^2 + \Phi^{\rm ext}\right]\delta\rho + \frac{\hbar^2}{2m}\delta_{(\rho)}\left[\left(\frac{\nabla\rho}{2\rho}\right)^2\rho\right], \tag{4.12b}$$

where

$$\delta_{(\rho)}\left[\left(\frac{\nabla\rho}{2\rho}\right)^2\rho\right] = \nabla\cdot\left(\frac{\nabla\rho}{2\rho}\delta\rho\right) - \nabla\cdot\left(\frac{\nabla\rho}{2\rho}\right)\delta\rho - \left(\frac{\nabla\rho}{2\rho}\right)\delta\rho\,. \tag{4.13}$$

Let us now assume, as usual, that, if $\rho$ is equal to zero at the boundary, then $\nabla\rho$ also goes to zero, thus maintaining $\nabla\rho/\rho$ finite. By integrating on the volume, the first term on the right-hand side of (4.13) gives a vanishing contribution.

Then, by inserting (4.12) in (4.11), one can see, by the standard argument, that the functional (4.10) is stationary if and only if (4.6) is satisfied.

Moreover, by integrating (4.6) on the volume and applying (4.12) again, we can also get the "mean Hamilton equations"

$$\int_{\rm Vol}\dot\rho(\mathbf{x},t)\delta S(\mathbf{x},t)d^3x = \int_{\rm Vol}[\delta_{(S)}\mathscr{H}(\rho,S)](\mathbf{x},t)d^3x\,, \tag{4.14a}$$

$$\int_{\rm Vol}\dot S(\mathbf{x},t)\delta\rho(\mathbf{x},t)d^3x = -\int_{\rm Vol}[\delta_{(\rho)}\mathscr{H}(\rho,S)](\mathbf{x},t)d^3x\,, \tag{4.14b}$$

where $\rho$ and $S$ play the role of conjugate canonical variables.

Let us now particularize $\delta S$ and $\delta\rho$ to

$$\delta S = \dot S\delta t + o(\delta t),\quad \delta\rho = \dot\rho\delta t + o(\delta t)\,. \tag{4.15}$$

Then, for any solution to (4.6),

$$\frac{d}{dt}\mathscr{E}\{E\}=\frac{d}{dt}\int_{\mathrm{Vol}}\mathscr{H}(\rho,S,\bar{\mathbf{A}})(\mathbf{x},t)d^3x$$

$$=\lim_{\delta t\to 0}\frac{1}{\delta t}\int_{\mathrm{Vol}}[\delta_{(S)}\mathscr{H}+\delta_{(\rho)}\mathscr{H}](\mathbf{x},t)d^3x$$

$$-\int_{\mathrm{Vol}}\left[\left(\frac{\nabla S-\bar{\mathbf{A}}}{m}\right)\dot{\cdot}\dot{\mathbf{A}}\rho\right](\mathbf{x},t)d^3x$$

$$+\int_{\mathrm{Vol}}\partial_t\Phi^{\mathrm{ext}}\rho\,d^3x$$

$$=\int_{\mathrm{Vol}}[\dot{\rho}\dot{S}-\dot{S}\dot{\rho}](\mathbf{x},t)d^3x$$

$$-\int_{\mathrm{Vol}}\left[\left(\frac{\nabla S-\bar{\mathbf{A}}}{m}\right)\dot{\cdot}\dot{\mathbf{A}}\rho\right](\mathbf{x},t)d^3x$$

$$+\int_{\mathrm{Vol}}\partial_t\Phi^{\mathrm{ext}}\rho\,d^3x\,. \tag{4.16}$$

Thus we can conclude that, if $\Phi^{\mathrm{ext}}$ is time independent, for any solution $(S,\rho)$ to (4.6) the corresponding mean energy evolves in time according to the equality

$$\frac{d}{dt}\mathscr{E}\{E\}=-\int_{\mathrm{Vol}}\left[\left(\frac{\nabla S-\bar{\mathbf{A}}}{m}\right)\dot{\cdot}\dot{\mathbf{A}}\rho\right](\mathbf{x},t)d^3x\,. \tag{4.17}$$

In particular, (4.17) must be true for $(S,\rho)=(\bar{S},\bar{\rho})$, which together with $\bar{\mathbf{A}}$, satisfies the system (3.9), and then in particular (3.9a), in the gauge $\Phi=\Phi^{\mathrm{ext}}$. So, by recalling the equalities

$$\nabla\bar{S}-\bar{\mathbf{A}}=m\mathbf{v},\quad \mathbf{v}_{\pm}=\mathbf{v}\pm\mathbf{u}\,, \tag{4.18}$$

one gets by (3.9b)

$$\left(\frac{\nabla\bar{S}-\bar{\mathbf{A}}}{m}\right)\dot{\cdot}\dot{\mathbf{A}}$$

$$=-m\mathbf{v}\dot{\cdot}\left[(\mathbf{v}\pm\mathbf{u})\wedge\nabla\wedge\mathbf{v}\pm\frac{\hbar}{2m}\nabla\wedge(\nabla\wedge\mathbf{v})\right]$$

$$=\mp m\mathbf{v}\dot{\cdot}\left[\left(\mathbf{u}+\frac{\hbar}{2m}\nabla\right)\wedge\nabla\wedge\mathbf{v}\right]. \tag{4.19}$$

Concluding, we can state the following extension of the energy theorem.[12]

**Theorem 2:** Let $(\rho,\mathbf{v})$ be any solution to (2.17) [resp. (2.17′)] and let us assume that if $\rho$ is zero at the boundary $\nabla\rho$ also goes to zero maintaining $\nabla\rho/\rho$ finite. By defining the energy function $E=\frac{1}{2}m\mathbf{v}^2+\frac{1}{2}m\mathbf{u}^2+\Phi^{\mathrm{ext}}$, with $u=(\hbar/2m)\nabla\ln\rho$ and time-independent external potential, then

$$\frac{d}{dt}\mathscr{E}\{E\}=\pm m\mathscr{E}\left\{\mathbf{v}\dot{\cdot}\left[\left(\mathbf{u}+\frac{\hbar}{2m}\nabla\right)\wedge\nabla\wedge\mathbf{v}\right]\right\}, \tag{4.20}$$

where $+$ refers to (2.17) and $-$ to (2.17′), respectively.

It is interesting to inquire into the consequences of (4.20) in the case when $(\rho,\mathbf{v})$ are solutions to (2.17′), which correspond to the equations for a flow with fixed initial velocity field.

Let us consider the particular case of a particle either in the free space or in the presence of perfectly smooth and reflecting barriers. As is well known, this implies, in quantum mechanics, that $\rho$ is zero at the boundary. Let us maintain this condition in the general case: by integrating by parts we obtain

$$\mathscr{E}\{\mathbf{v}\dot{\cdot}[(\hbar/2m)\nabla\wedge(\nabla\wedge\mathbf{v})]\}$$

$$=-\mathscr{E}\{\mathbf{v}\dot{\cdot}(\mathbf{u}\wedge(\nabla\wedge\mathbf{v}))\}+(\hbar/2m)\mathscr{E}\{(\nabla\wedge\mathbf{v})^2\}, \tag{4.21}$$

so that, for $(\rho,\mathbf{v})$ solutions to (2.17′), we have

$$\frac{d}{dt}\mathscr{E}\{E\}=-\frac{\hbar}{2}\mathscr{E}\{(\nabla\wedge\mathbf{v})^2\}\,. \tag{4.22}$$

Thus, in the assumption of Theorem 2 and if the density is zero at the boundary, the mean energy related to solutions to Eqs. (2.17′) has a negative time derivative, which is equal to zero if the current velocity field is irrotational. As a consequence, the set of solutions such that

$$\rho(\nabla\wedge\mathbf{v})^2=0 \tag{4.23}$$

works as a collection of attractors. Then we can also conclude that, in this case, if a solution is time independent it must be also irrotational, wherever $\rho$ is different from zero.

## V. DISCUSSION AND OUTLOOK

We have seen that the unrestricted Lagrangian variational principle in stochastic mechanics leads to a sort of (nonrelativistic) gauge structure and that, in the case of the flow described by (2.17′), the set of irrotational motions is an attracting set.

As a final remark we might also observe that the "quantum state" is, in such a case, represented by *four* scalar fields, in place of the two scalar fields defining the wave function in the standard case.

This would be the case of a quaternionic quantum mechanics, whose mathematical possibility, from an axiomatic point of view, was conjectured by Birkoff and Von Neumann since 1936.[13]

Recently a quaternionic field theory has been proposed by Adler,[14] who has also applied the results to the CP violation in $K$ meson decay.[15]

Possible connections between the unrestricted Lagrangian principle in stochastic mechanics and a quaternionic quantum dynamics could represent, to the authors' opinion, a challenging subject of future research.


## ACKNOWLEDGMENTS

One of us (MIL) began to work on this subject at the Centre de Physique Téorique, CNRS, Marseille; this is the right place to gratefully thank the Centre for the hospitality extended to her. Useful comments by Eric Carlen and Edward Nelson are also gratefully acknowledged.

This work was partially supported by the Ministero della Pubblica Istruzione and the Consiglio Nazionale delle Ricerche.



[1] F. Guerra and L. M. Morato, Phys. Rev. D **27**, 1774 (1983).

[2] E. Nelson, in *Séminaire de Probabilités XIX, Lecture Notes in Mathematics*, Vol. 1123, edited by J. Azéma and M. Yor (Springer, Berlin, 1985); *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1984).

[3] L. M. Morato, Phys. Rev. D **31**, 1982 (1985).

[4]L. M. Morato, in "*Stochastic Processes in Classical and Quantum Systems*," *Lecture Notes in Physics*, Vol. 262, edited by S. Albeverio, G. Casati, and D. Merlini (Springer, Berlin, 1986). [We warn the reader that formula (13′) contains a misprint.]
[5]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967).
[6]E. Carlen, Commun. Math. Phys. **94**, 293 (1984).
[7]F. Guerra, in "*Quantum probability and applications II*," *Lecture Notes in Mathematics*, Vol. 1136, edited by L. Accardi and W. Waldenfels (Springer, Berlin, 1985).
[8]P. Blanchard and S. Golin, Commun. Math. Phys. **109**, 421 (1987).
[9]U. G. Haussmann and E. Pardoux, Ann. Probab. **14**, 1188 (1986).
[10]J. Serrin, "Handbuch der physik VIII," *Fluid Dynamics* (Springer, Berlin, 1959), Vol. 1.
[11]W. F. Hughes and F. J. Young, *The Electromagnetodynamics of Fluids* (Wiley, New York, 1966).
[12]A similar expression of the time derivative of the mean energy has been independently guessed by R. Marra following a different line of thinking (private communication).
[13]G. Birkoff and J. Von Neumann, Ann. Math. **37**, 823 (1936).
[14]S. L. Adler, Phys. Rev. Lett. **55**, 8, 783 (1985).
[15]S. L. Adler, Phys. Rev. Lett. **57**, 2, 167 (1986).

# Global quantities in algebraic quantum mechanics of infinite systems: Classical observables or parameters?

Jürgen Pöttinger
*Laboratorium für Physiakalische Chemie, Eidgenössische Technische Hochschule, ETH-Zentrum, CH-8092 Zürich, Switzerland*

(Received 15 August 1987; accepted for publication 7 September 1988)

Local observables of an infinite system are described by a quasilocal $C^*$-algebra $A$. Traditionally classical observables are introduced as parameters that can be assigned to certain states on the local observables. As fundamental theories should be essentially free of parameters this can hardly be regarded as satisfactory. In the present paper classical observables are introduced explicitly as operators by a transitive system of imprimitivity of a kinematical group in the center of the $W^*$-algebra $\pi(A)''$ belonging to a representation $\pi$ of $A$. The states of the system are represented by linear functionals on $\pi(A)''$, which are dispersion-free on the center of $\pi(A)''$. For classical observables related to a kinematical group a $W^*$-description—with algebra $\pi(A)''$ and its associated linear functionals that are dispersion-free on the center and normal on the noncentral part of $\pi(A)''$—is equivalent to a $C^*$-description—with algebra $A$ and a distinguished family of linear functionals on $A$, determined uniquely by $\pi$. The implications of these results on the interpretation of quantum mechanics of infinite systems are discussed.

## I. INTRODUCTION

For infinite systems (systems with infinitely many degrees of freedom) it is important to distinguish between local and global observables. Global observables compatible with all other observables are called classical.

The many classical global observables of an infinite system allow a decomposition of the system into many independent autonomous descriptions of the system. A set of classical global observables characterizes the system in a special context. To select a particular context means to concentrate attention on a particular phenomenon and to ignore effects of secondary importance. In the context of equilibrium statistical mechanics one concentrates on the thermodynamical classical observables like temperature and chemical potential.

Infinite systems have an adequate description within the framework of algebraic quantum mechanics. In the traditional discussion classical global observables are not introduced explicitly but they are described by parameters that can be assigned to certain states on the local observables. In the present paper, classical observables are introduced explicitly by a kinematical group as elements of an algebra of observables, which are compatible with all other observables. A state of the system in a given context is described by an eigenstate of all classical observables referring to the given context. It is the main result of the present work that both descriptions of a given context are compatible.

In algebraic quantum mechanics, the basic objects of interest are the observables forming an algebra $A$ ($C^*$- or $W^*$-algebra). States of the system are represented by the positive linear functionals on $A$ normalized to 1. The expectation value of an observable $a = a^*$ in $A$ with respect to a state $\varphi$ is the number $\varphi(a)$. If $\varphi(a)^2 = \varphi(a^2)$ we say that the observable $a$ is dispersion-free with respect to the state $\varphi$ or equivalently $\varphi$ is an eigenstate of $a$. In this case, we interpret $\varphi(a)$ as the value the observable $a$ *has* in the state $\varphi$.

The *center* $Z(A) := \{a \in A \mid ab = ba, \text{ for all } b \in A\}$ of $A$ is a commutative subalgebra of $A$. The observables belonging to $Z(A)$ are called *classical* observables. In the usual individual realistic interpretation of classical physical theories, classical observables are those observables that are always actualized. Therefore, the functionals on $A$ that are dispersion-free on the center are taken as the *physical states* of the system.

In the framework of algebraic quantum mechanics of infinite systems, the quasilocal observables are described by a primitive $C^*$-algebra $A$ supporting a quasilocal structure (cf. Ref. 1): for any bounded open region $D$ of $\mathbb{R}^3$ there exists a subalgebra $A(D)$ of $A$ containing those local observables that can be performed inside of $D$. The algebra $A$ is the norm closure of the union of all subalgebras $A(D)$. A primitive algebra $A$ has a trivial center, i.e., $Z(A) = \{\mathbb{C}1\}$. This is expressed in the formalism that classical global observables correspond to operations that are performed outside of any bounded region. But one can think of global observables as averages of local observables over the whole space. To include such observables one can represent the abstract $C^*$-algebra $A$ faithfully on a Hilbert space $\mathscr{H}$ in a representation $\pi$. The weak closure $\pi(A)''$ of the $C^*$-algebra $\pi(A)$ is a $W^*$-algebra, which contains many new observables. It is especially true that $\pi(A)''$ can contain classical observables.

In the traditional description to each state of the system in a given context corresponds a state on the quasilocal algebra $A$ (cf. Ref. 2). In the context of equilibrium statistical mechanics a maximally specified equilibrium situation, which refers to particular values of the intensive thermodynamic quantities like the temperature or the chemical potential, is described by a KMS factor state. A given context is therefore described by a restriction of the state space of the quasilocal algebra to a family of states. A state belongs to this family if it describes a state of the system in the given context.

In the present paper a given context is described by a

 0022-2488/89/020361-08$02.50 © 1989 American Institute of Physics 

$W^*$-algebra $\pi(A)''$ of a representation $\pi$ of $A$ and its associated physical states. The classical global observables that define a given context are represented as elements of the center of $\pi(A)''$. A state of the system in the given context is described by an eigenstate $\varphi$ of all classical observables. In the state $\varphi$ of $\pi(A)''$, the classical global observables have a particular dispersion-free value. The nonclassical local observables are, e.g., not dispersion-free in the state $\varphi$.

In algebraic quantum mechanics, the conceptual characterization of the algebra of observables of a physical system is of central importance. Historically Bohr's principle of a correspondence between classical mechanics and quantum mechanics allowed one to handle special observables, but this recipe does not solve the conceptual problem. A deeper foundation of the correspondence principle can be related to the representation theory of an appropriate kinematical group, the Galilei group (cf. Refs. 3 and 4).

The important idea of this concept to look at observables is the use of a *kinematical group*. Kinematical groups can be considered as reflecting the idealizations and abstractions of a theory.[5,6] Every element of a kinematical group describes the admissible change between equivalent points of view in the context of a chosen abstraction. A well known example of a kinematical group is the Galilei group. But kinematical groups are not restricted to describing abstractions in time and space; elementary chemical kinetics (with the scaling group) is a theory formulated without reference to space.

Starting from a kinematical group observables can be introduced as operators transforming suitably under the group. This idea goes back to Weyl.[7] In the framework of algebraic quantum mechanics a symmetry group appears as a homomorphism into the automorphismgroup of the respective algebra of observables. An algebra $A$ and a homomorphism $\alpha$ of a kinematical group are said to *have observables* if there exists a system of imprimitivity for the algebra $A$ and the action $\alpha$ of the kinematical group (cf. Refs. 3, 8, and 9 and Sec. II). The theory of imprimitivity systems is sufficient for the treatment of observables such as position and momentum[4] and of classical observables. More general situations have been studied in Ref. 10.

In the present paper those classical observables are discussed that are related to a kinematical group. Given a quasilocal algebra $A$ and a continuous action $\alpha$ of a kinematical group the generation of classical observables in representations $\pi$ is discussed under the following three conditions: (1) the kinematical symmetry is not broken in the representation $\pi$; (2) the $W^*$-algebra $\pi(A)''$ has classical observables generated by the kinematical group; and (3) the system under description can be characterized by countably many experiments. In the algebraic formalism these conditions can be transferred to the following conditions for representations $\pi$ of $A$.

(1) Every automorphism $\alpha_g$, $g\in G$, of $A$ has an extension to an automorphism $\alpha_g''$ of $\pi(A)''$ such that $\alpha_g''\circ\pi = \pi\circ\alpha_g$.

(2) There exists a system of imprimitivity in the center of the $W^*$-algebra $\pi(A)''$ with respect to the action $\alpha''$ of the kinematical group.

(3) The physical state space of $\pi(A)''$ is separable.

A structure theory for representations $\pi$ satisfying the conditions (1)–(3) has been developed for separable $C^*$-algebras and separable, locally compact kinematical groups, which is reviewed in Sec. II. Using these results it is shown that a $C^*$-description of a given context is compatible with a $W^*$-description of the same context. Given a representation $\pi$ satisfying the conditions (1)–(3), it is shown that a $W^*$-description [with observable algebra $\pi(A)''$ and its associated physical states normal on the noncentral part of $\pi(A)''$] is one-to-one with a $C^*$-description (with observable algebra $A$ and a distinguished family of states of $A$, determined by $\pi$).

## II. INDUCED REPRESENTATIONS AND DISPERSION-FREE STATES

A *$C^*$-system* is a triple $(A,G,\alpha)$ consisting of a $C^*$-algebra $A$, a locally compact group $G$, and a continuous homomorphism $\alpha$ of $G$ into the group $\mathrm{Aut}(A)$ of $^*$-automorphisms of $A$ equipped with the topology of pointwise norm convergence. A *$W^*$-system* $(\mathcal{M},G,\alpha)$ is defined similarly: $\mathcal{M}$ is then a $W^*$-algebra and $\mathrm{Aut}(\mathcal{M})$ is equipped with the topology of pointwise $\sigma$-weak convergence.

A representation $(\pi,\mathcal{H})$ of a $C^*$-algebra $A$ is said to be *separable* if the representation Hilbert space is separable. A *quasi-invariant representation* of a $C^*$-system $(A,G,\alpha)$ is a quartet $(\pi,\mathcal{H},G,\alpha'')$, where $(\pi,\mathcal{H})$ is a representation of $A$ and $\alpha''$ a homomorphism of $G$ into $\mathrm{Aut}(\pi(A)'')$ such that $\alpha_g''\circ\pi = \pi\circ\alpha_g$, $g\in G$.

### A. Induced representations

In the following, $G$ will always denote a second countable, locally compact group and $H$ a closed subgroup of $G$. On $G/H$, the space of left cosets of $H$ in $G$, there exists a unique (up to quasiequivalence) probability measure $\nu$, quasi-invariant under the transitive representation

$$s_k\colon\ G/H\to G/H\,,\quad gH\to kgH\,,$$

$k,g\in G$, and a measurable cross section $r\colon G/H\to G$. (Denote by $p$ the projection mapping of $G$ onto $G/H$. A cross section $r$ is a map with the property $p\circ r = \mathrm{id}$.) The left regular representation $\lambda_{G/H}$ of $G$ on $L_2(G/H,\nu)$ is defined by

$$\{\lambda_{G/H}(k)\xi\}(\dot g) = \left(\frac{d\nu\circ s_{k^{-1}}}{d\nu}(\dot g)\right)^{1/2}\xi(s_{k^{-1}}(\dot g)),$$

$$\xi\in L_2(G/H,\nu),$$

where $d\nu\circ s_{k^{-1}}/d\nu$ denotes the Radon–Nykodym derivative of the translated measure $d\nu\circ s_{k^{-1}}$ with respect to $\nu$. In the following, elements $gH$ of $G/H$ will be denoted by $\dot g$.

Let $(A,G,\alpha)$ be a $C^*$-system and let $(\pi_0,\mathcal{H}_0,H,\gamma)$ be a separable quasi-invariant representation of the $C^*$-system $(A,H,\alpha)$. A separable quasi-invariant representation $(\pi,\mathcal{H},G,\alpha'')$ of $(A,G,\alpha)$ can be defined by

$$(\pi,\mathcal{H}) := \int_{G/H}^{\oplus}(\pi_0\circ\alpha_{r(\dot g)^{-1}},\mathcal{H}_0)\,d\nu(\dot g)\,,$$

$$\{\alpha_k''x\}(\dot g) := \gamma_{r(\dot g)^{-1}kr(s_{k^{-1}}(\dot g))}(x(s_{k^{-1}}(\dot g)))\,,$$

$$x\in\pi(A)''\,,\quad k\in G\,.$$

The representation $(\pi,\mathscr{H},G,\alpha'') := \mathrm{ind}_H^G(\pi_0,\mathscr{H}_0,H,\gamma)$ is called *induced up* from the representation $(\pi_0,\mathscr{H}_0,\mathscr{H},\gamma)$. $(\pi,\mathscr{H},G,\alpha'')$ does not depend (up to quasiequivalence) on the choice of the quasi-invariant measure $\nu$ and the measurable cross section $r$. $(\pi(A)'',G,\alpha'')$ is a $W^*$-system. (For proofs and more information, see Ref. 11.)

### B. Observables

Similar to Refs. 3, 9, and 12, we introduce the following *definition*: A $W^*$-system $(\mathscr{M},G,\alpha'')$ has *classical observables* (with respect to the kinematical group $G$) if the center of $\mathscr{M}$ admits a transitive system of imprimitivity. This is equivalent[13,14] to the existence of a closed subgroup $H$ of $G$ and a normal *-homomorphism $\tau$ from $L_\infty(G/H)$ into $Z(\mathscr{M})$ with the covariance property

$$(4)\quad \alpha_g''\circ\tau = \tau\circ\mathrm{ad}\,\lambda_{G/H}(g), \quad g\in G.$$

Here $L_\infty(G/H)$ is the $W^*$-algebra of equivalence classes of essentially bounded complex-valued $\nu$-measurable functions on $G/H$.

### C. Representations generating classical observables

As explained in Sec. I, we consider representations of a $C^*$-system $(A,G,\alpha)$ that satisfy the following conditions.

$\langle 1\rangle$ $(\pi,\mathscr{H},G,\alpha'')$ is a quasi-invariant representation of $(A,G,\alpha)$.

$\langle 2\rangle$ There exists a closed subgroup $H$ of $G$ and a normal *-homomorphism $\tau$ from $L_\infty(G/H)$ into $Z(\pi(A)'')$ satisfying (4).

$\langle 3\rangle$ $\mathscr{H}$ is separable.

If $A$ is a separable $C^*$-algebra, any representation satisfying $\langle 1\rangle$–$\langle 3\rangle$ is quasiequivalent to an induced representation.[11,13] More precisely, this means that there exists a quasi-invariant state $\varphi$ of $(A,H,\alpha)$ with representation $(\pi_\varphi,\mathscr{H}_\varphi,H,\gamma)$ and a *-isomorphism $\kappa$ from $\pi(A)''$ onto $\pi_\phi(A)''$, such that

$$\kappa\circ\pi = \pi_\phi,$$

$$\kappa\circ\alpha_g'' = \alpha_{\phi,g}''\circ\kappa, \quad g\in G,$$

$$\kappa(\tau(f)) = d_f, \quad f\in L_\infty(G/H),$$

where we used the following definitions:

$$\phi := \int_{G/H} \varphi\circ\alpha_{r(\dot g)^{-1}}\, d\nu(\dot g),$$

$$d_f := \int_{G/H}^{\oplus} f(\dot g)\, 1_{\mathscr{H}_\varphi}\, d\nu(\dot g),$$

$$(\pi_\phi,\mathscr{H}_\phi,G,\alpha_\phi'') := \mathrm{ind}_H^G(\pi_\varphi,\mathscr{H}_\varphi,H,\gamma).$$

The quasi-invariant representation $(\pi_\varphi,\mathscr{H}_\varphi,H,\gamma)$ is unique up to quasiequivalence.

These results allow one to consider only induced representations for the generation of classical observables in case $A$ is separable. The corresponding results are not available for nonseparable $C^*$-algebras. But any induced representation of a nonseparable $C^*$-algebra $A$ satisfying $\langle 2\rangle$ and $\langle 3\rangle$ generates classical observables.

All results derived in the next sections are valid for induced representations $\pi_\phi$ on arbitrary unital $C^*$-algebras $A$ satisfying $\langle 2\rangle$ and $\langle 3\rangle$. In the representation $\pi_\phi$, the $W^*$-algebra $\pi_\phi(A)''$ can be identified with the $W^*$-algebra $L_\infty(G/H)\,\overline{\otimes}\,\pi_\varphi(A)''$ that factorizes into a commutative algebra $L_\infty(G/H)$ generated by the classical observable, and into a noncommutative algebra $\pi_\varphi(A)''$.

### D. Examples

(a) Let $G$ denote the Galilei group and $H$ the closed subgroup generated by rotations, time translations, and space translations. Then the corresponding classical observable based on $G/H$ is a momentum operator.[15]

(b) Let $G$ denote the Galilei group and $H$ the closed subgroup generated by rotations, space translations, and velocity translations. Then the corresponding classical observable based on $G/H$ is a time operator.[12]

(c) Let $G$ denote the gauge group acting on the quasilocal algebra of an infinite free Bose gas. Then the classical observable corresponding to $G = G/\{e\}$ is a phase operator.[16]

### E. Dispersion-free states

For our purpose it suffices to consider $W^*$-algebras $\mathscr{M} = L_\infty(G/H)\,\overline{\otimes}\,\mathscr{F}$, $\mathscr{F}$ an arbitrary $W^*$-algebra, and states $\Omega$ on $\mathscr{M}$, which are dispersion-free on the $W^*$-subalgebra $L_\infty(G/H)\,\overline{\otimes}\,\{\mathbb{C}1\}$. The set of these states we denote by $D(\mathscr{M})$, i.e.,

$$D(\mathscr{M}) := \{\Omega\in S(\mathscr{M}) \mid \Omega(z^2) = \Omega(z)_2,$$

$$\forall z = z^*\in L_\infty(G/H)\,\overline{\otimes}\,\{\mathbb{C}1\}\}.$$

If $G/H$ is not countable as a set then any state in $D(\mathscr{M})$ is singular (finitely additive) (cf. Appendix B).

A state $\Omega$ on $\mathscr{M}$ is called a *product state* [with respect to the factorization $\mathscr{M} = L_\infty(G/H)\,\overline{\otimes}\,\mathscr{F}$] if there exists a state $\Omega_1$ of $L_\infty(G/H)$ and a state $\Omega_2$ of $\mathscr{F}$, such that $\Omega(x\otimes y) = \Omega_1(x)\Omega_2(y)$, for all $x\in L_\infty(G/H)$ and $y\in\mathscr{F}$. It is an immediate consequence of Ref. 17, IV.4.11), that any $\Omega\in D(\mathscr{M})$ is a product state where $\Omega_1$ is a multiplicative (hence pure) state of $L_\infty(G/H)$. On the other side, every product state $\Omega$ with a multiplicative state $\Omega_1$ belongs to $D(\mathscr{M})$. Therefore

$$D(\mathscr{M}) = \{\Omega\in S(\mathscr{M}) \mid \Omega \text{ is a product state and } \Omega_1 \text{ is pure}\}.$$

We use the following partition of $D(\mathscr{M})$:

$$D_f(\mathscr{M}) = \{\Omega\in D(\mathscr{M}) \mid \Omega_1 \restriction C_\infty(G/H) \neq 0\},$$

$$D_\infty(\mathscr{M}) = \{\Omega\in D(\mathscr{M}) \mid \Omega_1 \restriction C_\infty(G/H) = 0\}.$$

For states $\Omega$ belonging to $D_f(\mathscr{M})$, the classical observable generating $L_\infty(G/H)$ has a finite value. In the opposite case, $\Omega\in D_\infty(\mathscr{M})$, the classical observable has infinite value.

We interpret classical observables as those observables that are always actualized. Therefore, only those functionals of $D(\mathscr{M})$ describe a state of the system that is dispersion-free on the center of $\mathscr{M}$. If $\mathscr{F}$ is a factor, every functional of

$D(\mathcal{M})$ is dispersion-free on the center of $\mathcal{M}$. Especially in this case, we consider the set

$$\phi(\mathcal{M}) := \{\Omega \in D(\mathcal{M}) \mid \Omega_2 \text{ is normal}\}.$$

Here $\phi(\mathcal{M})$ contains those functionals of $D(\mathcal{M})$ which are, e.g., singular on the commutative (classical) part $L_\infty(G/H) \overline{\otimes} \{\mathbb{C}1\}$ and normal on the strictly noncommutative (purely quantum mechanical) part $\{\mathbb{C}1\} \otimes \mathcal{F}$. We use the partition of $\phi(\mathcal{M})$ in $\phi_f(\mathcal{M}) = \phi(\mathcal{M}) \cap D_f(\mathcal{M})$ and $\phi_\infty(\mathcal{M}) = \phi(\mathcal{M}) \cap D_\infty(\mathcal{M})$.

## III. EXTENSION OF STATES TO DISPERSION-FREE STATES

In this section we consider the extension of an arbitrary state $\omega$ on a $C^*$-algebra $A$ to a state $\Omega$ on $\mathcal{M} = \pi_\phi(A)''$, where $\phi$ denotes an induced state on $A$. The results of this section are valid under the following assumptions for the induced representation $\pi_\phi$ and the associated induced state

$$\phi = \int_{G/H} \varphi \circ \alpha_{r(\dot{g})^{-1}} \, d\nu(\dot{g}).$$

*Assumption 1:* $r$ is a continuous cross section.

*Assumption 2:* The states $\varphi \circ \alpha_{r(\dot{g})^{-1}}$, $\dot{g} \in G/H$, are mutually disjoint.

These assumptions are satisfied in every known physical example (cf. Sec. II). Note that Assumption 2 is always satisfied for the states $\varphi \circ \alpha_{r(\dot{g})^{-1}}$, $\dot{g} \in G/H - N$, where $N$ is a $\nu$-null Borel set of $G/H$. If $H$ is a normal subgroup or $r \circ p$ is a homomorphism and $HgH = Hg$, $g \in G$, the null set $N$ can be shown to be empty (cf. Lemma 2.6 in Ref. 11). There are no general conditions known to the author implying the continuity of the measurable cross section $r$.

*Lemma 3.1:* Let $(\pi_\nu, L_2(G/H,\nu))$ denote the multiplication representation of the $C^*$-algebra $C_b(G/H)$, the bounded complex-valued continuous functions on $G/H$, defined by

$$\{\pi_\nu(f)\xi\}(\dot{g}) = f(\dot{g})\xi(\dot{g}),$$

$$\xi \in L_2(G/H,\nu), \quad f \in C_b(G/H).$$

Then $\pi_\nu$ is a faithful representation.

*Proof:* Use that $\nu(u) > 0$ for any nonempty open subset $u$ of $G/H$ (cf. Ref. 18, Lemma 1.3 and Ref. 19, 9.2.2). □

Identifying $\mathcal{M} = \pi_\phi(A)''$ with $L_\infty(G/H) \overline{\otimes} \pi_\varphi(A)''$, it is important to consider the following $C^*$-subalgebras of $\mathcal{M}$:

$$\mathcal{N} := \pi_\nu(C_b(G/H)) \otimes \pi_\varphi(A)'',$$

$$\mathcal{N}_\infty := \pi_\nu(C_\infty(G/H)) \otimes \pi_\varphi(A)'',$$

$$\mathcal{R} := \pi_\nu(C_b(G/H)) \otimes \pi_\varphi(A),$$

$$\mathcal{R}_\infty := \pi_\nu(C_\infty(G/H)) \otimes \pi_\varphi(A).$$

Here $C_\infty(G/H)$ denotes the $C^*$-algebra of bounded complex-valued continuous functions on $G/H$ vanishing at infinity. If and only if $G/H$ is compact, $C_\infty(G/H) = C_b(G/H)$ and $\mathcal{N} = \mathcal{N}_\infty$, $\mathcal{R} = \mathcal{R}_\infty$.

Let $B$ be a $C^*$-algebra. By $C_b(G/H,B)$ we understand the $C^*$-algebra of bounded continuous functions from $G/H$ to $B$, and by $C_\infty(G/H,B)$ the subalgebra of functions $x$ vanishing at infinity, i.e., the function $\dot{g} \to \|x(\dot{g})\|$ belongs to $C_\infty(G/H)$. It follows from Ref. 20, Chap. 1.22, that we can identify

$$C_b(G/H,B) \cong C_b(G/H) \otimes B,$$

$$C_\infty(G/H,B) \cong C_\infty(G/H) \otimes B.$$

Furthermore, we can identify

$$C_b(G/H) \otimes B \cong \pi_\nu(C_b(G/H)) \otimes B,$$

$$C_\infty(G/H) \otimes B \cong \pi_\nu(C_\infty(G/H)) \otimes B,$$

under the $^*$-isomorphism $\pi_\nu \otimes \mathrm{id}$ (due to Lemma 3.1).

*Lemma 3.2:* If $r$ is continuous, $\pi_\phi(a)$, $a \in A$, belongs to $\mathcal{R}$.

*Proof:* The direct integral operator

$$\pi_\phi(a) = \int_{G/H}^{\oplus} \pi_\varphi(\alpha_{r(\dot{g})^{-1}}(a)) d\nu(\dot{g})$$

can be regarded as measurable mapping $\{\dot{g} \to \pi_\varphi(\alpha_{r(\dot{g})^{-1}}(a))\}$ from $G/H$ into $\pi_\varphi(A)$. If $r$ is continuous, the map $\{\dot{g} \to \pi_\varphi(\alpha_{r(\dot{g})^{-1}}(a))\}$ belongs to $C_b(G/H, \pi_\varphi(A))$. Identifying $C_b(G/H, \pi_\varphi(A))$ with $\mathcal{R}$, the result follows. □

*Lemma 3.3:* Let $\psi$ be a pure state of $L_\infty(G/H)$ and $\psi \restriction C_\infty(G/H) \neq 0$. Then there exists $\dot{g}_0 \in G/H$, such that $\psi(f) = f(\dot{g}_0)$, for all $f \in C_b(G/H)$.

*Proof:* From Ref. 21, 2.3.21, and Ref. 22, (20.52), there exists $\dot{g}_0 \in G/H$, such that $\psi(f) = f(\dot{g}_0)$, for all $f \in C_\infty(G/H)$. As $C_\infty(G/H)$ is a hereditary $C^*$-subalgebra of $C_b(G/H)$, the state defined on $C_\infty(G/H)$ has a unique extension to a state of $C_b(G/H)$ (Ref. 23, 3.1.6). This extension is, of course, given by $\psi(f) = f(\dot{g}_0)$ for $f \in C_b(G/H)$. □

The following lemma shows that there exists states on $\mathcal{M}$, which are dispersion-free on $L_\infty(G/H)$ and which have an infinite value for the classical observable generating $L_\infty(G/H)$, if and only if $G/H$ is noncompact.

*Lemma 3.4:* If $G/H$ is noncompact there exists a pure state $\psi$ of $L_\infty(G/H)$ such that $\psi \restriction C_\infty(G/H) = 0$.

*Proof:* $C_\infty(G/H)$ is a closed two-sided ideal of $C_b(G/H)$. As $G/H$ is noncompact the quotient $C^*$-algebra $U := C_b(G/H)/C_\infty(G/H)$ is nontrivial. The set

$$S_\infty : \{\varphi \in S(L_\infty(G/H)) \mid \varphi \restriction C_\infty(G/H) = 0\}$$

contains a set that is affine linear isomorphic to $S(U)$, the state space of $U$. Here $S_\infty$ is a nonempty convex weak$^*$-compact set which has an extreme point $\psi$ (Krein–Milman). Assume that $\psi$ can be decomposed in $\psi = \lambda\psi_1 + (1-\lambda)\psi_2$, $0 < \lambda < 1$, and $\psi_1, \psi_2$ states of $L_\infty(G/H)$. Then $0 = \psi(a) = \lambda\psi_1(a) + (1-\lambda)\psi_2(a)$, $\forall a \in C_\infty(G/H)_+$. Hence $\psi_1 \restriction C_\infty(G/H) = \psi_2 \restriction C_\infty(G/H) = 0$. Therefore $\psi_1, \psi_2 \in S_\infty$ and $\psi = \psi_1 = \psi_2$. Hence $\psi$ is pure. □

Lemma 3.5 and Corollary 3.6 following connect the dispersion-free states on $\mathcal{M}$ having a finite value for the classical observables generating $L_\infty(G/H)$ with the points of the "phase space" $G/H$ and the states on $\pi_\varphi(A)''$.

*Lemma 3.5:* Let $\Omega$ be an element of $D_f(\mathcal{M})$. There exists $\dot{g}_0 \in G/H$ such that for any $x = \{\dot{g} \to x(\dot{g})\} \in \mathcal{N}$ we have $\Omega(x) = \Omega_2(x(\dot{g}_0))$.

*Proof:* By definition, $\Omega$ is a product state where $\Omega_2$ is a state of $\pi_\varphi(A)''$ and $\Omega_1$ a pure state of $L_\infty(G/H)$ which do not vanish on $C_\infty(G/H)$. It follows from Lemma 3.3 that there exists a point $\dot{g}_0 \in G/h$ such that $\Omega_1(f) = f(\dot{g}_0)$, for all $f \in C_b(G/H)$. For any $\{\dot{g} \to x(\dot{g})\} \in C_b(G/H, \pi_\varphi(A)'')$ there exists a sequence

$$\left\{\sum_{m=1}^{N(n)} f_{nm}\otimes a_{nm}\right\}_{n\in\mathbb{N}}$$

in $\mathcal{N}$ such that

$$\lim\left\{\sup_{\dot g\in G/H}\|x(\dot g)-\sum_{m=1}^{N(n)} f_{nm}(\dot g)a_{nm}\|\right\}=0\,.$$

Now

$$\begin{aligned}\Omega(x)&=\lim\Omega\Big(\sum_{m=1}^{N(n)} f_{nm}\otimes a_{nm}\Big)\\&=\lim\sum_{m=1}^{N(n)} f_{nm}(\dot g_0)\Omega_2(a_{nm})\\&=\lim\Omega_2\Big(\sum_{m=1}^{N(n)} f_{nm}(\dot g_0)a_{nm}\Big)=\Omega_2(x(\dot g_0))\,.\end{aligned}\qquad\square$$

*Corollary 3.6:* (Let $r$ be continuous.) If $\Omega\in D_f(\mathcal{M})$, there exists $\dot g_0\in G/H$ such that

$$\Omega(\pi_\phi(a))=\Omega_2(\pi_\varphi(\alpha_{r(\dot g_0)^{-1}}(a)))\,,$$

for all $a\in A$.

The quasilocal $C^*$-algebras usually occurring in physics are simple, i.e., every representation of the respective algebra is faithful (with the exception of the noninteresting zero representation). The following Theorem 3.7 shows that every state $\omega$ on a simple $C^*$-algebra can be extended to a dispersion-free state on the generated $W^*$-algebra of an induced representation. If the state $\omega$ is pure then this extension may be chosen to be pure, too.

**Theorem 3.7:** Let $\pi_\varphi$ be faithful and let $\omega$ be a state of $A$. For every point $\dot g_0\in G/H$, there exists a state $\Omega\in D_f(\mathcal{M})$ given by $\Omega(x)=\Omega_2(x(\dot g_0))$, $x=\{\dot g\to x(\dot g)\}\in\mathcal{N}$, such that $\Omega\circ\pi=\omega$. If $\omega$ is a pure state of $A$ then $\Omega$ may be chosen to be a pure state of $\mathcal{M}$.

*Proof:* Let $\dot g_0\in G/H$. Denote by $\Omega_1$ a pure state of $L_\infty(G/H)$ such that $\Omega_1(f)=f(\dot g_0)$, for all $f\in C_b(G/H)$. Put $\Omega_2:=\omega\circ\alpha_{r(\dot g_0)}\circ\pi_\varphi^{-1}$. By the Hahn–Banach theorem, the state $\Omega_2$ of $\pi_\varphi(A)$ admits an extension to a state of $\pi_\varphi(A)''$ (Ref. 23, 3.1.6), which we denote by the same letter. The product state $\Omega=\Omega_1\otimes\Omega_2$ of $L_\infty(G/H)\otimes\pi_\varphi(A)''$ admits an extension to a state of $\mathcal{M}$ (Ref. 23, 3.1.6). By construction $\Omega\in D_f(\mathcal{M})$ and, according to Corollary 3.6,

$$\Omega(\pi_\varphi(a))=\Omega_2(\pi_\varphi(\alpha_{r(\dot g_0)^{-1}}(a)))=\omega(a),\quad\text{for all } a\in A\,.$$

If $\omega$ is pure then $\Omega_2:=\omega\circ\alpha_{r(\dot g)}\circ\pi_\varphi^{-1}$ is pure and $\Omega_2$ admits an extension to a pure state of $\pi_\varphi(A)''$ (Ref. 21, 2.3.24). The product state $\Omega_1\otimes\Omega_2$ of $L_\infty(G/H)\otimes\pi_\varphi(A)''$ is a pure state (Ref. 17, IV.4.13) and admits an extension to a pure state of $\mathcal{M}$ (Ref. 21, 2.3.24). $\square$

Theorems 3.9 and 3.10 following characterize the extension properties of states of a $C^*$-algebra $A$ to states of a $W^*$-algebra $\mathcal{M}=L_\infty(G/H)\,\overline{\otimes}\,\pi_\varphi(A)''$ that are dispersion-free on $L_\infty(G/H)$ and normal on $\pi_\varphi(A)''$. The physical meaning of these extension properties is described in the following section. Lemma 3.8 following is needed in the proofs of Theorems 3.9 and 3.10.

*Lemma 3.8:* Let $\psi$ be a state of $A$, $\pi_\psi$ the GNS representation of $\psi$, and $\pi$ an arbitrary representation of $A$. Consider the following conditions:

(1) $\psi$ is a $\pi$-normal state,

(2) $\pi_\psi$ is quasiequivalent to a subrepresentation of $\pi$ (in sign $\pi_\psi\leqslant\pi$),

(3) $\pi_\psi$ is quasiequivalent to $\pi$ (in sign $\pi_\psi\simeq\pi$).

It follows that (1) $\Leftrightarrow$ (2) $\Leftarrow$ (3). If, moreover, $\pi$ is a factor representation, then all conditions are equivalent.

*Proof:* (1) $\Rightarrow$ (2). We use the notations of Appendix A. The representation $\tau(a):=c(\pi)\pi_u(0)$, $a\in A$, is quasiequivalent to $\pi$. It follows from Appendix A that there exists a normal state $\omega$ of $\tau(A)''$, such that $\omega(\tau(a))=\psi(a)$, $a\in A$. At the same time, there exists a unit vector $\eta\in\mathcal{H}_u$ such that $(\eta|\pi_u(a)\eta)=\omega(\tau(a))$, $a\in A$. By normality this equation becomes $(\eta|x\eta)=\omega(c(\pi)x)$, $x\in A''$. By substituting $c(\pi)$ for $x$ we obtain

$$\|c(\pi)\eta\|^2=(\eta|c(\pi)\eta)=\omega(c(\pi))=\psi(1)=1=\|\eta\|^2\,.$$

Thus $\eta$ lies in the range of $c(\pi)$. Denote by $p$ the projection in $A'$ with range $[\pi_u(A)\eta]$. Thus $(\pi_u(\cdot)p,p\mathcal{H}_u,\eta)$ is the GNS representation of $\psi$. As $\eta$ lies in the range of $c(\pi)$, we have, from

$$c(\pi)\pi_u(a)\eta=\pi_u(a)c(\pi)\eta=\pi_u(a)\eta\,,\quad a\in A\,,$$

that $p\leqslant c(\pi)$. Therefore $p\leqslant c(\pi_\psi)\leqslant c(\pi)$ and $\pi_\psi\leqslant\pi$.

(2) $\Rightarrow$ (1). Follows from Appendix A and $\{\psi\}\subset N(\pi_\psi)\subset N(\pi)$. If $\pi$ is a factor representation, then $c(\pi)$ is an atom in the center of $A''$ (cf. Ref. 23, 3.8.13). It follows from the proof of (1) $\Rightarrow$ (2) that $c(\pi_\psi)=c(\pi)$. Hence (1) $\Rightarrow$ (3). $\square$

**Theorem 3.9** (existence of extension)**:** Let $\omega$ be a state of $A$ and $\pi_\omega$ the GNS representation of $\omega$. Consider the following conditions:

(1) there exists an $\Omega\in\phi_f(\mathcal{M})$ such that $\Omega\circ\pi_\phi=\omega$,

(2) there exists a $\dot g_0\in G/H$ such that $\pi_\omega\leqslant\pi_\varphi\circ\alpha_{r(\dot g_0)^{-1}}$,

(3) there exists a $\dot g_0\in G/H$ such that $\omega\simeq\varphi\circ\alpha_{r(\dot g_0)^{-1}}$.

It follows that (1) $\Leftrightarrow$ (2) $\Leftarrow$ (3). If, moreover, $\varphi$ is a factor state then all conditions are equivalent.

Let $G/H$ be noncompact. If, in addition, there exists a representation $(\pi_\infty,\mathcal{H}_\varphi)$ of $A$ such that $\|\pi_\varphi(\alpha_{r(\dot g)^{-1}}(a))-\pi_\infty(a)\|\to 0$, $a\in A$, for some net $\{\dot g\}$ running out of any compact subset of $G/H$, consider the following conditions:

(4) there exists $\Omega\in\phi_\infty(\mathcal{M})$ such that $\Omega\circ\pi_\phi=\omega$,

(5) $\pi_\omega\leqslant\pi_\infty$,

(6) $\pi_\omega\simeq\pi_\infty$.

It follows that (4) $\Leftrightarrow$ (5) $\Leftarrow$ (6). If, moreover, $\pi_\infty$ is a factor representation then all the conditions are equivalent.

*Proof:* (1) $\Rightarrow$ (2). According to Corollary 3.6, we have

$$\Omega(\pi_\phi(a))=\Omega_2(\pi_\varphi(\alpha_{r(\dot g_0)^{-1}}(a)))=\omega(a),\quad a\in A,$$

for some point $\dot g_0\in G/H$. Therefore, $\omega$ is a $\pi_\varphi\circ\alpha_{r(\dot g_0)^{-1}}$ normal state and $\pi_\omega\leqslant\pi_\varphi\circ\alpha_{r(\dot g_0)^{-1}}$ (Lemma 3.9).

(2) $\Rightarrow$ (1). According to Lemma 3.9, there exists a normal state $\Omega_2$ of $\pi_\varphi(A)''$ such that

$$\Omega_2(\pi_\varphi(\alpha_{r(\dot g_0)^{-1}}(a)))=\omega(a),\quad a\in A\,.$$

Denote by $\Omega_1$ a pure state of $L_\infty(G/H)$ such that $\Omega_1(f)=f(\dot g_0)$, $f\in C_b(G/H)$. By the Hahn–Banach theorem, the product state $\Omega_1\otimes\Omega_2$ of $L_\infty(G/H)\otimes\pi_\varphi(A)''$ admits an extension to a state of $\mathcal{M}$ (Ref. 23, 3.1.6). As $r$ is continuous, $\pi_\phi(A)\subset\mathcal{N}$ and $\Omega(\pi_\phi(a))=\omega(a)$ (Lemma 3.5). If $\varphi$ is a factor state, $\varphi\circ\alpha_{r(\dot g_0)^{-1}}$ is a factor state and (2) $\Rightarrow$ (3) follows from Lemma 3.9.

(4) $\Rightarrow$ (5). By definition, $\pi_\infty(A) \subset \pi_\varphi(A)$. Therefore, the constant function $\{\dot g \to \pi_\infty(a)\}$ is an element of $\mathcal{N}$ for all $a \in A$. From the definition of $\pi_\infty$ and from Ref. 24, 42.8, it follows that $\{\dot g \to (\pi_\varphi(\alpha_{r(\dot g)^{-1}}(a)) - \pi_\infty(a))\}$ belongs to $\mathcal{N}_\infty$. As $\Omega \in \phi_\infty(\mathcal{M})$, we have $\Omega \restriction \mathcal{N}_\infty = 0$. Therefore

$$\begin{aligned}\Omega(\pi_\phi(a)) &= \Omega(\{\dot g \to (\pi_\varphi(\alpha_{r(\dot g)^{-1}}(a)) \\ &\quad - \pi_\infty(a))\}) + \Omega(\{\dot g \to \pi_\infty(a)\}) \\ &= \Omega(\{\dot g \to \pi_\infty(a)\}) = \Omega_2(\pi_\infty(a)), \quad a \in A.\end{aligned}$$

$\Omega_2$ is a normal state of $\pi_\varphi(A)''$ and therefore a normal state of $\pi_\infty(A)''$. Hence $\omega$ is a $\pi_\infty$-normal state and $\pi_\omega \leqslant \pi_\infty$ (Lemma 3.9).

(5) $\Rightarrow$ (4). By Lemma 3.9 there exists a normal state $\Omega_2$ of $\pi_\infty(A)''$ such that $\Omega_2(\pi_\infty(a)) = \omega(a)$, $a \in A$. There exists a trace class operator $y$ on $\mathcal{H}_\varphi$ such that $\Omega_2(x) = \mathrm{tr}(xy)$, $x \in \pi_\infty(A)''$. Extend $\Omega_2$ to a normal state of $\pi_\varphi(A)''$ defining $\Omega_2(x) = \mathrm{tr}(xy)$, $x \in \pi_\varphi(A)''$. According to Lemma 3.4, there exists a pure state $\Omega_1$ on $L_\infty(G/H)$ such that $\Omega_1 \restriction C_\infty(G/H) = 0$. The product state $\Omega_1 \otimes \Omega_2$ on $L_\infty(G/H) \otimes \pi_\varphi(A)''$ admits an extension state of $\mathcal{M}$ (Ref. 23, 3.1.6). By construction $\Omega \in \phi_\infty(\mathcal{M})$. By the very same argument as in (4) $\Rightarrow$ (5) we have $\Omega(\pi_\phi(a)) = \Omega_2(\pi_\infty(a)) = \omega(a)$, $a \in A$. If $\pi_\infty$ is a factor representation, (5) $\Rightarrow$ (6) follows from Lemma 3.9. □

**Theorem 3.10** (uniqueness of extension): Let $\omega$ be a state of $A$. Suppose there exists $\Omega_i \in \phi_f(\mathcal{M})$, $i = 1,2$, such that $\Omega_i \circ \pi_\phi = \omega$. Then $\Omega_1 \restriction \mathcal{N} = \Omega_2 \restriction \mathcal{N}$.

Let $G/H$ be noncompact. Assume that there exists a representation $(\pi_\infty, \mathcal{H}_\varphi)$ of $A$ with the properties described in Theorem 3.10 and such that $\pi_\infty$ is disjoint to any $\varphi \circ \alpha_{r(\dot g)^{-1}}$, $\dot g \in G/H$. Suppose there exists $\Omega_i \in \phi(\mathcal{M})$, $i = 1,2$, such that $\Omega_i \circ \pi_\phi = \omega$. Then there remain two possibilities:

(a) $\Omega_i \in \phi_f(\mathcal{M})$ and then $\Omega_1 \restriction \mathcal{N} = \Omega_2 \restriction \mathcal{N}$,

(b) $\Omega_i \in \phi_\infty(\mathcal{M})$ and then

$$\Omega_1 \restriction \{\mathbb{C}1\} \bar\otimes \pi_\infty(A)'' = \Omega_2 \restriction \{\mathbb{C}1\} \bar\otimes \pi_\infty(A)''.$$

*Proof:* Let us first assume $\Omega_1, \Omega_2 \in \phi_f(\mathcal{M})$. According to Corollary 3.6, there exists $\dot g_i \in G/H$ and normal states $\Omega_{2i}$ of $\pi_\varphi(A)''$, such that $\Omega_i(\pi_\phi(a)) = \Omega_{2i}(\alpha_{r(\dot g_i)^{-1}}(a))$, $a \in A$. Therefore $\omega \in N(\pi_\varphi \circ \alpha_{r(\dot g_1)^{-1}}) \cap N(\pi_\varphi \circ \alpha_{r(\dot g_2)^{-1}})$. It follows from Assumption 2 that $\dot g_3 \neq \dot g_4$ implies $\pi_\varphi \circ \alpha_{r(\dot g_3)^{-1}} \not\simeq \pi_\varphi \circ \alpha_{r(\dot g_4)^{-1}}$ or equivalently $N(\pi_\varphi \circ \alpha_{r(\dot g_3)^{-1}}) \cap N(\pi_\varphi \circ \alpha_{r(\dot g_4)^{-1}}) = \varnothing$ (cf. Appendix A). Therefore $\dot g_1 = \dot g_2$ and $\Omega_{21}(\pi_\varphi(a)) = \Omega_{22}(\pi_\varphi(a))$, $a \in A$. As $\Omega_{2i}$ are normal states of $\pi_\varphi(A)''$, we have $\Omega_{21} = \Omega_{22}$. Therefore the product states $\Omega_{11} \otimes \Omega_{21}$ and $\Omega_{21} \otimes \Omega_{22}$ of $\mathcal{N}$ coincide.

Now let $\Omega_1 \in \phi_\infty(\mathcal{M})$ and $\Omega_2 \in \phi(\mathcal{M})$. Then $\Omega_1 \restriction \mathcal{N}_\infty = 0$. By the very same argument as in the proof of Theorem 3.10, (4) $\Rightarrow$ (5), we have $\Omega_1(\pi_\phi(a)) = \Omega_{12}(\pi_\infty(a))$, $a \in A$.

*Case 1:* $\Omega_2 \in \phi_f(\mathcal{M})$. It follows as above $\omega \in N(\pi_\infty) \cap N(\pi_\varphi \circ \alpha_{r(\dot g_2)^{-1}})$. By assumption $\pi_\infty \not\simeq \pi_\varphi \circ \alpha_{r(\dot g_2)^{-1}}$. This is a contradiction.

*Case 2:* $\Omega_2 \in \phi_\infty(\mathcal{M})$. Then $\Omega_2(\pi_\phi(a)) = \Omega_{22}(\pi_\infty(a))$, $a \in A$. As $\Omega_{2i}$ are normal states of $\pi_\varphi(A)''$, we have $\Omega_{21} \restriction \pi_\infty(A)'' = \Omega_{22} \restriction \pi_\infty(A)''$. □

## IV. COMPATIBILITY OF $C^*$- AND $W^*$-ALGEBRAIC QUANTUM MECHANICS

A $W^*$-algebraic description of an infinite system in a particular context has to fulfill compatibility conditions with its underlying $C^*$-algebraic description. The compatibility conditions we introduce are

(1) only certain states $\omega$ on the quasilocal $C^*$-algebra $A$ can be extended to states on $\mathcal{M} = \pi_\phi(A)''$ belonging to $D_f(\mathcal{M})$,

(2) states $\Omega \in D(\mathcal{M})$ extending a given state $\omega$ cannot have different values for the classical observables generating $L_\infty(G/H)$.

Condition (1) states that not all states on $A$ can describe physical situations corresponding to values of the classical global observables generating $L_\infty(G/H)$. Condition (2) excludes that a state $\omega$ on $A$ can at the same time describe two physical situations that differ with respect to an infinite number of degrees of freedom.

The quasilocal $C^*$-algebras usually occurring in physics in the description of the local context-independent features of an infinite quantum system are simple. Therefore Theorem 3.7 states that any (pure) state $\omega$ on a quasilocal $C^*$-algebra $A$ can be extended to a (pure) state $\Omega$ on $\mathcal{M}$ belonging to $D_f(\mathcal{M})$. Furthermore, a (pure) state $\omega$ on $A$ admits an extension to a (pure) state $\Omega$ belonging to $D_f(\mathcal{M})$ having any point of the "phase space" $G/H$ as value for the classical observables generating $L_\infty(G/H)$. Hence the compatibility conditions (1) and (2) can be maximally violated with states in $D_f(\mathcal{M})$.

The corresponding questions for infinite classical systems are not settled. The quasilocal $C^*$-algebras of an infinite classical system are, of course, never simple. This makes it more difficult to prove the existence of suitable faithful representations as presupposed in Theorem 3.7.

In the case where all classical global observables of a given context are generated by a kinematical group the $W^*$-algebra $\mathcal{M}$ factorizes into a classical part, the center $L_\infty(G/H)$ of $\mathcal{M}$, and a purely quantum mechanically part, a factor algebra. Every state on $\mathcal{M}$ that is dispersion-free on the center of $\mathcal{M}$ factorizes in the same way. Theorems 3.9 and 3.10 are tailored for this situation. They show that a $W^*$-description of a given context—with algebra $\mathcal{M}$ and the subset $\phi_f(\mathcal{M}) = S_{\mathcal{M}}$ of all states on $\mathcal{M}$—is equivalent to a $C^*$-description—with algebra $A$ and a subset

$$S_A = \{\omega \in S(A) \mid \omega \text{ is quasiequivalent to } \varphi \circ \alpha_{r(dg)^{-1}}, dg \in G/H\}$$

of all factor states on $A$.

Theorem 3.10 shows that any state $\omega$ in $S_A$ admits an extension to a state $\Omega$ in $S_{\mathcal{M}}$ and that any state $\omega$ not in $S_A$ does not admit such an extension.

Theorem 3.11 shows that two states $\Omega_1$ and $\Omega_2$ in $S_{\mathcal{M}}$ extending a state $\omega$ in $S_A$ have the same value for the classical observables generating the center of $\mathcal{M}$ and are equal on the purely quantum mechanically part of $\mathcal{M}$. Although the two states $\Omega_1$ and $\Omega_2$ can differ mathematically as functionals on $\mathcal{M}$ they describe physically identical states of the system in the given context. Two states $\Omega_1$ and $\Omega_2$ with these properties are henceforth called *physically identical.*

These results may be regarded as an *a posteriori* explanation of the fact that in the traditional description of infinite

systems in $C^*$-algebraic quantum mechanics factor states are used to describe a maximally specified physical situation referring to a given context and of the fact that only normal states on the corresponding factor algebra are considered.

Furthermore, these results open the possibility to introduce classical observables as parameters related to a kinematical group which can be assigned to certain states on a quasilocal $C^*$-algebra. This possibility does not depend on any limiting process involving the local $C^*$-algebras usually occurring in the parameter description of classical observables.

## V. INTERPRETATIONS OF QUANTUM MECHANICS OF INFINITE SYSTEMS

An interpretation of algebraic quantum mechanics of an infinite system in a particular context corresponds, on the one hand, to a subset $S_A$ of the set of all states on the quasilocal $C^*$-algebra $A$ and, on the other hand, to a subset $S_{\mathcal{M}}$ of the set of all states on the $W^*$-algebra $\mathcal{M} = \pi_\phi(A)''$ of observables. A minimal requirement that a particular interpretation is admissible is the compatibility of the $C^*$-algebra description and the $W^*$-algebra description as described in the previous section.

Therefore the results of the previous sections have the following consequences.

*Consequence 1:* An individual interpretation of an infinite system in a particular context whose respective $W^*$-algebra of observables is of type III or type II is not possible.

An individual interpretation of quantum mechanics presupposes the existence of "suitable" pure states. In the previous section we have seen that a $W^*$-algebra description corresponding to all pure states on the $W^*$-algebra of observables is not compatible with a $C^*$-algebra description. A $W^*$-algebra description corresponding to all pure states on the respective $W^*$-algebra of observables that are normal on the purely quantum mechanical part of the respective $W^*$-algebra is compatible with a suitable $C^*$-algebra description. But this description is only possible if the factor part of the respective $W^*$-algebra is a type I factor, since on a factor of type III and type II no pure normal states exist (cf. Lemma 2 in Appendix B).

The same result has also been established in Ref. 25 using completely different methods.

*Consequence 2:* An interpretation of algebraic quantum mechanics of an infinite system in a particular context where the classical observables of the respective $W^*$-algebra are interpreted as belonging to an individual system and the purely quantum part as belonging to an ensemble of systems equals the interpretation of the traditional description where the classical observables defining the given context are treated as parameters.

This consequence is established by the compatibility of the $W^*$-algebraic description $[\mathcal{M}, S_{\mathcal{M}} = \phi_f(\mathcal{M})]$ and the $C^*$-algebraic description $[A, S_A]$, where $S_A$ consists of the factor states on the $C^*$-algebra $A$ that are quasiequivalent to $\varphi \circ \alpha_{r(\dot{g})^{-1}}$, $\dot{g} \in G/H$.

## APPENDIX A

Let $(\pi_u, \mathcal{H}_u)$ denote the universal representation of the $C^*$-algebra $A$ and $c(\pi)$ the central cover of the representation $(\pi, \mathcal{H})$. The strong closure of $\pi_u(A)$ will henceforth be denoted by $A''$. The set of $\pi$-normal states of $A$ is defined by $N(\pi) := \{\omega \circ \pi | \omega$ a normal state of the weak closure of $\pi(A)\}$.

The concepts of quasiequivalence, disjointness, and quasisubequivalence are basic in the theory of representations (Ref. 26, Chap. 5). The implications (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3), (1′) $\Leftrightarrow$ (2′) $\Leftrightarrow$ (3′), (1″) $\Leftrightarrow$ (2″) $\Leftrightarrow$ (3″) between the following conditions are well known:

(1) $\pi_1$ is quasiequivalent to $\pi_2$,
(2) $c(\pi_1) = c(\pi_2)$,
(3) $N(\pi_1) = N(\pi_2)$,
(1′) $\pi_1$ is quasisubequivalent to $\pi_2$,
(2′) $c(\pi_1) \leqslant c(\pi_2)$,
(3′) $N(\pi_1) \subseteq N(\pi_2)$,
(1″) $\pi_1$ and $\pi_2$ are disjoint,
(2″) $c(\pi_1) \perp c(\pi_2)$,
(3″) $N(\pi_1) \cap N(\pi_2) = \emptyset$.

*Proof:* (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) (Ref. 17, III.2.12).

(1′) $\Leftrightarrow$ (2′) (Ref. 27, 10.5.42).

(1″) $\Leftrightarrow$ (2″) (Ref. 23, 3.8.10). The representation $\tau_i$ defined by $\tau_i(a) = c(\pi_i)\pi_u(a)$, $a \in A$, is a quasiequivalent to $\pi_i$ (Ref. 23, 3.8.2). Therefore $N(\tau_i) = N(\pi_i)$. From the proof of Lemma 3.9 it follows that

$$N(\tau_i) = \{(\xi | \pi_u(\cdot)\xi) | \xi \text{ a unit vector in Ran } c(\pi_i)\}.$$

(2′) $\Leftrightarrow$ (3′).

$$c(\pi_1) \leqslant c(\pi_2) \Leftrightarrow \text{Ran } c(\pi_1) \subseteq \text{Ran } c(\pi_2) \Leftrightarrow N(\pi_1) \subseteq N(\pi_2).$$

(2″) $\Leftrightarrow$ (3″).

$$c(\pi_1) \perp c(\pi_2) \Leftrightarrow \text{Ran } c(\pi_1) \cap \text{Ran } c(\pi_2) = \phi \Leftrightarrow N(\pi_1) \cap N(\pi_2) = \emptyset. \quad \square$$

## APPENDIX B

*Lemma 1:* Let $\nu$ denote the probability measure on $G/H$ introduced in Sec. II. Then $(G/H, \nu)$ is an atom-free measure space iff $G/H$ is not countable as a set.

*Proof:* $G/H$ is a second countable, locally compact Hausdorff space. Therefore $G/H$ becomes a countably separated measure space (Ref. 19, 8.6.11). It follows from Ref. 28, p. 589, that $(G/H, \nu)$ is nonatomic iff $\nu(\{\dot{g}\}) = 0$, for all $\dot{g} \in G/H$. The assertion is then an immediate consequence. $\square$

*Lemma 2:* A normal state $\varphi$ of a $W^*$-algebra $\mathcal{M}$ is pure iff its support projection $s(\varphi)$ is an atom in $\mathcal{M}$.

*Proof:* $L_\varphi := \{a \in \mathcal{M} | \varphi(a^*a) = 0\} = \mathcal{M}(1 - s(\varphi))$ (cf. Ref. 20, 1.14.2). If $s(\varphi)$ is not an atom in $\mathcal{M}$, then $L_\varphi$ is not a regular maximal left ideal of $\mathcal{M}$, and $\varphi$ is not pure (Ref. 23, 3.13.6). If $s(\varphi)$ is an atom in $\mathcal{M}$, then $s(\varphi)as(\varphi) = \varphi(a)s(\varphi)$, $a \in \mathcal{M}$. Suppose $\psi$ is a state of $\mathcal{M}$ majorized by $\varphi$. Then $\psi(s(\varphi)) = 1$ and

$$\psi(a) = \psi(s(\varphi)as(\varphi)) = \psi(\varphi(a)s(\varphi)) = \varphi(a), \quad a \in \mathcal{M}.$$

Hence $\varphi$ is pure. $\square$

[1]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[2]D. Kastler, "Equilibrium states of matter and operator algebras," in *Symposia Mathematica*, Vol. 20 (Academic, London, 1976).

[3]J. M. Jauch, *Foundations of Quantum Mechanics* (Addision–Wesley, Reading, MA, 1968).
[4]J.-M. Lévy-Leblond, "Galilei group and Galilei invariance," in *Group Theory and its Applications*, Vol. 2, edited by E. M. Loebl (Academic, New York, 1971).
[5]H. Primas, "Kinematical symmetries in molecular quantum mechanics," in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 79 (Springer, Berlin, 1978).
[6]H. Primas, *Chemistry, Quantum Mechanics and Reductionism, Lecture Notes in Chemistry*, Vol. 24 (Springer, Berlin, 1981).
[7]H. Weyl, Z. Phys. **46**, 1 (1927).
[8]G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[9]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, MA, 1976).
[10]A. Amann, Fortschr. Phys. **34**, 167 (1986).
[11]J. Pöttinger, Ph.D. dissertation, ETH Zürich, 1988.
[12]A. Amann, Helv. Phys. Acta **60**, 384 (1987).
[13]M. Takesaki, Acta Math. **119**, 273 (1967).
[14]G. W. Mackey, Acta Math. **99**, 265 (1958).
[15]A. Amann and U. Müller-Herold, Helv. Phys. Acta **59**, 1311 (1986).
[16]H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).
[17]M. Takesaki, *Theory of Operator Algebras I* (Springer, New York, 1979).
[18]G. W. Mackey, Acta Math. **55**, 101 (1952).
[19]D. L. Cohn, *Measure Theory* (Birkhäuser, Boston, 1980).
[20]S. Sakai, *$C^*$-Algebras and $W^*$-Algebras* (Springer, Berlin, 1971).
[21]O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, New York, 1979 and 1981), Vols. I and II.
[22]E. Hewitt and K. Stromberg, *Real and Abstract Analysis* (Springer, New York, 1965).
[23]G. K. Pedersen, *$C^*$-Algebras and Their Automorphism Groups* (Academic, London, 1979).
[24]H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie* (de Gruyter, Berlin, 1974).
[25]A. Amann, J. Math. Phys. **28**, 2384 (1987).
[26]J. Dixmier, *$C^*$-Algebras* (North-Holland, Amsterdam, 1977).
[27]R. V. Kadison and J. R. Ringrose, *Fundamentals of the Theory of Operator Algebras* (Academic, Orlando, 1983 and 1986), Vols. I and II.
[28]P. R. Halmos and J. v. Neumann, Ann. Math. **43**, 332 (1942).

# Additive functionals and operators on a quaternionic Hilbert space

C. S. Sharma and D. F. Almeida
*Mathematical Research Unit, Birkbeck College, University of London, 43 Gordon Square, London WC1H 0PD, England*

(Received 26 July 1988; accepted for publication 28 September 1988)

It is shown that the structure of functionals and operators on a quaternionic Hilbert space is much richer than is generally appreciated, but one has to work simultaneously with the usual definition and an unusual one of multiplication by scalar for functionals—the different definitions, of course, give rise to different vector spaces. A generalized version of the Riesz representation theorem for quaternionic Hilbert spaces is proved along with the basic theorem on the algebra of additive operators on such a space.

## I. INTRODUCTION

It has become abundantly clear that quaternionic vector spaces have important roles to play in celestial mechanics,[1,2] in quantum mechanics,[3,4] and in relativity.[5,6] For a study of a quaternionic Hilbert space as a possible model for quantum phenomena, Jauch and co-workers[7,8] have done some pioneering work on the theory of operators on a quaternionic Hilbert space. Our involvement with quaternionic spaces started with an attempt[9] to clarify and simplify the work of Jauch and co-workers. The referee of Ref. 9 raised a number of interesting questions and this led us to consider the structures of complex and quaternionic vector spaces at a very basic level.[10] The insight gained in Ref. 10 was used to develop a theory of semilinear operators[11] on a complex Hilbert space and to demonstrate that this theory had both expected and unexpected applications in physics. In the present work we return to quaternionic Hilbert spaces at a very basic level and study additive functionals and operators on such a space. It is shown that in this context additivity has a somewhat unexpected but nevertheless simple yet rich structure. We will see that for a full exploitation of the structure of additive functionals it becomes necessary to work with more than one definition of multiplication by scalars.

In Sec. II we establish our notation and state with indications of a proof a lemma that collects some of the properties of quaternionic numbers we are going to need in later sections.

In Sec. III we study additive functionals on a quaternionic Hilbert space. The collection of additive functionals on such a space forms a vector space with the usual definition of multiplication by scalars, but the Riesz representation therorem[3] has a very restricted form. We define a Riesz set of additive functionals that is a proper subset of the underlying set of the space of additive functionals but is not a subspace. We endow this subset with a vector space structure by defining a somewhat unusual multiplication by scalars and then we prove a generalized version of the Riesz representation theorem that establishes a linear vector space isomorphism between the vector space of bounded Riesz functionals and a direct sum of two copies of the quaternionic Hilbert space. For a real Hilbert space of dimension $n$ the space of bounded additive functionals is just the space of linear functionals on the space and has the same dimension, for a complex Hilbert space of dimension $n$ the space of bounded additive functionals is isomorphic to a direct sum of two copies of the Hilbert space and thus has dimension $2n$. We prove in this work that in a quaternionic Hilbert space of dimension $n$ the space of bounded additive functionals is isomorphic to a direct sum of four copies of the Hilbert space and, therefore, has dimension $4n$.

In Sec. IV we use the insight gained in the previous section to deduce various formulas for the adjoints of additive operators on a quaternionic Hilbert space and then state and prove the basic theorem on the algebra of additive operators on a quaternionic Hilbert space.

In Sec. V we conclude with a few remarks.

## II. FORMALITIES

We denote the fields of real and complex numbers by $\mathbb{R}$ and $\mathbb{C}$, respectively, and the skew field of quaternionic numbers by $\mathbb{H}$. Elementary properties of quaternions are described in Ref. 9. We state briefly the properties we are going to need in this work. Quaternions form a normed associative division algebra over $\mathbb{R}$ and are best described with the help of three distinct linearly independent abstract square roots of $-1$, which are denoted by symbols $i, j$, and $k$ and whose products are defined by

$$i^2 = j^2 = k^2 = -1, \tag{2.1}$$

$$ij = -ji = k, \tag{2.2}$$

$$jk = -kj = i, \tag{2.3}$$

$$ki = -ik = j. \tag{2.4}$$

It is easy to verify that $\mathbb{H}$ is a four-dimensional vector space over $\mathbb{R}$ where $1, i, j$, and $k$ are members of a basis. Thus any $\gamma \in \mathbb{H}$ has a unique representation as

$$\gamma = \gamma_0 + \gamma_1 i + \gamma_2 j + \gamma_3 k, \tag{2.5}$$

with $\gamma_0, \gamma_1, \gamma_2, \gamma_3 \in \mathbb{R}$. Quaternionic conjugation is defined by

$$1^* = 1, i^* = -i, j^* = -j, k^* = -k. \tag{2.6}$$

It is easy to verify that $\mathbb{H}$ is a normed algebra with the norm defined by

$$\|\gamma\| = (\gamma^*\gamma)^{1/2}. \tag{2.7}$$

In addition to the axioms of the norm, the norm satisfies, as in the complex case,

$$\|\gamma_1\gamma_2\| = \|\gamma_1\| \, \|\gamma_2\|. \tag{2.8}$$

 0022-2488/89/020369-07$02.50 © 1989 American Institute of Physics 

The remaining properties of quaternionic numbers that we need are collected together as the following lemma.

*Lemma 2.1:* Let $\gamma\in\mathbb{H}$. Then

(i) $\gamma = 0 \Leftrightarrow \gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = 0$,

where $\gamma_0$, $\gamma_1$, $\gamma_2$, and $\gamma_3$ are as in (2.5).

(ii) If the imaginary square root of $-1$ in $\mathbb{C}$ is identified with the quaternionic $i$ then $\gamma$ has a unique representation as

$$\gamma = c_0 + c_1 j, \tag{2.9}$$

with $c_0, c_1 \in \mathbb{C}$ and further

$$i\gamma i^* = -i\gamma i = c_0 - c_1 j. \tag{2.10}$$

Furthermore, in the preceding assertion $i$ can be replaced by $j$ (resp. $k$) and $j$ by $k$ (resp. $i$).

$$\text{(iii)} \sum_{\tau = i,j,k} \tau\gamma\tau = -\gamma - 2\gamma^*. \tag{2.11}$$

*Proof:* Statement (i) follows from the fact that $1, i, j$, and $k$ form a basis in $\mathbb{H}$ as a vector space over $\mathbb{R}$ and (ii) and (iii) follow as a result of straightforward calculations using (2.1)–(2.6). □

Let $\mathscr{H}$ be a vector space over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. We define a positive definite Hermitian form on $\mathscr{H}$ by

$$\langle\ ,\ \rangle : \mathscr{H}\times\mathscr{H} \to \mathbb{F},$$

$$\langle pu, qv\rangle = p\langle u,v\rangle q^*, \tag{2.12}$$

$$\langle u+v, w\rangle = \langle u,w\rangle + \langle v,w\rangle, \tag{2.13}$$

$$\langle u,v\rangle^* = \langle v,u\rangle, \tag{2.14}$$

$$\langle u,u\rangle = 0 \quad \text{only if } u = 0, \tag{2.15}$$

where $p^* = p$ if $\mathbb{F}$ is real, $p^*$ is the *complex conjugate* of $p$ if $\mathbb{F}$ is complex, and $p^*$ is the *quaternionic conjugate of* $p$ if $\mathbb{F}$ is quaternionic.

Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be Hilbert spaces over $\mathbb{F}$. We say that a map $A: \mathscr{H}_1 \to \mathscr{H}_2$ is *additive* if and only if for all $u, v \in \mathscr{H}_1$

$$A(u+v) = A(u) + A(v). \tag{2.16}$$

If, in addition, the map $A$ satisfies

$$A(pu) = pA(u), \tag{2.17}$$

for all $p\in\mathbb{F}$ and all $u\in\mathscr{H}_1$, then it is called *linear*.

If, on the other hand, $A$ satisfies

$$A(pu) = p^*A(u), \tag{2.18}$$

for all $p\in\mathbb{F}$ and all $u\in\mathscr{H}_1$, then it is called *semilinear*. It was shown in Ref. 10 that semilinear maps according to this particular definition do not exist for the quaternionic case, where it is necessary to define three different kinds of semilinearities called *i, j,* and *k semilinearity* thus: an $i$-semilinear map $A$ from $\mathscr{H}_1$ to $\mathscr{H}_2$ is an additive map that satisfies

$$\begin{aligned} &A(ru) = rAu, \quad A(iu) = iAu, \\ &A(ju) = -jAu, \quad A(ku) = -kAu, \end{aligned} \tag{2.19}$$

for all $r\in\mathbb{R}$ and all $u\in\mathscr{H}_1$ and $j$ and $k$ semilinearities have analogous definitions. It was shown by Coulson[12] that any additive map $A$ can be written as a sum of four maps $A_0$, $A_1$, $A_2$, and $A_3$, where $A_0$ is linear and $A_1$, $A_2$, and $A_3$ are, respectively, $i$, $j$, and $k$ semilinear and that this decomposition is unique. In this decomposition $A_0$, $A_1$, $A_2$, and $A_3$ are given by

$$A_0u = \tfrac{1}{4}[Au - iA(iu) - jA(ju) - kA(ku)], \tag{2.20a}$$

$$A_1u = \tfrac{1}{4}[Au - iA(iu) + jA(ju) + kA(ku)], \tag{2.20b}$$

$$A_2u = \tfrac{1}{4}[Au + iA(iu) - jA(ju) + kA(ku)], \tag{2.20c}$$

$$A_3u = \tfrac{1}{4}[Au + iA(iu) + jA(ju) - kA(ku)]. \tag{2.20d}$$

It was proved by Pian and Sharma[13] that in the complex case every additive continuous map from $\mathscr{H}_1$ to $\mathscr{H}_2$ is a direct sum of a linear and a semilinear continuous map from $\mathscr{H}_1$ to $\mathscr{H}_2$. (Reference 13 was in the more general context of Banach spaces of which Hilbert spaces are particular cases.)

A map from a Hilbert space $\mathscr{H}$ to itself will be called an operator on $\mathscr{H}$.

Let $A$ be an additive operator on $\mathscr{H}$. A norm of $A$ denoted by $\|A\|$ is defined by the formula

$$\|A\| = \sup_{\|x\|=1} \|Ax\|. \tag{2.21}$$

Let $\mathscr{N}_1 \oplus \mathscr{N}_2$ be the direct sum of two normed spaces $\mathscr{N}_1$ and $\mathscr{N}_2$. An element of $\mathscr{N}_1 \oplus \mathscr{N}_2$ is a pair $(x_1, x_2)$ with $x_1\in\mathscr{N}_1$ and $x_2\in\mathscr{N}_2$. A norm of $(x_1, x_2)$ denoted by $\|(x_1, x_2)\|$ can be defined by the formula

$$\|(x_1,x_2)\| = \|x_1\| + \|x_2\|. \tag{2.22}$$

Let $A$ be a bounded additive operator on a complex Hilbert space $\mathscr{H}$. Then by the theorem of Pian and Sharma[13] $A$ belongs to the direct sum of the spaces of bounded linear and semilinear operators on $\mathscr{H}$ and hence by (2.22) its norm is defined by

$$\|A\| = \|A_1\| + \|A_2\|, \tag{2.23}$$

where $A_1$ and $A_2$ are, respectively, the linear and semilinear components of $A$.

Let $A$ be an operator or a matrix. Then $A$ is said to be *involutive*[14] if and only if

$$A^2 = I. \tag{2.24}$$

Let $\mathscr{A}$ be an algebra. An *involution* $*$ on $\mathscr{A}$ is an involutive operator $\mathscr{A}$ that takes $A$ to $A^*$ and satisfies the following properties: (i) $*$ is a homomorphism of the additive group in the algebra, that is,

$$(A+B)^* = A^* + B^* \tag{2.25}$$

for all $A, B\in\mathscr{A}$; (ii) $*$ is product reversing, that is,

$$(AB)^* = B^*A^* \tag{2.26}$$

and being involutive, of course, means that it satisfies

$$A^{**} = A, \tag{2.27}$$

for all $A\in\mathscr{A}$. This definition is a generalization of the one given by Rudin.[15] Here, unlike Rudin, we do not require $*$ to be semilinear but we require it to be additive.

## III. ADDITIVE FUNCTIONALS ON A QUATERNIONIC HILBERT SPACE

A deeper study of additive functionals reveals that they have a much richer structure than has been realized in earlier studies,[7–12] but one has to work with two different kinds of multiplication by scalars. For a map $f$ that takes values in any vector space over a field $\mathbb{F}$, multiplication by a scalar $s$ (a scalar is any member of $\mathbb{F}$) is normally defined to give a map $sf$ with the property

$$(sf)(u) = s(f(u)). \tag{3.1}$$

It was pointed out in Ref. 9 that for vector spaces over the quaternionic field scalar multiples of linear maps are not linear. However, for the special case of functionals on such vector spaces, it is possible to define an unusual multiplication by scalars in such a way that it turns the set of linear functionals into a vector space over the quaternions.

*Definition 3.1:* Let $F$ be an additive functional on a vector space $\mathscr{V}$ over $\mathbb{H}$. Multiplication of $F$ by $s\in\mathbb{H}$ on the left is defined by

$$(s\cdot F)(x) = F(x)s^* \quad \forall x\in\mathscr{V}. \tag{3.2}$$

It is easy to verify that this definition satisfies all the necessary axioms and the collection of additive functionals on a vector space is a vector space over $\mathbb{H}$ whether we use the usual definition of multiplication by scalars or this one, but in any vector space only one kind of multiplication by scalars is allowed and thus the two definitions lead to two different vector spaces though they share the same underlying set. We shall retain the bullet whenever we use the new definition to emphasize that we are using the alternative definition. We shall denote the vector space of bounded additive functionals on $\mathscr{H}$ by $\mathscr{A}(\mathscr{H})$ when the usual definition of multiplication by scalars is used and by $\mathscr{A}\cdot(\mathscr{H})$ when our alternative definition is used. The collection of bounded linear functionals with the alternative definition of multiplication by scalars is a subspace of $\mathscr{A}\cdot(\mathscr{H})$ and will be denoted by $\mathscr{L}\cdot(\mathscr{H})$.

Given a linear functional $F$, it is easy to verify that $\tau F$ ($\tau = i,j,k$) is $\tau$ semilinear. The multiplication by scalars defined through (3.2) not only turns the collection of bounded linear functionals into a subspace but the collection of bounded $\tau$-semilinear functionals that we denote by $\mathscr{S}\mathscr{L}_\tau(\mathscr{H})$ for each value of $\tau$ also becomes a subspace; which, as in the linear case, we denote by $\mathscr{S}\mathscr{L}_\tau\cdot(\mathscr{H})$. Multiplication by $\tau$ on the left in $\mathscr{A}\cdot(\mathscr{H})$ is then a linear operator that satisfies

$$\tau(p\cdot F) = p\cdot(\tau F), \quad \forall p\in\mathbb{H} \text{ and } \forall F\in\mathscr{A}\cdot(\mathscr{H}), \tag{3.3}$$

but one must carefully remember that in this structure

$$\tau(p\cdot F) \neq (\tau p)\cdot F. \tag{3.4}$$

In the next lemma we summarize some properties of $\mathscr{A}\cdot(\mathscr{H})$ in this context.

*Lemma 3.2:* The vector space $\mathscr{A}\cdot(\mathscr{H})$ has the decomposition

$$\mathscr{A}\cdot(\mathscr{H}) = \mathscr{L}\cdot(\mathscr{H}) \oplus \mathscr{S}\mathscr{L}_i\cdot(\mathscr{H}) \oplus \mathscr{S}\mathscr{L}_j\cdot(\mathscr{H}) \oplus \mathscr{S}\mathscr{L}_k\cdot(\mathscr{H}). \tag{3.5}$$

Multiplication on the left by each $\tau$ ($\tau = i,j,k$) in $\mathscr{A}\cdot(\mathscr{H})$ is a linear automorphism that takes each of the subspaces in the decomposition (3.5) to another in the same family.

*Proof:* The decomposition follows from (2.20a)–(2.20d) and the observation that the collection of linear and $\tau$-semilinear functionals in this structure are subspaces. Linearity of multiplication on the left by $\tau$ follows by definition (3.3). That it is a bijection and therefore an automorphism follows from (2.1) and a straightforward calculation using definitions (2.17) and (2.19) shows that $\tau$ maps bijectively $\mathscr{L}\cdot(\mathscr{H})$ and $\mathscr{S}\mathscr{L}_\tau\cdot(\mathscr{H})$ into each other and the remaining two subspaces into each other. □

As was stated in the previous section and proved in Ref. 10, in a quaternionic Hilbert space $\mathscr{H}$, an additive map $A$ with the property

$$A(px) = p^*Ax, \quad \forall p\in\mathbb{H} \text{ and } \forall x\in\mathscr{H}, \tag{3.6}$$

does not exist. However, in the special case of functionals it is possible to define an *antilinearity* by the following definition.

*Definition 3.3:* Let $F$ be an additive functional on a quaternionic Hilbert space $\mathscr{H}$. Here $F$ is said to be *antilinear* if and only if it satisfies

$$F(px) = (F(x))p^*, \quad \forall p\in\mathbb{H} \text{ and } \forall x\in\mathscr{H}. \tag{3.7}$$

It is easily verified that a scalar multiple (with the usual definition) of an antilinear functional is antilinear and the collection of bounded antilinear functionals on $\mathscr{H}$ is a subspace of $\mathscr{A}(\mathscr{H})$; this subspace will be denoted by $\mathscr{S}(\mathscr{H})$.

We have seen that every additive functional can be decomposed into a sum of a linear and an $i$-, a $j$-, and a $k$-semilinear functional according to Eqs. (2.20a)–(2.20d). For a decomposition of an additive functional in which antilinearity plays a role corresponding to that played by linearity in the decomposition just referred to, it is necessary to define three kinds of semiantilinearity.

*Definition 3.4:* Let $F$ be an additive functional on a quaternionic Hilbert space $\mathscr{H}$. Then $F$ is said to be *i semiantilinear* if and only if it has the following properties:

$$F(ru) = rF(u), \quad F(iu) = -F(u)i, \quad F(ju) = F(u)j,$$
$$F(ku) = F(u)k, \quad \forall r\in\mathbb{R} \text{ and } \forall u\in\mathscr{H}. \tag{3.8}$$

There are analogous definitions of *j* and *k semiantilinearities.*

We can now state the following lemma.

*Lemma 3.5:* Let $F$ be an additive functional on a quaternionic Hilbert space $\mathscr{H}$. Then $F$ has the following decomposition:

$$F = F_0 + F_1 + F_2 + F_3, \tag{3.9}$$

where

$$F_0(u) = \tfrac{1}{4}[F(u) + F(iu)i + F(ju)j + F(ku)k], \tag{3.10a}$$
$$F_1(u) = \tfrac{1}{4}[F(u) + F(iu)i - F(ju)j - F(ku)k], \tag{3.10b}$$
$$F_2(u) = \tfrac{1}{4}[F(u) - F(iu)i + F(ju)j - F(ku)k], \tag{3.10c}$$
$$F_3(u) = \tfrac{1}{4}[F(u) - F(iu)i - F(ju)j + F(ku)k]. \tag{3.10d}$$

In this decomposition $F_0$ is antilinear and $F_1$, $F_2$, and $F_3$ are, respectively, $i$, $j$, and $k$ semiantilinear.

*Proof:* By adding together (3.10a)–(3.10d) we get (3.9). The rest of the assertion follows by equally straightforward computation. □

We can now observe that under the usual multiplication by scalars in $\mathscr{A}(\mathscr{H})$ not only the collection of bounded antilinear functionals denoted by $\mathscr{S}(\mathscr{H})$ but also the collections of bounded $\tau$-semiantilinear functionals, denoted by $\mathscr{A}\mathscr{S}_\tau(\mathscr{H})$ for each value of $\tau$, constitute subspaces. It is now possible to define an operator $R_\tau$, multiplication on the right by $\tau$ in $\mathscr{A}(\mathscr{H})$, by

$$(R_\tau F)(x) = F(x)\tau, \quad \forall x\in\mathscr{H}, \tag{3.11}$$

similar to multiplication on the left by $\tau$ in $\mathscr{A}\cdot(\mathscr{H})$. We now have a lemma that is merely the dual of Lemma 3.2.

*Lemma 3.6:* The vector space $\mathscr{A}(\mathscr{H})$ has the decomposition

$$\mathscr{A}(\mathscr{H}) = \mathscr{S}(\mathscr{H}) \oplus \mathscr{A}\mathscr{S}_i(\mathscr{H}) \oplus \mathscr{A}\mathscr{S}_j(\mathscr{H}) \oplus \mathscr{A}\mathscr{S}_k(\mathscr{H}). \tag{3.12}$$

Multiplication on the right by each $\tau$ ($\tau = i,j,k$) in $\mathscr{A}(\mathscr{H})$ is a linear automorphism that takes each of the subspaces in the decomposition (3.12) to another in the same family.

*Proof:* The proof is similar to that of Lemma 3.2. □

We define next a very obvious conjugation for additive functionals.

*Definition 3.7.* Let $F$ be an additive functional on a quaternionic Hilbert space $\mathscr{H}$. The conjugate of $F$ denoted by $F^*$ is defined by the formula

$$F^*(x) = (F(x))^*, \quad \forall x \in \mathscr{H}. \tag{3.13}$$

We state next the obvious properties of conjugation of functionals as a lemma.

*Lemma 3.8:* Conjugation is a norm-preserving linear isomorphism between $\mathscr{A}(\mathscr{H})$ and $\mathscr{A}^\cdot(\mathscr{H})$ in which the subspaces $\mathscr{S}(\mathscr{H})$ and $\mathscr{L}^\cdot(\mathscr{H})$ correspond to each other and so do subspaces $\mathscr{A}\mathscr{S}_\tau(\mathscr{H})$ and $\mathscr{S}\mathscr{L}_\tau{}^\cdot(\mathscr{H})$ for $\tau = i,j,k$.

*Proof:* Let $F \in \mathscr{A}(\mathscr{H})$. Additivity of $F^*$ follows from the additivity of $F$ and the additivity of conjugation of quaternionic numbers. Linearity of conjugation follows from its evident additivity and the following calculation:

$$(sF)^*(x) = (sF(x))^* = (F(x))^* s^* = F^*(x)s^* = (s \cdot F^*)(x). \tag{3.14}$$

The involutive property of conjugation proves that it is a bijection. The norm being preserved follows from the fact that conjugation of quaternionic numbers is norm preserving. Next suppose $F$ is linear. Then the calculation

$$F^*(px) = (F(px))^* = (pF(x))^* = (F(x))^* p^* = F^*(x)p^* \tag{3.15}$$

shows that $F^*$ is antilinear. A similar calculation shows the correspondence between $\tau$-semilinear and $\tau$-semiantilinear maps. □

Before proceeding further we should point out that in Refs. 7 and 8 multiplication by a scalar in a quaternionic vector space was defined on the right while operators and maps acted from the left. If that convention is used then the usual multiplication by a scalar of a linear functional will yield a linear functional and thus linear functionals will naturally form a subspace of the space of additive functionals. On the other hand, one will have to define an unusual multiplication by scalars to turn the collection of antilinear functionals (these will have a somewhat different definition in the notations of Ref. 7) into a vector space. The advantage of that approach is that linear functionals that many workers regard as more important have the usual and more familiar multiplication by scalars. The advantage of our approach is that it reminds us of the fact that additive functionals are a very special case of an additive map between two quaternionic spaces and only in this special case can linear functionals be given a vector space structure by a special construct and that the usual multiplication by scalars gives a vector space structure to antilinear functionals that have no counterparts in the more general case.

In Ref. 10 it was proved that for the complex case the most general version of the Riesz representation theorem establishes an additive norm-preserving isomorphism between the space of additive functionals and the direct sum of two copies of the Hilbert space. In order to get a similar theorem for the quaternionic case we have to introduce some more definitions and some more new structure.

*Definition 3.9:* An additive functional $R$ on a quaternionic Hilbert space $\mathscr{H}$ is said to be a *Riesz functional* if and only if $R$ has the following decomposition:

$$R = R_L + R_A, \tag{3.16}$$

where $R_L$ is linear and $R_A$ is antilinear. The collection of Riesz functionals on $\mathscr{H}$ will be called a *Riesz set* and will be denoted by $\mathscr{R}(\mathscr{H})$.

It is obvious that the decomposition when it exists is unique. We establish the necessary and sufficient condition that an additive functional on $\mathscr{H}$ is a Riesz functional.

*Proposition 3.10:* Let $F$ be an additive functional on a quaternionic Hilbert space $\mathscr{H}$. The necessary and sufficient condition that $F$ is a Riesz functional is that it has the following property:

$$3F(x) = \sum_{\tau = i,j,k,} (F(\tau x)\tau - \tau F(\tau x) - \tau F(x)\tau), \quad \forall x \in \mathscr{H}. \tag{3.17}$$

*Proof: Necessity:* From the definitions of linearity and antilinearity it immediately follows that if $F$ is either linear or antilinear

$$F(x) = F(\tau x)\tau - \tau F(\tau x) - \tau F(x)\tau, \quad \forall \tau \in \{i,j,k\} \text{ and } \forall x \in \mathscr{H}. \tag{3.18}$$

Suppose that $F$ is a Riesz functional. Necessity of (3.17) now follows by adding the three equations (3.18) for the linear and antilinear components of $F$.

*Sufficiency:* Suppose that the additive functional $F$ satisfies condition (3.17). Define $F_l$ and $F_a$ by

$$4F_l(x) = F(x) - \sum_{\tau = i,j,k} \tau F(\tau x), \quad \forall x \in \mathscr{H} \tag{3.19}$$

and

$$4F_a(x) = F(x) + \sum_{\tau = i,j,k} F(\tau x)\tau, \quad \forall x \in \mathscr{H}. \tag{3.20}$$

Comparison of (3.19) with (2.20a) and (3.20) with (3.10a) immediately shows that $F_l$ is linear and $F_a$ is antilinear. With the help of (2.11), (3.19), and (3.20) the condition (3.17) takes the form

$$F = \tfrac{1}{2}F^* + F_l + F_a. \tag{3.21}$$

By taking the conjugate of (3.21) we obtain

$$F^* = \tfrac{1}{2}F + F_l^* + F_a'. \tag{3.22}$$

By eliminating $F^*$ between (3.21) and (3.22) $F$ can be written as

$$F = F_L + F_A, \tag{3.23}$$

where

$$F_L = \tfrac{4}{3}F_l + \tfrac{2}{3}F_a^* \tag{3.24}$$

being a sum of linear functionals is linear and, similarly,

$$F_A = \tfrac{4}{3}F_a + \tfrac{2}{3}F_l^* \tag{3.25}$$

is antilinear. □

Since a sufficient condition always implies the necessary conditions and a look at the proof will show that the necessary condition could have had a somewhat different formulation, we have, in passing, proved the following corollary.

*Corollary 3.11:* Let $F$ be an additive functional on a quaternionic Hilbert space that satisfies condition (3.17). Then it also satisfies (3.18).

While the set of bounded members of $\mathscr{R}(\mathscr{H})$ is a subset of $\mathscr{A}(\mathscr{H})$, it is a subspace of neither $\mathscr{A}(\mathscr{H})$ nor $\mathscr{A}^{\cdot}(\mathscr{H})$, but its definition (3.16) indicates that it is a vector space, namely the direct sum of $\mathscr{L}^{\cdot}(\mathscr{H})$ and $\mathscr{S}(\mathscr{H})$. From this it is easy to deduce the prescription for multiplication by scalars in $\mathscr{R}(\mathscr{H})$ that gives it a vector space structure.

*Lemma 3.12:* Let $\mathscr{H}$ be a quaternionic Hilbert space and let $\mathscr{R}(\mathscr{H})$ be the Riesz subset of the space $\mathscr{A}(\mathscr{H})$ of bounded additive functionals on $\mathscr{H}$. The prescription for multiplication by a scalar on the left which makes $\mathscr{R}(\mathscr{H})$ a vector space is the following.

Let $R\in\mathscr{R}(\mathscr{H})$. Then $R$ is first decomposed into the sum of its linear and antilinear components thus

$$R = R_L + R_A, \tag{3.26}$$

then multiplication of $R$ by a scalar $p\in\mathbb{H}$ on the left denoted by $p*R$ is given by

$$p*R = p\cdot R_L + pR_A. \tag{3.27}$$

We shall denote the bounded members of the Riesz set with this definition of multiplication by a scalar by $\mathscr{R}^*(\mathscr{H})$ and call it the *Riesz space* of the quaternionic Hilbert space $\mathscr{H}$. We can now state and prove the generalized version of the Riesz representation theorem for a quaternionic Hilbert space.

*Proposition 3.13:* Let $\mathscr{H}$ be a quaternionic Hilbert space. Then there exists a norm-preserving linear isomorphism between the Riesz space $\mathscr{R}^*(\mathscr{H})$ of $\mathscr{H}$ and $\mathscr{H}\oplus\mathscr{H}$.

*Proof:* We recall that

$$\mathscr{R}^*(\mathscr{H}) = \mathscr{L}^{\cdot}(\mathscr{H}) \oplus \mathscr{S}(\mathscr{H}). \tag{3.28}$$

By repeating each step of the proof of the Riesz representation theorem for bounded semilinear functionals on a complex Hilbert space as given in Ref. 10 where none of the arguments are altered by the changed circumstances: $\mathscr{H}$ is now quaternionic and $\Phi$ is now antilinear, we prove that a norm-preserving linear isomorphism exists between $\mathscr{S}(\mathscr{H})$ and $\mathscr{H}$. By Lemma 3.8 conjugation is a norm-preserving linear isomorphism between $\mathscr{S}(\mathscr{H})$ and $\mathscr{L}^{\cdot}(\mathscr{H})$, by taking the composition of this isomorphism with the isomorphism of the preceding sentence we prove that there exists a norm-preserving linear isomorphism between $\mathscr{L}^{\cdot}(\mathscr{H})$ and $\mathscr{H}$. By taking the direct sum of the two isomorphisms we get the norm-preserving linear isomorphism between $\mathscr{R}^*(\mathscr{H})$ and $\mathscr{H}\oplus\mathscr{H}$. □

In the complex case, the isomorphism between $\mathscr{A}(\mathscr{H})$ and $\mathscr{H}\oplus\mathscr{H}$ was merely additive: one of its components was linear and the other semilinear. Here it has been necessary to arrange things in such a way that both components are linear. We could have contrived the situation in the complex case too in such a way that both components of the Riesz isomorphism were linear: it would have been necessary to define multiplication by scalars for linear functionals in the complex Hilbert space by (3.2) remembering that all complex numbers commute.

We can now state and prove our final proposition on additive functionals.

*Proposition 3.14:* Both $\mathscr{A}(\mathscr{H})$ and $\mathscr{A}^{\cdot}(\mathscr{H})$ are isomorphic with $\mathscr{H}\oplus\mathscr{H}\oplus\mathscr{H}\oplus\mathscr{H}$. Given a bounded additive functional $F$ on $\mathscr{H}$ and a definite choice of $i$, $j$, and $k$ in $\mathbb{H}$, there exist four unique vectors $u_0$, $u_1$, $u_2$, and $u_3$ in $\mathscr{H}$ such that

$$F(x) = \langle x,u_0\rangle + i\langle x,u_1\rangle + j\langle x,u_2\rangle + k\langle x,u_3\rangle, \quad \forall x\in\mathscr{H}, \tag{3.29}$$

and

$$F^*(x) = \langle u_0,x\rangle - \langle u_1,x\rangle i - \langle u_2,x\rangle j - \langle u_3,x\rangle k, \quad \forall x\in\mathscr{H}. \tag{3.30}$$

*Proof:* The proof immediately follows from Lemma 3.2 and isomorphisms found in Lemma 3.8 and Proposition 3.13. □

An easy corollary of this proposition is the following.

*Corollary 3.15:* Let $\mathscr{H}$ be a quaternionic Hilbert space of dimension $n$. Then both $\mathscr{A}(\mathscr{H})$ and $\mathscr{A}^{\cdot}(\mathscr{H})$ have dimension $4n$.

*Proof:* In view of Proposition 3.14, the proof is obvious. □

Before concluding this section, we give a nice characterization of $\tau$ semilinearity and $\tau$ semiantilinearity that is coordinate-free in both $\mathscr{H}$ and $\mathbb{H}$.

*Definition 3.16:* Let $\tau$ be any quaternionic square root of $-1$, then (a) an additive map $f$ from a quaternionic Hilbert space $\mathscr{H}_1$ to another such space $\mathscr{H}_2$ is said to be *$\tau$ semilinear* if and only if

$$f(\alpha x) = \tau\alpha\tau^* f(x), \tag{3.31}$$

for all $\alpha\in\mathbb{H}$ and for all $x\in\mathscr{H}_1$, and (b) an additive functional $F$ on a quaternionic Hilbert space $\mathscr{H}$ is said to be *$\tau$ semiantilinear* if and only if

$$F(\alpha x) = F(x)\tau\alpha^*\tau^*, \tag{3.32}$$

for all $\alpha\in\mathbb{H}$ and for all $x\in\mathscr{H}$.

Note that (a) is valid for any additive map from $\mathscr{H}_1$ to $\mathscr{H}_2$ including a functional ($\mathscr{H}_2 = \mathbb{H}$) and an operator ($\mathscr{H}_2 = \mathscr{H}_1$) while (b) is valid only for functionals. It is easy to verify that the present definitions are completely equivalent to the previous ones. Note also that if $\tau$ is allowed to be a square root of either $-1$ or $+1$, then $\tau$ semilinearity corresponding to $\tau = \sqrt{+1} = 1$ is just linearity.

## IV. THE ALGEBRA OF ADDITIVE OPERATORS ON A QUATERNIONIC HILBERT SPACE

Let $\mathscr{H}$ be a quaternionic Hilbert space. The existence and elementary properties of the adjoint of a bounded linear operator $\mathscr{H}$ were established by Horwitz and Biedenharn.[3] Our main aim in the section is to establish the existence of the adjoint of any bounded additive operator on $\mathscr{H}$ and to prove the basic theorem on the algebra of bounded additive opera-

tors on $\mathscr{H}$ similar to the one that was proved for such operators on a complex Hilbert space in Ref. 16.

It was shown in Ref. 10 that the existence of the adjoint of a linear operator in a quaternionic Hilbert space could also be demonstrated by treating the quaternionic Hilbert space as a complex space. The same method will be used here to establish the existence and fundamental properties of $\tau$-semilinear operators ($\tau = i,j,k$). It is therefore necessary to recapitulate briefly the steps that are taken in treating a quaternionic Hilbert space as a complex one and then reverting back to the original quaternionic structure. First, we make a choice of $i$, $j$, and $k$, and then we choose one of these, say (without loss of generality), $i$. We identify quaternionic numbers of the form $r_0 + r_1 i$ with $\mathbb{C}$, now every quaternionic number $\gamma$ has the unique representation of the form

$$\gamma = c_1 + jc_2, \text{ with } c_1, c_2 \in \mathbb{C}. \tag{4.1}$$

Next we treat $\mathscr{H}$ as a complex Hilbert space with vectors $u$ and $ju$ for each $u \in \mathscr{H}$ regarded as not only linearly independent but also perpendicular. The relation between $u$ and $ju$ is no longer that of a multiplication by a scalar on the left: $j$ in this representation is not a complex number. However, $j$ does have interpretation as an antiunitary operator whose square is the negative of the identity operator $I$ on $\mathscr{H}$ as a complex Hilbert space: in this context we denote $j$ by a bold Roman letter $\mathbf{j}$ to remind us that it is an operator and $\mathscr{H}$ by $\mathscr{H}_c$ to remind us that we are in a complex space. From the properties described in the preceding sentence we deduce

$$\mathbf{j}^* = -\mathbf{j}. \tag{4.2}$$

In the representation (4.1) we call $c_1$ the complex part of $q$ and for the inner product $\langle u,v\rangle_c$ of two vectors $u$ and $v$ in $\mathscr{H}_c$ we use the formula

$$\langle u,v\rangle_c = \text{complex part of } \langle u,v\rangle, \tag{4.3}$$

where $\langle u,v\rangle$ is the inner product in the quaternionic Hilbert space $\mathscr{H}$. Linear operators on $\mathscr{H}$ continue to be linear on $\mathscr{H}_c$ and have the property that they commute with $\mathbf{j}$; $i$-semilinear operators on $\mathscr{H}$ are also linear on $\mathscr{H}_c$ but they anticommute with $\mathbf{j}$. There are also appropriate representations for $j$- and $k$-semilinear operators, but we do not need them. Finally we can revert to the original quaternionic structure by identifying the operator $\mathbf{j}$ with the quaternionic number $j$, regarding $u$ and $ju$ linearly dependent for all $u \in \mathscr{H}$ and computing the quaternionic inner product $\langle u,v\rangle$ by the formula

$$\langle u,v\rangle = \langle u,v\rangle_c - j\langle \mathbf{j}\, u,v\rangle_c. \tag{4.4}$$

These are all the rules we need; further details and explanations are given in Ref. 10. We also recapitulate the basic theorem on the algebra of bounded additive operators on a complex Hilbert space which was proved in Ref. 16 and whose counterpart in quaternionic Hilbert spaces we seek.

*Proposition 4.1:* The algebra of bounded additive operators on a complex Hilbert space is the smallest algebra containing both linear and semilinear bounded operators and in this algebra $A \mapsto A^*$ is a norm-preserving involution. Furthermore, in this algebra it is not generally true that

$$(\alpha A)\cdot(\beta B) = (\alpha\beta)(AB). \tag{4.5}$$

The bounded linear operators form a subalgebra of this algebra and for this subalgebra (4.5) is always valid.

Every bounded additive operator on a quaternionic Hilbert space $\mathscr{H}$ is also an additive operator on $\mathscr{H}_c$ that is $\mathscr{H}$ regarded as a complex Hilbert space; hence the existence and the algebraic properties of the adjoint are obtained merely by taking the operators into $\mathscr{H}_c$ and then bringing them back to $\mathscr{H}$. It has already been shown by this method[10] and by a direct method by Horwitz and Biedenharn[3] that for a bounded linear operator $A$ on $\mathscr{H}$ the adjoint $A^*$ exists, is linear, and satisfies the familiar equation

$$\langle u,Av\rangle = \langle A^*u,v\rangle, \quad \forall u,v \in \mathscr{H}. \tag{4.6}$$

The most important difference between the complex and the quaternionic cases is that we no longer have just one semilinear operator but for any choice of $i$, $j$, and $k$ three different kinds of semilinear operators called $\tau$-semilinear ($\tau = i,j,k$) operators. Our next task is to establish a formula corresponding to (4.6) for a $\tau$-semilinear operator, which we do in our next proposition.

*Proposition 4.2:* Let $A$ be a $\tau$-semilinear operator ($\tau = i$, $j$, or $k$) on quaternionic Hilbert space $\mathscr{H}$. Then the adjoint $A^*$ of $A$ is a $\tau$-semilinear operator that satisfies

$$\langle u,Av\rangle = -\tau\langle A^*u,v\rangle\tau, \quad \forall u,v \in \mathscr{H}, \tag{4.7}$$

and

$$\|A^*\| = \|A\|. \tag{4.8}$$

*Proof:* Since $A$ is $\tau$ semilinear, an easy calculation involving the definitions shows that $B = \tau A$ is linear and that $B$ and $A$ commute with $\tau$. By taking everything into $\mathscr{H}_c$, that is, $\mathscr{H}$ treated as a complex Hilbert space, and using Proposition 4.1 and then bringing everything back to $\mathscr{H}$, we know that the adjoint of $\tau$ is $\tau^*$, which is the same as its quaternionic conjugate and satisfies

$$\tau^* = -\tau \tag{4.9}$$

and the adjoint of $\tau A$ is given by

$$(\tau A)^* = (A\tau)^* = -\tau A^*, \tag{4.10}$$

and is linear and therefore $A^*$ is $\tau$ semilinear. We can now write

$$\langle u,Av\rangle = \langle u,\tau Av\rangle\tau = -\langle \tau A^*u,v\rangle\tau = -\tau\langle A^*u,v\rangle\tau, \quad \forall u,v,\in \mathscr{H}. \tag{4.11}$$

From (4.11) with the help of (2.8) it follows that

$$\|\langle u,Av\rangle\| = \|\langle A^*u,v\rangle\| \quad \forall u,v \in \mathscr{H}, \tag{4.12}$$

which is all we need to prove (4.8); the steps are exactly the same as in the proof of the corresponding result for a linear operator on a Hilbert space on any field. □

We can now define the adjoint of any additive operator $A$ on $\mathscr{H}$: we write $A$ as a sum of linear and three $\tau$-semilinear (with $\tau = i$, $j$, and $k$) operators where the various components are given by (2.20a)–(2.20d) and $A^*$ is then defined to be the sum of the adjoints of the four operators in this decomposition. It is clear that taking adjoints preserves the norm of each of the four components of an additive operator. We can now write down the generalization of Proposition 4.1 to quaternionic Hilbert spaces.

*Proposition 4.3:* In the algebra of bounded additive operators on a quaternionic Hilbert space the correspondence between an operator $A$ and its adjoint $A^*$ is an additive invo-

lution that preserves the norm of each component of the decomposition of $A$ as a sum of a linear and three $\tau$-semilinear operators corresponding to $\tau = i,j,k$. Furthermore, in this algebra it is not generally true that

$$(\alpha A)\cdot(\beta B) = (\alpha\beta)(AB). \tag{4.13}$$

Unlike the case of bounded additive operators on a complex Hilbert space, bounded linear operators (nor for that matter bounded $\tau$-semilinear operators for any value of $\tau$) on a quaternionic Hilbert space do not form a subalgebra of the algebra of bounded additive operators on the space.

## V. CONCLUDING REMARKS

We have presented here an architecture of additive functionals and operators on a quaternionic Hilbert space, in which a great many nice structures are neatly intertwined; as a consequence none of the propositions we have put forward have required more than a few lines in proof. We would like to warn the reader that though the simplicity is quite genuine, while working in this area one false turn leads one into a maze where one can go round and round in circles or do pages of complicated calculations to prove one of our propositions with a few lines of proof. We would like to confess that the earlier versions of some of our proofs were very long and messy. A little reflection at this stage enables one to appreciate why the structure of a quaternionic vector space is called symplectic[10] (from $\sigma\nu\mu\pi\lambda\epsilon\kappa\tau\iota\kappa\acute{o}\varsigma$ in Greek meaning twining or plaiting together). We hope to develop this theory further and in the light of the insight gained we propose to study also the applications of quaternions in physics.


[1]M. D. Vivarelli, Celestial Mech. **29**, 45 (1983).
[2]M. D. Vivarelli, Celestial Mech. **36**, 45 (1985).
[3]L. P. Horwitz and L. C. Biedenharn, Ann Phys. (NY) **157**, 432 (1984).
[4]S. Adler, Commun. Math. Phys. **104**, 611 (1986).
[5]P. Rastall, Rev. Mod. Phys. **36**, 820 (1964).
[6]E. Y. Rocher, J. Math. Phys. **13**, 1919 (1972).
[7]D. Finkelstein, J. M. Jauch, and D. Speiser, in *The Logico-Algebraic Approach to Quantum Mechanics. Volume II: Contemporary Consolidation*, edited by C. A. Hooker (Reidel, Dordrecht, 1979).
[8]D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962).
[9]C. S. Sharma and T. J. Coulson, J. Math. Phys. **28**, 1941 (1987).
[10]C. S. Sharma, J. Math. Phys. **29**, 1069 (1988).
[11]C. S. Sharma and D. F. Almeida, J. Math. Phys. **29**, 2411 (1988).
[12]T. J. Coulson, Ph. D. thesis, London University, 1986.
[13]J. Pian and C. S. Sharma, Int. J. Theor. Phys. **22**, 107 (1983).
[14]P. M. Cohn, *Algebra 1* (Wiley, Chichester, 1982).
[15]W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973).
[16]C. S. Sharma, Nuovo Cimento B **100**, 291 (1987).

# A harmonic oscillator system with the generalized commutation relations

Shuji Watanabe
*Department of Physics, Nagoya University, Nagoya 464, Japan*

(Received 16 December 1987; accepted for publication 5 October 1988)

The system of a one-dimensional harmonic oscillator is discussed with the generalized commutation relations obtained by Wigner. The Hamiltonian is shown to have self-adjoint extensions. The domain of self-adjointness is explicitly specified in some cases. The proof is carried out by the use of the Lax–Milgram lemma. A suitable rigged Hilbert space is found for this system to reformulate the earlier arguments. The main difficulty is that the momentum and Hamiltonian operators contain the reflection $R\psi(x) = \psi(-x)$ as well as singular terms.

## I. INTRODUCTION

The canonical commutation relations are postulated together with equations of motion in nonrelativistic quantum mechanics. About 40 years ago Wigner[1] investigated the problem of whether the commutation relations are uniquely determined by the equations of motion. In his matrix-mechanical treatment of the system of a one-dimensional harmonic oscillator, the answer is in the negative. The ground-state energy $E_0$ is found to be arbitrary but positive, and is not necessarily equal to $\frac{1}{2}$ in suitable units. Yang[2] discussed the same problem in the wave-mechanical representation. His conclusion is that if wave functions of this system satisfy some stringent conditions, then $E_0$ is specified by discrete values of $c$, a characteristic parameter of the theory, and consequently only discrete values of $E_0$ are allowed. This is not the case in the matrix-mechanical representation, and so there naturally arises a question about the equivalence of the two representations. Ohnuki and Kamefuchi[3] were the first to settle this question, on the natural assumption that wave functions are subject to the probabilistic interpretation. They dealt with wave functions in the sense of generalized functions and solved the eigenvalue problems of the momentum and Hamiltonian operators to find two classes of eigenfunctions. Mukunda, Sudarshan, Sharma, and Mehta[4] showed that they are mutually unitarily equivalent, i.e., the ground states are related to one another by a unitary transformation.

In this paper we shall discuss self-adjointness of the Hamiltonian, and then reformulate the earlier arguments on this system by constructing a suitable rigged Hilbert space. The dynamics of this system should be given by a unitary group whose infinitesimal generator is the Hamiltonian. Stone's theorem[5] states that the generator is always self-adjoint. We thus need to prove self-adjointness of the Hamiltonian in order to guarantee unitarity of time evolution of the system. As mentioned above, generalized functions are used by Ohnuki and Kamefuchi to deal with wave functions. Also, generalized functions called distributions are used systematically to construct rigged Hilbert spaces,[6] and consequently quantum mechanics can be beautifully formulated. We are now interested in constructing a rigged Hilbert space for our system.

In Sec. II we restrict the domain of the Hamiltonian operator to some dense subset of the Hilbert space $L^2(R^1)$, and then show that it has self-adjoint extensions. In Sec. III the domain of self-adjointness is explicitly specified for $|c| > 1$ using the Lax–Milgram lemma, which is a consequence of the Riesz representation theorem and plays an essential role in our approach. A rigged Hilbert space is found in Sec. IV to reformulate the earlier arguments on this system in the wave-mechanical representation. Here the space including eigenfunctions of not only the position operator but the momentum operator is presented.

## II. EXISTENCE OF SELF-ADJOINT EXTENSIONS

Canonical variables $x$ and $p$ of our system are assumed to satisfy the following generalized commutation relations proposed by Wigner[1]:

$$ip = [x,H] \quad \text{and} \quad -ix = [p,H], \tag{G}$$

where the Hamiltonian $H$ is of the classical form $(p^2 + x^2)/2$.

We shall begin by showing the existence of self-adjoint extensions of the Hamiltonian. To this end it is necessary to find out concrete expressions for the relevant dynamical variables. Suppose that $x$ is an operator of multiplication by $x$. Then the corresponding expression for $p$, when $x$ and $p$ satisfy (G), is derived by Yang[2] as follows:

$$p = -i\frac{d}{dx} + i\frac{c}{x}R,$$

where $R\psi(x) = \psi(-x)$ and $c$ is an arbitrary real constant. From this we obtain

$$H = \frac{1}{2}\left\{-\frac{d^2}{dx^2} + \frac{c}{x^2}(c - R) + x^2\right\}.$$

The ground-state energy $E_0$ is related to $c$ linearly.[3]

**Theorem 1:** For any $c$, the Hamiltonian has self-adjoint extensions.

*Proof:* Denote by $D(\Omega)$ the totality of $C^\infty$ functions with compact support in the set $\Omega = R^1 - \{0\}$, and restrict the domain of the Hamiltonian $H$ to $D(\Omega)$. It is easy to verify that $H$ is symmetric on $D(\Omega)$ and commutes with complex conjugation. Thus the von Neumann theorem[7] applies. Q.E.D.

## III. THE DOMAIN OF SELF-ADJOINTNESS

In this section we describe explicitly the domain of self-adjointness of $H$ in the case of $|c| > 1$. We use here the Lax–Milgram lemma, cited below, as a useful tool. This is a consequence of the Riesz representation theorem, and has been applied to guarantee existence of solutions of elliptic partial

 0022-2488/89/020376-04$02.50 © 1989 American Institute of Physics 

differential equations. We consider it reasonable to use it in our problem as a sort of elliptic operator $H$. Our method indeed enables us to find out clearly and easily the domain of self-adjointness as a concrete subset of the usual Sobolev space $H^1(R^1)$.

Let us consider the subset

$$V=\{\psi\in H^1(R^1): x\psi,\psi/x\in L^2(R^1)\}$$

of $L^2(R^1)$, and the sesquilinear form

$$h(\psi,\phi)=\frac{1}{2}\int_{R^1}\left\{D\psi^* D\phi + \frac{\psi^*}{x}\frac{c}{x}(c-R)\phi + x\psi^* x\phi\right\}dx$$

on the product space $V\times V$, where $D$ and $*$ denote the $L^2$ derivative and the complex conjugate, respectively. It is expected that $H$ is represented as $h(\psi,\phi)=(H\psi,\phi)$ with the $L^2$ inner product $(\cdot,\cdot)$.

First we examine the property of the set $V$.

*Lemma 2:* The set $V$ is a Hilbert space with the inner product

$$(\psi,\phi)_V=(\psi,\phi)+(D\psi,D\phi)+(x\psi,x\phi)+(c^2-|c|)(\psi/x,\phi/x)\,.$$

*Proof:* Clearly $V$ is a linear subspace of $L^2(R^1)$. So it suffices to show the completeness of it under the norm $|\cdot|_V=((\cdot,\cdot)_V)^{1/2}$.

Suppose $|\psi_m-\psi_n|_V\to 0$ as $m,n\to\infty$. Then from the completeness of $L^2(R^1)$ and $H^1(R^1)$ we can obtain without difficulty that $\psi_n\to\psi$ in $H^1(R^1)$ for some $\psi\in H^1(R^1)$, and that $\psi_n\to\phi$, $x\psi_n\to x\phi$, and $\psi_n/x\to\phi/x$ in $L^2(R^1)$ for some $\phi\in L^2(R^1)$, with $x\phi\in L^2(R^1)$ and $\phi/x\in L^2(R^1)$. But $H^1(R^1)$ is continuously imbedded in $L^2(R^1)$ and $\psi_n\to\psi$ in $L^2(R^1)$. Thus $\phi$ coincides with $\psi$, and hence $\psi\in V$ and $|\psi_n-\psi|_V\to 0$. Q.E.D.

Next we apply the following to the form $h(\cdot,\cdot)$.

*The Lax–Milgram Lemma:* Let $W$ be a Hilbert space with inner product $(\cdot,\cdot)_W$ and norm $|\cdot|_W$, and let $s(\cdot,\cdot)$ be a sesquilinear form on $W\times W$ satisfying, for $\psi,\phi\in W$,

(i) $|s(\psi,\phi)|\leqslant K_1|\psi|_W|\phi|_W$,

(ii) $s(\psi,\psi)\geqslant K_2|\psi|_W^2$,

with positive constants $K_1$ and $K_2$. Then for each given $\chi\in W$ there exists a unique $\psi\in W$ such that $s(\psi,\phi)=(\chi,\phi)_W$ for all $\phi\in W$.

For the proof of this lemma see Ref. 8, for example.

*Lemma 3:* For each $\chi\in V$ there exists a unique $\psi\in V$ satisfying

$$h(\psi,\phi)+\tfrac{1}{2}(\psi,\phi)=(\chi,\phi)_V,\quad \text{for all } \phi\in V.$$

*Proof:* We have only to apply the Lax–Milgram lemma to the sesquilinear form $t(\cdot,\cdot)=h(\cdot,\cdot)+\frac{1}{2}(\cdot,\cdot)$ on $V\times V$ for $V$, which has been shown to be a Hilbert space by Lemma 2. A simple calculation shows

(i) $|t(\psi,\phi)|\leqslant c^2(c^2-|c|)^{-1}|\psi|_V|\phi|_V$,

(ii) $t(\psi,\psi)\geqslant\frac{1}{2}|\psi|_V^2$,

for $\psi,\phi\in V$. Q.E.D.

We are now in a position to define $H$ as an operator in $L^2(R^1)$ associated with $h(\cdot,\cdot)=t(\cdot,\cdot)-\frac{1}{2}(\cdot,\cdot)$.

The Riesz representation theorem[9] implies that for an arbitrary $f\in L^2(R^1)$ there exists a unique $\chi\in V$ such that

$$(\chi,\phi)_V=(f,\phi),\quad \text{for all } \phi\in V,$$

since $(f,\cdot)$ is a bounded linear functional on $V$. Combining this equality with that in Lemma 3, we obtain $\psi\in V$ with

$$t(\psi,\phi)=(f,\phi),\quad \text{for all } \phi\in V.$$

Clearly $V$ is dense in $L^2(R^1)$, and hence there is a one-to-one correspondence between $\psi$ and $f$. Thus we can define an operator $H+\frac{1}{2}$ in $L^2(R^1)$ as follows:

$\psi$ belongs to $D(H)$ and $(H+\frac{1}{2})\psi=f$

if and only if (D)

$h(\psi,\phi)+\frac{1}{2}(\psi,\phi)=(f,\phi)$, for all $\phi\in V$.

This operator $H$ is the desired operator. Indeed, it is easy to verify that $D(\Omega)$ is included in $D(H)$, and

$$H\psi=\frac{1}{2}\left\{-\frac{d^2}{dx^2}+\frac{c}{x^2}(c-R)+x^2\right\}\psi,\quad \text{for } \psi\in D(\Omega)\,.$$

The self-adjointness of $H$ follows from that of $H+\frac{1}{2}$. As is well known, the operator associated with the adjoint sesquilinear form $\bar{t}(\psi,\phi)=t(\phi,\psi)^*$ is nothing but the adjoint operator[10] of $H+\frac{1}{2}$. An elementary calculation shows $\bar{t}(\psi,\phi)=t(\psi,\phi)$, which implies the self-adjointness of $H+\frac{1}{2}$ and hence of $H$.

Thus we have proved our main result of this section, which is as follows.

**Theorem 4:** If $|c|>1$, then the Hamiltonian with domain $D(\Omega)$ is extended to the self-adjoint operator $H$, whose domain $D(H)$ is specified by (D).

Our method based on the Lax–Milgram lemma is also available for eigenvalue problems. Let us study eigenvalues of this self-adjoint operator $H$. Ohnuki and Kamefuchi[3] have obtained the two classes of eigenfunctions of the Hamiltonian of the original form as follows:

(i) $c>-\frac{1}{2}$, $E_n=\frac{1}{2}+c+n$,

$$\psi_n(x)=\begin{cases}K^m_{c+1/2}|x|^c L^{c-1/2}_m(x^2)\exp(-x^2/2) & (n=2m)\\ K^m_{c+3/2}\varepsilon(x)|x|^{c+1}L^{c+1/2}_m(x^2)\exp(-x^2/2) & (n=2m+1),\end{cases}$$

(ii) $c<\frac{1}{2}$ and $c\neq 0$, $E_n=\frac{1}{2}-c+n$,

$$\psi_n(x)=\begin{cases}K^m_{-c+1/2}\varepsilon(x)|x|^{-c}L^{-c-1/2}_m(x^2)\exp(-x^2/2) & (n=2m),\\ K^m_{-c+3/2}|x|^{-c+1}L^{-c+1/2}_m(x^2)\exp(-x^2/2) & (n=2m+1),\end{cases}$$

where $L_m^a(x)$ denotes the generalized Laguerre polynomial, and $K_b^m$ denotes the normalization constant.

*Proposition 5:* If $|c| \geqslant 2$, then the self-adjoint operator $H$ has the eigenvalues $E_n$ $(n = 0,1,2,\ldots)$.

*Proof:* As for the class (i), clearly $\psi_n$ belongs to $V$. Noting that $d^2\psi_n/dx^2$ is continuous and belongs to $L^2(R^1)$ for $|c| \geqslant 2$, and recalling (D), we have that for all $\phi \in V$,

$$h(\psi_n,\phi) = \left(\frac{1}{2}\left\{-\frac{d^2}{dx^2} + \frac{c}{x^2}(c-R) + x^2\right\}\psi_n,\phi\right)$$
$$= (E_n\psi_n,\phi) .$$

Thus $\psi_n$ belongs to $D(H)$, and every $E_n$ becomes an eigenvalue of $H$. The same reasoning goes for the class (ii). Q.E.D.

## IV. A RIGGED HILBERT SPACE FOR THE SYSTEM

In this section we shall construct a rigged Hilbert space for this system to reformulate the earlier arguments in the wave-mechanical representation. Roberts' method of construction[11] is quite effective in systems whose potentials are multiplication operators, even if they are singular and/or discontinuous. His method is not applicable directly to our system whose potential is not a multiplication operator but contains the reflection operator $R$. However, his idea is still useful essentially for our problem, as is exhibited below.

*Lemma 6:* $D(\Omega)$ is a dense subspace of $L^2(R^1)$ and is invariant under the operators $x$, $p$, and $H$.

The proof is simple and hence may be omitted.

It is known that $D(\Omega)$ becomes a linear topological space when equipped with the usual inductive limit topology.[12]

*Lemma 7:* (i) Each of the operators $x$, $p$, and $H$ is continuous in the topology of $D(\Omega)$.

(ii) $D(\Omega)$ is continuously imbedded in $L^2(R^1)$, and is nuclear.

*Proof:* (i) Let $\{\psi_n\}$ be a sequence in $D(\Omega)$ which tends to 0 as $n \to \infty$. Equivalently there exists a compact subset $K$ of $\Omega$ satisfying

$$\text{supp } \psi_n \subset K$$

and (S)

$$\sup_{x\in K} |\psi_n^{(\alpha)}(x)| \to 0, \quad \text{as } n \to \infty ,$$

where $\phi^{(\alpha)} = (d/dx)^\alpha\phi$. As for the operator $x$ we have

$$\text{supp } x\psi_n \subset K$$

and

$$\sup_{x\in K} |(x\psi_n(x))^{(\alpha)}| \to 0, \quad \text{as } n \to \infty ,$$

which is evident from $(x\psi_n(x))^{(\alpha)} = x\psi_n^{(\alpha)}(x) + \alpha\psi_n^{(\alpha-1)}(x)$. Thus $x\psi_n$ vanishes as $n \to \infty$, i.e., $x$ is continuous. As for the reflection operator $R$, it is easy to verify

$$\text{supp } R\psi_n \subset K'$$

and

$$\sup_{x\in K'} |(R\psi_n)^{(\alpha)}(x)| \to 0, \quad \text{as } n \to \infty ,$$

with $K' = K \cup (-K)$, which implies the continuity of $R$. Showing similarly the continuity of the operators $1/x$ and $d/dx$, we can prove the continuity of $p$ and $H$.

(ii) The topology of $D(\Omega)$ is finer than that induced by $L^2(R^1)$, and therefore the imbedding is continuous. For nuclearity of spaces see Ref. 13, for example. Q.E.D.

In view of the above lemmas we see that $D(\Omega)$ fulfils all the requirements imposed on the smallest space of triplets.

Thus we have obtained our desired result.

**Theorem 8:** A rigged Hilbert space,

$$D(\Omega) \subset L^2(R^1) \subset D^*(\Omega) ,$$

is constructed for our system, where $D^*(\Omega)$ is the dual of $D(\Omega)$.

*Remark:* It is well known that $S(R^1) \subset L^2(R^1) \subset S^*(R^1)$ is constructed in the case of $c = 0$, where $S(R^1)$ is the usual Schwartz space. This rigged Hilbert space indeed corresponds to a harmonic oscillator system with the canonical commutation relations.

We shall finally discuss our system based on the above framework. Ohnuki and Kamefuchi[3] have solved the eigenvalue problem of the momentum operator to obtain the two classes of eigenfunctions.

(i) For $c > -\frac{1}{2}$,

$$\psi_k(x) = |kx|^{1/2}\{J_{c-1/2}(|kx|) + i\varepsilon(kx)J_{c+1/2}(|kx|)\}/2 .$$

(ii) For $c < \frac{1}{2}$ and $c \neq 0$,

$$\psi_k(x) = |kx|^{1/2}\{J_{-c+1/2}(|kx|) - i\varepsilon(kx)J_{-c-1/2}(|kx|)\}/2 ,$$

where $J_\nu$ denotes the Bessel function, and $\psi_k$ is the eigenfunction corresponding to an eigenvalue $k$.

*Proposition 9:* If $k \neq 0$, then $\psi_k$ belongs to $D^*(\Omega)$.

*Proof:* As for the class (i), we put

$$\langle\psi|\psi_k\rangle = \int_{R^1} \psi^*(x)\psi_k(x)dx, \quad \text{for every } \psi \in D(\Omega) .$$

Evidently this is an antilinear functional on $D(\Omega)$. So it suffices to prove its continuity. Let $\{\psi_n\}$ be a sequence that vanishes in $D(\Omega)$, for instance, satisfying (S). Then denoting by $\mu(\cdot)$ the Lebesgue measure, we have

$$|\langle\psi_n|\psi_k\rangle| \leqslant C \cdot \sup_{x\in K} |\psi_n(x)| ,$$

with

$$C = \sup_{x\in K}|kx|^{1/2}\{\sup_{x\in K}|J_{c-1/2}(|kx|)| + \sup_{x\in K}|J_{c+1/2}(|kx|)|\}\mu(K)/2 ,$$

a constant independent of $n$, which implies $\langle\psi_n|\psi_k\rangle \to 0$ as $n \to \infty$. The same reasoning goes for the class (ii). Q.E.D.

*Remark:* The set $\{\delta_\xi : -\infty < \xi < \infty\}$ is a complete system of $D^*(\Omega)$. Here $\delta_\xi$, defined by $\delta_\xi(\psi) = \psi(\xi)$ for every $\psi \in D(\Omega)$, corresponds to Dirac's delta function $\delta(x-\xi)$, and is, moreover, the eigenfunction corresponding to an eigenvalue $\xi$ of the position operator $x$.


## ACKNOWLEDGMENTS

I would like to thank Professor Y. Ohnuki for initiating my interest in this subject and for his suggestions during the

preparation of this paper. I also thank Dr. C. Lee for stimulating discussions about rigged Hilbert spaces. I am grateful to Professor M. Watanabe, Professor of Mathematics at Niigata University, for suggesting the treatment of the Hamiltonian through the Lax–Milgram lemma.

[1]E. P. Wigner, Phys. Rev. **77**, 711 (1950).
[2]L. M. Yang, Phys. Rev. **84**, 788 (1951).
[3]Y. Ohnuki and S. Kamefuchi, J. Math. Phys. **19**, 67 (1978); see also Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (University of Tokyo/Springer, Tokyo/Berlin, 1982), Chap. 23; Z. Phys. C **2**, 367 (1979).
[4]N. Mukunda, E. C. G. Sudarshan, J. K. Sharma, and C. L. Mehta, J. Math. Phys. **21**, 2386 (1980); see also J. K. Sharma, C. L. Mehta, N. Mukunda, and E. C. G. Sudarshan, *ibid.* **22**, 78 (1981); J. K. Sharma, C. L. Mehta, and E. C. G. Sudarshan, *ibid.* **19**, 2089 (1978).
[5]See, for example, K. Yosida, *Functional Analysis* (Springer, Berlin, 1980), 6th ed., p. 345.
[6]See, for example, I. M. Gel'fand, G. E. Shilov, and N. Ya. Vilenkin, *Generalized Functions*, Vols. 1–4 (Academic, New York, 1964); A. Böhm, *Lectures in Theoretical Physics, Vol. IX A: Mathematical Methods of Theoretical Physics*, edited by W. E. Brittin, A. O. Barut, and M. Guenin (Gordon and Breach, New York, 1967); *The Rigged Hilbert Space and Quantum Mechanics, Lecture Notes in Physics*, Vol. 78 (Springer, Berlin, 1978); J. E. Roberts, J. Math. Phys. **7**, 1097 (1966); Commun. Math. Phys. **3**, 98 (1966); J. P. Antoine, J. Math. Phys. **10**, 53, 2276 (1969).
[7]See, for example, M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. II (Academic, New York, 1975), p. 143.
[8]See Ref. 5, p. 92.
[9]See Ref. 5, p. 90.
[10]See, for example, H. Tanabe, *Equations of Evolution* (Pitman, London, 1979), pp. 24–29, for adjoint sesquilinear forms and associated adjoint operators.
[11]See J. E. Roberts, Ref. 6.
[12]See Ref. 5, pp. 28–30.
[13]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967), Part III.

# The quantum relativistic two-body bound state. II. The induced representation of SL(2,C)

R. Arshansky and L. P. Horwitz
*School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel*

(Received 4 November 1987; accepted for publication 24 August 1988)

It was shown in I [J. Math. Phys. **30**, 66 (1989)] that the eigenfunctions for the reduced motion of the quantum relativistic bound state with O(3,1) symmetric potential have support in an O(2,1) invariant subregion of the full spacelike region. They form irreducible representations of SU(1,1) [in the double covering of O(2,1)] parametrized by the unit spacelike vector $n_\mu$, taken in I as the direction of the $z$ axis (the spectrum is independent of this choice), for which this O(2,1) is the stabilizer. Lorentz transformations move these representations on an orbit whose range is the single-sheeted hyperboloid covered by this spacelike vector, providing a set of induced representations of SL(2,C). From linear combinations of functions from the irreducible representations of SU(1,1), the representations of the SU(2) subgroup of SL(2,C) on the orbit are extracted and the differential equations that are the eigenvalue equations for the Casimir operators of SL(2,C) are solved. It is found that these SU(2) representations form a basis for the principal series in the canonical representations of Gel'fand. There is a natural scalar product, obtained from group integration on SL(2,C), for which the canonical basis forms an orthogonal set, and the representation is unitary. Since the scalar product [over the O(2,1) invariant measure space] of SU(1,1) irreducible representations is invariant under the action of the little group, the remaining group measure [on the coset space SL(2,C)/SU(1,1)] is the volume on the hyperboloidal spacelike hypersurface $d\mu_n = d^4n\,\delta(n^2-1)$. The family of Hilbert spaces ($\mathscr{H}_n$) that carries the representations of O(3,1) is therefore embedded in a larger Hilbert space $\mathscr{H}$ with measure $d^4y\,d\mu_n$, where the $\{y\}$ are the space-time coordinates of the restricted region associated with $n_\mu$. The representations with nonrelativistic limit coinciding with the known Schrödinger solutions for corresponding spherically symmetric potential problems are in the double covering (half-integer values for the lowest $L$ level) of O(3,1).

## I. INTRODUCTION

In a previous paper[1] (to be called I), we studied the two-body relativistic quantum-mechanical bound-state problem for which the dynamical evolution operator $K$ is the sum of a part describing free two-body motion and a part describing the interaction, represented by a potential function $V(\rho)$, of the invariant (spacelike) interval between the two events. Since the part describing the free two-body motion is the sum of two terms quadratic in the four-momenta, the center of mass motion can be separated as for the nonrelativistic problem. The remaining part, called $K_{\rm rel}$, of the evolution operator contains a term describing free relative motion and the interaction term. The existence of such a potential function implies, classically, a correlation between events along the world histories of the two events, determining which point along one world line, $x_1{}^\mu$, say, corresponds to a point $x_2{}^\mu$ on the second world line. Quantum mechanically, this correlation is described by the joint distribution represented by the two-body wave function. In nonrelativistic classical and quantum theory, the potential $V(\mathbf{x}_1-\mathbf{x}_2)$ refers to points on two particle trajectories at a given time $t$; in the relativistic framework that we shall use,[2] correlation is achieved by parametrizing the evolution of the system with a universal invariant "historical time" $\tau$.

For the case of a Coulomb-type potential proportional to $1/\rho$, where

$$\rho=\sqrt{(x_1{}^\mu-x_2{}^\mu)(x_{1\mu}-x_{2\mu})}\equiv\sqrt{(x_1-x_2)^2}, \qquad (1.1)$$

it was found in I that when the support of the wave functions on the relative coordinates $(x_1{}^\mu-x_2{}^\mu)$ is restricted to an O(2,1) invariant subregion of the full spacelike region, one finds a lower mass ground state than for the case in which the support is on the full spacelike region. This subregion corresponds to the part of the Minkowski space exterior (in a spacelike direction) to two hyperplanes tangent to the light cone and oriented along an axis which, in this paper, we shall specify by a unit spacelike vector $n_\mu$ (in I this direction was taken to be the $z$ axis). We shall call this subregion the "restricted Minkowski space," or RMS ($n_\mu$). In Fig. 1 it is represented in a three-dimensional form in which we have folded the axes $x_1$,$x_2$ together to form $x_\perp=\sqrt{x_1{}^2+x_2{}^2}$. The two hyperplanes then become planes and intersect along the $n_\mu$ axis. One can similarly use the projective space of Fig. 2 in I by tilting the $Z$ ($=z/t$) axis to the direction of $n_\mu$. This is the subregion RMS($n_\mu$). The fact that this restricted subregion admits a lower mass ground state than the full spacelike region constitutes a spontaneous symmetry breaking of the O(3,1) invariance of the dynamical equations.

The restriction of the solutions to RMS corresponds to a class of admissible correlations between the events associated with the world histories of the two particles. We have assumed that these correlations are characteristic of the bound state, and that they are maintained in all excited

 0022-2488/89/020380-13$02.50 © 1989 American Institute of Physics 

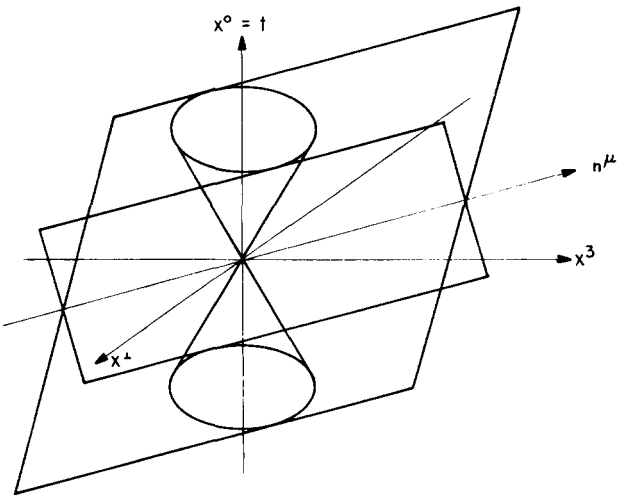



FIG. 1. The restricted Minkowski space RMS($n_\mu$), oriented according to an arbitrary spacelike direction $n_\mu$, taken for the support of the eigenvalue equation for the two-body bound state in relative coordinates. The spatial coordinates $x_1$ and $x_2$ are folded into a single axis in this figure ($x_\perp$); in 3 + 1 dimensions the RMS is connected. Coordinates in the system with its third spatial axis along the direction of $n_\mu$ are called $y_\mu$.

bound states as well. The mass spectrum of the bound states then coincides with the nonrelativistic Schrödinger energy spectrum, for every corresponding central potential; when the binding is small compared to the particle masses, the mass spectrum (bounded below) is well approximated by the results of the nonrelativistic theory.

The operator $K$ used in I is explicitly a Lorentz (as well as a Poincaré) scalar, i.e., its form is independent of the reference frame chosen to represent it in coordinate space. The bound state eigenfunctions, however, form a set of representations of SU(1,1) [the double covering of O(2,1)]. In this paper, we shall study the motion of the measure space and the transformations of these functions induced by the action of Lorentz transformations to construct irreducible representations of SL(2,$C$).

In I, in addition to the d'Alembertian, we made use of the set of differential operators

$$M^{\mu\nu} = -i\left(x^\mu \frac{\partial}{\partial x_\nu} - x^\nu \frac{\partial}{\partial x_\mu}\right), \tag{1.2}$$

which formally satisfy the algebra of O(3,1) and commute with $K_{\rm rel}$. We shall designate the set of elements of this formal algebra as $\mathfrak{o}(3,1)$. The measure space of the $L^2(\mathbb{R}^4 \subseteq \mathrm{RMS}(n_\mu))$, which contains the bound-state eigenfunctions, is invariant only under an O(2,1) subalgebra (the stabilizer of $n_\mu$) of $\mathfrak{o}(3,1)$. The RMS ($n_\mu$) is also stable under the action of the Casimir operator $\Lambda = (\mathbf{L}^2 - \mathbf{A}^2)$ of $\mathfrak{o}(3,1)$.

The operators $\Lambda$, $L_3$, and $N^2 \equiv L_3^2 - A_1^2 - A_2^2$, the Casimir operator of the representation of the O(2,1) subalgebra of $\mathfrak{o}(3,1)$, were diagonalized in I along with $K_{\rm rel}$. The eigenfunctions constitute irreducible representations of SU(1,1).

Under the action of *physical* Lorentz transformations, i.e., the mapping of space-time coordinates of events in an initial frame $F$ to their space-time coordinates observed in a moving frame $F'$, the RMS($n_\mu$) is mapped to an RMS($n'_\mu$), which is, in general, different. The operators of $\mathfrak{o}(3,1)$ do not induce changes in $n_\mu$, and hence, except for its O(2,1) subalgebra, which is the stabilizer of $n_\mu$, $\mathfrak{o}(3,1)$ cannot represent the action of the physical Lorentz group O(3,1). We shall identify this O(2,1) subalgebra of $\mathfrak{o}(3,1)$ as a representation on $\mathbf{L}^2(\mathbb{R}^4 \subseteq \mathrm{RMS}(n_\mu))$ of the subalgebra of the Lorentz group that leaves $n_\mu$ invariant, and for which the RMS($n'_\mu$) in $F'$ is identical to the RMS($n_\mu$) in $F$.

The RMS($n_\mu$) in each frame (or corresponding to each $n_\mu$) is isomorphic to any other, and the eigenfunctions of the dynamical evolution operator for the relative motion $K_{\rm rel}$ form a complete set for the subspace (of the Hilbert space associated with $n_\mu$) belonging to the discrete spectrum. The wave function in any Lorentz frame is therefore a linear superposition of the solutions obtained in some standard frame (or orientation). We shall label this standard frame by $(\mathring{n}_\mu) = (0,0,0,1)$. As we shall show, changes in these linear superpositions correspond to the action of SU(1,1); this action, together with the change in orientation of the RMS($n_\mu$), provides an induced representation of SL(2,$C$) with SU(1,1) little group.[3]

Since the motion along the orbit does not change the value of the SU(1,1) Casimir operator, or that of the $\mathfrak{o}(3,1)$ Casimir [this second-order operator is well defined and Hermitian in the RMS($n_\mu$) for any $n_\mu$], the linear combinations change their structure only in regard to the degeneracy quantum number $k$ of SU(1,1), corresponding to the action of the SU(1,1) little group. The coefficients therefore play the role of the Wigner $D$ functions in the induced representation of relativistic particles with spin.

Along with the infinitesimal operators of O(2,1) generating transformations in the $y \in \mathrm{RMS}(n_\mu)$ dependence [where the coordinates $\{y_\mu\}$ are defined in an accompanying frame for the RMS($n_\mu$) with $y_3$ along the $n_\mu$ axis] of the eigenfunctions, the generators of the Lorentz group contain infinitesimal operators generating transformations in the $n_\mu$ dependence of the coefficients of the linear combinations of elements of the SU(1,1) representations. Because of the fact that the Lorentz group generators contain these two pieces, as for induced representations for systems with spin, the second Casimir $\hat{c}_2 = \mathbf{L}(n)\cdot\mathbf{A}(n)$ of SL(2,$C$) is not identically zero and must be considered along with the first, $\hat{c}_1 = \mathbf{L}(n)^2 - \mathbf{A}(n)^2$. The operators $\mathbf{L}(n)^2$ and $L_1(n)$ commute with these, and we decompose the functions along the orbits in terms of eigenfunctions of these operators from SU(2) $\subset$ SL(2,$C$). This decomposition provides the basis for irreducible ladder representations (the "canonical" representations of Gel'fand) of the Lorentz group. It is carried out by constructing and solving the differential eigenvalue equations for the Casimir operators $\hat{c}_1$ and $\hat{c}_2$. A consistency condition for the solutions relates the possible values of $c_1, c_2$ of these Casimir operators. We find that the basis sets $\xi_{Lq}$ [$\mathbf{L}(n)^2 \to L(L+1)$, $L_1(n) \to q$] belong to the *principal series* of Gel'fand,[4] corresponding to $L_1$ pure imaginary, $L_0$ integer or half-integer, where $c_2 = -iL_0L_1$, $-c_1$

$= 1 - L_0^2 - L_1^2$ ($L_0$ is the lowest value of $L$ in the ladder).

The representations that correspond to spectra and wave functions with nonrelativistic limit coinciding with those of the nonrelativistic Schrödinger equation are those with *half-integer* values for the lowest $L$ level. Although the quantum number $L$ corresponds to the Casimir operator of a rotation subgroup of the Lorentz group, the partial wave expansions in scattering theory[5] (for the continuous spectrum of $K_{\rm rel}$ under the same conditions of correlation), for the most symmetric case, depend on $l$, associated with the Casimir of $o(3,1)$, and $n$, associated with the Casimir of SU(1,1), and not on $L$ and $q$. The quantum numbers $l$ and $n$ play the role of orbital angular momentum and magnetic quantum number of the nonrelativistic theory, and coincide with these quantities in the nonrelativistic limit.

In the nonrelativistic limit, the $o(3,1)$ algebra is deformed to an O(3) algebra, and its O(2,1) subalgebra is deformed to an O(2) subalgebra. In this limit, the induced representations of SL(2,$C$) go over to induced representations of SU(2) with U(1) little group.[6] We shall discuss this deformation procedure elsewhere.

The Gel'fand-type representation that we have obtained is unitary under integration over the SL(2,$C$) group. The basis used in its construction is a set of functions from the Hilbert spaces $\mathscr{H}_n = L^2{}_n(\mathbb{R}^4 \subset \mathrm{RMS}(n_\mu))$ with measure $d^4y$, for all $n_\mu$ on $n_\mu n^\mu = 1$. The scalar product of these functions, for each $n_\mu$, is SU(1,1) invariant. What remains of the group integration is an integral over the single-sheeted hyperboloid with measure $d^4n\,\delta(n^2-1)$ [the coset space SL(2,$C$)/SU(1,1)]. This procedure constitutes an embedding of $\{\mathscr{H}_n\}$ into $\mathscr{H}$, a large Hilbert space on $\mathbb{R}^4(y) \otimes \mathbb{R}^4{}_{(n)}|_{n^2=1}$, with a natural scalar product.

In Sec. II we construct the abstract formulation of the framework for the induced representation of SL(2,$C$) by defining a standard set of frames associated with every value of the spacelike vector $n_\mu$. In Sec. III we discuss and construct the Lie algebra for the induced representation as a commutator algebra among differential operators, and in Sec. IV a specific choice is made for the parametrization of a set of standard frames on the orbit in order to explicitly construct the generators. In Sec. V the Casimir operators for the Lorentz group are constructed as differential operators. It is shown there that there is a Pauli–Lubanski operator $W_\mu(n)$ for the SU(1,1) little group, for which $W^\mu(n)W_\mu(n)$ is the invariant SU(1,1) Casimir operator. In the next section, the differential equations requiring $\mathbf{L}^2(n)$, $L_1(n)$, and the Casimir operators $\hat{c}_1$, $\hat{c}_2$ to be diagonal are solved. A consistency relation, $-c_1 = 1 - \hat{n}^2 - c_2^2/\hat{n}^2$, where $c_1$,$c_2$ are the values of $\hat{c}_1$,$\hat{c}_2$ and $\hat{n}$ is associated with the value of the Casimir operator for SU(1,1), is found. In Sec. VII we show, by applying Gel'fand's formulas (which follow from the Lorentz algebra) on the one hand and our differential operators on the other to the resulting basis functions extracted from the induced representation, that this set of functions forms a Gel'fand canonical basis that is unitary under integration over SL(2,$C$). The consistency relation corresponds to the well-known relations between the Casimir operators and the pair of numbers parametrizing the Gel'fand ladder. Finally, Sec. VIII contains a summary and conclusions.

## II. INDUCED REPRESENTATIONS BASED ON THE SU(1,1) LITTLE GROUP

The representations of SU(1,1) discussed in I are defined on what we shall call here the RMS($\mathring{n}_\mu$), oriented according to the direction of the (arbitrarily chosen) $z$ axis. If we view these states from a moving frame, they will appear to be distorted by Lorentz transformations. The transformed wave functions, however, have support in the corresponding transform of the RMS, which is oriented according to the transformed $z$ direction. If we represent the state of the system in every Lorentz frame by the configuration of the wave functions in the corresponding RMS, the action of the Lorentz group will be represented by its action on these wave functions and on the orientation of the RMS. The representations of SU(1,1) that were obtained in I are carried on an orbit by the action of the Lorentz group, and SU(1,1) acts as a little group in the construction of representations of SL(2,$C$).

The representations that we shall obtain belong to the principal series of Gel'fand, and are unitary in a Hilbert space with scalar product that is defined through integration over SL(2,$C$). This includes an integration over the measure space of SU(1,1), carried out in the scalar product of $L^2(\mathbb{R}^4 \subseteq \mathrm{RMS}(n_\mu))$, for each $n_\mu$ (corresponding to the arbitrary orientation of the $z$ axis), and an integration over the measure of the coset space SL(2,$C$)/SU(1,1), the hyperboloid defined by $n^\mu n_\mu$. The structure of this scalar product, and of the associated norm, implies the existence of a probability measure on $\mathbb{R}^7$, with Lebesgue measure $d^4y\,d^4n\,\delta(n^2-1)$, where $y_\mu \in \mathrm{RMS}(n_\mu)$. We shall therefore consider the wave functions as functions of both $y_\mu$ and $n_\mu$; the coordinate description of the *quantum state* therefore corresponds to an ensemble of events lying in a set of RMS($n_\mu$)'s over all possible $\{n_\mu\}$.

We shall first study the set of elements $\{\Psi_n \in \mathscr{H}_n\}$ and their coordinate representation,

$$\phi_n(x) = {}_n\langle x|\Psi_n\rangle, \tag{2.1}$$

defined by means of the spectral representation of the space-time coordinate operator in each of the Hilbert spaces $\mathscr{H}_n$. A coordinate system oriented with its $z$ axis along the direction of $n_\mu$ can be constructed by means of a coordinate transformation of Lorentz type,

$$y_\mu = L(n)_\mu{}^\nu x_\nu, \tag{2.2}$$

where

$$L(n)_\mu{}^\nu n_\nu = \mathring{n}_\mu = (0,0,0,1). \tag{2.3}$$

For example, taking $x_\mu$ parallel to $n_\mu$, i.e., $x_\mu = \lambda n_\mu$, the corresponding $y_\mu$ vector has the form

$$y_\mu = L(n)_\mu{}^\nu(\lambda n_\nu) = \lambda \mathring{n}_\mu. \tag{2.4}$$

The family of Lorentz frames defined by $L(n)_\mu{}^\nu$ depends on its particular form. [According to (2.3), they may differ by right multiplication by an element of the O(2,1) stability group of $n_\mu$, and multiplication on the left by an element of the O(2,1) stability group of $\mathring{n}_\mu$]. We shall specify this matrix function of $n_\mu$ later.

The variables $y_\mu$ described by the standard coordinatization used in I,

$$y^0=\rho\sin\theta\sinh\beta,\quad y^2=\rho\sin\theta\sin\phi\cosh\beta,$$
$$y^1=\rho\sin\theta\cos\phi\cosh\beta,\quad y^3=\cos\theta, \tag{2.5}$$

provide a complete characterization of the configuration space in the RMS $(n_\mu)$ that is universal in the sense that it is of the same form in every Lorentz frame. It is therefore convenient to define the functions

$$\psi_n(y)\equiv\phi_n(L^T(n)y)=\phi_n(x). \tag{2.6}$$

To study the Lorentz transformation properties of these wave functions, we define the unitary (in the larger space) mapping $U(\Lambda)$: $\mathscr{H}_n\to\mathscr{H}_{\Lambda n}$ such that the state vectors are related by

$$\Psi^\Lambda{}_{\Lambda n}=U(\Lambda)\Psi_n \tag{2.7}$$

and $\|\Psi^\Lambda{}_{\Lambda n}\|_{\Lambda n}=\|\Psi_n\|_n$. In the new Lorentz frame [with $y=L(\Lambda n)x$],

$$\begin{aligned}\phi^\Lambda{}_{\Lambda n}(x)&={}_{\Lambda n}\langle x|\Psi^\Lambda{}_{\Lambda n}\rangle\\&={}_{\Lambda n}\langle x|U(\Lambda)\Psi_n\rangle=\phi^\Lambda{}_{\Lambda n}(L^T(\Lambda n)y)\\&\equiv\psi^\Lambda{}_{\Lambda n}(y).\end{aligned} \tag{2.8}$$

If $\phi_n(x)$ is a scalar under Lorentz transformation, so that (with no additional phase)

$$\phi^\Lambda{}_{\Lambda n}(\Lambda x)=\phi_n(x), \tag{2.9}$$

it follows from (2.8) that

$$U(\Lambda)|x\rangle_n=|\Lambda x\rangle_{\Lambda n}. \tag{2.10}$$

The wave function $\phi^\Lambda{}_{\Lambda n}(x)$ describes a system in a Lorentz frame in motion with respect to the frame in which the state is described by $\phi_n(x)$, and for which the support is in the RMS$((\Lambda n)_\mu)$. The value of this function at $x$ in the new frame is determined by its value at $\Lambda^{-1}x$ in the original frame; moreover, the subensemble associated with $n_\mu$ in the new frame is determined by the subensemble associated with $(\Lambda^{-1}n)_\mu$ in the old frame. We define the description of the state of the system in the new frame in terms of the set (over $\{n_\mu\}$) of transformed wave functions

$$\begin{aligned}\psi^\Lambda{}_n(y)\equiv\phi_n{}^\Lambda(x)&=\phi_{\Lambda^{-1}n}(\Lambda^{-1}x)\\&=\psi_{\Lambda^{-1}n}(D^{-1}(\Lambda,n)y),\end{aligned} \tag{2.11}$$

where we have used the relation (2.2) (the transformed function has support oriented with $n_\mu$) and defined the (pseudo) orthogonal matrix

$$D(\Lambda,n)=L(n)\Lambda L^T(\Lambda^{-1}n). \tag{2.12}$$

The transformation $D^{-1}(\Lambda,n)$ stabilizes $\mathring{n}_\mu$ and hence is in the O(2,1) subgroup of the group of Lorentz transformations O(3,1) that act on space-time. [As we shall show, the wave functions transform in the double covering SU(1,1) of this O(2,1)]. Equation (2.11) defines an induced representation of SL(2,$C$).

## III. LIE ALGEBRA OF THE INDUCED REPRESENTATION

In this section, we shall study the representation of the Lorentz group that is provided by wave functions of the form $\psi^\Lambda{}_n(y)$, as defined in (2.11). The transformed wave function reflects, along with an O(2,1) transformation of its coordinate dependence, a motion along the hyperboloidal orbit parametrized by $n_\mu$. We shall construct the differential operators that generate these transformations by considering the action of the infinitesimal Lorentz transformation

$$\Lambda\simeq 1+\lambda, \tag{3.1}$$

for which

$$\psi^{1+\lambda}{}_n(y)=\psi_{n-\lambda n}(D^{-1}(1+\lambda,n)y). \tag{3.2}$$

To first order, the little-group transformation is

$$D^{-1}(1+\lambda,n)\simeq 1-(d_n(\lambda)L(n))L^T(n)-L(n)\lambda L^T(n), \tag{3.3}$$

where (the derivative with respect to $n_\mu$ is taken holding $y_\mu$ fixed)

$$d_n(\lambda)=\lambda_\mu{}^\nu n_\nu\frac{\partial}{\partial n_\mu}. \tag{3.4}$$

From the orthogonality property $L(n)L^T(n)=1$, it follows that

$$(d_n(\lambda)L(n))L^T(n)=-L(n)(d_n(\lambda)L^T(n)), \tag{3.5}$$

and hence (3.3) can be equivalently written

$$\begin{aligned}D^{-1}(1+\lambda,n)&\simeq 1+L(n)(d_n(\lambda)L^T(n)-\lambda L^T(n))\\&\equiv 1-G_n(\lambda).\end{aligned} \tag{3.6}$$

From (3.2), we obtain

$$\psi^{1+\lambda}{}_n(y)\simeq\psi_n(y)-(d_n(y)+g_n(\lambda))\psi_n(y), \tag{3.7}$$

where

$$g_n(\lambda)=G_n(\lambda)_\mu{}^\nu y_\nu\frac{\partial}{\partial y_\mu}. \tag{3.8}$$

With the help of the identity (3.5), one finds that

$$\begin{aligned}[G_n(\lambda_2),G_n(\lambda_1)]&=G_n([\lambda_2,\lambda_1])+d_n(\lambda_2)G_n(\lambda_1)\\&\quad-d_n(\lambda_1)G_n(\lambda_2).\end{aligned} \tag{3.9}$$

Let

$$h_n(\lambda)=-i(d_n(\lambda)+g_n(\lambda)). \tag{3.10}$$

It then follows from

$$\begin{aligned}[g_n(\lambda_2),g_n(\lambda_1)]&=-[G_n(\lambda_2),G_n(\lambda_1)]_\mu{}^\nu y_\nu\frac{\partial}{\partial y_\mu},\\[g_n(\lambda_2),g_n(\lambda_1)]&=(d_n(\lambda_2)G_n(\lambda_1)_\mu{}^\nu)y_\nu\frac{\partial}{\partial y_\mu}\end{aligned} \tag{3.11}$$

that [note that it is precisely the motion along the orbit, generated by $d_n$, that compensates for the differentiated terms in (3.9)],

$$\begin{aligned}&[h_n(\lambda_2),h_n(\lambda_1)]\\&\quad=d_n([\lambda_2,\lambda_1])+G_n([\lambda_2,\lambda_1])_\mu{}^\nu y_\nu\frac{\partial}{\partial y_\mu}\\&\quad=ih_n([\lambda_2,\lambda_1]).\end{aligned} \tag{3.12}$$

## IV. EXPLICIT FORM FOR THE GENERATORS

It is our purpose to investigate the representations that can be constructed from the set of wave functions $\{\psi_n(y)\}$ that form the bound states of the O(3,1) invariant two-body

relativistic potential problem. The spectrum of the Casimir operators and a complete set of commuting generators from the algebra (3.12) provides a characterization of these representations. In order to construct these operators, we shall need an explicit form for the O(2,1) little-group generators $G_n(\lambda)$ that act on the RMS ($n_\mu$) variables $y_\mu$.

Let us choose for the construction of $L^T(n)$ (which takes $\mathring{n}$ to $n$) the product of a boost along the three-direction with parameter $\alpha$ in $(-\infty,\infty)$, followed by a rotation about the two-axis with angle $\omega$ in $(-\pi/2,\pi/2)$ and then about the one-axis with angle $\gamma$ in $(0,2\pi)$ (in the sense of Euler "rotations"):

$$L^T(n)=\begin{pmatrix} \cosh\alpha & 0 & 0 & \sinh\alpha \\ -\sin\omega\sinh\alpha & \cos\omega & 0 & -\sin\omega\cosh\alpha \\ \sin\gamma\cos\omega\sinh\alpha & \sin\gamma\sin\omega & \cos\gamma & \sin\gamma\cos\omega\cosh\alpha \\ \cos\gamma\cos\omega\sinh\alpha & \sin\omega\cos\gamma & -\sin\gamma & \cos\gamma\cos\omega\cosh\alpha \end{pmatrix}, \tag{4.1}$$

where the matrix indices are assigned to be $L^T(n)_\mu{}^\nu$. The general spacelike vector that we obtain from (2.3) in this way is

$$n_\mu=\begin{pmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \end{pmatrix}=\begin{pmatrix} \sinh\alpha \\ -\sin\omega\cosh\alpha \\ \sin\gamma\cos\omega\cosh\alpha \\ \cos\gamma\cos\omega\cosh\alpha \end{pmatrix}. \tag{4.2}$$

It then follows that (note that $\cos\omega\geqslant 0$ in the range of $\omega$; it will, furthermore, always be clear when $n$ stands for the magnitude of the space part of $n_\mu$)

$$\begin{aligned} &\sinh\alpha=n_0, \quad \cosh\alpha=\sqrt{n_1{}^2+n_2{}^2+n_3{}^2}\equiv n, \\ &\sin\omega=-n_1/n, \quad \cos\omega=n_{23}/n, \\ &\sin\gamma=n_2/n_{23}, \quad \cos\gamma=n_3/n_{23}, \end{aligned} \tag{4.3}$$

where

$$n_{23}=\sqrt{n_2{}^2+n_3{}^2}=\cos\omega\cosh\alpha, \tag{4.4}$$

and hence

$$L^T(n)=\begin{pmatrix} n & 0 & 0 & n_0 \\ \dfrac{n_0n_1}{n} & \dfrac{n_{23}}{n} & 0 & n_1 \\ \dfrac{n_0n_2}{n} & -\dfrac{n_1n_2}{nn_{23}} & \dfrac{n_3}{n_{23}} & n_2 \\ \dfrac{n_0n_3}{n} & -\dfrac{n_1n_3}{nn_{23}} & -\dfrac{n_2}{n_{23}} & n_3 \end{pmatrix}. \tag{4.5}$$

We remark that the matrix $L(n)$ has just three parameters, $n_1,n_2,n_3$ (or $\alpha,\omega,\gamma$). It is of the form of a Lorentz transformation, since it leaves $n_\mu n^\mu$ invariant, but has only the role of providing a "standard" set of coordinate frames corresponding to every $n_\mu$ on the spacelike hyperboloid. The $L(n)$, in fact, are a representative set of elements from the coset $O(3,1)/O(2,1)$.[7]

The explicit form of the generators $G_n(\lambda)$ of the little group can now be computed from (3.6). Defining the matrices $\lambda^{\alpha\beta}$ according to

$$M^{\alpha\beta}=-i(\lambda^{\alpha\beta})_\mu{}^\nu x_\nu\frac{\partial}{\partial x_\mu}=-i\left(x^\alpha\frac{\partial}{\partial x_\beta}-x^\beta\frac{\partial}{\partial x_\alpha}\right), \tag{4.6}$$

we obtain [since the infinitesimal Lorentz transformations preserve the relation $n^2-n_0{}^2=1$, the $n_\mu$ in $L^T(n)$ may be differentiated freely]

$$\begin{aligned} &G_n(\lambda_1)=0, \quad G_n(\lambda_2)=\frac{nn_2}{nn_{23}{}^2}\lambda_3, \quad G_n(\lambda_3)=\frac{nn_3}{n_{23}}\lambda_3, \\ &G_n(\lambda^{01})=\frac{n_{23}}{n^2}\lambda^{01}, \\ &G_n(\lambda^{02})=\frac{n_3}{nn_{23}}\lambda^{02}-\frac{n_1n_2}{n^2n_{23}}\lambda^{01}-\frac{n_0n_1n_3}{nn_{23}{}^2}\lambda_3, \\ &G_n(\lambda^{03})=\frac{n_0n_1n_2}{nn_{23}{}^2}\lambda_3-\frac{n_1n_3}{n^2n_{23}}\lambda^{01}-\frac{n_2}{nn_{23}}\lambda^{02}, \end{aligned} \tag{4.7}$$

where we have called, e.g., $\lambda^{12}\equiv\lambda_3$, and omitted the infinitesimal factor multiplying each generator. Note that the generators $G_n(\lambda)$ are all expressed as linear combinations of the generators $\lambda_3,\lambda^{01},\lambda^{02}$ belonging to the little group in the standard coordinate frame (for which $n_\mu=\mathring{n}_\mu$); these matrices act on the $y_\mu$ that belong to the RMS referred to the $y_3$ axis, and this measure space remains invariant under the action of the little group. In particular, for $n_\mu\to\mathring{n}_\mu$, we see that the $G_n(\lambda)$ for $\lambda=\lambda_1,\lambda_2$, and $\lambda^{03}$ vanish, and

$$G_n(\lambda^{01})\to\lambda^{01}, \quad G_n(\lambda^{02})\to\lambda^{02}, \quad G_n(\lambda_3)\to\lambda_3. \tag{4.8}$$

Extracting these infinitesimal actions on the $y$ coordinates as differential operators on the wave functions [with the definition (3.8)], these become $A_1$, $A_2$, and $L_3$.

Expressing the differential operators $d_n(\lambda)$ in terms of the parameters $\alpha,\omega,\gamma$ of $n_\mu$, we obtain the following expressions for the generators of $SL(2,C)$ on the wave functions $\psi_n(y)$:

$$\begin{aligned} h_n(\lambda_1)&=i\frac{\partial}{\partial\gamma} \\ h_n(\lambda_3)&=i\left(\cos\gamma\frac{\partial}{\partial\omega}+\tan\omega\sin\gamma\frac{\partial}{\partial\gamma}\right)+\frac{\sin\gamma}{\cos\omega}L_3, \\ h_n(\lambda_3)&=-i\left(\sin\gamma\frac{\partial}{\partial\omega}-\tan\omega\cos\gamma\frac{\partial}{\partial\gamma}\right)+\frac{\cos\gamma}{\cos\omega}L_3, \\ h_n(\lambda^{01})&=-i\left(\sin\omega\frac{\partial}{\partial\alpha}+\cos\omega\tanh\alpha\frac{\partial}{\partial\omega}\right) \\ &\quad+\frac{\cos\omega}{\cosh\alpha}A_1, \end{aligned}$$

$$h_n(\lambda^{02})$$
$$= -i\Big( -\sin\gamma\cos\omega\frac{\partial}{\partial\alpha} + \sin\omega\sin\gamma\tanh\alpha\frac{\partial}{\partial\omega} - \frac{\cos\gamma\tanh\alpha}{\cos\omega}\frac{\partial}{\partial\gamma}\Big) + \frac{\cos\gamma}{\cosh\alpha}A_2 + \frac{\sin\omega\sin\gamma}{\cosh\alpha}A_1 + \tanh\alpha\tan\omega\cos\gamma L_3, \tag{4.9}$$

$$h_n(\lambda^{03})$$
$$= -i\Big( -\cos\gamma\cos\omega\frac{\partial}{\partial\alpha} + \tanh\alpha\sin\omega\cos\gamma\frac{\partial}{\partial\omega} + \frac{\tanh\alpha\sin\gamma}{\cos\omega}\frac{\partial}{\partial\gamma}\Big) - \tanh\alpha\tan\omega\sin\gamma L_3 + \frac{\sin\omega\cos\gamma}{\cosh\alpha}A_1 - \frac{\sin\gamma}{\cosh\alpha}A_2.$$

The Pauli–Lubanski-type vector (orthogonal to $n_\mu$)

$$W_\mu(n) = \tfrac{1}{2}\epsilon_{\mu\nu\lambda\sigma}n^\sigma h_n(\lambda^{\nu\lambda}) = -(i/2)\epsilon_{\mu\nu\lambda\sigma}n^\sigma g_n(\lambda^{\nu\lambda}) \tag{4.10}$$

extracts the little-group operators from the generators of the total Lie algebra. In fact, using the expressions (4.7), we have

$$W_0(n) = -inL_3, \quad W_1(n) = i\Big(\frac{n_{23}}{n}A_2 - \frac{n_0n_1}{n}L_3\Big),$$
$$W_2(n) = -i\Big(\frac{n_1n_2}{nn_{23}}A_2 + \frac{n_3}{n_{23}}A_1 + \frac{n_0n_2}{n}L_3\Big), \tag{4.11}$$
$$W_3(n) = i\Big( -\frac{n_1n_3}{nn_{23}}A_2 + \frac{n_2}{n_{23}}A_1 - \frac{n_0n_3}{n}L_3\Big).$$

In the standard coordinate system where $n\to\mathring{n}$, the vector $W_\mu(\mathring{n})$ reduces to

$$W(\mathring{n}) = -i\begin{pmatrix} L_3 \\ -A_2 \\ A_1 \\ 0 \end{pmatrix}. \tag{4.12}$$

Equation (4.11) can be reconstructed with the help of (4.5) as

$$W_\mu(n)W^\mu(n) = L^T(n)_\mu{}^\nu W_\nu(\mathring{n}). \tag{4.13}$$

This vector property of $W_\mu$ follows from the fact that the quantities $g_n(\lambda^{\mu\nu})$ form a tensor up to an additive term that vanishes under the formation of the Pauli–Lubanski vector.

The square of the operator-valued vector $W_n(n)$ is

$$W_\mu(n)W^\mu(n) = W_\mu(\mathring{n})W^\mu(\mathring{n}) = L_3{}^2 - A_1{}^2 - A_2{}^2 = N^2, \tag{4.14}$$

the Casimir operator for SU(1,1)

## V. THE CASIMIR OPERATORS OF SL(2,C)

In the previous section, we obtained an explicit form for the differential operators representing the generators of SL(2,C) on the family of Hilbert spaces $\{\mathscr{H}_n\}$, i.e., on the orbit of the induced representation. We shall represent the state of the system along the orbit induced by Lorentz transformations ($K_{\rm rel}$ is degenerate on the entire orbit) by a linear combination of degenerate bound-state eigenfunctions with definite values of $K_{\rm rel}$, the Casimir operator of the SU(1,1) algebra stabilizing $n_\mu$ (also invariant along the orbit), and the invariant Casimir of $\mathfrak{o}(3,1)$, with coefficients depending on $n_\mu$. We shall then obtain the irreducible representations of the Lorentz group by using these differential operators to construct the differential eigenvalue equations corresponding to the action of a complete set of operators, including the Casimir operators of SL(2,C), and solve them. These operators act on the $n_\mu$ dependent coefficients of these linear combinations, as well as (in the sense of the action of the little group) the coordinate dependence of the bound-state eigenfunctions in the accompanying RMS ($n_\mu$).

The Casimir operators are[8]

$$\hat{c}_1 = \mathbf{L}(n)^2 - \mathbf{A}(n)^2, \qquad \hat{c}_2 = \mathbf{L}(n)\cdot\mathbf{A}(n), \tag{5.1}$$

where, for $i = 1,2,3$,

$$\mathbf{L}(n) = \{h_n(\lambda_i)\} \tag{5.2}$$

and

$$\mathbf{A}(n) = \{h_n(\lambda^{0i})\}, \tag{5.3}$$

the space-space and space-time parts of the antisymmetric tensor $M^{\mu\nu}(n)$ defined by (4.6) and (4.9).

We shall diagonalize the Casimir operator for the SU(2) subgroup of SL(2,C),

$$\mathbf{L}(n)^2 = \sum_{i=1}^{3} h_n(\lambda_i)^2, \tag{5.4}$$

and, as is convenient with our parametrization for $n_\mu$,

$$L_1(n) = h_n(\lambda_1), \tag{5.5}$$

along with the Casimir operators (5.1). Adding the squares of the operators $h_n(\lambda_i)$, as indicated in (5.4), one obtains

$$\mathbf{L}(n)^2 = -\frac{\partial^2}{\partial\omega^2} + \tan\omega\frac{\partial}{\partial\omega} + \frac{1}{\cos^2\omega}\Big\{ -\frac{\partial^2}{\partial\gamma^2} + 2iL_3\sin\omega\frac{\partial}{\partial\gamma} + L_3{}^2\Big\}, \tag{5.6}$$

and, according to (5.5),

$$L_1(n) = i\frac{\partial}{\partial\gamma}. \tag{5.7}$$

We now let

$$z = \sin\omega \tag{5.8}$$

and obtain

$$\mathbf{L}(n)^2 = -\frac{\partial}{\partial z}\Big((1-z^2)\frac{\partial}{\partial z}\Big) + \frac{1}{1-z^2}\Big\{ -\frac{\partial^2}{\partial\gamma^2} + 2izL_3\frac{\partial}{\partial\gamma} + L_3{}^2\Big\}. \tag{5.9}$$

Adding the squares of $h_n(\lambda^{0i})$, one obtains

$$\mathbf{A}(n)^2 = -\frac{\partial}{\partial\alpha^2} - 2\tanh\alpha\frac{\partial}{\partial\alpha} + \tanh^2\alpha\,\mathbf{L}(n)^2 - \frac{1}{\cosh^2\alpha}(N^2 - 2L_3{}^2) - L_3{}^2 - 2i\frac{\tanh\alpha}{\cosh\alpha}\left\{A_1\frac{\partial}{\partial\omega} - A_2\frac{1}{\cos\omega}\frac{\partial}{\partial\gamma} + iA_2L_3\tan\omega\right\}, \tag{5.10}$$

hence the first Casimir operator $\hat{c}_1$ is

$$\hat{c}_1 = \mathbf{L}(n)^2 - \mathbf{A}(n)^2 = \frac{\partial^2}{\partial\alpha^2} + 2\tanh\alpha\frac{\partial}{\partial\alpha} + \frac{1}{\cosh^2\alpha}\times\{\mathbf{L}(n)^2 + N^2 - 2L_3{}^2\} + L_3{}^2 + 2i\frac{\tanh\alpha}{\cosh\alpha}\left\{A_1\frac{\partial}{\partial\omega} - A_2\frac{1}{\cos\omega}\frac{\partial}{\partial\gamma} + iA_2L_3\tan\omega\right\}, \tag{5.11}$$

where we have used the commutation relations of the little-group algebra,

$$[L_3,A_2] = iA_1, \quad [L_3,A_1] = iA_2, \tag{5.12}$$

and remark in passing that

$$[A_1,A_2] = iL_3. \tag{5.13}$$

The second Casimir operator $\hat{c}_2$ is given by

$$\hat{c}_2 = \mathbf{L}(n)\cdot\mathbf{A}(n) = \frac{1}{\cosh\alpha}\left\{iA_2\frac{\partial}{\partial\omega} + iA_1\frac{1}{\cos\omega}\frac{\partial}{\partial\gamma} + A_1L_3\tan\omega\right\} + iL_3\left\{\frac{\partial}{\partial\alpha} + \tanh\alpha\right\}. \tag{5.14}$$

## VI. FUNCTIONS ON THE ORBIT WITH DEFINITE VALUES OF $\hat{c}_1$, $\hat{c}_2$, $L(n)^2$, AND $L_1(n)$

We now turn to the structure of wave functions $\psi_n(y)$ which have definite values of $\mathbf{L}(n)^2$, $L_1(n)$ and the SL(2,C) Casimir operators. Since $\rho$ is Lorentz invariant, and $y_3 = \rho\cos\theta$ is invariant along the orbit [this coordinate is a distance along the $n_\mu$ direction, which moves with the Lorentz transformation; alternatively, one sees that $\partial/\partial\theta$ does not occur in the $\{h_n(\lambda)\}$], the only possibility for *alteration* of the structure of the linear combination of bound-state wave functions belonging to a particular value of $K_{\rm rel}$, $N^2$, and the Casimir of the $\mathfrak{o}(3,1)$ algebra [with value $l(l+1) - \frac{3}{4}$, as given in Eq. (4.16) of I] is a change in the coefficients of the linear combinations over the SU(1,1) degeneracy index $k$ of the $\beta,\phi$ dependent part [defined in Eq. (A2) of I],

$$\chi_{n+k}{}^{-n}(\zeta,\phi) = B_{n+k,n}(\beta)\Phi_{n+k}(\phi) = (1-\zeta^2)^{1/4}\hat{B}_{n+k,n}(\zeta)\Phi_{n+k}(\phi), \tag{6.1}$$

where

$$\zeta = \tanh\beta.$$

The parameters $\rho,\theta,\beta,\phi$ are defined in the parametrization (3.1) of I given for the restricted Minkowski space; these are applicable, as indicated in (2.5), for any $n_\mu$.

The Casimir operator of $\mathfrak{o}(3,1)$ (with differential operators now defined in terms of the coordinates $y_\mu$), defined in Eq. (2.10) of I, is a second-order differential operator depending only on the $y_\mu$ and derivatives with respect to $y_\mu$. It is invariant, in particular, under O(2,1) transformations applied to the $y_\mu$. The derivatives that the generators $h_n(\lambda)$ contain with respect to $n_\mu$ hold $y_\mu$ fixed, and the derivatives with respect to $y_\mu$ are contained only in generators of the O(2,1) which stabilizes $n_\mu$. It therefore follows that the $\mathfrak{o}(3,1)$ Casimir, the Casimir operator of its O(2,1) subalgebra $N^2$ which stabilizes $n_\mu$, $\mathbf{L}(n)^2$, $L_1(n)$, $\hat{c}_1$, and $\hat{c}_2$ all commute and can be diagonalized simultaneously. We shall therefore study the functions[9]

$$\psi_n{}^Q(y) = R_l(\rho)\Theta_l{}^n(\theta)\xi^Q(n_\mu,\beta,\phi), \tag{6.2}$$

where

$$\xi^Q(n_\mu,\beta,\phi) = \sum_k \mathscr{D}_k{}^Q(\alpha,\omega,\gamma)\chi_{n+k}{}^{-n}(\beta,\phi), \tag{6.3}$$

the functions $\mathscr{D}_k{}^Q$ are coefficients that we must determine, and

$$Q = \{l,n,L,q,c_1,c_2\}, \tag{6.4}$$

where the $n$ in the second place labels the Casimir of the stabilizing O(2,1). We have labeled the eigenvalue of $\mathbf{L}(n)^2 \to L(L+1)$ by $L$, and denoted the eigenvalue of $L_1(n)$ by $q$.

We now impose the condition that $\xi^Q$ be an eigenfunction with definite value $L(L+1)$ of $\mathbf{L}(n)^2$ and $q$ of $L_1(n)$. If we apply the operator $\mathbf{L}(n)^2$ to this function, $L_3$ and $L_3{}^2$ in the expression (5.9) take on values specified by the factor $\Phi_{n+k}(\phi)$ of $\chi_{n+k}{}^{-n}$. Requiring that $\mathbf{L}(n)^2$ take the value $L(L+1)$ and using the orthogonality of the functions $\chi_{n+k}{}^{-n}$, we obtain the differential equation

$$\left\{-\frac{\partial}{\partial z}\left((1-z^2)\frac{\partial}{\partial z}\right) + \frac{1}{1-z^2}\left(-\frac{\partial^2}{\partial\gamma^2} + 2iM_kz\frac{\partial}{\partial\gamma} + M_k{}^2\right)\right\}\mathscr{D}_k{}^Q(\alpha,\omega,\gamma) = L(L+1)\mathscr{D}_k{}^Q(\alpha,\omega,\gamma), \tag{6.5}$$

where we have defined

$$M_k = n + k + \tfrac{1}{2}. \tag{6.6}$$

Imposing the condition

$$L_1(n)\mathscr{D}_k{}^Q(\alpha,\omega,\gamma) = i\frac{\partial}{\partial\gamma}\mathscr{D}_k{}^Q(\alpha,\omega,\gamma) = q\mathscr{D}_k{}^Q(\alpha,\omega,\gamma), \tag{6.7}$$

separation of variables occurs in Eq. (6.5), and we obtain a solution of the form[10]

$$\mathscr{D}_k{}^Q(\alpha,\omega,\gamma) = \Xi^Q{}_k(\alpha)P^L{}_{q,-M_k}(z)e^{-iq\gamma}, \tag{6.8}$$

where the functions $P^L{}_{a,b}(z)$ satisfy[11]

$$(1-z^2)\frac{\partial^2 P^L{}_{a,b}}{\partial z^2}-2z\frac{\partial P^L{}_{a,b}}{\partial z}$$
$$-\frac{1}{1-z^2}[b^2+a^2-2abz]P^L{}_{a,b}=-L(L+1)P^L{}_{a,b}, \tag{6.9}$$

and are given by[12]

$$P^L{}_{a,b}(z)=\frac{(-1)^{L-b}i^{b-a}}{2^L(L-b)!}\sqrt{\frac{(L-b)!\,(L+a)!}{(L+b)!\,(L-a)!}}$$
$$\cdot(1-z)^{(b-a)/2}(1+z)^{-(a+b)/2}$$
$$\times\frac{d^{L-a}}{dz^{L-a}}[(1-z)^{L-b}(1+z)^{L+b}]. \tag{6.10}$$

In terms of Jacobi polynomials $P_k{}^{(\alpha,\beta)}$ they are[13]

$$P^L{}_{ab}(z)=\frac{i^{a-b}}{2^a}\sqrt{\frac{(L-a)!\,(L+a)!}{(L-b)!\,(L+b)!}}(1-z)^{(a-b)/2}$$
$$\times(1+z)^{(a+b)/2}P_{L-a}{}^{(a-b,a+b)}(z). \tag{6.11}$$

For $L,a,b$ together integer or half-integer, and $a,b=-L,\ -L+1,...L$, the functions $P_{ab}{}^L$ are finite[14] at $z=\pm 1$ and, with the $\gamma$ and $\phi$ dependence, form an irreducible representation with angular momentum $L$ of $SU(2)\subset SL(2,C)$. This implies that the indices of $P^L{}_{q,-M_k}$ must be in the range

$$q,M_k=-L,-L+1,...,+L, \tag{6.12}$$

and hence, since $k\geqslant 0$,[15]

$$\hat{n}\equiv n+\tfrac{1}{2}\leqslant M_k\leqslant L, \tag{6.13}$$

or

$$0\leqslant k\leqslant L-\hat{n}. \tag{6.14}$$

We shall now impose the condition that the Casimir operators $\hat{c}_1$, $\hat{c}_2$ be numerical valued (with values $c_1,c_2$) on the functions

$$\xi^{Q'}{}_{Lq}=\sum_k \Xi^{Q'}{}_{Lq,k}(u)P^L{}_{q,-M_k}(z)e^{iq\gamma}\chi_{n+k}{}^{-n}(\beta,\phi), \tag{6.15}$$

where we have introduced the variable

$$u=\tanh\alpha \tag{6.16}$$

and written the quantum numbers $L,q$ explicitly [leaving $Q'$ to denote the remainder of the set (6.4)]. The differential operators (5.11) and (5.14) corresponding to the Casimir operators $\hat{c}_1$ and $\hat{c}_2$ become, in terms of the variables $u,z$,

$$\hat{c}_1=(1-u^2)^2\frac{\partial^2}{\partial u^2}+(1-u^2)\{\mathbf{L}(n)^2+N^2-2L_3{}^2\}$$
$$+L_3{}^2+2iu\sqrt{1-u^2}\left\{A_1\sqrt{1-z^2}\frac{\partial}{\partial z}\right.$$
$$\left.-A_2\frac{1}{\sqrt{1-z^2}}\frac{\partial}{\partial\gamma}+iA_2L_3\frac{z}{\sqrt{1-z^2}}\right\} \tag{6.17}$$

and

$$\hat{c}_2=iL_3\left((1-u^2)\frac{\partial}{\partial u}+u\right)+\sqrt{1-u^2}\left\{iA_2\sqrt{1-z^2}\frac{\partial}{\partial z}\right.$$
$$\left.+iA_1\frac{1}{\sqrt{1-z^2}}\frac{\partial}{\partial\gamma}+A_1L_3\frac{z}{\sqrt{1-z^2}}\right\}. \tag{6.18}$$

Applying the Casimir operators to the functions occurring in Eq. (6.15), with the help of the relations (A8) and (A9) of I and the properties[16]

$$\left(\sqrt{1-z^2}\frac{\partial}{\partial z}-\frac{a-zb}{\sqrt{1-z^2}}\right)P^L{}_{ab}$$
$$=-i\sqrt{(L-b)(L+b+1)}P^L{}_{a,b+1} \tag{6.19}$$

and

$$\left(\sqrt{1-z^2}\frac{\partial}{\partial z}+\frac{a-zb}{\sqrt{1-z^2}}\right)P^L{}_{ab}$$
$$=-i\sqrt{(L+b)(L-b+1)}P^L{}_{a,b-1}, \tag{6.20}$$

we obtain, for the condition that $\hat{c}_1\to c_1$,

$$c_1\Xi^{Q'}{}_{Lq,k}(u)=(1-u^2)^2\frac{\partial^2}{\partial u^2}\Xi^{Q'}{}_{Lq,k}(u)$$
$$+\{(1-u^2)[L(L+1)+(n^2-\tfrac{1}{4})$$
$$-2M_k{}^2]+M_k{}^2\}\Xi^{Q'}{}_{Lq,k}(u)$$
$$+iu\sqrt{1-u^2}\{a_k\Xi^{Q'}{}_{Lq,k-1}(u)$$
$$-a_{k+1}\Xi^{Q'}{}_{Lq,k+1}(u)\}. \tag{6.21}$$

In (6.21), we have used the fact that $N^2$ takes on the value $n^2-\frac{1}{4}$, and we have defined

$$a_k=\sqrt{k(2n+k)}\sqrt{(L-M_k+1)(L+M_k)}$$
$$=\sqrt{k(2n+k)}\sqrt{L(L+1)-M_k(M_k-1)}. \tag{6.22}$$

For the condition that $\hat{c}_2\to c_2$, we obtain [with the help of (19) and (20)]

$$c_2\Xi^{Q'}{}_{Lq,k}(u)=iM_k\left((1-u^2)\frac{\partial}{\partial u}+u\right)\Xi^{Q'}{}_{Lq,k}(u)$$
$$+\tfrac{1}{2}\sqrt{1-u^2}(a_k\Xi^{Q'}{}_{Lq,k-1}(u)$$
$$+a_{k+1}\Xi^{Q'}{}_{Lq,k+1}(u)). \tag{6.23}$$

Since $0\leqslant k\leqslant L-n$, Eqs. (6.21) and (6.23) form a finite system of coupled equations for each $L$. These equations can be solved by dividing out the factor

$$(1-u^2)^{-M_k/2+1/2}$$

and studying the special cases $k=0$ for the second-order equation and $L=\hat{n}$ (its lowest bound) for the first-order equation to obtain the consistency condition

$$-c_1=1-\hat{n}^2+c_2{}^2/\hat{n}^2. \tag{6.24}$$

Since the SU(1,1) Casimir (and hence $\hat{n}$) and $c_1,c_2$ are invariant under the action of the Lorentz group, this relation remains valid for all $\{k,L,q\}$ that are connected to these spe-

cial cases by the action of the Lorentz group. The solution to the system of equations (6.21) and (6.23) is given by

$$\Xi^{Q'}{}_{Lq,k} = (-1)^k \sqrt{\Gamma(2\hat{n}+k)/\Gamma(2\hat{n})\Gamma(k+1)}\, N^{Q'}{}_{Lq} \times (1-u^2)^{-\hat{n}/2+1/2} P^{L}{}_{-ic_2/\hat{n},\hat{n}+k}(u), \quad (6.25)$$

where $N^{Q'}{}_{Lq}$ is a normalization constant. In the next section, we shall show that it is independent of $q$, and that its $L$ dependence (for $L \geqslant \hat{n}$) is determined by requiring that the $\xi^{Q'}{}_{Lq}$ transform as a Gel'fand basis.

## VII. THE GEL'FAND–NAIMARK CANONICAL REPRESENTATION

We have imposed the condition on $\psi^{Q}{}_{n}(y)$ that it have definite values of $\mathbf{L}(n)^2$ (labeled by $L$) and $L_1(n)$ (labeled by $q$), along with $l$, $n$, and the Casimir operators $\hat{c}_1$ and $\hat{c}_2$. The action of the Lorentz group changes the values of $L$ and $q$, keeping $c_1$, $c_2$ (and $l,n$) fixed; this action corresponds to the infinitesimal motion of our representation of SU(1,1) along the orbit labeled by $c_1,c_2$. The resulting representation of SL(2,$C$) is the unitary canonical representation of Gel'fand and Naimark.[4,8]

To study the structure of this representation, we shall apply the generators of SL(2,$C$) to the basis vectors

$$\xi^{Q'}{}_{Lq}(n_\mu,\beta,\phi) = \sum_{k=0}^{L-\hat{n}} \Xi^{Q'}{}_{Lq,k}(u) P^{L}{}_{q,-M_k}(z) e^{iq\gamma} \chi_{n+k}{}^{-n}(\beta,\phi). \quad (7.1)$$

Since we have diagonalized $L_1(n) = i\partial/\partial\gamma$, we define

$$L_{\pm}(n) = L_2(n) \pm iL_3(n) = e^{\mp i\gamma}\left(i\sqrt{1-z^2}\frac{\partial}{\partial z} \mp \frac{z}{\sqrt{1-z^2}}\frac{\partial}{\partial\gamma} \pm \frac{i}{\sqrt{1-z^2}}L_3\right). \quad (7.2)$$

We now show that the transformation properties of $\xi_{Lq}$ (we suppress the superscript $Q'$ in the following) as a multiplet under the Lie algebra of SU(2) $\subset$ SL(2,$C$) imply that $\Xi_{Lq,k}$ does not depend on $q$. As a representation of SU(2), we require that [it follows directly from (7.1) that $L_1(n)\xi_{Lq} = q\xi_{Lq}$]

$$L_{\pm}(n)\xi_{Lq} = \sqrt{(L \mp q)(L \pm q + 1)}\,\xi_{L,q\pm 1}. \quad (7.3)$$

Applying the differential operators (7.2) to the function $\xi_{Lq}$ defined by (7.1), we obtain (evaluating $L_3$ on $\chi_{n+k}{}^{-n}$ and using the recursion relations for the functions $P^{L}{}_{q,-M_k}$)

$$L_{\pm}\xi_{Lq} = \sqrt{(L \mp q)(L \pm q + 1)} \times \sum_k \Xi_{Lq,k}(u) P^{L}{}_{q\pm 1,-M_k}(z) \times e^{-i(q\pm 1)\gamma}\chi_{n+k}{}^{-n}. \quad (7.4)$$

Comparing this result with Eq. (7.3), using (7.1) for $\xi_{L,q\pm 1}$, one concludes that $\Xi_{Lq,k} \equiv \Xi_{L,k}$ is independent of $q$.

We now turn to the action of the noncompact generators, in particular, $A_1(n)$. The action of $A_2(n)$, $A_3(n)$ can then be obtained by commutation with $L_3(n)$. According to Gel'fand,[17] we must have

$$A_1(n)\xi_{Lq} = C^{L}\sqrt{L^2-q^2}\,\xi_{L-1,q} - A_L q\xi_{Lq} - C_{L+1}\sqrt{(L+1)^2-q^2}\,\xi_{L+1,q}, \quad (7.5)$$

where

$$C_L = (i/L)\sqrt{(L^2-L_0^2)(L^2-L_1^2)/4L^2-1)}, \quad A_L = iL_0L_1/L(L+1), \quad (7.6)$$

and $q = -L, -L+1, \ldots, +L$, $L = L_0, L_0+1, \ldots$. In general, unitary representations exist[18] in the cases (i) $L_1$ purely imaginary, $L_0$ arbitrary integer or half-integer, which form the principal series, (ii) $L_0 = 0$ and $-1 \leqslant L_1 \leqslant 1$, which belong to the complementary series, and (iii) a one-dimensional representation for $L_0 = 0$, $L_1 = 1$. The representations we shall study here are from case (i), the principal series.

One can now apply $A_1(n)$ to the functions of $u,z,\gamma,\beta,\phi$ appearing in (7.1). For this purpose, we write the operator in the form

$$A_1(n) = -i\left(z(1-u^2)\frac{\partial}{\partial u} + u(1-z^2)\frac{\partial}{\partial z}\right) + \sqrt{1-z^2}\sqrt{1-u^2}A_1. \quad (7.7)$$

Using the relations (A8) and (A9) of I, (6.19), and (6.20), a recurrence relation given in the Bateman series[19] for Jacobi functions converted to one for the functions $P^{L}{}_{ab}(x)$ by Eq. (6.11) [the application of the differential operator produces combinations of the functions $P^{L}{}_{ab}$ over values of $a,b$, but with $L$ fixed; according to the relation (7.5) of Gel'fand, $L \to L \pm 1$],

$$P^{L-1}{}_{ab}(x) = \frac{L(2L+1)}{\sqrt{(L^2-b^2)(L^2-a^2)}}xP^{L}{}_{ab}(x) - \frac{ab(2L+1)}{(L+1)\sqrt{(L^2-b^2)(L^2-a^2)}}P^{L}{}_{ab}(x) - \frac{L\sqrt{((L+1)^2-a^2)((L+1)^2-b^2)}}{(L+1)\sqrt{(L^2-a^2)(L^2-b^2)}} \times P^{L+1}{}_{ab}(x), \quad (7.8)$$

and a relation given by Vilenkin,[20]

$$\sqrt{(L+1)^2-a^2}P^{L+1}{}_{ab}(x) = \sqrt{(L+1)^2-b^2}xP^{L}{}_{ab}(x) + (i/2)\sqrt{1-x^2}\left\{\sqrt{(L-b)(L-b+1)}P^{L}{}_{a,b+1}(x) + \sqrt{(L+b)(L+b+1)}P^{L}{}_{a,b-1}\right\}, \quad (7.9)$$

the result agrees with the form (7.5) if

$$\frac{N^{Q'}{}_{L+1,q}}{N^{Q'}{}_{Lq}} = -\frac{i}{C_{L+1}}\sqrt{(L+1)^2+\frac{c_2^2}{\hat{n}^2}}\left(\frac{L-\hat{n}+1}{(2L+1)(L+1)}\right), \quad A_L = c_2/L(L+1), \quad (7.10)$$

and

$$C_L = (i/L)\sqrt{(L^2-\hat{n}^2)(L^2+c_2^2/\hat{n}^2)/(4L^2-1)}. \tag{7.11}$$

Comparing these results with Eqs. (7.6), we see that

$$c_2 = -iL_0L_1, \quad L_0 = \hat{n}, \quad L_1 = +ic_2/\hat{n}, \tag{7.12}$$

and that the relations for the values of the Casimir operators, given by Naimark,[21] which are the first of (7.12) and

$$-c_1 = 1 - L_0^2 - L_1^2, \tag{7.13}$$

are, with (6.22), clearly satisfied. The wave functions given by (7.1) therefore form a canonical basis for $L \geqslant \hat{n}$. Since $\hat{c}_2$ is Hermitian in the large Hilbert space $\mathscr{H}$, which we shall discuss below, and since $L_0$ is real, $L_1$ must be pure imaginary. We therefore, as remarked above, are dealing with the principal series (the restriction $L_0 \geqslant \frac{1}{2}$ corresponds to $n \geqslant 0$).

The action of the operators

$$A_{\pm}(n) = A_2(n) \pm iA_3(n) \tag{7.14}$$

on the basis $\xi_{Lq}$ is determined by (7.5), the commutators

$$[L_{\pm}(n), A_1(n)] = \mp A_{\pm}(n), \tag{7.15}$$

and the relations (7.3), to be of the form given by Gel'fand.[17]

With Eq. (7.11), the ratio of normalization coefficients for $L$ values differing by one, Eq. (7.9), can be used to relate $N^{Q'}{}_L$ to the coefficient for the lowest element of the ladder, i.e., for which $L = \hat{n}$:

$$N^{Q'}{}_L = (-1)^{L-\hat{n}}\sqrt{(2L+1)/(2\hat{n}+1)}\,(L-\hat{n})!\,N^{Q'}{}_{\hat{n}}. \tag{7.16}$$

Under the action of the Lorentz group, vectors defined on the space spanned by the basis $\{\xi_{Lq}\}$ have a natural scalar product with respect to integration over the group SL(2,C).[4] This scalar product results in an embedding of the Hilbert spaces $\mathscr{H}_n$ in a larger Hilbert space $\mathscr{H}$. In what follows we shall construct this embedding.

Since, in this case, the Casimir operator $\hat{c}_2$ has continuous spectrum, the basis elements for each value of $c_2$ are not normalizable with respect to the norm associated with this scalar product; they are distributions, as for wave functions of sharp momentum. Denoting these functions by $\xi^{c_2}{}_{Lq}(n)$, to make the $c_2$ dependence explicit, let us define

$$\begin{aligned}\xi(n) &= \int dc_2 \sum_{Lq} a_{Lq}(c_2)\xi^{c_2}{}_{Lq}(n),\\ \eta(n) &= \int dc_2 \sum_{Lq} b_{Lq}(c_2)\xi^{c_2}{}_{Lq}(n),\end{aligned} \tag{7.17}$$

where $a_{Lq}(c_2)$ and $b_{Lq}(c_2)$ are square-integrable functions of $c_2$ with measure $dc_2$. The functions $\xi(n)$, $\eta(n)$ are elements of $\mathscr{H}_n$. We now embed $\mathscr{H}_n$ in the Hilbert space $\mathscr{H}$ with the scalar product

$$(\xi,\eta) = \int dg(T_g(n)\xi(n), T_g(n)\eta(n))_{\Lambda^{-1}{}_g n}, \tag{7.18}$$

where $(\xi(n),\eta(n))_n$ is the scalar product defined in $\mathscr{H}_n$, and the linear operator $T_g(n)$ is defined by the irreducible action of $g$ on each component $\xi^{c_2}{}_{Lq}(n)$ of the Gel'fand basis [for which the infinitesimal action is given by Eqs. (4.9), (5.2), and (5.3)]. In what follows we shall omit the factors $R^{Q'}{}_l(\rho)$ and $\Theta^n{}_l(\theta)$ from the functions that form a complete basis for representing the component of the quantum-mechanical state in $\mathscr{H}_n$, as well as corresponding integrations on $\rho,\theta$, since these variables are invariant under SL(2,C).

The transformation $T_g(n)$ acts both on $n_\mu$ and on the space-time coordinates $y$ through the SU(1,1) little group as indicated in the relation (2.12),

$$T_g(n)\xi(n,y) = \xi(\Lambda^{-1}{}_g n, D^{-1}(\Lambda_g,n)y), \tag{7.19}$$

and since $d^4y$—and hence $\cosh\beta\, d\beta\, d\phi$—is invariant under SU(1,1), the scalar product in $\mathscr{H}_n$ is given by

$$\begin{aligned}&\int \cosh\beta\, d\beta\, d\phi (T_g(n)\xi(n,y))^*(T_g(n)\eta(n,y))\\ &\quad = \int \cosh\beta\, d\beta\, d\phi\, \xi(\Lambda^{-1}{}_g n,y)^*\eta(\Lambda^{-1}{}_g n,y).\end{aligned} \tag{7.20}$$

Since $n_\mu$ is stabilized by the part of $\Lambda$ that belongs to SU(1,1), the integral over this part of the group acts trivially; hence the scalar products defined by Eq. (7.18) contain a common factor of the volume of this SU(1,1). We shall therefore omit this (unbounded) part of the integration; what remains is only an integration over the hyperboloid defined by $n^2 = 1$ with measure

$$\begin{aligned}&d^4n\,\delta(n^2-1)\\ &\quad = \mathbf{n}^2\, d|\mathbf{n}|\cos\omega\, dw\, d\gamma (1/2|\mathbf{n}|)\delta\left(|\mathbf{n}| - \sqrt{1+n_0^2}\right)dn_0,\end{aligned}$$

and, carrying out the $d|\mathbf{n}|$ integration, we obtain the invariant measure

$$d\mu = \tfrac{1}{2}\cosh^2\alpha\cos\omega\, d\gamma\, d\alpha, \tag{7.21}$$

where $|\mathbf{n}| = \cosh\alpha$.

We remark that the integration over $n_\mu$ [which is linked to the coordinate transformations defined by $L(n)$] corresponds to an integration over the coset space of SL(2,C)/SU(1,1); the integration in the scalar product in $\mathscr{H}_n$, as we have seen above, has the effect of integration over the SU(1,1) subgroup. In this way, the integration over the entire group is taken into account.

In Eq. (7.20), we can change variables according to $n \to \Lambda_g n$ to obtain

$$\langle\xi,\eta\rangle = \int d\mu(\xi(n),\eta(n)). \tag{7.22}$$

The norm squared is

$$\|\xi\|^2 = \int d\mu \|\xi(n)\|^2, \tag{7.23}$$

and may be given the following quantum-mechanical interpretation. The norm in $\mathscr{H}_n$, $\|\xi(n)\|^2$, corresponds to the probability per unit volume, with measure $d\mu$ on the hypersurface $n^2 = 1$, that the system is in a state for which the support is in the RMS($n_\mu$). The density $|\psi_n(y)|^2$ corresponds to the probability per unit volume in $\mathbb{R}^7$, with measure $d^4y\, d\mu$, that an event may be found at the space-time point $y$ [with coordinates measured in the system associated with the RMS($n_\mu$)], conditioned by the orientation, according to $n_\mu$, of its restricted domain in space-time. The com-

plete description of the state of the system is then represented in the large space $\mathscr{H}$, with elements $\xi = \{\xi(n)\}$, over all unit spacelike vectors $n_\mu$.

Since the set of functions we have chosen to span the spaces $\mathscr{H}_n$ form a representation of SL(2,C), the action of the Lorentz group results in a natural embedding of the transformed vector, an element of $\mathscr{H}_{n'}$, for $n \to n'$ under the transformation, back into the space $\mathscr{H}_n$. As we have seen, the action of the infinitesimal operators of the Lorentz group on the basis elements of $\mathscr{H}_n$ induce linear combinations over $L,q$, according to the formulas of Gel'fand; the result remains an element of $\mathscr{H}_n$.

Denoting the operators $\mathbf{L}(n)$, $\mathbf{A}(n)$ collectively by $M^{\mu\nu}(n)$, we define the operators

$$\hat{M}^{\mu\nu} = \{M^{\mu\nu}(n)\}$$

on $\mathscr{H}$ according to

$$\hat{M}^{\mu\nu}\xi = \{M^{\mu\nu}(n)\xi(n)\}.$$

The group element $g \in \mathrm{SL}(2,C)$ has the action defined by (7.19) in taking $\mathscr{H}_n \to \mathscr{H}_{\Lambda^{-1}{}_g n}$. This action is represented on the Gel'fand basis as a linear combination of elements in $\mathscr{H}_n$; we may therefore define the representation $\hat{T}_g$ of $g$ on $\mathscr{H}$ according to

$$\hat{T}_g\xi = \{T_g(n)\xi(n)\}, \tag{7.24}$$

where

$$T_g(n)\xi(n) \in \mathscr{H}_n. \tag{7.25}$$

The scalar product $\langle \eta, \hat{T}\xi \rangle$ is therefore well defined, and is of the form of a direct integral over $n_\mu$ on the measure $d\mu$.

The operators $\hat{M}^{\mu\nu}$ are clearly symmetric with respect to the scalar product (7.22). We shall now show that they are self-adjoint by showing that the operators $\hat{T}_g$ are unitary. Using the general form of the vectors $\eta,\xi$ given by (7.17), and the definition (7.24), one finds

$$\begin{aligned}\langle \hat{T}_g\eta, \hat{T}_g\xi \rangle &= \sum_{Lq,L'q'} \int d\mu \int dc_2\, dc_2' \cdot b^*_{Lq}(c_2) a_{L'q'}(c_2') \\ &\times (T^{c_2}{}_g(n)\xi^{c_2}{}_{Lq}(n), T^{c_2'}{}_g(n)\xi^{c_2'}{}_{L'q'}(n)).\end{aligned} \tag{7.26}$$

The scalar product in the right-hand side of (7.26) involves representations for the values $c_2$ and $c_2'$ of the second Casimir operator. However, as we shall show explicitly below, elements of Gel'fand representations with different values of $c_2$ are orthogonal. It then follows from the fact that Gel'fand's representation is unitary for each $n_\mu$ and $c_2$ that the representations $\hat{T}_g$ are unitary as well. The operators $\hat{M}^{\mu\nu}$ are therefore self-adjoint.

We shall now show that

$$\langle \xi^{c_2}{}_{Lq}, \xi^{c_2'}{}_{L'q'} \rangle = \int d\mu \left( \xi^{c_2}{}_{Lq}(n), \xi^{c_2'}{}_{L'q'}(n) \right) = 0 \tag{7.27}$$

for $c_2 \neq c_2'$, $L \neq L'$, $q \neq q'$. As evident from (7.1), the scalar product (7.27) vanishes for $q \neq q'$. What remains is the integral

$$\sum_{k=0}^{\min(L,L')-\hat{n}} \pi \int \frac{du\, dz}{(1-u^2)^2} \Xi^{Q'(c_2)*}{}_{L,k}(u) \Xi^{Q(c_2')}{}_{L',k}(u) \times P^L{}_{q,-M_k}(z) P^{L'}{}_{q,-M_m}(z). \tag{7.28}$$

The $z$ integration, corresponding to the integral on the angle $\omega$, vanishes for $L \neq L'$ since[22]

$$\int_{-1}^{1} dz\, P^L{}_{ab}(z) P^{L'}{}_{ab}(z) = \frac{2}{2L+1}\delta_{LL'}. \tag{7.29}$$

We now investigate the remaining integral

$$\langle \xi^{c_2}{}_{Lq}, \xi^{c_2'}{}_{Lq} \rangle = \frac{2\pi}{2L+1} \sum_{k=0}^{L-\hat{n}} \int \frac{du}{(1-u^2)^2} \Xi^{Q(c_2)}{}_{L,k}(u)^* \Xi^{Q(c_2')}{}_{L,k}(u). \tag{7.30}$$

With the help of Eq. (6.23), one obtains

$$(c_2 - c_2') \sum_{k=0}^{L-\hat{n}} \int \Xi^{Q(c_2)}{}_{L,k}(u)^* \Xi^{Q(c_2')}{}_{L,k}(u) \frac{du}{(1-u^2)^2} = \sum_{k=0}^{L-\hat{n}} \frac{-iM_k}{1-u^2} \Xi^{Q(c_2)}{}_{L,k}(u)^* \Xi^{Q(c_2')}{}_{L,k}(u) \Big|_{u=-1}^{u=1}. \tag{7.31}$$

It follows from Eq. (6.9) that, in the neighborhood of $u = \pm 1$,

$$P^L{}_{(-ic_2/\hat{n}),\hat{n}+k}(u) \sim (1-u^2)^{(\hat{n}+k+ic_2/\hat{n})/2}, \tag{7.32}$$

and hence

$$\Xi^{Q(c_2)}{}_{L,k}(u) \sim (1-u^2)^{(k+1+ic_2/\hat{n})/2}. \tag{7.33}$$

The boundary terms therefore vanish for $k \geqslant 1$. For $k = 0$, these limits have the form

$$\lim_{u \to \pm 1} \exp\{-i[(c_2 - c_2')/\hat{n}] \ln(1-u^2)\}, \tag{7.34}$$

which is zero as a distribution on spaces of smooth functions of $c_2 - c_2'$. Hence, in the sense of distributions,

$$(c_2 - c_2')(\xi^{c_2}{}_{Lq}, \xi^{c_2'}{}_{Lq}) = 0, \tag{7.35}$$

from which it follows that

$$\langle \xi^{c_2}{}_{Lq}, \xi^{c_2'}{}_{Lq} \rangle = \delta(c_2 - c_2'), \tag{7.36}$$

where we have chosen the normalization coefficient arbitrarily as unity.

## VIII. SUMMARY AND CONCLUSIONS

We have shown that the set of eigenfunctions, on the RMS based on a standard (arbitrary) inertial coordinate frame, for the discrete spectrum of the dynamical evolution operator $K_{\mathrm{rel}}$ for relative motion for the two-body relativistic bound state, forms the basis for a class of induced representations from which the principal series of Gel'fand's unitary canonical representations can be constructed. The coordinates $y_\mu$ of the standard RMS are transported along the orbit generated by the action of the Lorentz group by defining $y_\mu = L_\mu{}^\nu(n) x_\nu$, where the matrix $L_\mu{}^\nu(n)$ takes $n_\nu \to \mathring{n}_\nu = (0,0,0,1)$; the variables $y$ undergo an action of the

SU(1,1) [which acts on the $y$'s as O(2,1)] little group, which leaves the RMS invariant. Wave functions in $\mathscr{H}_n$ with support in this RMS($n_\mu$) are represented on the complete set of irreducible representations of SU(1,1) that span this space. In each subspace characterized by a given value of the SU(1,1) Casimir operator, the action of the little group, induced by Lorentz transformations, transforms the coefficients to a new set in the new frame.

A fibration of this representation was then constructed by extracting irreducible representations of the SU(2) subgroup of SL(2,$C$) as the solutions of differential equations (in $n_\mu$ as well as $y_\mu$) on the orbit. The resulting eigenfunctions of $\mathbf{L}(n)^2$ and $L_1(n)$ then form the basis for the construction of irreducible representations of Gel'fand's canonical type for SL(2,$C$).

The solutions of the differential equations for the Casimir operators $\hat{c}_1 = \mathbf{L}(n)^2 - \mathbf{A}(n)^2$ and $\hat{c}_2 = \mathbf{L}(n)\cdot\mathbf{A}(n)$ impose a condition between $c_1$ and $c_2$, i.e., that

$$-c_1 = 1 - \hat{n}^2 + c_2^2/\hat{n}^2, \tag{8.1}$$

where $\hat{n}$ is related to the (invariant) value of the Casimir operator of SU(1,1). Comparing the action of the differential operators representing infinitesimal transformations of the Lorentz group, on this basis, with the formulas of Gel'fand, one finds for the coefficients

$$C_L = (i/L)\sqrt{(L^2 - L_0^2)(L^2 - L_1^2)/(4L^2 - 1)}\,,$$
$$A_L = iL_0L_1/L(L+1) \tag{8.2}$$

that

$$L_0 = \hat{n}, \quad L_1 = ic_2/\hat{n}. \tag{8.3}$$

The SL(2,$C$) Casimirs are related to these parameters by

$$-c_1 = 1 - L_0^2 - L_1^2, \quad c_2 = iL_0L_1. \tag{8.4}$$

For square integrability of the irreducible representations of SU(1,1) that are eigenfunctions of $K_{\rm rel}$ in the RMS, both the label $l$ of the Casimir operator of the algebra of differential operators $\mathfrak{o}(3,1)$ defined on the RMS and the label $n$ of the Casimir operator of its associated SU(1,1) must be integer or half-integer together. If our solutions are to be put into correspondence in the nonrelativistic limit with the solutions of the nonrelativistic Schrödinger equation, for which $l$ and $n$ go over continuously to the orbital angular momentum and magnetic quantum number, they must be integer valued. It is clear from the structure of the induced representations that we found for SL(2,$C$), e.g., from the relation

$$0 \leqslant k \leqslant L - \hat{n}, \tag{8.5}$$

where $k$ is the degeneracy index of SU(1,1), that in this case $L$, and hence $L_0$, must be half-integer. Our representations are therefore in the double covering of O(3,1), i.e., they are representations of SL(2,$C$).

It is usual to think of $O(3) \subset O(3,1)$ as a rotation subgroup associated with a nonrelativistic interpretation. This interpretation, in the case of a relativistic particle, is consistent with spin, where the little group for inducing representations of the Poincaré group is O(3), and its representations [in fact, for SU(2)] are defined in the "rest frame," or $\mathbf{p} = 0$ state, of the particle. Since the Galilean group is not a subgroup of the Poincaré group, however, it is not necessary that the O(3) subgroup of O(3,1) which we have used have a direct Galilean interpretation; it is a relativistic O(3), associated with *relative* motion for which there is no rest frame. The representations of its double covering, expressed in a nonunitary finite-dimensional form, may serve as a composite model for spin[23] as an intrinsic property of a particle associated with the center of mass. The nonrelativistic limit of our result arrives at the nonrelativistic O(3) in a completely different way. The algebra of differential operators $\mathfrak{o}(3,1)$ goes over to the algebra of O(3) ($\sin u$, the relative time and relative energy go to zero in this limit), and the O(2,1) subgroup becomes O(2) (corresponding to the generator $L_3$; both pieces $n \geqslant 0$ and $n \leqslant 0$ must be taken into account). It is remarkable that the quantum numbers $l,n$ describe the states of a relativistic system in a way closely analogous to their application to nonrelativistic systems and, in this sense, satisfy a correspondence principle.

We finally wish to comment on the structure of our representation of SL(2,$C$) on a Hilbert space $\mathscr{H}$ that is a direct sum of Hilbert spaces associated with each point of the spacelike hyperboloid for which $n_\mu n^\mu = 1$. The operators of the Lorentz group map each of the spaces $\mathscr{H}_n$ into itself, and therefore the matrix elements of these operators are of direct integral form as well. The wave function $\psi_n(y)$ corresponds to the amplitude for a probability per unit volume in $\mathbb{R}^7$, with measure $d^4y\,d\mu$, that an event may be found at the space-time point $y_\mu$ [with coordinates measured in the system associated with the RMS($n_\mu$)], conditioned by the orientation, according to $n_\mu$, of its restricted domain in space-time. The complete description of the state of the system is then represented in the large space $\mathscr{H}$, with elements $\hat{\psi} = \{\psi_n\}$, over all unit spacelike vectors $n_\mu$.

## ACKNOWLEDGMENTS

We wish to thank V. Rittenberg for his encouragement in the early stages of this work, and I. Pietetsky-Shapiro, H. Bacry, and L. C. Biedenharn for useful discussions. One of us (L.H.) wishes to thank his colleagues at Syracuse University, in particular, F. Rohrlich, and at the University of Texas, in particular, I Prigogine and W. C. Schieve, for the opportunity to visit and for discussions of this work.

[1] R. Arshansky and L. P. Horwitz, J. Math. Phys. **30**, 66 (1989).
[2] E. C. G. Stueckelberg, Helv. Phys. Acta **14**, 372, 588 (1941); L. P. Horwitz and C. Piron, Helv. Phys. Acta **46**, 316 (1973); L. P. Horwitz and Y. Lavie, Phys. Rev. D **26**, 819 (1982), and references listed there.
[3] E. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann, *ibid.* **48**, 568 (1947).
[4] I. M. Gel'fand, R. A. Minlos, and P. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Application* (Pergamon, New York, 1963).
[5] R. Arshansky and L. P. Horwitz, to be published in J. Math. Phys.
[6] N. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, RI, 1986), p. 106.
[7] We thank H. Bacry for a discussion of this point.

[8]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).

[9]As we have pointed out in I, in addition to $n+k\geqslant 0$ there is a set of solutions with $n+k\leqslant 0$; since the action of the Lorentz group affects these indices only through the generators of the SU(1,1) little group, the two sequences are disconnected [the corresponding analysis based on $\chi_{n+k}{}^{-n}$ gives rise to unitarily inequivalent representations of SL(2,$C$)].

[10]See Ref. 6, p. 144.

[11]See Ref. 6, p. 150.

[12]See Ref. 6, p. 129.

[13]See Ref. 6, p. 133.

[14]This condition replaces the requirements of square integrability, since we do not consider these functions as elements of a Hilbert space as yet. We shall carry out such an embedding in a later section.

[15]We remark that for the solutions with $n+k\leqslant 0$ corresponding to $\chi_{n+k}{}^{-n}$, $M_k\rightarrow -M_k$ in Eq. (6.7).

[16]See Ref. 6, p. 143.

[17]See Ref. 4, p. 194; see Ref. 8, p. 104.

[18]See Ref. 4, p. 206.

[19]*Higher Transcendental Functions*, Vol. 2, Bateman Manuscript Project, California Institute of Technology, edited by A. Erdelyi (McGraw-Hill, New York, 1953).

[20]See Ref. 6, p. 158.

[21]See Ref. 8, p. 144.

[22]See Ref. 6, p. 168.

[23]R. Arshansky, Ph.D. thesis, Tel Aviv University, 1987, p. 160.

# The generalization of the binomial theorem

J. Morales[a] and A. Flores-Riveros
*Instituto Mexicano del Petróleo, Investigación Básica de Procesos, Apartado Postal 14-805 07730, Mexico D. F., Mexico*

(Received 24 May 1988; accepted for publication 7 September 1988)

As is well known, the binomial theorem is a classical mathematical relation that can be straightforwardly proved by induction or through a Taylor expansion, albeit it remains valid as long as $[A,B]=0$. In order to generalize such an important equation to cases where $[A,B]\neq 0$, an algebraic approach based on Cauchy's integral theorem in conjunction with the Baker–Campbell–Hausdorff series is presented that allows a partial extension of the binomial theorem when the commutator $[A,B]=c$, where $c$ is a constant. Some useful applications of the new proposed generalized binomial formula, such as energy eigenvalues and matrix elements of power, exponential, Gaussian, and arbitrary $f(\hat{x})$ functions in the one-dimensional harmonic oscillator representation are given. The results here obtained prove to be consistent in comparison to other analytical methods.

## I. INTRODUCTION

A common task in quantum mechanical problems is the calculation of expectation values and transition amplitudes that typically relate to physical information experimentally obtained. The numerical or analytical evaluation of such quantities is directly connected to the computation of matrix elements of potential functions in some suitable representation.

In a second quantization formalism, ladder operators enter as the new variables through which one expresses matrix elements of the form $\langle m|f(\hat{x})|n\rangle$, where $f(\hat{x})$ is an arbitrary function of $\hat{x}$ given in terms of two or more operators that in general do not commute. The expansion of $f(\hat{x})$ in powers of $\hat{x}$ leads consequently to expressions of nonstandard form for the matrix elements. Indeed, while boson systems are based on the commutation rule $[A,B]=1$, the algebraic structure of fermion systems is built up on the anticommutation prescription $[A,B]_+=1$. Commutator expansions are essential in different branches of physics; within time-dependent phenomena in quantum mechanics the evolution operator has been the subject of classical studies dating back to the early papers by Dyson[1] and Feynman,[2] who derived expressions for the chronological ordering operator. Magnus[3] was able to deduce an infinite series in terms of multiple commutators by iterating an integral equation. Such a series has been later utilized and refined to derive an exponential form of time-displacement operators.[4] The celebrated Baker–Campbell–Hausdorff (BCH) series,[5] originally given by an iterative procedure of noncommuting quantities, laid the theoretical grounds for addressing a continuous version of the BCH formula that yields a general algorithm to construct the evolution operator.[6]

Recently, Morales *et al.*[7–12] have undertaken the problem of evaluating one- and two-center integrals by resorting to different approaches: the second quantization formalism along with Cauchy's integral formula for a complex variable and the BCH expansion series in commutators,[7,8] the hypervirial theorem combined with second quantization,[9,10] and parameter differentiation.[11] In the present work, with the purpose of going beyond the binomial theorem, we present a similar approach to the one given in Refs. 7 and 8 that properly accounts for the noncommuting nature of the involved operators, which specifically transform according to

$$[A,B]=c=\text{const}\,, \tag{1}$$

where $c$ in general is complex. In Sec. II a detailed derivation of the generalized binomial theorem is provided and some connections with earlier investigations are pointed out. It is shown how the pertinent expression brings about interesting analytical results that relate to orthogonal polynomials. Section III discusses some applications specialized to matrix elements that involve exponential, Gaussian, and more general functions framed within the one-dimensional harmonic oscillator (HO) representation. A few general remarks and conclusions are left for Sec. IV.

## II. GENERALIZED BINOMIAL FORMULA

The binomial theorem has become a widely used mathematical expression that holds provided that the involved quantities commute. The theorem asserts that

$$(A+B)^N=\sum_{r=0}^{N}\frac{N!}{r!(N-r)!}A^{N-r}B^r\,, \tag{2}$$

which can be readily generalized to the so-called multinomial theorem[13]

$$(a_1+a_2+a_3+\cdots+a_m)^n=\sum_{\{n_i\}}\frac{n!}{n_1!n_2!n_3!\cdots n_m!}a_1^{n_1}a_2^{n_2}a_3^{n_3}\cdots a_m^{n_m},\quad \sum_{i=1}^{m}n_i=n\,, \tag{3}$$

where in either case it is assumed that all variables $a_i$ commute to one another, i.e., $[a_i,a_j]=0$ $(i,j=1,2,\ldots,m)$. The above formulas can easily be proved by induction or by means of a Taylor expansion under the assumption, in both cases, that a commutative algebra is being considered.

In an attempt to transcend the scope of the binomial theorem to more general cases dealing with non-Abelian al-

[a] Also at the Universidad Autónoma Metropolitana, CBI-Física, Azcapotzalco, Mexico.

 0022-2488/89/020393-05$02.50 © 1989 American Institute of Physics 

gebra elements, we address here a formulation that considers a pair of operators following the commutation rule given by Eq. (1). Such a condition leads to

$$[A,[A,B]]=[B,[B,A]]=0\,. \tag{4}$$

By using Cauchy's integral formula

$$(A+B)^N=\frac{N!}{2\pi i}\oint\frac{dt}{t^{N+1}}e^{(A+B)t}\,, \tag{5}$$

which assumes a contour enclosing the origin once, in conjunction with the relation[14]

$$e^{A+B}=e^Ae^Be^{-(1/2)[A,B]}\,, \tag{6}$$

one can obtain

$$(A+B)^N=\frac{N!}{2\pi i}\oint\frac{dt}{t^{N+1}}e^{At}e^{Bt}e^{-(1/2)ct^2}\,. \tag{7}$$

It is important to notice that by virtue of condition (1), formula (6) is rendered as a particular case of the BCH series of multiple commutators,[6]

$$e^{A\#B}=\exp\{A+B+\tfrac{1}{2}[A,B]+\tfrac{1}{12}([A,[A,B]]+[B,[B,A]])+\cdots\}\,, \tag{8}$$

where $A$ and $B$ are elements of an associative but noncommutative algebra, and where

$$\exp(A\#B)=\exp(A)\exp(B)\,.$$

Relation (7) can be reduced by making a Taylor expansion of the exponentials, a change of variable, and use of the following identity for the Hermite polynomial of zero argument:

$$H_n(0)=\frac{n!}{2\pi i}\oint\frac{ds}{s^{n+1}}e^{-s^2}\,. \tag{9}$$

Thus

$$(A\pm B)^N=\sum_{\alpha=0}^{N}\sum_{\beta=0}^{N-\alpha}\frac{N!A^\alpha B^\beta}{\alpha!\beta!\rho!}\left(-\frac{c}{2}\right)^\rho\,, \tag{10}$$

where $\rho=(N-\alpha-\beta)/2$. In a similar way, by choosing the antinormal product, one gets the dual relationship

$$(A\pm B)^N=\sum_{\alpha=0}^{N}\sum_{\beta=0}^{N-\alpha}\frac{N!B^\alpha A^\beta}{\alpha!\beta!\rho!}\left(\frac{c}{2}\right)^\rho\,. \tag{11}$$

The above two forms arise as a natural consequence of the commutation rule and give rise to a normal and antinormal ordered product in each case.

A more tractable expression is obtained by slight manipulation of Eq. (10),

$$\begin{aligned}(A+B)^N&=\sum_{\alpha=0}^{N}\frac{N!A^\alpha B^\beta}{\alpha!(N-\alpha)!}+\sum_{\alpha=0}^{N-2}\frac{N!a^2A^\alpha B^{N-2-\alpha}}{\alpha!(N-2-\alpha)!}\\&\quad+\cdots+\sum_{\alpha=0}^{N-2k}\frac{N!a^{2k}A^\alpha B^{N-2k-\alpha}}{\alpha!(N-2k-\alpha)!k!}+\cdots+B^N\\&=\sum_{k=0}^{[N/2]}\frac{N!}{k!(N-2k)!}\left(-\frac{c}{2}\right)^k(\bar A+\bar B)^{N-2k}\,,\\&\quad a=i\sqrt{c/2}\,,\end{aligned} \tag{12}$$

where

$$[N/2]=\begin{cases}N/2, & N\ \text{even},\\(N-1)/2, & N\ \text{odd}.\end{cases}$$

The notation $\bar A,\bar B$ is introduced to distinguish the "ordinary" binomial formula, $(\bar A+\bar B)^{N-2k}$, provided we keep track of the product $A^nB^m$. We notice *en passant* that only the first term contributes when $c$ vanishes and therefore the binomial original expression is recovered:

$$(A+B)^N=(\bar A+\bar B)^N\,. \tag{13}$$

As a straightforward application, it is worthwhile to note how Eq. (12) may prompt a direct identification with the Hermite polynomials[15] given in the form

$$H_n(x)=\sum_{s=0}^{[n/2]}\frac{(-1)^s n!}{(n-2s)!s!}(2x)^{n-2s}\,, \tag{14}$$

which is readily obtained by using the corresponding generating function.

On the other hand, one can also realize that there exists a connection between binomial expressions containing noncommuting variables and orthogonal polynomials. Indeed, if we take advantage of the property $[2x,-d/dx]=2$ it is immediately inferred that the binomial $(2x-d/dx)$ acting on a given function actually gives rise to an algebraic expression in powers of $x$. Quite simply one can obtain

$$\begin{aligned}&\left(2x-\frac{d}{dx}\right)^n\mathbf{1}\\&\quad=\left[\sum_{k=0}^{[n/2]}\frac{n!(-1)^k}{k!(n-2k)!}(2x)^{n-2k}\left(-\frac{d}{dx}\right)^k\right]\mathbf{1}\\&\quad=H_n(x)\end{aligned} \tag{15}$$

by direct application of Eq. (12). It should be noted that proof of the preceding formula would otherwise require the usual inductive method.

Although this merely represents one sole example, it suggests an alternative procedure for generating known algebraic functions in terms of condensed binomial expressions in the manner just shown. This could serve as a mathematical device to provide new analytical results.

## III. APPLICATIONS

### A. Power operators

In the HO representation, the coordinate operator $\hat x$ is expressed in terms of the creation ($\hat a^\dagger$) and annihilation ($\hat a$) operators

$$\hat x=(\hbar/2\mu\omega)^{1/2}(\hat a+\hat a^\dagger)\,, \tag{16}$$

fulfilling the condition

$$[\hat a,\hat a^\dagger]=1\,, \tag{17}$$

with properties

$$\hat a^\dagger|n\rangle=\sqrt{n+1}\,|n+1\rangle,\quad \hat a|n\rangle=\sqrt{n}\,|n-1\rangle\,. \tag{18}$$

Accordingly, the associated matrix elements can straightforwardly be evaluated in terms of our formulation yielding

$$\langle m|\hat{x}^k|n\rangle = \left(\frac{\hbar}{2\mu\omega}\right)^{k/2} \sum_{\alpha=0}^{k} \sum_{\beta=0}^{k-\alpha} \frac{k!}{\alpha!\beta!\rho!}\left(\frac{1}{2}\right)^{\rho} \times \left(\frac{m!n!}{(m-\alpha)!(n-\beta)!}\right)^{1/2} \langle m-\alpha|n-\beta\rangle . \quad (19)$$

We notice that the overlap integral $\langle m-\alpha|b-\beta\rangle$ conveys the orthogonality condition

$$\langle m-\alpha|n-\beta\rangle = \delta_{m-\alpha,n-\beta} , \quad (20)$$

thereby implying $\beta = n-m+\alpha$, and thus reducing the double summation to

$$\langle m|\hat{x}^k|n\rangle = \frac{k!(2^{n+m}m!n!)^{1/2}}{2^k} \sum_{\alpha=0}^{[m,n]} \left(\frac{1}{2}\right)^{\alpha} \times \frac{1}{\alpha!(m-\alpha)!(n-\alpha)![(k-n-m)/2+\alpha]!} , \quad (21)$$

where $[m,n]$ denotes the smaller of $m$ and $n$ and natural units, $\hbar = \mu = \omega = 1$, have been assumed. The use of the identities

$$(t-s)! = (-1)^s t!/(-t)_s , \quad (t+s)! = (t+1)_s t! \quad (22)$$

allows us to obtain the final compact expression

$$\langle m|\hat{x}^k|n\rangle = \frac{k!}{2^k}\left(\frac{2^{n+m}}{m!n!}\right)^{1/2} \frac{1}{((k-m-n)/2)!} \times \mathbf{F}\left(-m,-n;\frac{k-n-m}{2}+1;\frac{1}{2}\right), \quad (23)$$

which gives rise to many other equivalent formulas owing to the properties of the hypergeometric function $\mathbf{F}(a,b;c;z)$. In particular, the latter expression corresponds to Eq. (6) in Ref. 7 and also coincides with Eq. (20) in Wilcox's paper[16] upon appropriate identification of units and indices. The above formula can also provide an explicit way to evaluate anharmonic corrections to the HO within perturbation theory at any order of anharmonicity. For example, let us consider the Hamiltonian

$$\mathbf{H} = \hat{\mathbf{p}}^2/2\mu + (\mu\omega^2/2)\hat{\mathbf{x}}^2 + \lambda\hat{\mathbf{x}}^4 , \quad (24)$$

where $\hat{p} = -i(\hbar\mu\omega/2)(\hat{a}-\hat{a}^\dagger)$ is the momentum operator. Assume that $\lambda$ is small enough ($\ll\hbar\omega$) that first-order perturbation theory can be safely applied, treating $\lambda\hat{x}^4$ as a perturbation of the HO Hamiltonian. The perturbed energy levels are then

$$E_n = (n+\tfrac{1}{2})\hbar\omega + \Delta n , \quad (25)$$

where

$$\Delta n = \lambda(\hbar/2\mu\omega)^2\langle n|(\hat{a}+\hat{a}^\dagger)^4|n\rangle . \quad (26)$$

We thus have a particular case of Eq. (21) that, when applied to the above expectation value, yields[17]

$$\Delta n = \lambda(\hbar/2\mu\omega)^2(3+6n+6n^2), \quad n\geqslant 2 . \quad (27)$$

In addition, Eq. (23) also relates to a general algebraic relationship for calculating multipole moments[18] in the HO representation. In particular, the transition moments $\langle m|\hat{x}^{2k}|n\rangle$ are given as

$$\langle m|\hat{x}^{2k}|n\rangle = \frac{(2k)!}{2^{2k}}\left(\frac{2^{n+m}}{m!n!}\right)^{1/2} \frac{1}{((2k-n-m)/2)!} \times \mathbf{F}\left(-m,-n;\frac{2k-n-m}{2}+1;\frac{1}{2}\right), \quad (28)$$

where the diagonal elements can be identified with the so-called even moments discussed in Ref. 18,

$$\langle m|\hat{x}^{2k}|m\rangle = \frac{1}{2^{2k-m}}\frac{(2k)!}{m!(k-m)!} \times \mathbf{F}\left(-m,-m;k-m+1;\frac{1}{2}\right). \quad (29)$$

The use of the linear transformation

$$\mathbf{F}(a,b;c;z) = \frac{1}{(1-z)^a}\mathbf{F}\left(a,c-b;c;\frac{z}{z-1}\right) \quad (30)$$

enables us to obtain

$$\langle m|\hat{x}^{2k}|m\rangle = \frac{1}{2^{2k}}\frac{(2k)!}{k!}\sum_{\alpha=0}^{[m,k]}\frac{(k+m-\alpha)!}{\alpha!(m-\alpha)!(k-\alpha)!} , \quad (31)$$

which coincides with the formula recently derived by Namias[19] by solving a finite difference equation generated by a recursive relation of the multipole moments.

### B. General $f(\hat{x})$ operators

In order to discuss matrix elements that entail arbitrary $f(\hat{x})$ functions, we consider a Taylor expansion and the result for $\langle m|\hat{x}^k|n\rangle$. Thus

$$\langle m|f(\hat{x})|n\rangle = \sum_{k=0}^{\infty}\sum_{\alpha=0}^{[m,(k+m-n)/2]}\frac{f^{(k)}(0)}{k!}\left(\frac{\hbar}{2\mu\omega}\right)^{k/2}\frac{k!2^\alpha\sqrt{m!n!}}{2^{(k+m-n)/2}[(k+m-n)/2-\alpha]!(m-\alpha)!(n-m+\alpha)!\alpha!} . \quad (32)$$

Replacement of the index $k+m-n-2\alpha$ by $2r$, which follows from the condition of having integer factorial arguments, readily leads to[20]

$$\langle m|f(\hat{x})|n\rangle = \left(\frac{m!n!}{2^{n-m}}\right)^{1/2}\sum_{r=0}^{\infty}\sum_{\alpha=0}^{m}\frac{f^{(2r+n-m+2\alpha)}(0)}{2^{2r+\alpha}(m-\alpha)!(n-m+\alpha)!\alpha!r!} , \quad (33)$$

where, as before, we have used natural units.

### C. Exponential operators

The preceding formula allows us to find the corresponding matrix elements automatically once the function is differentiated and evaluated at zero. So, for $f(\hat{x}) = \exp(-\beta\hat{x})$, we obtain

$$\langle m|\exp(-\beta\hat{x})|n\rangle = \left(\frac{m!}{n!}\right)^{-1}\left(-\frac{\beta}{\sqrt{2}}\right)^{n-m} L_m^{n-m}\left(-\frac{\beta^2}{2}\right)\exp\left(\frac{\beta^2}{4}\right), \quad n \geqslant m, \tag{34}$$

which has previously been derived by employing the hypervirial theorem and the parameter differentiation technique.[11]

### D. Gaussian integrals

In this case, for a Gaussian-type operator, the $N$th derivative

$$\frac{\partial^N}{\partial\hat{x}^N} e^{-\beta\hat{x}^2}\Big|_{\hat{x}=0} = \frac{(-\beta)^{N/2}N!}{(N/2)!}, \quad N = 2r + 2\alpha + n - m, \tag{35}$$

in Eq. (33), enables us to obtain

$$\langle m|\exp(-\beta\hat{x}^2)|n\rangle = \left(\frac{m!n!}{2^{n-m}}\right)^{1/2} \sum_{r=0}^{\infty}\sum_{\alpha=0}^{m} \frac{(-\beta)^{(2r+2\alpha+n-m)/2}(2r+2\alpha+n-m)!}{((2r+n-m+2\alpha)/2)!2^{2r+\alpha}(m-\alpha)!(n-m+\alpha)!r!\alpha!}. \tag{36}$$

Thus it immediately follows that $n - m$ must be an even number, which gives rise to a well known selection rule for the transition matrix elements involving Gaussian functions.

After slight manipulation, Eq. (36) becomes

$$\langle m|\exp(-\beta\hat{x}^2)|n\rangle = (m!n!)^{1/2} \sum_{\alpha=0}^{m} \frac{2^{(n-m+2\alpha)/2}}{\alpha!(m-\alpha)!(n-m+\alpha)!} \times \sum_{s=(n-m+2\alpha)/2}^{\infty} \frac{(-\beta)^s(2s)!}{2^{2s}s!(s-(n-m+2\alpha)/2)!}, \tag{37}$$

which has also been derived earlier by means of an entirely different formulation.[11] After identification of the infinite sum, one gets

$$\langle m|\exp(-\beta\hat{x}^2)|n\rangle = \left(\frac{m!n!}{1+\beta}\right)^{1/2} \sum_{\alpha=0}^{m}\left[-\frac{\beta}{2(1+\beta)}\right]^{(n-m+2\alpha)/2} \times \frac{(n-m+2\alpha)!}{(m-\alpha)!\alpha!(n-m+\alpha)!((n-m)/2+\alpha)!}. \tag{38}$$

This equation can already be seen to match Eq. (27) in Wilcox's paper[16] and the corresponding formulas of Sack[21] and of Chan and Stelman[22] when translated into their notation. In addition, Eq. (38) can be further simplified through the properties of the hypergeometric function. In fact, by using the identity[16]

$$(2t+2s)! = (2t)!2^{2s}(t+\tfrac{1}{2})_s(t+1)_s, \tag{39}$$

one is led to

$$\langle m|\exp(-\beta\hat{x}^2)|n\rangle = \frac{1}{((n-m)/2)!}\left(\frac{n!}{m!}\right)^{-1}\left(-\frac{\beta}{2}\right)^{(n-m)/2} \times\left(\frac{1}{1+\beta}\right)^{(n-m+1)/2} \times \mathbf{F}\left(-m, \frac{n-m+1}{2}; n-m+1; \frac{2\beta}{1+\beta}\right). \tag{40}$$

It is interesting to note that due to the particular relationship existing between the arguments $b$ and $c$ the hypergeometric function can admit quadratic transformations. In particular, if we make use of the transformation

$$\mathbf{F}(a,b;2b;z) = \left(1-\frac{z}{2}\right)^{-a}\mathbf{F}\left(\frac{a}{2}, \frac{1}{2}+\frac{a}{2}; b+\frac{1}{2}; \frac{z^2}{(2-z)^2}\right), \tag{41}$$

one readily gets

$$\langle m|\exp(-\beta\hat{x}^2)|n\rangle = \frac{1}{((n-m)/2)!}\left(\frac{n!}{m!}\right)^{-1}\left(-\frac{\beta}{2}\right)^{(n-m)/2} \times\left(\frac{1}{1+\beta}\right)^{(n+m+1)/2} \times \mathbf{F}\left(-\frac{m}{2}, \frac{1}{2}-\frac{m}{2}; \frac{n-m}{2}+1; \beta^2\right), \tag{42}$$

which, after using the identity

$$\left(-\frac{m}{2}\right)_\alpha = (-1)^\alpha \frac{1}{2^\alpha}\frac{m!!}{(m-2\alpha)!!}, \tag{43}$$

explicitly becomes

$$\langle m|\exp(-\beta\hat{x}^2)|n\rangle = \left(-\frac{\beta}{2}\right)^{(n-m)/2}\left(\frac{m!n!}{(1+\beta)^{n+m+1}}\right)^{1/2} \times \sum_{\alpha=0}^{[m/2]}\left(\frac{\beta}{2}\right)^{2\alpha}\frac{1}{\alpha!(m-2\alpha)!((n-m)/2+\alpha)!}. \tag{44}$$

An obvious advantage of the above relation with respect to its equivalent, Eq. (38), is that it only needs half the number of terms in the evaluation of the corresponding matrix elements, which constitutes an important improvement in connection with computational calculations. Moreover, the approach addressed here should be appreciated when compared to the Fourier transform method proposed by Wilcox; our formula simply requires identification of the

$N$th derivative of the corresponding function evaluated at $x = 0$. We therefore avoid additional complications that very often arise in integration-technique-based methods.

## IV. CONCLUDING REMARKS

An algebraic formulation has been presented that resorts to Cauchy's integral theorem combined with the BCH series expansion for noncommuting variables, specialized to the case $[A,B] = c = \text{const}$, which allows us to derive an expression viewed as a partial generalization of the binomial theorem. Straightforward applications of the new formula were carried out to evaluate algebraically matrix elements of power and more general functions in the one-dimensional harmonic oscillator representation. The consistency of the results thus obtained is put forward in light of the various expressions that have been derived in the literature through different techniques as well as the advantages underlying our formulation when compared to those methods. The present algebraic procedure can additionally be extended to the calculation of two-center HO integrals. The treatment of more general cases can gradually be incorporated into our mathematical framework, which would allow the appropriate handling of operators that obey more complex commutation rules.

## ACKNOWLEDGMENT

We are grateful to Professor Muneer A. Rashid at the Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria, for having pointed out to us the convenience of using a quadratic transformation of the hypergeometric function in order to obtain Eq. (44), which has improved the results presented here.

[1]F. J. Dyson, Phys. Rev. **75**, 486 (1949).
[2]R. P. Feynman, Phys. Rev. **84**, 108 (1951).
[3]W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[4]P. Pechukas and J. C. Light, J. Chem. Phys. **44**, 3897 (1966).
[5]J. E. Campbell, Proc. London Math. Soc. **29**, 14 (1898); H. F. Baker, Proc. London Math. Soc. **34**, 347 (1902); **35**, 333 (1903); **36**, 24 (1904). F. Hausdorff, Ber. Verh. Saechs. Akad. Wiss. Leipzig Math. Naturwis. Kl. **58**, 19 (1906).
[6]B. Mielnik and J. Plebanski, Ann. Inst. H. Poincare **XII**, 215 (1970).
[7]J. Morales, A. Palma, and M. Berrondo, Int. J. Quant. Chem. Symp. **18**, 57 (1984).
[8]A. Palma, L. Sandoval, and J. Morales, Int. J. Quant. Chem. Symp. **21**, 729 (1987).
[9]J. Morales, A. Palma, and L. Sandoval, Int. J. Quant. Chem. **XXIX**, 211 (1986).
[10]J. Morales, L. Sandoval, and A. Palma, J. Math. Phys. **27**, 2966 (1986).
[11]J. Morales, J. Lopez-Bonilla, and A. Palma, J. Math. Phys. **28**, 1032 (1987).
[12]J. Morales, L. Sandoval, A. Palma, and J. Lopez-Bonilla, Chem. Phys. Lett. **135**, 149 (1987).
[13]A. Beiser, *Concepts of Modern Physics* (McGraw-Hill, Kogakusha, Tokyo, 1967).
[14]E. Vilallonga and D. A. Micha, J. Chem. Phys. **86**, 750 (1987).
[15]G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1967).
[16]R. M. Wilcox, J. Chem. Phys. **45**, 3312 (1966).
[17]R. P. Feynman, *Statistical Mechanics. A Set of Lectures* (Benjamin, New York, 1972). See also S. Flugge, *Practical Quantum Mechanics, Part I* (Springer, Berlin, 1974).
[18]J. L. Richardson and R. Blankenbecler, Phys. Rev. D **19**, 496 (1979).
[19]V. Namias, Am. J. Phys. **55**, 1008 (1987).
[20]This formula is equivalent to Eq. (4.3) in Ref. 12; however, it should be pointed out that a misprint has been detected in that paper: $(2j + n - m - 2r)$ is to be replaced by $(2j + n - m + 2r)$.
[21]R. A. Sack, Philos. Mag. **3**, 497 (1958).
[22]S. I. Chan and D. Stelman, J. Chem. Phys. **39**, 545 (1963).

# Recurrence relations for two-center harmonic oscillator integrals and their solutions using generating functions

Muneer A. Rashid
*Mathematics Department, Ahmadu Bello University, Zaria, Nigeria*

(Received 6 June 1988; accepted for publication 10 August 1988)

Using the algebraic second quantization formalism, the general recurrence relations for some two-center harmonic oscillator integrals are obtained. The various known special cases for one-center integrals are evident in the formalism. Generating functions for these integrals are obtained and most of them are solved to find explicit expressions for polynomial, exponential, and Gaussian operator integrals. For these cases, *all* the generating integrals needed for utilizing the recurrence relations are also calculated in their simplest forms.

## I. INTRODUCTION

Since the linear harmonic oscillator problem is a classical problem in quantum mechanics, everything seems to be known about it. In fact, the energy levels and the corresponding wave functions are explicitly known and therefore, in principle, we can calculate the matrix elements of any function of $x$ though the calculations may not be straightforward generally. For example, even to verify the orthonormality properties of the wave functions involves a lot of calculations. In order to evaluate a large number of integrals, the use of explicit expressions usually becomes cumbersome. The knowledge of recurrence relations for such matrix elements has thus an advantage of its own.

Recently Morales *et al.*[1-3] have used various methods to derive recurrence relations for the two-center harmonic integrals ${}_G\langle m|f(a_E, a_E^\dagger)|n\rangle_E$. (The notation will be explained in the next section.) In particular, in Ref. 1, they have utilized some hypervirial theorems to obtain these recurrence relations. They have also obtained recurrence relations for the cases where

$$f(a_E, a_E^\dagger) = 1, \quad x^k, \quad e^{-\rho x}, \quad e^{-\rho x^2}$$

and evaluated some generating integrals that are needed to utilize the recurrence relations. In this paper, we use purely algebraic methods to obtain the general recurrence relations and their special cases. We believe that our derivations are simple and straightforward and can be followed comparatively more easily. We have also calculated *all* the generating integrals that are needed for the special cases discussed in the paper. In addition, we have obtained generating functions for the matrix elements

$${}_G\langle m|x^k|n\rangle_E, \quad {}_G\langle m|e^{-\rho x}|n\rangle_E, \quad {}_G\langle m|e^{-\rho x^2}|n\rangle_E,$$

and using these generating functions, evaluated explicitly the matrix elements mentioned before. The special known cases of one-center harmonic-oscillator integrals where the two oscillators have the same masses and frequencies are evident in our results.

## II. NOTATIONS

We follow closely the notation used in Ref. 1 in order to facilitate easy comparison.

Let us consider two harmonic oscillators with masses $\mu_E$ and $\mu_G$ and the corresponding frequencies $\omega_E$ and $\omega_G$. Here $G$ and $E$ refer to the ground (${}_G\langle$) and excited ($\rangle_E$) states. The corresponding creation and annihilation operators are given by $a_E^\dagger$ and $a_E$ for operations on $\rangle_E$ and by $a_G$ and $a_G^\dagger$ for operations on ${}_G\langle$. Their operations on these ket and bra states are given in the well-known relations

$$a_E|n\rangle_E = \sqrt{n}\,|n-1\rangle_E, \tag{1a}$$

$$a_E^\dagger|n\rangle_E = \sqrt{n+1}\,|n+1\rangle_E, \tag{1b}$$

$$\langle m|a_G = \sqrt{m+1}\,{}_G\langle m+1|, \tag{2a}$$

$$\langle m|a_G^\dagger = \sqrt{m}\,{}_G\langle m-1|. \tag{2b}$$

We assume that the two oscillators are displaced by a length $l$ with respect to each other, i.e.,

$$x_G - x_E = l, \tag{3}$$

where $x_G(x_E)$ is the average position of the oscillator represented by $G(E)$.

The operators $a_E$, $a_E^\dagger$, $a_G$, $a_G^\dagger$ are given in terms of $x$, $l$, and $p$ by

$$a_E = \frac{\mu_E\omega_E x + ip}{\sqrt{2\hbar\mu_E\omega_E}}, \quad a_E^\dagger = \frac{\mu_E\omega_E x - ip}{\sqrt{2\hbar\mu_E\omega_E}},$$

$$a_G = \frac{\mu_G\omega_G(x+l) + ip}{\sqrt{2\hbar\mu_G\omega_G}}, \quad a_G^\dagger = \frac{\mu_G\omega_G(x+l) - ip}{\sqrt{2\hbar\mu_G\omega_G}}. \tag{4}$$

Thus

$$x = x_E = (\alpha_E/\sqrt{2})(a_E + a_E^\dagger)$$

$$= (\alpha_E\beta/\sqrt{2})(a_G + a_G^\dagger - \sqrt{2}\gamma), \tag{5}$$

and

$$(1+\beta^2)a_G = \sqrt{2}\beta^2\gamma + (1-\beta^2)a_G^\dagger + 2\beta a_E, \tag{6a}$$

$$(1+\beta^2)a_E^\dagger = -\sqrt{2}\beta\gamma - (1-\beta^2)a_E + 2\beta a_G^\dagger, \tag{6b}$$

where we have defined the quantities

$$\alpha_E = \sqrt{\hbar/\mu_E\omega_E}, \quad \beta = \sqrt{\mu_E\omega_E/\mu_G\omega_G}, \quad \gamma = \sqrt{\mu_G\omega_G/\hbar}\,l. \tag{7}$$

Note that

$$x + l = (\alpha_G/\sqrt{2})(a_G + a_G^\dagger) \tag{8}$$

 0022-2488/89/020398-08$02.50 © 1989 American Institute of Physics 

as expected, where

$$\alpha_G = \sqrt{\hbar/\mu_G \omega_G}. \tag{9}$$

In the sequel, we shall also use another quantity $\eta$ defined by

$$\eta = 1 + \beta^2 + 2\alpha_E^2 \beta^2 \rho, \tag{10}$$

where the parameter $\rho$ appears in the matrix elements

$${}_G\langle m|e^{-\rho x}|n\rangle_E \quad \text{and} \quad {}_G\langle m|e^{-\rho x^2}|n\rangle_E.$$

Note also that we have not added a caret on the operators $x$, $p$, $a_E$, $a_E^\dagger$, $a_G$, $a_G^\dagger$ as in Ref. 1. The text makes the situation clear.

Since

$$[a_E, a_E^\dagger] = 1 = [a_G, a_G^\dagger], \tag{11}$$

using expansions of $f_E = f(a_E, a_E^\dagger)$ in powers of $a_E, a_E^\dagger$ for any function $f_E$, we have

$$[a_E, f_E] = \frac{\partial f_E}{\partial a_E^\dagger} \tag{12a}$$

and

$$[a_E^\dagger, f_E] = -\frac{\partial f_E}{\partial a_E}. \tag{12b}$$

We shall often be concerned with the symmetry (called $S$ in the sequel)

$$\begin{aligned} &E \leftrightarrow G, \quad m \leftrightarrow n, \quad \beta \leftrightarrow 1/\beta, \\ &\gamma \leftrightarrow -\beta\gamma, \quad \alpha_E \leftrightarrow \alpha_E \beta = \alpha_G, \quad l \leftrightarrow -l. \end{aligned} \tag{13}$$

## III. DERIVATION OF THE GENERAL RECURRENCE RELATIONS FOR ${}_G\langle m|\, f_E|n\rangle_E$

Note that we have defined

$$f_E = f(a_E, a_E^\dagger). \tag{14}$$

We have

$${}_G\langle m|\, f_E|n\rangle_E = (1/\sqrt{n})\, {}_G\langle m|\, f_E a_E^\dagger|n-1\rangle_E$$

using (1a). But

$$\begin{aligned} f_E a_E^\dagger &= a_E^\dagger f_E - [a_E^\dagger, f_E] \\ &= \frac{1}{1+\beta^2}\left[-\sqrt{2}\beta\gamma - (1-\beta^2)a_E + 2\beta a_G^\dagger\right] f_E \\ &\quad + \frac{\partial f_E}{\partial a_E} \end{aligned}$$

[using Eqs. (6b) and (12b)]

$$\begin{aligned} &= \frac{1}{1+\beta^2}\left[-\sqrt{2}\beta\gamma + 2\beta a_G^\dagger\right] f_E - \frac{1-\beta^2}{1+\beta^2}[a_E, f_E] \\ &\quad - \frac{1-\beta^2}{1+\beta^2} f_E a_E + \frac{\partial f_E}{\partial a_E} \\ &= \frac{1}{1+\beta^2}\left[-\sqrt{2}\beta\gamma + 2\beta a_G^\dagger\right] - \frac{1-\beta^2}{1+\beta^2} f_E a_E \\ &\quad + \frac{\partial f_E}{\partial a_E} - \frac{1-\beta^2}{1+\beta^2}\frac{\partial f_E}{\partial a_E^\dagger} \end{aligned}$$

[using Eq. (12b)].

Now we can obtain the recurrence relation using Eqs. (1) and (2) to arrive at

$$\begin{aligned} {}_G\langle m|\, f_E|n\rangle_E = &-\left(\frac{2}{n}\right)^{1/2} \frac{\beta\gamma}{1+\beta^2}\, {}_G\langle m|\, f_E|n-1\rangle_E \\ &-\left(\frac{n-1}{n}\right)^{1/2} \frac{1-\beta^2}{1+\beta^2}\, {}_G\langle m|\, f_E|n-2\rangle_E \\ &+\left(\frac{m}{n}\right)^{1/2} \frac{2\beta}{1+\beta^2}\, {}_G\langle m-1|\, f_E|n-1\rangle_E \\ &+\frac{1}{\sqrt{n}}\, {}_G\left\langle m\left|\frac{\partial f_E}{\partial a_E}\right|n-1\right\rangle_E \\ &-\frac{1}{\sqrt{n}}\frac{1-\beta^2}{1+\beta^2}\, {}_G\left\langle m\left|\frac{\partial f_E}{\partial a_E^\dagger}\right|n-1\right\rangle_E, \\ &\text{for } m \geqslant 0,\ n \geqslant 1. \end{aligned} \tag{15}$$

Similarly starting from

$${}_G\langle m|\, f_E|n\rangle_E = (1/\sqrt{m})\, {}_G\langle m-1|\, a_G f_E|n\rangle_E,$$

we use Eq. (6a) to write

$$a_G f_E = [1/(1+\beta^2)]\left[\sqrt{2}\beta^2\gamma + (1-\beta^2)a_G^\dagger + 2\beta a_E\right] f_E,$$

wherein

$$a_E f_E = [a_E, f_E] + f_E a_E = \frac{\partial f_E}{\partial a_E^\dagger} + f_E\, a_E$$

from Eq. (12a).

These manipulations finally result in

$$\begin{aligned} {}_G\langle m|\, f_E|n\rangle_E = &\left(\frac{2}{m}\right)^{1/2} \frac{\beta^2\gamma}{1+\beta^2}\, {}_E\langle m-1|\, f_E|n\rangle_G \\ &+\left(\frac{m-1}{m}\right)^{1/2} \frac{1-\beta^2}{1+\beta^2}\, {}_G\langle m-2|\, f_E|n\rangle_E \\ &+\left(\frac{n}{m}\right)^{1/2} \frac{2\beta}{1+\beta^2}\, {}_G\langle m-1|\, f_E|n-1\rangle_E \\ &+\frac{1}{\sqrt{m}}\frac{2\beta}{1+\beta^2}\, {}_G\left\langle m-1\left|\frac{\partial f_G}{\partial a_E^\dagger}\right|n\right\rangle_E, \\ &\text{for } m \geqslant 1, \quad n \geqslant 0. \end{aligned} \tag{16}$$

The above equations (15) and (16) are the right-hand and the left-hand general recurrence relations given in Eqs. (2.12) and (2.24) in Ref. 1. Note that our procedure is simple and straightforward. This remark can be appreciated if one examines the manipulations and the use of virial type theorems in Morales *et al.*[1]

The recurrence relations given in Eqs. (15) and (16) above are useful in increasing the values of $n$ and $m$, respectively. Except for the terms involving the derivatives $\partial f_E/\partial a_E$, $\partial f_E/\partial a_E^\dagger$ of $f_E$, the two recursion relations go into one another under the symmetry $S$ defined in Eq. (13) and complex conjugation provided $f$ is a Hermitian operator.

In the sequel, we shall be concerned with the special cases where

$$f_E = f(a_E, a_E^\dagger) = f_E(x)$$

and then

$$\frac{\partial f_E}{\partial a_E} = \frac{\partial f_E}{\partial a_E^\dagger} = \frac{\alpha_E}{\sqrt{2}} \frac{df_E}{dx}$$

using Eq. (5). The recurrence relations (15) and (16) thus become

$$\begin{aligned} {}_G\langle m|f_E|n\rangle_E = &-\left(\frac{2}{n}\right)^{1/2} \frac{\beta\gamma}{1+\beta^2} {}_G\langle m|f_E|n-1\rangle_E \\ &-\left(\frac{n-1}{n}\right)^{1/2} \frac{1-\beta^2}{1+\beta^2} {}_G\langle m|f_E|n-2\rangle_E \\ &+\left(\frac{m}{n}\right)^{1/2} \frac{2\beta}{1+\beta^2} {}_G\langle m-1|f_E|n-1\rangle_E \\ &+\left(\frac{2}{n}\right)^{1/2} \alpha_E \frac{\beta^2}{1+\beta^2} {}_G\left\langle m\left|\frac{df_E}{dx}\right|n-1\right\rangle_E, \end{aligned} \tag{15'}$$

$$\begin{aligned} {}_G\langle m|f_E|n\rangle_E = &\left(\frac{2}{m}\right)^{1/2} \frac{\beta^2\gamma}{1+\beta^2} {}_G\langle m-1|f_E|n\rangle_E \\ &+\left(\frac{m-1}{m}\right)^{1/2} \frac{1-\beta^2}{1+\beta^2} {}_G\langle m-2|f_E|n\rangle_E \\ &+\left(\frac{n}{m}\right)^{1/2} \frac{2\beta}{1+\beta^2} {}_G\langle m-1|f_E|n-1\rangle_E \\ &+\left(\frac{2}{m}\right)^{1/2} \alpha_E \frac{\beta}{1+\beta^2} {}_G\left\langle m-1\left|\frac{df_E}{dx}\right|n\right\rangle_E. \end{aligned} \tag{16'}$$

These recurrence relations do go over into one another under the symmetry $S$ and complex conjugation.

## IV. RECURRENCE RELATIONS AND THEIR SOLUTIONS FOR THE EXPONENTIAL MATRIX ELEMENTS

Taking

$$f_E(x) = e^{-\rho x}, \quad \frac{df_E}{dx} = -\rho e^{-\rho x}, \tag{17}$$

thus Eqs. (15) and (16) take the form

$$\begin{aligned} &{}_G\langle m|e^{-\rho x}|n\rangle_E \\ &= -\left(\frac{2}{n}\right)^{1/2} \frac{\beta(\gamma+\alpha_E\beta\rho)}{1+\beta^2} {}_G\langle m|e^{-\rho x}|n-1\rangle_E \\ &\quad -\left(\frac{n-1}{n}\right)^{1/2} \frac{1-\beta^2}{1+\beta^2} {}_G\langle m|e^{-\rho x}|n-2\rangle_E \\ &\quad +\left(\frac{m}{n}\right)^{1/2} \frac{2\beta}{1+\beta^2} {}_G\langle m-1|e^{-\rho x}|n-1\rangle_E \end{aligned}$$

for $m \geqslant 0$, $n \geqslant 1$, (18)

and

$$\begin{aligned} &{}_G\langle m|e^{-\rho x}|n\rangle_E \\ &= \left(\frac{2}{m}\right)^{1/2} \frac{\beta(\beta\gamma-\alpha_E\rho)}{1+\beta^2} {}_G\langle m-1|e^{-\rho x}|n\rangle_E \\ &\quad +\left(\frac{m-1}{m}\right)^{1/2} \frac{1-\beta^2}{1+\beta^2} {}_G\langle m-2|e^{-\rho x}|n\rangle_E \\ &\quad +\left(\frac{n}{m}\right)^{1/2} \frac{2\beta}{1+\beta^2} {}_G\langle m-1|e^{-\rho x}|n-1\rangle_E \end{aligned}$$

for $m \geqslant 1$, $n \geqslant 0$. (19)

As remarked earlier, these recursion relations go over into one another under the symmetry $S$.

To obtain the generating function for the matrix elements ${}_G\langle m|e^{-\rho x}|n\rangle_E$ we define

$$E_{m,n}(\rho) = {}_G\langle m|e^{-\rho x}|n\rangle_E/\sqrt{m!n!} \tag{20}$$

and the generating function

$$E(\rho; y,z) = \sum_{m,n=0}^{\infty} E_{m,n}(\rho) y^m z^n. \tag{21}$$

Replacing $m$ by $(m+1)$, $n$ by $n+2$, in Eq. (18) and multiplying by $\sqrt{n+2}/\sqrt{(m+1)!(n+1)!}$, we obtain

$$\begin{aligned} &(n+2)E_{m+1,n+2}(\rho) \\ &= -\sqrt{2}\frac{\beta(\gamma+\alpha_E\beta\rho)}{1+\beta^2} E_{m+1,n+1}(\rho) \\ &\quad -\frac{1-\beta^2}{1+\beta^2} E_{m+1,n}(\rho) + \frac{2\beta}{1+\beta^2} E_{m,n+1}(\rho). \end{aligned} \tag{22}$$

We multiply the terms in the above equation by $y^m z^n$ and sum over $m$ and $n$ from 0 to $\infty$ to obtain the differential equation

$$\begin{aligned} &\frac{\partial E}{\partial z}(\rho;y,z) \\ &= -\left[\sqrt{2}\frac{\beta(\gamma+\alpha_E\beta\rho)}{1+\beta^2} \right. \\ &\quad \left. +\frac{1-\beta^2}{1+\beta^2} z - \frac{2\beta}{1+\beta^2} y\right] E(\rho; y,z), \end{aligned} \tag{23}$$

*wherein the "boundary terms" cancel out* when we compare coefficients of various powers of $y$ or $z$ and use the recurrence relation (18) for *special values* of $m$ and $n$.

The above differential equation can be solved immediately to give

$$E(\rho;y,z) = E(\rho;y,0)\exp\left[\frac{-(1-\beta^2)z^2 + 4\beta yz - 2\sqrt{2}\beta(\gamma+\alpha_E\beta\rho)z}{2(1+\beta^2)}\right]. \tag{24}$$

To obtain $E(\rho;y,0)$ we use the left-hand recurrence relation [Eq. (19)] *which is valid for* $n=0$ and note that

$$E(\rho;y,0) = \sum_{m=0}^{\infty} E_{m,0}(\rho) y^m$$

to obtain the differential equation

$$\frac{\partial}{\partial y} E(\rho;y,0) = \left[\frac{\sqrt{2}\beta(\beta\gamma-\alpha_E\rho)}{1+\beta^2} + \frac{1-\beta^2}{1+\beta^2} y\right] E(\rho;y,0) \tag{25}$$

with the solution

$$E(\rho;y,0) = E(\rho;0,0)\exp\left[\frac{(1-\beta^2)y^2 + 2\sqrt{2}\beta(\beta\gamma - \alpha_E\rho)y}{2(1+\beta^2)}\right], \tag{26}$$

wherein

$$E(\rho;0,0) = E_{0,0}(\rho) = {}_G\langle 0|e^{-\rho x}|0\rangle_E \tag{27}$$

is the generator matrix element to be calculated. We call it $E(\rho)$ and to calculate it, we differentiate $E(\rho)$ with respect to the *parameter* $\rho$ to write

$$\frac{dE(\rho)}{d\rho} = -{}_G\langle 0|e^{-\rho x}x|0\rangle_E = -\frac{\alpha_E}{\sqrt{2}}\,{}_G\langle 0|e^{-\rho x}|1\rangle_E, \tag{28}$$

since $a_E|0\rangle_E = 0$, $a_E^\dagger|0\rangle_E = |1\rangle_E$, and $x = (\alpha_E/\sqrt{2})(a_E + a_E^\dagger)$. But the recurrence relation (18) for $m = 0$, $n = 1$ implies

$${}_G\langle 0|e^{-\rho x}|1\rangle_E = -\sqrt{2}\,\frac{\beta(\gamma + \alpha_E\beta\rho)}{1+\beta^2}\,{}_G\langle 0|e^{-\rho x}|0\rangle_E. \tag{29}$$

Thus

$$\frac{dE(\rho)}{d\rho} = \frac{\alpha_E\beta}{1+\beta^2}(\gamma + \alpha_E\beta\rho)E(\rho), \tag{30}$$

which gives the solution

$$E(\rho) = {}_G\langle 0|0\rangle_E \exp\left[\frac{2\alpha_E\beta\gamma\rho + \alpha_E^2\beta^2\rho^2}{2(1+\beta^2)}\right]. \tag{31}$$

Finally we require the overlap integral ${}_G\langle 0|0\rangle_E$. This is the only one we have to evaluate using integration (nonalgebraic) techniques. Indeed, from the *ground-state normalized* wave function for the linear harmonic oscillator,

$$\psi_E(x) = (\pi\alpha_E^2)^{1/4}e^{-(1/2)(x^2/\alpha_E^2)}, \tag{32}$$

$$\psi_G(x) = (\pi\alpha_G^2)^{1/4}e^{-(1/2)[(x+l)^2/\alpha_G^2]}, \tag{33}$$

we find

$$\begin{aligned}{}_G\langle 0|0\rangle_E &= \frac{1}{\sqrt{\pi\alpha_E\alpha_G}} \times\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\frac{x^2}{\alpha_E^2} - \frac{1}{2}\frac{(x+l)^2}{\alpha_G^2}\right)dx \\ &= \left(\frac{2\beta}{1+\beta^2}\right)^{1/2}\exp\left(-\frac{1}{2}\frac{\beta^2\gamma^2}{1+\beta^2}\right)\end{aligned} \tag{34}$$

on using the definitions in Eqs. (7) and (9). Note that when we consider the single harmonic-oscillator integrals $G = E$, $\beta = 1$ and $\gamma = 0$ and ${}_G\langle 0|0\rangle_E = 1$ as expected. Also this explicit answer has the symmetry $S$. We shall require the above expression for ${}_G\langle 0|0\rangle_E$ in the next section also.

Combining Eqs. (24), (26), (31), and (34), we finally arrive at

$$\begin{aligned}E(\rho;y,z) &= \sum_{m,n=0}^{\infty}\frac{{}_G\langle m|e^{-\rho x}|n\rangle_E}{\sqrt{m!n!}}y^m z^n \\ &= \left(\frac{2\beta}{1+\beta^2}\right)^{1/2} \times\exp\left[\frac{(1-\beta^2)(y^2-z^2) + 4\beta yz + 2\sqrt{2}\beta\{(\beta\gamma - \alpha_E\rho)y - (\gamma + \alpha_E\beta\rho)z\} + \beta^2(\alpha_E^2\rho^2 - \gamma^2) + 2\alpha_E\beta\gamma\rho}{2(1+\beta^2)}\right].\end{aligned} \tag{35}$$

Except for the factor $\exp(\alpha_E\beta\gamma\rho/1+\beta^2)$, the above generating function remains invariant under the transformation

$\beta \leftrightarrow 1/\beta$,

$\gamma \leftrightarrow -\gamma\beta$,

$y \leftrightarrow z$ (this corresponds to $m \leftrightarrow n$ in $S$),

$\alpha_E \leftrightarrow \gamma_G = \alpha_E\beta$.

Under this transformation

$$\exp\left(\frac{\alpha_E\beta\gamma\rho}{1+\beta^2}\right) \to \exp\left(-\frac{\alpha_E\beta^3\gamma\rho}{1+\beta^2}\right) = \exp\left(\frac{\alpha_E\beta\gamma\rho}{1+\beta^2} - lp\right)$$

as expected since $e^{-\rho x} \to e^{-\rho(x+l)} = e^{-\rho x}\times e^{-lp}$. In fact, had we used a *symmetrical* operator

$$e^{-(1/2)\rho(x_E + x_G)} = e^{-\rho(x+(1/2)l)},$$

we would have obtained a *completely symmetrical* expression for the generating function.

Finally we expand $E(\rho; y,z)$ in powers of $y$ and $z$ and pick up the coefficient of $y^m z^n/\sqrt{m!n!}$ to obtain ${}_G\langle m|e^{-\rho x}|n\rangle_E$. This results in

$$\begin{aligned}{}_G\langle m|e^{-\rho x}|n\rangle_E &= \left(m!n!\,\frac{2\beta}{1+\beta^2}\right)^{1/2}\exp\left[\frac{\beta^2(\alpha_E^2\rho^2 - \gamma^2) + 2\alpha_E\beta\gamma\rho}{2(1+\beta^2)}\right] \\ &\times\sum_{\gamma_1,\gamma_2,\gamma_3}(-1)^{n+\gamma_2+\gamma_3}2^{(1/2)(m+n)-2\gamma_1-2\gamma_2}\frac{\beta^{m+n-2\gamma_1-2\gamma_2-2\gamma_3}}{(1+\beta^2)^{m+n-\gamma_1-\gamma_2-\gamma_3}} \\ &\times\frac{(1-\beta^2)^{\gamma_1+\gamma_2}(\beta\gamma - \alpha_E\rho)^{m-2\gamma_1-\gamma_3}(\gamma + \alpha_E\beta\rho)^{n-2\gamma_2-\gamma_3}}{\gamma_1!\gamma_2!\gamma 3!(m-2\gamma_1-\gamma_3)!\,(n-2\gamma_2-\gamma_3)!}.\end{aligned} \tag{36}$$

As mentioned earlier, this expression has also the required symmetry $S$ except for the factor $\exp[\alpha_E\beta\gamma\rho/1+\beta^2]$. Again both $\gamma_1$ and $\gamma_2$ summations could be replaced by a Laguerre polynomial for each, but the actual expressions in each case depend upon whether $m-\gamma_3, n-\gamma_3$ are even or odd. Thus our answer will not be very elegant and not any more useful then Eq. (36) above. In place of this, we replace the $\gamma_3$ summation by a Laguerre polynomial using the definition[4]

$$L_m^{\alpha}(x) = \sum_{m=0}^{n} \frac{\Gamma(n+\alpha+1)}{m!(n-m)!\Gamma(m+\alpha+1)}(-x)^m$$
$$= \sum_{m=0}^{n} \frac{\Gamma(n+\alpha+1)}{m!(n-m)!\Gamma(n-m+\alpha+1)}(-x)^{n-m}$$

given in Eq. (8.970(1)), p. 1037 in Ref. 5. Thus we arrive at

$$ {}_G\langle m|e^{-\rho x}|n\rangle_E = \left(m!n!\frac{2\beta}{1+\beta^2}\right)^{1/2}\exp\left[\frac{\beta^2(\alpha_E^2\rho^2-\gamma^2)+2\alpha_E\beta\gamma\rho}{2(1+\beta^2)}\right]$$
$$\times\sum_{\gamma_1,\gamma_2}(-1)^{n+m+\gamma_2}2^{(1/2)(m+n)-2\gamma_1-2\gamma_2}\frac{\beta^{n-2\gamma_2}(1-\beta^2)^{\gamma_1+\gamma_2}(\gamma+\alpha_E\beta\rho)^{n-m+2\gamma_1-2\gamma_2}}{\gamma_1!\gamma_2!(n-2\gamma_2)!(1+\beta^2)^{n+\gamma_1-\gamma_2}}$$
$$\times L_{m-2\gamma_1}^{n-m+2\gamma_1-2\gamma_2}\left(\frac{\beta(\beta\gamma-\alpha_E\rho)(\gamma+\alpha_E\beta\rho)}{1+\beta^2}\right). \tag{37}$$

The range of summations in Eq. (36) is not indicated. In general, it is dictated by the non-negativity of the arguments of the factorials present and will be omitted whenever this rule applies. In Eq. (37), $0\leqslant\gamma_1\leqslant[n/2]$, $0\leqslant\gamma_2\leqslant[m/2]$, which ensures that the argument of $(n-2\gamma_2)!$ is non-negative and the lower index $m-2\gamma_1$ of the Laguerre polynomial is a non-negative integer which is required for it to be a *polynomial*. In such a situation also, the range of summation variables will be omitted.

For the single harmonic oscillator, the equations corresponding to Eqs. (35)–(37) can be obtained by taking $\beta=1, \gamma=0$ writing $\alpha$ for $\alpha_E$ and omitting the subscripts $G$ and $E$ from the states in the matrix elements. Thus we have

$$E(\rho;y,z) = \sum_{m,n=0}^{\infty}\frac{\langle m|e^{-\rho x}|n\rangle}{\sqrt{m!n!}}y^m z^n$$
$$= \exp\left[yz-(\alpha/\sqrt{2})\rho(y+z)+\tfrac{1}{4}\alpha^2\rho^2\right]. \tag{35'}$$

Again

$$\langle m|e^{-\rho x}|n\rangle$$
$$= \sqrt{m!n!}\left(-\frac{\alpha\rho}{\sqrt{2}}\right)^{m+n}$$
$$\times\exp\left(\frac{1}{4}\alpha^2\rho^2\right)\sum\left(\frac{2}{\alpha^2\rho^2}\right)^{\gamma}\frac{1}{\gamma!(m-\gamma)!(n-\gamma)!} \tag{36'}$$
$$= \left(\frac{m!}{n!}\right)^{1/2}\left(-\frac{\alpha\rho}{\sqrt{2}}\right)^{-m+n}$$
$$\times\exp\left(\frac{1}{4}\alpha^2\rho^2\right)L_m^{n-m}\left(-\frac{\alpha^2\rho^2}{2}\right). \tag{37'}$$

A symmetrical recursion relation can be obtained from Eqs. (18) and (19) by taking $\beta=1$, $\gamma=0$ and eliminating the term $\langle m-1|e^{-\rho x}|n-1\rangle$. This results in

$$(m-n)\langle m|e^{-\rho x}|n\rangle = \alpha\rho[\sqrt{(n/2)}\langle m|e^{-\rho x}|n-1\rangle$$
$$-\sqrt{(m/2)}\langle m-1|e^{-\rho x}|n\rangle] \tag{38}$$

with the generator

$$\langle 0|e^{-\rho x}|0\rangle = \exp(\tfrac{1}{4}\alpha^2\rho^2) \tag{39}$$

from Eq. (31) since $\langle 0|0\rangle = 1$.

The above Eq. (37') has recently been derived by Morales *et al.* using hypervirial type of results and parametric differentiation.[6] Also Eqs. (38) and (39) appear as Eqs. (3.21) and (3.22) in Ref. 1 and elsewhere.

## V. RECURRENCE RELATIONS FOR THE MATRIX ELEMEMTS ${}_G\langle m|x^k|n\rangle_E$

We take

$$f_E = x^{k+1} = (\alpha_E/\sqrt{2})(a_E+a_E^{\dagger})^{k+1} \tag{40}$$

which gives

$$\frac{\partial f_E}{\partial a_E} = \frac{\partial f_E}{\partial a_E^{\dagger}} = (k+1)\frac{\alpha_E}{\sqrt{2}}x^k. \tag{41}$$

Thus the right-hand recurrence relation Eq. (15) becomes

$$ {}_G\langle m|x^{k+1}|n\rangle_E$$
$$= -\left(\frac{2}{n}\right)^{1/2}\frac{\beta\gamma}{1+\beta^2}\,{}_G\langle m|x^{k+1}|n-1\rangle_E$$
$$-\left(\frac{n-1}{n}\right)^{1/2}\frac{1-\beta^2}{1+\beta^2}\,{}_G\langle m|x^{k+1}|n-2\rangle_E$$
$$+\left(\frac{m}{n}\right)^{1/2}\frac{2\beta}{1+\beta^2}\,{}_G\langle m-1|x^{k+1}|n-1\rangle_E$$
$$+(k+1)\alpha_E\frac{\beta^2}{1+\beta^2}\,{}_G\langle m|x^k|n-1\rangle_E$$
$$\text{for } m\geqslant 0, \quad n\geqslant 1. \tag{42}$$

If we wish to have a recurrence relation that involves the matrix elements of $x^k$ only (and not of $x^{k+1}$ and $x^k$ as in the above relation), we simply have to replace

$$x^{k+1} \text{ by } x^k\cdot x = (\alpha_E/\sqrt{2})x^k(a_E+a_E^{\dagger}) \tag{43}$$

and operate by $a_E+a_E^{\dagger}$ on the ket vectors. Then we find

$$
{}_G\langle m|x^k|n\rangle_E = -\left(\frac{2}{n}\right)^{1/2}\frac{\beta\gamma}{1+\beta^2}\,{}_G\langle m|x^k|n-1\rangle_E + \frac{\beta^2(2k+1)-2n+3}{(1+\beta^2)\sqrt{n(n-1)}}\,{}_G\langle m|x^k|n-2\rangle_E
$$
$$
-\left(\frac{2(n-2)}{n(n-1)}\right)^{1/2}\frac{\beta\gamma}{1+\beta^2}\,{}_G\langle m|x^k|n-3\rangle_E - \frac{1-\beta^2}{1+\beta^2}\left(\frac{(n-2)(n-3)}{n(n-1)}\right)^{1/2}\,{}_G\langle m|x^k|n-4\rangle_E
$$
$$
+2\left(\frac{m}{n}\right)^{1/2}\frac{\beta}{1+\beta^2}\,{}_G\langle m-1|x^k|n-1\rangle_E + 2\left(\frac{m(n-2)}{n(n-1)}\right)^{1/2}\frac{\beta}{1+\beta^2}\,{}_G\langle m-1|x^k|n-3\rangle_E . \tag{44}
$$

This right-hand recurrence relation is valid for $m\geqslant 0$, $n\geqslant 2$, whereas the one in Eq. (42) is valid for $m\geqslant 0$, $n\geqslant 1$. Both can be used to increase $n$. Note that Eq. (44) requires two generating matrix elements, namely, ${}_G\langle 0|x^k|0\rangle_E$, ${}_G\langle 0|x^k|1\rangle_E$. Similar manipulations on the bra vectors and simplifications result in the left-hand recurrence relation from Eq. (16),

$$
{}_G\langle m|x^k|n\rangle_E = \left(\frac{2}{m}\right)^{1/2}\frac{\gamma(1+2\beta^2)}{1+\beta^2}\,{}_G\langle m-1|x^k|n\rangle_E
$$
$$
-\frac{1}{\sqrt{m(m-1)}}\,\frac{(2k+1)-\beta^2(2m-3+2\gamma^2)}{1+\beta^2}\,{}_G\langle m-2|x^k|n\rangle_E
$$
$$
+\left(\frac{2(m-2)}{m(m-1)}\right)^{1/2}\frac{\gamma(2\beta^2-1)}{1+\beta^2}\,{}_G\langle m-3|x^k|n\rangle_E + \left(\frac{(m-2)(m-3)}{m(m-1)}\right)^{1/2}\frac{1-\beta^2}{1+\beta^2}\,{}_G\langle m-4|x^k|n\rangle_E
$$
$$
+2\left(\frac{n}{m}\right)^{1/2}\frac{\beta}{1+\beta^2}\,{}_G\langle m-1|x^k|n-1\rangle_E - 2\left(\frac{2n}{m(m-1)}\right)^{1/2}\frac{\beta\gamma}{1+\beta^2}\,{}_G\langle m-2|x^k|n-1\rangle_E
$$
$$
+2\left(\frac{n(m-2)}{m(m-1)}\right)^{1/2}\frac{\beta}{1+\beta^2}\,{}_G\langle m-3|x^k|n-1\rangle_E , \tag{45}
$$

which is valid for $m\geqslant 2$, $n\geqslant 0$.

We define a generating function

$$
X(k;y,z) = \sum_{m,n=0}^{\infty}\frac{{}_G\langle m|x^k|n\rangle_E}{\sqrt{m!n!}}\,y^m z^n . \tag{46}
$$

From the definition of $E(\rho;y,z)$ in Eqs. (20) and (21), $X(k;y,z) = (-1)k!\times$coefficient of $\rho^k$ in the generating function $E(\rho;y,z)$, which is given in Eq. (35) explicitly. The manipulations required to arrive at the answer should by now be familiar. In order to express the answer in terms of a Laguerre polynomial, we have to distinguish the two cases $k =$ even and $k =$ odd. We shall also require the use of the duplication formula[7]

$$
\Gamma(2z) = (2^{2z-1}/\sqrt{\pi})\Gamma(z)\Gamma(z+1) \tag{47}
$$

for the gamma function. We finally obtain

$$
X(k;y,z) = \left(\frac{2\beta}{1+\beta^2}\right)^{1/2}\left(\frac{k}{2}\right)!\left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)k} L_{(1/2)k}^{-1/2}\left(-\frac{(y+\beta z-\gamma/\sqrt{2})^2}{1+\beta^2}\right)
$$
$$
\times\exp\left[\frac{(1-\beta^2)(y^2-z^2)+4\beta yz+2\sqrt{2}\beta\gamma(\beta y-z)-\beta^2\gamma^2}{2(1+\beta^2)}\right], \tag{48}
$$

for $k$ even, and

$$
X(k;y,z)
$$
$$
= -\frac{2\beta}{1+\beta^2}\left(\frac{k-1}{2}\right)!\frac{\alpha_E\beta}{1+\beta^2}\left(\frac{2\alpha_E^2\rho^2}{1+\beta^2}\right)^{(1/2)(k-1)}(\gamma-\sqrt{2}(y+\beta z))\,L_{(1/2)(k-1)}^{1/2}\left(-\frac{(y+\beta z-(1/\sqrt{2})\gamma)^2}{1+\beta^2}\right)
$$
$$
\times\exp\left[\frac{(1-\beta^2)(y^2-z^2)+4\beta yz+2\sqrt{2}\beta\gamma(\beta y-z)-\beta^2\gamma^2}{2(1+\beta^2)}\right], \tag{49}
$$

for $k$ odd.

Since the recurrence relations in Eqs. (44) and (45) are valid for $m\geqslant 0$, $n\geqslant 2$ and $m\geqslant 2$, $n\geqslant 0$, respectively, to use these, we require the matrix elements ${}_G\langle 0|x^k|0\rangle_E$, ${}_G\langle 0|x^k|1\rangle_E$, and ${}_G\langle 1|x^k|0\rangle_E$ both for $k =$ even and odd integers. The matrix elements ${}_G\langle 0|x^k|0\rangle_E$ can be easily obtained from the function $E(\rho) = {}_G\langle 0|e^{-\rho x}|0\rangle_E$ given explicitly in Eq. (31) by taking $(-1)^k k!$ times the coefficient of $\rho^k$ in $E(\beta)$. Thus we obtain

$$
{}_G\langle 0|x^k|0\rangle_E = {}_G\langle 0|0\rangle_E\left(\frac{k}{2}\right)!\left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)k}
$$
$$
\times L_{(1/2)k}^{-1/2}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right), \tag{50}
$$

for $k$ even, and

$$
\begin{aligned}
&{}_G\langle 0|x^k|0\rangle_E \\
&\quad = -{}_G\langle 0|0\rangle_E \left(\frac{k-1}{2}\right)! \left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)(k-1)} \\
&\qquad \times \frac{\alpha_E\beta\gamma}{1+\beta^2} L^{1/2}_{(1/2)(k-1)}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right), \qquad (51)
\end{aligned}
$$

for $k$ add.

Again using recurrence relations of the type given in Eq. (42) taking $m=0$, $n=1$ or *directly from*

$$
\begin{aligned}
{}_G\langle 0|x^{k+1}|0\rangle_E &= (\alpha_E/\sqrt{2})\ {}_G\langle 0|x^k(a_E+a_E^\dagger)|0\rangle_E \\
&= (\alpha_E/\sqrt{2}){}_G\langle 0|x^k|1\rangle_E
\end{aligned}
$$

i.e.,

$$
{}_G\langle 0|x^k|1\rangle_E = (\sqrt{2}/\alpha_E)\langle 0|x^{k+1}|0\rangle \qquad (52)
$$

and we arrive at, using the expressions in Eqs. (50) and (51),

$$
\begin{aligned}
{}_G\langle 0|x^k|1\rangle_E &= -{}_G\langle 0|0\rangle_E \left(\frac{k}{2}\right)! \left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)k} \\
&\quad \times \frac{\sqrt{2}\beta\gamma}{1+\beta^2} L^{1/2}_{(1/2)k}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right), \qquad (53)
\end{aligned}
$$

for $k$ even, and

$$
\begin{aligned}
{}_G\langle 0|x^k|1\rangle_E &= {}_G\langle 0|0\rangle_E \left(\frac{k+1}{2}\right)! \left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)(k+1)} \\
&\quad \times \frac{\sqrt{2}}{\alpha_E} L^{-1/2}_{(1/2)(k+1)}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right), \qquad (54)
\end{aligned}
$$

for $k$ odd.

Similarly we have

$$
\begin{aligned}
&{}_G\langle 1|x^k|0\rangle_E \\
&\quad = -{}_G\langle 0|0\rangle_E \left(\frac{k}{2}\right)! \left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)k} \\
&\qquad \times\sqrt{2}\gamma\left[\frac{1}{1+\beta^2} L^{1/2}_{(1/2)k}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right)\right. \\
&\qquad \left. - L^{-1/2}_{(1/2)k}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right)\right], \qquad (55)
\end{aligned}
$$

for $k$ even, and

$$
\begin{aligned}
&{}_G\langle 1|x^k|0\rangle_E \\
&\quad = {}_G\langle 0|0\rangle_E \left(\frac{2\alpha_E^2\beta^2}{1+\beta^2}\right)^{(1/2)(k-1)} \left(\frac{k-1}{2}\right)! \frac{\sqrt{2}\alpha_E\beta}{1+\beta^2} \\
&\qquad \times\left[(k+1)L^{-1/2}_{(1/2)(k+1)}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right)\right. \\
&\qquad \left. - \gamma^2 L^{1/2}_{(1/2)(k-1)}\left(-\frac{\gamma^2}{2(1+\beta^2)}\right)\right], \qquad (56)
\end{aligned}
$$

for $k$ odd.

## VI. RECURRENCE RELATIONS AND THEIR SOLUTIONS FOR THE GAUSSIAN MATRIX ELEMENTS ${}_G\langle m|e^{-\rho x^2}|n\rangle_E$

The recurrence relations for the Gaussian matrix elements ${}_G\langle m|e^{-\rho x^2}|n\rangle_E$ can be obtained as for the exponential matrix elements from the general recurrence relations by using

$$
f_E = e^{-\rho x^2}
$$

for which

$$
\frac{\partial f_E}{\partial a_E} = \frac{\partial f_E}{\partial a_E^\dagger} = -\alpha_E^2\rho\, e^{-\rho x^2}(a_E + a_E^\dagger). \qquad (57)
$$

Thus the right-hand recurrence relation takes the form

$$
\begin{aligned}
&{}_G\langle m|e^{-\rho x^2}|n\rangle_E \\
&\quad = -\left(\frac{2}{n}\right)^{1/2}\frac{\beta\gamma}{\eta}\,{}_G\langle m|e^{-\rho x^2}|n-1\rangle_E \\
&\qquad -\left(\frac{n-1}{n}\right)^{1/2}\frac{\eta-2\beta^2}{\eta}\,{}_G\langle m|e^{-\rho x^2}|n-2\rangle_E \\
&\qquad +\left(\frac{m}{n}\right)^{1/2}\frac{2\beta}{\eta}\,{}_G\langle m-1|e^{-\rho x^2}|n-1\rangle_E, \qquad (58)
\end{aligned}
$$

which is valid for $m\geqslant 0$, $n\geqslant 1$, and where $\eta = 1+\beta^2 + 2\alpha_E^2\beta^2\rho$. For the left-hand recurrence relations, we have similarly

$$
\frac{\partial f_E}{\partial a_E^\dagger} = -\alpha_E^2\beta\rho\,(-\sqrt{2}\gamma + a_G + a_G^\dagger)e^{-\rho x^2}, \qquad (59)
$$

which results in

$$
\begin{aligned}
&{}_G\langle m|e^{-\rho x^2}|n\rangle_E \\
&\quad = \left(\frac{2}{m}\right)^{1/2}\frac{\beta^2\gamma(1+2\alpha_E^2\rho)}{\eta}\,{}_G\langle m-1|e^{-\rho x^2}|n\rangle_E \\
&\qquad -\left(\frac{m-1}{m}\right)^{1/2}\frac{\eta-2}{\eta}\,{}_G\langle m-2|e^{-\rho x^2}|n\rangle_E \\
&\qquad +\left(\frac{n}{m}\right)^{1/2}\frac{2\beta}{\eta}\,{}_G\langle m-1|e^{-\rho x^2}|n-1\rangle_E, \qquad (60)
\end{aligned}
$$

which is valid for $m\geqslant 1$, $n\geqslant 0$.

To obtain the generating function, we define

$$
G(\rho;x,y) = \sum_{m,n=0}^{\infty} \frac{{}_G\langle m|e^{-\rho x^2}|n\rangle_E}{\sqrt{m!n!}}\, y^m z^n. \qquad (61)
$$

Then the right-hand recurrence relation Eq. (58) leads to the differential equation

$$
\begin{aligned}
&\frac{\partial}{\partial z} G(\rho;y,z) \\
&\quad = \left[-\frac{\sqrt{2}\beta\gamma}{\eta} - \frac{\eta-2\beta^2}{\eta} z + \frac{2\beta}{\eta} y\right] G(\rho;y,z) \qquad (62)
\end{aligned}
$$

with the solution

$$
\begin{aligned}
G(\rho;y,z) &= G(\rho;y,0) \\
&\quad \times\exp\left[-\frac{\sqrt{2}\beta\gamma}{\eta} z - \frac{\eta-2\beta^2}{2\eta} z^2 + \frac{2\beta}{\eta} yz\right], \qquad (63)
\end{aligned}
$$

where

$$G(\rho; y,0) = \sum_{m=0}^{\infty} \frac{{}_G\langle m|e^{-\rho x^2}|0\rangle_E}{\sqrt{m!}} y^n \tag{64}$$

and a differential equation for it is obtained from the left-hand recurrence relation Eq. (60) by taking $n = 0$. This differential equation is

$$\frac{\partial}{\partial y} G(\rho; y,0) = \left( \frac{\sqrt{2}\beta^2\gamma(1+2\alpha_E^2\rho)}{\eta} - \frac{\eta-2}{\eta} y \right) G(\rho; y,0), \tag{65}$$

which gives

$$G(\rho; y,0) = G(\rho;0,0) \times \exp\left[ \frac{\sqrt{2}\beta^2\gamma(1+2\alpha_E^2\rho)}{\eta} y - \frac{\eta-2}{2\eta} y^2 \right], \tag{66}$$

where

$$G(\rho;0,0) = {}_G\langle 0|e^{-\rho x^2}|0\rangle_E \tag{67}$$

is the generating integral for these Gaussian matrix elements. To compute it, we call it $G(\rho)$ and we have

$$\frac{dG(\rho)}{d\rho} = -\frac{\alpha_E^2}{2} {}_G\langle 0|e^{-\rho x^2}(a_E + a_E^\dagger)^2|0\rangle_E = -\frac{\alpha_E^2}{2} {}_G\langle 0|e^{-\rho x^2}(a_E a_E^\dagger + a_E^{\dagger 2})|0\rangle_E,$$

since

$$a_E|0\rangle_E = 0 = -(\alpha_E^2/\sqrt{2})\left[ {}_G\langle 0|e^{-\rho x^2}|0\rangle_E + \sqrt{2}\,{}_G\langle 0|e^{-\rho x^2}|2\rangle_E \right]. \tag{68}$$

Now we use repeatedly the right-hand recurrence relation in Eq. (58) to evaluate ${}_G\langle 0|e^{-\rho x^2}|2\rangle_E$ in terms of ${}_G\langle 0|e^{-\rho x^2}|0\rangle_E$. This finally gives the differential equation

$$\frac{dG(\rho)}{d\rho} = -\alpha_E^2\beta^2\left[ \frac{1}{\eta} + \frac{\gamma^2}{\eta^2} \right] G(\rho), \tag{69}$$

with the solution

$$G(\rho) = (C/\eta^{1/2}) e^{(1/2)(\gamma^2/\eta)} \tag{70}$$

on noting that $\eta = 1 + \beta^2 + 2\alpha_E^2\beta^2\rho$. To obtain $C$, put $\rho = 0$, which gives

$$C = \sqrt{1+\beta^2}\, e^{-(1/2)\gamma^2/1+\beta^2}\, {}_G\langle 0|0\rangle_E. \tag{71}$$

Thus

$$G(\rho) = {}_G\langle 0|e^{-\rho x^2}|0\rangle_E = \left( \frac{1+\beta^2}{\eta} \right)^{1/2} \exp\left[ -\frac{\alpha_E^2\beta^2\gamma^2\rho}{(1+\beta^2)\eta} \right] {}_G\langle 0|0\rangle_E \tag{72}$$

$$= \left( \frac{2\beta}{\eta} \right)^{1/2} \exp\left[ -\frac{\beta^2\gamma^2(1+2\alpha_E^2\rho)}{2\eta} \right] \tag{73}$$

on substituting for ${}_G\langle 0|0\rangle_E$ from Eq. (34). Combining Eqs. (63), (66), and (73) we finally arrive at

$$G(\rho; y,z) = \left( \frac{2\beta}{\eta} \right)^{1/2} \times \left[ \exp\left( -\frac{(\eta-2)y^2 + (\eta-2\beta^2)z^2 - 4\beta yz - 2\sqrt{2}\beta^2\gamma(1+2\alpha_E^2\rho)y + 2\sqrt{2}\beta\gamma z + \beta^2\gamma^2(1+2\alpha_E^2\rho)}{2\eta} \right) \right]. \tag{74}$$

Finally on computing the coefficient of $y^m z^n$ in the above, we find

$$\begin{aligned} {}_G\langle m|e^{-\rho x^2}|n\rangle_E = {} & \left( m!n!\,\frac{2\beta}{\eta} \right)^{1/2} \exp\left[ -\frac{\beta^2\gamma^2(1+2\alpha_E^2\rho)}{2\eta} \right] \\ & \times \sum_{\gamma_1,\gamma_2} (-1)^{n-m+\gamma_1+\gamma_2} \frac{(\eta-2)^{\gamma_1}(\eta-2\beta^2)^{\gamma_2}\beta^{\eta-2\gamma_2}\gamma^{n-m+2\gamma_1-2\gamma_2}}{\eta^{n+\gamma_1-\gamma_2}\gamma_1!\gamma_2!(n-2\gamma_2)!} \\ & \times 2^{(n+m)/2-2\gamma_1-2\gamma_2} L_{m-2\gamma_1}^{n-m+2\gamma_1-2\gamma_2}\left( \frac{\beta^2\gamma^2(1+2\alpha_E^2\rho)}{\eta} \right), \end{aligned} \tag{75}$$

which for the single harmonic oscillator becomes

$$\langle m|e^{-\rho x^2}|n\rangle = \frac{\sqrt{n!m!}}{(1+\alpha^2\rho)^{(1/2)(n+m+1)}} \left( -\frac{1}{2}\alpha^2\rho \right)^{(n-m)/2} \sum_\gamma \frac{(\alpha^2\rho/2)^\gamma}{\gamma!((n-m)/2-\gamma)!(m-2\gamma)!} \tag{76}$$

when $n - m$ = even and is 0 otherwise.

An expression equivalent to the one in Eq. (76) was obtained by Morales *et al.* in Ref. 3 and appears as Eq. (3.10) in that reference. The equivalence requires a quadratic transformation of the hypergeometric functions.


[1] J. Morales, L. Sandoval, and A. Palma, J. Math. Phys. **27**, 2966 (1986).
[2] J. Morales, A. Palma, and M. Berrondo, Int. J. Chem. **S18**, 57 (1985).
[3] J. Morales, J. Zopez-Bonilla, and A. Palma, J. Math. Phys. **28**, 1032 (1987).

[4] Note that the definition of Laguerre polynomials is not standardized. We are using the one in Ref. 5.

[5] I. M. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).

[6] See Eq. (2.15) in Ref. 3.
[7] See Eq. (8.335(1)) in Ref. 5.

# A Rodrigues formula approach to determining closed-form solutions to the Schrödinger equation for symmetric anharmonic oscillators

P. G. L. Leach
*Department of Computational and Applied Mathematics and Centre for Nonlinear Studies, The University of the Witwatersrand, Wits Post Office, 2050, Johannesburg, Republic of South Africa*

G. P. Flessas
*Division of Applied Mathematics and Mechanics, Department of Engineering Science, University of Patras, GR 261 10 Patras, Greece*

V. M. Gorringe
*Department of Computational and Applied Mathematics and Centre for Nonlinear Studies, The University of the Witwatersrand, Wits Post Office, 2050, Johannesburg, Republic of South Africa*

(Received 1 June 1988; accepted for publication 21 September 1988)

A natural generalization to the Rodrigues formula for the Hermite polynomials leads to the definition of sets of polynomials. Under certain constraints they are used to explain the existence of closed-form solutions to the Schrödinger equation for symmetric anharmonic potential wells and to obtain further closed-form solutions.

## I. INTRODUCTION

The anharmonic oscillator has been of interest in quantum mechanics for many years. In particular, in the study of molecular spectra of simple models such as ammonia and hydrogen-bonded solids, the potential is modeled as a double-minimum well.[1] Generally it is not possible to solve the Schrödinger equation exactly and resort must be made to numerical approaches that have attracted a vast literature.[2]

In a series of papers[3] exact (in the sense of closed-form) solutions for certain anharmonic oscillators were found by *ad hoc* methods. These *ad hoc* methods have been systematized for various anharmonic systems in both one[4] and more than one dimension.[5] These exact solutions exhibit two features. The first is that only a limited number (often only one) of wave functions is obtained and the second is that these exist only for certain potentials within a class of potentials. For example,[4(a)] the Schrödinger equation

$$\psi''(x) + (\lambda - \omega(x))\psi(x) = 0, \tag{1.1}$$

with $x \in (-\infty, \infty)$ and $\psi(\pm\infty) = 0$, where

$$w(x) = (ax^2 + b)^2 x^2 - kax^2,$$

has exact solutions only if

$$k = 2(2M + \epsilon) + 3,$$

where $M$ is an integer and $\epsilon = 0$ or 1. The wave functions obtained are for the first $M + 1$ even ($\epsilon = 0$) or odd ($\epsilon = 1$) states.

Although only a limited number of wave functions and energy levels are obtained, they do provide a method for checking the accuracy and suitability of numerical algorithms and this must be regarded as the sole practical reason for this investigation. However, the *ad hoc* method does not explain how it is that these exact solutions arise. It is the purpose of this paper to provide an explanation of how they arise. This explanation is based on the well-accepted physical concept of the annihilation operator. The vehicle used is the symmetric anharmonic potential of highest degree $4n + 2$, $n$ an integer, as this, in the case $n = 1$, has been seen[4(a)] to provide the richest results.

## II. ANNIHILATION OPERATORS

In the case of the harmonic oscillator with Schrödinger equation

$$\psi'' + (\lambda - x^2)\psi = 0, \tag{2.1}$$

it is well known that the operators $a = 2^{-1/2}(d/dx + x)$ and $a^\dagger = 2^{-1/2}(-d/dx + x)$ act as annihilation and creation operators on the wave function, i.e.,

$$a\psi_n = \sqrt{n}\,\psi_{n-1}, \quad a^\dagger\psi_n = \sqrt{n+1}\,\psi_{n+1}. \tag{2.2}$$

In particular, the ground state is found from $a\psi_0 = 0$. In terms of coordinates this is

$$2^{-1/2}\left(\frac{d}{dx} + x\right)\psi_0 = 0 \tag{2.3}$$

so that

$$\psi_0(x) = N_0 e^{-(1/2)x^2}, \tag{2.4}$$

where $N_0$ is the normalization constant. It is instructive to rewrite (2.3) as

$$\left(\frac{1}{x}\frac{d}{dx} + 1\right)\psi_0 = 0 \Leftrightarrow \left(\frac{d}{d(x^2)} + \frac{1}{2}\right)\psi_0 = 0, \tag{2.5}$$

or, on setting $x^2 = u$,

$$\left(\frac{d}{du} + \frac{1}{2}\right)\psi_0 = 0 \tag{2.6}$$

and for a general harmonic oscillator potential (2.6) would be written as

$$\left(\frac{d}{du} + \frac{1}{2}\omega\right)\psi_0 = 0. \tag{2.7}$$

One can recognize that (2.7) is a special case of the more general definition of an annihilation operator

 0022-2488/89/020406-07$02.50 © 1989 American Institute of Physics 

$$\left(\frac{d}{du}+v(u)\right)\psi_0=0, \tag{2.8}$$

which yields the ground state wave function

$$\psi_0(u)=N_0\exp\left\{-\int^u v(s)ds\right\}. \tag{2.9}$$

For (2.9) to be physically meaningful the argument of the exponential would have to be such that

$$\int_0^\infty \exp\left\{-2\,\mathrm{Re}\int^u v(s)ds\right\}du<\infty. \tag{2.10}$$

Also, for $\psi_0$ to be the ground state $v(u)$ cannot be of the form such that

$$\exp\left\{-\int^u v(s)ds\right\}=p(u)\exp\{-z(u)\}, \tag{2.11}$$

where $p(u)$ has zeros in $u\in(0,\infty)$.

In the notation of (2.8), the creation operator would have the form $-d/du+v(u)$, but this would be an incorrect form for the harmonic oscillator as it stands and would just regenerate $\psi_0(u)$. However, for other $v(u)$, different functions are generated and it is this factor that we exploit in searching for closed-form expressions for the wave functions of polynomial anharmonic oscillators. To make our considerations more precise, in the next section we consider the even eigenstates for potentials associated with the ground state wave function

$$\psi_0(x)=N_0\exp\left[-\tfrac{1}{4}ax^4-\tfrac{1}{2}bx^2\right]. \tag{2.12}$$

## III. EVEN EIGENSTATES FOR $w(x)=(ax^2+b)^2x^2-kax^2$

The potential associated with $\psi_0(x)$ as given by (2.12) is

$$V(x)=\tfrac{1}{2}\{(ax^2+b)^2x^2-3ax^2\}.$$

In this section we take $w(x)$ to be given by

$$w(x)=(ax^2+b)^2x^2-kax^2. \tag{3.1}$$

Analogous to the definition of Hermite polynomials, we define a set of polynomials $f_n(x)$ by means of the Rodrigues-like formula

$$f_n(x)=\frac{1}{2^n n!}e^{(1/2)ax^4+bx^2}\left(-\frac{1}{x}\frac{d}{dx}\right)^n e^{-(1/2)ax^4-bx^2}. \tag{3.2}$$

These polynomials span the set of even functions over $(-\infty,\infty)$, but in general are not orthogonal with respect to the weight function $\exp\left[-\tfrac{1}{2}ax^4-bx^2\right]$. However, they have the following recurrence relations:

$$(n+1)f_{n+1}=(ax^2+b)f_n-af_{n-1}, \tag{3.3}$$

$$f_n'=2x(ax^2+b)f_n-2x(n+1)f_{n+1}=2axf_{n-1}, \tag{3.4}$$

and have the general form

$$f_n(x)=\sum_{j=0}^{[n/2]}\frac{(-a)^j(ax^2+b)^{n-2j}}{(n-2j)!\,j!\,2^j}, \tag{3.5}$$

where, as usual, $[n/2]$ means the integer part of $n/2$. The $f_n$ also satisfy the second-order differential-difference equation

$$f_n''-2(ax^2+b)xf_n'-4nax^2f_n=2af_{n-1}. \tag{3.6}$$

If we define a set of functions $\{y_n,\ n=0,1,\ldots\}$ such that

$$y_n=f_n\exp\left[-\tfrac{1}{4}ax^4-\tfrac{1}{2}b^2\right], \tag{3.7}$$

which, just as $\{f_n,\ n=0,1,\ldots\}$, constitutes a set of linearly independent functions, from (3.6) we find that

$$y_n''=[(ax^2+b)^2x^2-(4n+3)ax^2-b\,]y_n+2ay_{n-1}. \tag{3.8}$$

As our Schrödinger equation we take

$$\psi''+[\lambda-(ax^2+b)^2x^2+kax^2]\psi=0, \tag{3.9}$$

which, since the potential is even, has odd and even wave functions, and we assume an expansion for the even wave functions in terms of the $y_n$ functions, viz.,

$$\psi=\sum_{n=0}^\infty C_n y_n. \tag{3.10}$$

We substitute (3.10) into (3.9) to obtain

$$\sum_{n=0}^\infty C_n\{[\lambda-b+(k-4n-3)ax^2]y_n+2ay_{n-1}\}=0. \tag{3.11}$$

From the recurrence relation (3.3),

$$(ax^2+b)y_n=(n+1)y_{n+1}+ay_{n-1}.$$

We substitute this into (3.11) to obtain

$$\sum_{n=0}^\infty C_n\{(k-4n-3)(n+1)y_{n+1}+[\lambda+(4n+2-k)b\,]y_n+(k-4n-1)ay_{n-1}\}=0.$$

This leads to the recurrence relation for the coefficients $C_n$:

$$(k-4n+1)nC_{n-1}+[\lambda+(4n+2-k)b\,]C_n+(k-4n-5)aC_{n+1}=0, \tag{3.12}$$

when $n=0,1,\ldots$. Equation (3.12) may be written as an infinite tridiagonal matrix equation

$$\begin{bmatrix}
\lambda+(2-k)b & (k-5)a & & & & \\
(k-3).1 & \lambda+(6-k)b & (k-9)a & & & \\
 & (k-7).2 & \lambda+(10-k)b & (k-13)a & & \\
 & & (k-11).3 & \lambda+(14-k)b & (k-17)b & \\
 & & & \ddots & \ddots & \ddots
\end{bmatrix}\begin{bmatrix}C_0\\C_1\\C_2\\C_3\\\vdots\end{bmatrix}=0. \tag{3.13}$$

We observe that, if the lower diagonal term in the $(M+2)$th, i.e., $n=M+1$, row of this matrix is zero, the determinant of the infinite matrix factors into the product of the determinent of an $(M+1)\times(M+1)$ matrix and that

of the remaining terms and the eigenvector can be taken to be $(C_0,C_1,...,C_M,0,0,...)^T$. The infinite eigenvalue problem is reduced to a finite one. We obtain $(M+1)$ eigenvalues and $(M+1)$ eigenvectors. The value of $k$ for which this occurs is $k=4M+3$. For this value of $k$ the product of the $(i+1,j)$ and $(i,j+1)$ elements of the matrix is either zero or positive so that the matrix can be symmetrized. Hence, as one would expect, the eigenvalues are real. Furthermore the eigenvectors are orthogonal and so we obtain $(M+1)$ independent wave functions. This explains the result obtained by Leach[4(a)] that for this value of $k$ one obtains the first $(M+1)$ even eigenvalues and wave functions for the potential given in (3.1).

## IV. ODD EIGENSTATES FOR $w(x)=(ax^2+b)^2x^2-kax^2$

The polynomials $f_n$ defined by (3.2) cover the set of even functions on $(-\infty,\infty)$. It is not readily apparent how to define a covering set for the odd functions. However, the results of the previous section provide a clue. The value of $k$ for which exact results were obtained is positive and so the potential is a double well or a triple well for which there is a tendency for the lower eigenvalues to be grouped in pairs or triples (the latter only if the three minima are comparable). This indicates that the square of the wave function for an odd eigenstate will be similar to that of the square of that for an even state [see Ref. 2(f) for computed comparisons]. (Strictly speaking it should be mod squared, but we are dealing with real wave functions.)

We define a complete set of polynomials covering odd functions as

$$g_n = xf_n$$

$$= \frac{x}{2^n n!}\, e^{(1/2)ax^4+bx^2}\left(-\frac{1}{x}\frac{d}{dx}\right)^n e^{-(1/2)ax^4-bx^2}, \tag{4.1}$$

with the properties

$$(n+1)g_{n+1} = (ax^2+b)g_n - ag_{n-1}, \tag{4.2}$$

$$xg_n' = [2x^2(ax^2+b)+1]g_n - 2x^2(n+1)g_{n+1}$$

$$= g_n + 2ax^2 g_{n-1}, \tag{4.3}$$

and

$$g_n'' = 2x(ax^2+b)g_n' - [(4n+2)ax^2+2b]g_n + 6ag_{n-1}. \tag{4.4}$$

If we let

$$z_n = g_n \exp[-\tfrac{1}{4}ax^4 - \tfrac{1}{2}bx^2], \tag{4.5}$$

where the set of functions $\{z_n, n=0,1,...\}$ constitutes a linearly independent set,

$$z_n'' = [(ax^2+b)^2x^2 - (4n+5)ax^2 - 3b]z_n + 6az_{n-1}. \tag{4.6}$$

To determine the odd wave functions of the Schrödinger equation (1.1) with $w(x)$ as in (3.1), we make the expansion

$$\psi = \sum_{n=0}^{\infty} d_n z_n. \tag{4.7}$$

Proceeding in the same fashion as in Sec. III we find that the coefficients $d_n$ satisfy the general recurrence relation

$$(k-4n-1)nd_{n-1} + [\lambda + (4n+2-k)b]d_n + (k-4n-3)ad_{n+1} = 0, \tag{4.8}$$

which leads to the infinite matrix equation

$$\begin{bmatrix} \lambda+(2-k)b & (k-3)a & & & & \\ (k-5).1 & \lambda+(6-k)b & (k-7)a & & & \\ & (k-9).2 & \lambda+(10-k)b & (k-11)a & & \\ & & (k-13).3 & \lambda+(14-k)b & (k-15)a & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ d_3 \\ \vdots \end{bmatrix} = 0. \tag{4.9}$$

As in the case of the even wave functions, the determinant of this infinite matrix may be factored into the product of two determinants if $k=4M+5$. As in the even case we obtain $M+1$ eigenvalues and $M+1$ orthogonal eigenvectors of the form $(d_0,d_1,...,d_M,0,0,...)^T$. If we rewrite $k$ as $k=2(2M+1)+3$, we see that we have recovered the odd case results obtained by Leach.[4(a)]

## V. THE GENERAL CASE FOR SYMMETRIC ANHARMONIC POTENTIALS

The function

$$\psi_0(x) = N_0 \exp\left[-\sum_{i=1}^{M}\frac{a_i}{2i}x^{2i}\right], \quad a_M>0, \tag{5.1}$$

is the ground state solution of the Schrödinger equation

$$\psi'' - \left\{\left(\sum_{i=0}^{M-1} a_{i+1}x^{2i}\right)^2 x^2 - \sum_{i=0}^{M-1}(2i+1)a_{i+1}x^{2i}\right\}\psi = 0, \tag{5.2}$$

i.e., the ground state eigenvalue is $a_1$ and the scaled potential is

$$w(x) = \left(\sum_{i=0}^{M-1} a_{i+1}x^{2i}\right)^2 x^2 - \sum_{i=1}^{M-1}(2i+1)a_{i+1}x^{2i}. \tag{5.3}$$

As in Sec. III we define a set of polynomials $f_n(x)$ according to

$$f_n(x) = \frac{1}{2^n n!}\exp\left[\sum_{i=1}^{M}\frac{a_i}{i}x^{2i}\right] \times\left[-\frac{1}{x}\frac{d}{dx}\right]^n \exp\left[-\sum_{i=1}^{M}\frac{a_i}{i}x^{2i}\right]. \tag{5.4}$$

For $M \geqslant 3$ these polynomials do not cover all even functions. However, they do have the recurrence relations

$$(n+1)f_{n+1} = \sum_{j=0}^{M-1} (-1)^j \sum_{i=j}^{M-1} \binom{i}{j} a_{i+1} x^{2i-2j} f_{n-j}, \tag{5.5}$$

$$f_n' = 2x \sum_{i=0}^{M-1} a_{i+1} x^{2i} f_n - 2x(n+1) f_{n+1}$$

$$= 2x \sum_{j=1}^{M-1} (-1)^{j+1} \sum_{i=j}^{M-1} \binom{i}{j} a_{i+1} x^{2i-2j} f_{n-j}. \tag{5.6}$$

Let

$$y_n = f_n \exp\left[ -\sum_{i=1}^{M} \frac{a_i}{2i} x^{2i} \right]. \tag{5.7}$$

Then $y_n$ satisfies the nonhomogeneous second-order differential equation

$$y_n'' = \left\{ \left( \sum_{i=0}^{M-1} a_{i+1} x^{2i} \right) x^2 - a_1 - \sum_{i=1}^{M-1} [(4n+2)i+1] a_{i+1} x^{2i} \right\} y_n$$
$$- 2 \sum_{j=1}^{M-1} \sum_{i=j}^{M-1} (-1)^j \binom{i}{j} a_{i+1} x^{2i-2j} y_{n-j}$$
$$- 4(n+1) \sum_{j=1}^{M-2} \sum_{i=j+1}^{M-1} \times (-1)^j \binom{i}{j+1} a_{i+1} x^{2i-2j} y_{n-j}. \tag{5.8}$$

The Schrödinger equation to be considered is

$$\psi'' + \left\{ \lambda - \left( \sum_{i=0}^{M-1} a_{i+1} x^{2i} \right)^2 x^2 + \sum_{i=1}^{M-1} k_i x^{2i} \right\} \psi = 0. \tag{5.9}$$

We expand $\psi$ as

$$\psi = \sum_{n=0}^{\infty} C_n y_n \tag{5.10}$$

and substitute this and (5.8) into (5.9) to obtain

$$\sum_{n=0}^{\infty} C_n \left\{ \left[ \lambda - a_1 - \sum_{i=1}^{M-1} [(4n+2)i+1] a_{i+1} - k_i ] x^{2i} \right] y_n \right.$$
$$- 2 \sum_{j=1}^{M-1} \sum_{i=j}^{M-1} (-1)^j \binom{i}{j} a_{i+1} x^{2i-2j} y_{n-j} - 4(n+1)$$
$$\left. \times \sum_{j=1}^{M-2} \sum_{i=j+1}^{M-1} (-1)^j \binom{i}{j+1} a_{i+1} x^{2i-2j} y_{n-j} \right\} = 0. \tag{5.11}$$

From the recurrence relation (5.5) we may write

$$a_M x^{2M-2} y_n = (n+1) y_{n+1} - \sum_{i=0}^{M-2} a_{i+1} x^{2i} y_n$$
$$- \sum_{j=1}^{M-1} \sum_{i=j}^{M-1} (-1)^j \binom{i}{j} a_{i+1} x^{2i-2j} y_{n-j}. \tag{5.12}$$

When (5.12) is substituted into (5.11), we have

$$\sum_{n=0}^{\infty} C_n \left[ \left\{ \lambda - a_1 - \sum_{i=1}^{M-2} [((4n+2)i+1) a_{i+1} - k_i] x^{2i} \right\} y_n \right.$$
$$- \left[ (4n+2)(M-1) + 1 - \frac{k_{M-1}}{a_M} \right] \left\{ (n+1) y_{n+1} \right.$$
$$- \sum_{i=0}^{M-2} a_{i+1} x^{2i} y_n$$
$$\left. - \sum_{j=1}^{M-1} \sum_{i=j}^{M-1} (-1)^j \binom{i}{j} a_{i+1} x^{2i-2j} y_{n-j} \right\}$$
$$- 2 \sum_{j=1}^{M-1} (-1)^j \sum_{i=j}^{M-1} \binom{i}{j} a_{i+1} x^{2i-2j} y_{n-j} - 4(n+1)$$
$$\left. \times \sum_{j=1}^{M-2} \sum_{i=j+1}^{M-1} (-1)^j \binom{i}{j+1} a_{i+1} x^{2i-2j} y_{n-j} \right] = 0. \tag{5.13}$$

The "recurrence relation" (the reason for the " " is apparent immediately below) for the coefficients $C_n$ is

$$- \left[ (4n-2)(M-1) + 1 - \frac{k_{M-1}}{a_M} \right] n C_{n-1}$$
$$+ \left[ \lambda + \left[ (4n+2)(M-1) - \frac{k_{M-1}}{a_M} \right] a_1 \right.$$
$$+ \sum_{i=1}^{M-2} \left\{ \left[ (4n+2)(M-1-i) - \frac{k_{M-1}}{a_M} \right] a_{i+1} \right.$$
$$\left. \left. - k_i \right\} x^{2i} \right] C_n$$
$$+ \sum_{j=1}^{M-1} \sum_{i=j}^{M-1} (-1)^j \left\{ (M-1)[4(n+j)+2] \right.$$
$$\left. - 1 - \frac{k_{M-1}}{a_M} \right\} \binom{i}{j} a_{i+1} x^{2i-2j} C_{n+j}$$
$$- 4 \sum_{j=1}^{M-2} \sum_{i=j+1}^{M-1} (-1)^j (n+j+1) \binom{i}{j+1}$$
$$\times a_{i+1} x^{2i-2j} C_{n+j} = 0. \tag{5.14}$$

In contrast to the case $M = 2$, when $M \geqslant 3$, it is not possible to remove all of the powers of $x$ using the recurrence relation (5.5). We must further separate the terms in (5.14) by coefficients of different powers of $x$ to obtain a set of recurrence relations which must all be satisfied simultaneously. They are

$$- \left[ (4n-2)(M-1) + 1 - \frac{k_{M-1}}{a_M} \right] n C_{n-1}$$
$$+ \left\{ \lambda + \left[ (4n+2)(M-1) - \frac{k_{M-1}}{a_M} \right] a_1 \right\} C_n$$
$$+ \sum_{j=1}^{M-1} \left\{ [4(n+j)+2](M-1) - 1 - \frac{k_{M-1}}{a_M} \right\}$$
$$\times (-1)^j a_{j+1} C_{n+j} = 0 \tag{5.15}$$

from the coefficient of $x^0$ and

$$\left\{\left[(4n+2)(M-1-i)-\frac{k_{M-1}}{a_M}\right]a_{i+1}-k_i\right\}C_n$$
$$+\sum_{j=1}^{M-1-i}(-1)^j\left[\left\{[4(n+j)+2](M-1)\right.\right.$$
$$\left.-1-\frac{k_{M-1}}{a_M}\right\}\binom{i+j}{j}$$
$$\left.-4(n+j+1)\binom{i+j}{j+1}\right]a_{i+j+1}C_{n+j}=0, \qquad (5.16)$$

where $i=1,2,\ldots,M-2$, from the other terms.

The change from the case $M=2$ to $M\geqslant 3$ is quite clear. For $M=2$ there was just one eigenvalue problem to be solved. For $M\geqslant 3$ there are $M-2$ sets of linear equations for the coefficients. For $M>3$ we would not expect the solution of one set to be the solution of the other set(s). Indeed, for general values of the coefficients $k$ and $a$ we would not expect any nontrivial solution. However, in the spirit of finding closed-form solutions we could set all but a number of the expansion coefficients equal to zero. For Eqs. (5.16) to yield a nontrivial solution constraints must be imposed on the coefficients $k_i$ and $a_i$. Assuming that Eqs. (5.16) can be satisfied, Eqs. (5.15) provide a further set of constraints since the coefficients are now no longer independent. Let us consider the case $M=3$ for which Eqs. (5.15) and (5.16) are

$$-\left(8n-3-\frac{k_2}{a_3}\right)nC_{n-1}+\left[\lambda+\left(8n+4-\frac{k_2}{a_3}\right)a_1\right]C_n$$
$$+\sum_{j=1}^{2}\left[8(n+j)+3-\frac{k_2}{a_3}\right](-1)^j a_{j+1}C_{n+j}=0, \qquad (5.17)$$

$$\left[\left(4n+2-\frac{k_2}{a_3}\right)a_2-k_1\right]C_n$$
$$-\left(12n+14-\frac{2k_2}{a_3}\right)a_3C_{n+1}=0. \qquad (5.18)$$

If the expansion is to terminate with $C_N$, setting $n=N+1$ in (5.17) and then $n=N$ in (5.18) we find that

$$k_2=(8N+5)a_3, \quad k_1=-(4N+3)a_2,$$

and (5.17) and (5.18) now become

$$8(N+1-n)nC_{n-1}+[\lambda-(8(N-n)+1)a_1]C_n$$
$$+[8(N-n)-6]a_2C_{n+1}-[8(N-n)-14]$$
$$\times a_3C_{n+2}=0, \qquad (5.17')$$
$$(N-n)a_2C_n=(4N-3n-1)a_3C_{n+1}. \qquad (5.18')$$

From (5.18') it is evident that

$$C_{N-j}=\left[\prod_{k=1}^{j}(N+3k-1)\right]\left(\frac{a_3}{a_2}\right)^j\frac{C_N}{j!}, \quad j=1,N. \qquad (5.19)$$

Equations (5.17') now consist of $N+1$ constraints if $C_n$ is to be nonzero. The bottom part of the matrix equation is

$$\begin{bmatrix} \ddots & \ddots & \ddots & & \\ \lambda-33a_1 & 26a_2 & -18a_3 & & \\ 32(N-3) & \lambda-25a_1 & 18a_2 & -10a_3 & \\ & 24(N-2) & \lambda-17a_1 & 10a_2 & 0 \\ 0 & & 16(N-1) & \lambda-9a_1 & -2a_3 \\ & & & 8N & 2a_2 \\ & & & & \lambda-a_1 \end{bmatrix}\begin{bmatrix} \vdots \\ C_{N-4} \\ C_{N-3} \\ C_{N-2} \\ C_{N-1} \\ C_N \end{bmatrix}=0, \qquad (5.20)$$

in which $C_{N-j}$ is expressible in terms of $C_N$ through (5.19). The last of (5.20) using (5.19) gives

$$\lambda=a_1-8N(N+2)a_3/a_2.$$

The next requires $a_1$ to be expressible in terms of $N$, $a_2$, and $a_3$. The next requires $a_2$ to be expressed in terms of $N$ and $a_3$ and the one after that gives $a_3$ in terms of $N$. (The expressions are very messy and are not given here for general $N$.) For $n>3$ the additional constraints would not be expected to be satisfied. In fact the situation is even worse than that due to the appearance of the ratio $a_3/a_2$. For $N=3$ the fourth and third of (5.20) give

$$\lambda=a_1-120\left(\frac{a_3}{a_2}\right), \quad a_1=\frac{a_3}{a_2}+\frac{1}{20}\frac{a_2^2}{a_3},$$

respectively. Using these in the second and first of (5.20) we obtain the inconsistent results that

$$a_2=(30a_3^2)^{1/3}, \quad a_2=(\tfrac{95840}{333}a_3^2)^{1/3}.$$

We conclude that it is only possible to obtain a closed-form solution when $M=3$ only if $N\leqslant 2$.

From the considerations of the $M=3$ case for which (5.16) provided just one constraint we would not expect to find much for $M\geqslant 4$. However, we are able to obtain some general results as we shall see below. If in (5.15) we put $n=N+1$ and in (5.16) we set $n=N$ and require that $C_n=0$ for $n\geqslant N+1$, we find that

$$k_{M-1}=[(4N+2)(M-1)+1]a_M,$$
$$k_i=[(4N+2)i+1]a_{i+1}, \quad i=1,M-2. \qquad (5.21)$$

With these values for the $k_i$, $i=1,M-1$, substituted back into (5.15) and (5.16) we set $i=M-2$ and $i=M-3$ (valid since $m\geqslant 4$) and $n=N-1$ and $n=N-2$ (on the assumption that $N\geqslant 2$) to obtain the following relations:

$$(M-2,N-1): \; C_{N-1}=\tfrac{1}{2}(M-1)$$
$$\times[N(M-2)+M-1](a_M/a_{M-1})C_N, \qquad (5.22)$$

$$(M-3,N-1):\ C_{N-1}=\tfrac{1}{4}(M-2)$$
$$\times[N(M-3)+M-2](a_{M-1}/a_{M-2})C_N, \quad (5.23)$$
$$(M-2,N-2):\ C_{N-2}=\tfrac{1}{4}(M-1)$$
$$\times[N(M-2)+2M-1](a_M/a_{M-1})C_{N-1}, \quad (5.24)$$
$$(M-3,N-2):\ 16a_{M-2}C_{N-1}-2(M-2)$$
$$\times[N(M-3)+2M-1]a_{M-1}C_{N-1}$$
$$+\tfrac{1}{3}(M-1)(M-2)[2N(M-3)+2M-3]$$
$$\times a_M C_N=0. \quad (5.25)$$

In (5.25) we use (5.23) and (5.24) to express $C_{N-2}$ in terms of $C_N$ and (5.22) to express $C_{N-1}$ in terms of $C_N$. After a certain amount of elementary algebra (5.25) reduces to

$$-\tfrac{4}{3}(M-1)(M-2)a_M C_N(N+1)M=0,$$

from which it follows that $N=-1$! We conclude that, for $M\geqslant 4$, $N$ may at most take the values 0 and 1.

In the case $N=0$, (5.16) vanishes identically and we have from (5.21) that the coefficients $k_i$ are

$$k_{M-1}=(2M-1)a_M,\quad k_i=-(2i+1)a_{i+1},$$
$$i=1,M-2.$$

Thus (5.15) gives

$$\{\lambda+[2M-2-(2M-1)]a_1\}C_0=0,$$

i.e., $\lambda=a_1$ for all $M$.

There are no constraints on the coefficients $a_i$. In the case $N=1$, (5.15) reduces to the two equations

$$[\lambda-(4M-3)a_1]C_0+2a_2C_1=0,$$
$$4(M-1)C_0+(\lambda-a_1)C_1=0, \quad (5.26)$$

and (5.16) to the $M-2$ equations

$$4(M-1-i)a_{i+1}C_0-(i+1)(2i+1)a_{i+2}C_1=0. \quad (5.27)$$

For (5.27) to have a nontrivial solution the coefficients $a_i$ must satisfy the relations

$$\frac{2}{(M-1)(2M-3)}\frac{a_{M-1}}{a_M}=\frac{4}{(M-2)(2M-5)}\frac{a_{M-2}}{a_{M-1}}$$
$$=\cdots=\frac{2(M-3)}{15}\frac{a_3}{a_4}$$
$$=\frac{2(M-2)}{4}\frac{a_2}{a_3},$$

so that, once $a_M$ and $a_{M-1}$ are specified, $a_{M-2}$ through $a_2$ are also specified. (Note that $a_{M-1}$ cannot be zero.) From (5.26) we find

$$a_1=-\frac{(M-1)(2M-3)a_M}{2a_{M-1}}$$
$$+\frac{a_2a_{M-1}}{(M-1)^2(2M-3)a_m},$$
$$\lambda=-\frac{(M-1)(2M-3)(4M-3)a_M}{2a_{M-1}}$$
$$+\frac{a_2a_{M-1}}{(M-1)^2(2M-3)a_M}.$$

Finally the coefficients $k_i$ are given by

$$k_{M-1}=(6M-5)a_M,\quad k_i=-(6i+1)a_{i+1},$$
$$i=1,M-2.$$

We conclude that for $M\geqslant 4$ all Schrödinger equations of the form (5.9) with the symmetric potential it contains have a ground state solution $y_0$ provided the coefficients $k_i$ take specific values. The coefficients $a_i$ $(i<M)$ may take any real value and $a_M$ any positive value. For a solution of the form $C_0y_0+C_1y_1$ all of the $a_i$'s and $k_i$'s are specified in terms of $a_M$ and $a_{M-1}$ and cannot be zero. In this case the wave function is

$$\psi=C_1\left\{\frac{(M-1)(2M-3)a_M}{2a_{M-1}}+\sum_{i=1}^{M}a_ix^{2i-2}\right\}$$
$$\times\exp\left[-\sum_{i=1}^{M}\frac{a_i}{2i}x^{2i}\right].$$

If $a_{M-1}>0$, the even polynomial has all coefficients positive and so has no real zeros and the wave function represents the ground state. If $a_{M-1}<0$, the polynomial can have an even number of zeros (since there is at least one positive root for $x^2$) and so represents an even excited state.

In contrast to the case $M=2$ in which the polynomials $f_n$ and $g_n$ covered the even and odd functions, respectively, when $M\geqslant 3$ the polynomials $f_n$ do not even cover the even functions. To provide such a cover one could contemplate introducing additional sets of even polynomials defined by

$$h_n^i=x^{2i}f_n,\quad i=1,M-2.$$

Likewise the odd wave functions could be covered by the sets of polynomials $g_n^i$ defined by

$$g_n^i=x^{2i-1}f_n,\quad i=1,M-1.$$

We shall not discuss this idea any further to avoid largely repetitious material. However, we do point out that, in the case the $h_n^1$ polynomials when $M=3$, the recurrence relations obtained when the expansion (5.10) (with $h_n^1$ replacing $f_n$) is substituted into the Schrödinger equation (5.9) contain $f$'s as well as $h$'s and so lead to additional constraints on the expansion coefficients.

## VI. DISCUSSION

We have seen how it is possible to explain the existence of a number (generically one for nonsextic potentials) of closed-form eigenstates for symmetric potentials via the mechanism of polynomials defined by a generalized Rodrigues formula. Apart from the intrinsic interest in obtaining closed-form solutions to the Schrödinger equation, the exact eigenvalues so found provide a check on numerical algorithms for the determination of eigenvalues for such problems. One such algorithm is to expand the wave function on, say, a harmonic oscillator basis. One could imagine doing the same using the functions defined by (3.7) and (4.5). One obtains a tridiagonal matrix for the sextic potential. However, numerical experiments have indicated that the rate of convergence is extremely poor and does not bear comparison with the more standard methods such as those presented by Hautot and Magnus.[2(d)]

As a final remark we point out that the generalized Rodrigues formula (5.4) can be generalized even further. We can define a set of polynomials by

$$f_n = \frac{1}{[2(m+1)]^n n!} \exp\left[\sum_{i=1}^{M} \frac{a_i x^{2(K+i)(m+1)}}{(K+i)(m+1)}\right] \times\left(-\frac{1}{x^{2m+1}}\frac{d}{dx}\right)^n \times\exp\left[-\sum_{i=1}^{M} \frac{a_i x^{2(K+i)(m+1)}}{(K+i)(m+1)}\right], \tag{6.1}$$

where $m$ and $K$ are positive integers. However, these suffer from the same defect as the polynomials defined in (5.4) in that, in addition to the eigenvalue equations, one obtains subsidiary constraints. In general one would expect to obtain only one eigenvalue for each potential for which the constraints are satisfied.

## ACKNOWLEDGMENTS

G.P.F. expresses his thanks for the warm hospitality he received during his stay at the University of the Witwatersrand.

V.M.G. thanks the University of the Witwatersrand, Johannesburg and the Foundation for Research and Development of the CSIR of South Africa for financial support whilst this work was being undertaken.

[1] See, for example, D. M. Dennison and G. E. Uhlebeck, Phys. Rev. **41**, 313 (1932); F. J. Wall and G. Glockler, J. Chem. Phys. **5**, 314 (1937); R. L. Somorjai and D. F. Hornig, *ibid.* **36**, 1980 (1962).

[2] (a) For a small sampling of which see S. N. Biswas, H. Datta, R. P. Saxena, P. H. Srivastova, and V. S. Varma, Phys. Rev. D **4**, 3617 (1971); (b) F. J. Hioe and E W. Montroll, J. Math. Phys. **16**, 1945 (1975); (c) F. J. Hioe, D. MacMillan, and E. W. Montroll, *ibid.* **17**, 1320 (1976); (d) A. Hautot and A. Magnus, J. Comp. Appl. Math. **3**, 3 (1979); (e) V. M. Gorringe and P. G. L. Leach, *Proceedings of the 12th South African Symposium on Numerical Mathematics*, edited by T. J. Ypma (University of Natal, Durban, 1986), pp. 57–74; (f) V. M. Gorringe and P. G. L. Leach, S. A. J. Phys. **11**, 1 (1988).

[3] G. P. Flessas, Phys. Lett. A **27**, 289 (1979); G. P. Flessas and K. P. Das, *ibid.*, **78**, 19 (1980); G.P. Flessas, *ibid.*, **81**, 17 (1981); G. P. Flessas, J. Phys. A **14**, 209 (1981); A. Khare, Phys. Lett. A **83**, 237 (1981); E. Magyari, *ibid.*, **81** 116 (1981).

[4] (a) P. G. L. Leach, Physica D **17**, 331 (1985); M. H. Blecher and P. G. L. Leach, J. Phys. A **20**, 5923 (1987); (b) M. H. Blecher and P. G. L. Leach, *Proceedings of the 13th South African Symposium on Numerical Mathematics*, edited by P. J. Vermeulen (University of Natal, Durban, 1987), pp. 5–17.

[5] P. G. L. Leach, J. Math. Phys. **25**, 2974 (1984); D. R. Taylor and P. G. L. Leach, "Exact solutions of the Schrödinger equation for nonseparable anharmonic potentials in two dimensions," submitted to Physica D.

# On the power-series construction of Schrödinger bound states. II. The effective Hill determinants

M. Znojil
*Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež, Czechoslovakia*

(Received 2 January 1988; accepted for publication 28 September 1988)

For a class of potentials, the radial differential Schrödinger equation for wave functions $\psi(r)$ of bound states may often be replaced by an equivalent matrix equation for coefficients $p_n$ in an associated power-series ansatz. In Paper I [J. Math. Phys. **29**, 1433 (1988)], a rigorous foundation of such a transition (sometimes called the Hill-determinant method) was presented. Now, it will be shown that and how the Feshbach–Löwdin projection-operator idea may help one to reduce the underlying infinite-dimensional diagonalization to a mere finite-dimensional eigenvalue problem.

## I. INTRODUCTION

In the preceding paper (I),[1] we have considered the Schrödinger bound-state problem

$$\left[-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+V(r)\right]\psi(r)=E\psi(r)\,,\qquad \psi\in L_2(0,\infty),\quad l=0,1,\ldots, \tag{1.1}$$

with a class of potentials equivalent, up to a change of variables, to the polynomials

$$V(r)=\sum_{n=1}^{2q+1} a_n r^{2n},\quad a_{2q+1}=a^2>0\,. \tag{1.2}$$

In essence, we have proved an equivalence of the differential equation (1.1) (with the standard boundary conditions[2]) to an algorithm based on an infinite-dimensional limiting transition in the $(N+1)$-dimensional matrix equation

$$Qz=Ez,\quad z=\begin{pmatrix}p_0\\ p_1\\ \vdots\\ p_N\end{pmatrix}, \tag{1.3}$$

with a "quasi-Hamiltonian" matrix $Q$ possessing $2q+2$ diagonals and Hessenberg structure,

$$Q=Q^{(N)}=\begin{pmatrix} A_0 & -B_0 & 0 & \cdots & & & \cdots & 0\\ C_1^{(1)} & A_1 & -B_1 & 0 & \cdots & & \cdots & 0\\ \cdots & & & & & & & \cdots\\ 0 & \cdots & 0 & D^{(q)} & D^{(q-1)} & \cdots & D^{(1)} & C_N^{(q)} & \cdots & A_N \end{pmatrix} \tag{1.4}$$

[cf. Eqs. (5.3)–(5.5) in I].

In the literature,[3–16] the whole approach is usually called the method of Hill determinants. It is based on the ansatz

$$\psi(r)=\exp(-g(r))\sum_{n=0}^{\infty} p_n r^{2n+l+1}\,, \tag{1.5}$$

with some (polynomial) $g(r)$. An insertion of (1.5) in (1.1) leads immediately to an explicit form of matrix elements in Eq. (1.3). Vice versa, each numerical solution of the truncated Eq. (1.3) has been proposed[3–6] to convert Eq. (1.5) into an approximate wave function, which becomes exact in the $N\to\infty$ limit.

Recently, the above Hill-determinant (HD) method has been criticized[7] and shown to have a restricted validity,[8–13] which, essentially, depends on our choice of $g(r)$. In I, the latter point has been clarified for all the potentials (1.2). We proposed a use of polynomials

$$g(r)=\frac{a}{2q+2}r^{2q+2}+\sum_{j=1}^{q}\frac{\beta_j}{2j+2}r^{2j},\qquad a>0,\quad a^2=a_{2q+1}, \tag{1.6}$$

which reflect the correct WKB $r\to\infty$ asymptotics of $\psi(r)$. All the possible failures (cf., e.g., their $q=1$ examples and analysis in Refs. 8–13) were eliminated by means of the simple restriction

$$a_{2q}/2a=\gamma_q>-\beta_q\,, \tag{1.7}$$

which has to be added to (1.6).

In the resulting well-founded form, the HD technique is of interest as a possible new type of resummation of the divergent perturbation series.[4] The recent increase of interest in it also reflects its simplicity. Indeed, mere $2q+2$ diagonals in $Q$ replace the $4q+3$ nonzero diagonals of the Hamiltonian when represented as a matrix in an ordinary harmonic oscillator basis.[11]

Numerically, the latter variational diagonalization converges extremely quickly. Thus some authors also emphasize an improvement of efficiency of the purely numerical HD algorithms.[10,12,14] The latter effort proved successful: The extreme simplicity of the matrix elements (1.4) enables one to reduce the infinite-dimensional (exact) HD secular equation

 0022-2488/89/020413-07$02.50 © 1989 American Institute of Physics 

$$\det(Q-E)=0 \tag{1.8}$$

to a mere transcendental equation

$$1/F_0=0\,, \tag{1.9}$$

where $F_n$ denotes a certain "extended" analytic continued fraction[6] (and coincides with the ordinary continued fraction in the simple example of Ref. 4). This arrangement makes full use of the specific Hessenberg structure of $Q$ and simplifies significantly the numerical infinite-dimensional limiting transition $N\to\infty$.

Geometrically, the asymptotically dominant components of $F_n$, $n\gg 1$, may be identified with the so-called fixed points (FP's) of the underlying continued-fractional mappings.[6] Thus our algebra proceeds one step further and we can replace the recurrently defined quantities $F_n$, $n\gg 1$, by certain FP expansions in roots of algebraic equations of degree $q+1$ (see Ref. 6).

The first numerical test of the above FP asymptotic expansion idea has already been performed in Ref. 14, for the potential $V(r)=r^2+gr^4$ and with the first- and second-order FP corrections taken into account. The convergence still remains comparatively slow. Fortunately, an incorporation of the higher-order corrections makes the combined FP–HD algorithm fully competitive with the other methods. On a few examples, this was demonstrated numerically by Tater and co-workers.[10,15]

An unpleasant shortcoming of the rather universal FP technique lies in a tedious algebra pertaining to the derivation of the explicit FP corrections. In a way, these complications stem from the "too geometric" character of the original algorithms. In the present paper based on a close analogy between the Hermitian and non-Hermitian cases, we recall an experience gained in the former case[17] and describe a simplified construction of the FP expansions.

The core of our contribution is formulated in Sec. II. In place of the $N=\infty$ eigenvalue condition (1.8), we propose use of the finite-dimensional rigorous formula

$$\det(Q^{\text{eff}}-E)=0\,. \tag{1.10}$$

This will complement and complete the results of Refs. 1 and 6. In Sec. III, we change our point of view, reinterpret Eq. (1.3) as a difference equation, and derive a new formula for the unknown matrix elements of $Q^{\text{eff}}$. Finally, an illustration of technicalities is provided by Sec. IV where both the algebraic and numerical aspects of our eigenvalue condition (1.10) are discussed for the $q=1$ example. Section V is a summary.

## II. THE FINITE-DIMENSIONAL "EFFECTIVE" HILL DETERMINANTS

### A. The extended continued fractions (ECF)

In the spirit of Ref. 6, we may postulate that

$$Q-E=U\cdot L\,, \tag{2.1}$$

where the matrix $U$ has only two nonzero diagonals,

$$U_{nn}=1,\quad U_{nn+1}=-B_nF_{n+1},\quad n\geqslant 0\,, \tag{2.2}$$

while $L$ is a lower triangular matrix with $2q+1$ diagonals, in general:

$$L=\begin{pmatrix} Z_0 & & 0 & 0 & \cdots & & & & & & \\ Y_1^{(1)} & & Z_1 & 0 & \cdots & & & & & & \\ & \cdots & & & & & & & & & \\ 0 & \cdots & 0 & X_n^{(q)} & \cdots & X_n^{(1)} & Y_n^{(q)} & \cdots & Y_n^{(1)} & Z_n & 0 & \cdots \\ \cdots & & & & & & & & & & & \end{pmatrix}. \tag{2.3}$$

The new matrix elements can be computed, from definition (2.1), by recurrences. Their general form may be found elsewhere[14]: For the present purposes, we restrict our attention to the first nontrivial example with $q=1$ and, abbreviating

$$G_n=C_n^{(1)}+B_nF_{n+1}D^{(1)},\quad n\geqslant 0\,, \tag{2.4}$$

write the definitions

$$Y_n^{(1)}=G_n,\quad X_n^{(1)}=D^{(1)},\quad Z_n=1/F_n\,,$$

and recurrences

$$F_n=1/(A_n-E+B_nF_{n+1}G_{n+1}),\quad n=0,1,\ldots\,. \tag{2.5}$$

For $D^{(1)}=0$, the latter ECF rule defines simply the ordinary continued fractions.[4]

### B. The Löwdin–Feshbach projection operators in the HD non-Hermitian case

Let us assume that the $N\to\infty$ problem of convergence has been settled[6] and fix some large dimension $N<\infty$. Next, for reasons that will become obvious immediately, let us also introduce a pair of projectors $\hat{p}$ and $\hat{q}=1-\hat{p}$ such that

$$\hat{p}z=\hat{p}\begin{pmatrix} p_0\\ p_1\\ \vdots\\ p_N\end{pmatrix}=\begin{pmatrix} p_0\\ p_1\\ \vdots\\ p_M\\ 0\\ \vdots\\ 0\end{pmatrix},$$
$$\hat{q}z=\begin{pmatrix} 0\\ \vdots\\ 0\\ p_{M+1}\\ p_{M+2}\\ \vdots\\ p_N\end{pmatrix},\quad M\geqslant 0,\quad M<N\,. \tag{2.6}$$

Then, Eq. (1.3) may be divided by the regular operator $U$ from the left, giving $Lz=0$, i.e.,

$$L\hat{p}z = -L\hat{q}z\,. \tag{2.7}$$

Next, we use the fact that $\hat{p}L\hat{q} = \hat{p}L\hat{p}\hat{q} = 0$ (identically) and, applying $\hat{p}U\hat{p}$ on both sides of Eq. (2.7), we get

$$\begin{aligned}&(Q^{\text{eff}} - E)\hat{p}z = 0\,,\\ &Q^{\text{eff}} - E = \hat{p}U\hat{p}\cdot\hat{p}L\hat{p}\,.\end{aligned} \tag{2.8}$$

This may be interpreted as a Feshbach or Löwdin model-space equation which defines the first $M+1$ components of the exact wave function coefficients $z$.

In the $N\to\infty$ limit, Eq. (2.8) remains valid in the same form, although its derivation must be performed more carefully since the division by $U$ is not permitted anymore.[6,10] Fortunately, we may understand simply its first $M$ rows as equivalent to the explicit determinantal definitions of $p_n$'s since $Q^{\text{eff}}_{mn} = Q_{mn}$, for all $n<M$,

$$\begin{aligned}p_{n+1} &= (p_0/B_0B_1\cdots B_n)\det(Q^{(n)} - E),\\ &\qquad n = 0,1,\ldots,M-1\,.\end{aligned} \tag{2.9}$$

Then, only the remaining (last) row of Eq. (2.8) [cf. also Eq. (2.3)],

$$Z_M p_M + \sum_{l=1}^{q} Y_M^{(l)} p_{M-l} + \sum_{l=1}^{q} X_M^{(l)} p_{M-q-l} = 0\,, \tag{2.10}$$

has the ECF coefficients that contain all the relevant information about the missing part of the full matrix $Q$. Their recurrent evaluation is therefore equivalent to a return to the original Eq. (1.3) with $N=\infty$.[14]

In essence, our present main purpose lies just in a non-ECF specification of these coefficients. After achieving this purpose [cf. Eq. (3.15) below], we shall be able to treat Eq. (2.10) as a transcendental equation that specifies the binding energies numerically.

## III. THE DIFFERENCE-EQUATION CONSTRUCTION OF $Q^{\text{eff}}$

### A. The boundary conditions

As shown in Ref. 6, the potentials (1.2) with the "superconfining property"

$$\lim_{r\to\infty}\left[V(r) - a_{2q+1}r^{4q+2}\right]/r^{2q} = +\infty \tag{3.1}$$

enable us to choose the parameters $\beta_k$ in (1.6) in such a way that all the coefficients $D^{(j)}$ in $Q$ vanish. In the general case, let us now introduce an integer $t\in[0,q]$ such that $D^{(t)}\neq 0$ while $D^{(t+1)} = D^{(t+2)} = \cdots = D^{(q)} = 0$. Then, we may write each row of our infinite-dimensional matrix equation (1.3) as a difference equation for some function $h_n$,

$$B_n h_{n+1} - \sum_{j=0}^{q} C_n^{(j)} h_{n-j} - \sum_{k=1}^{t} D^{(k)} h_{n-q-k} = 0$$

$$(C_n^{(0)} \equiv A_n - E)\,, \tag{3.2}$$

and put $h_n = p_n$ only after an incorporation of the corresponding boundary conditions.

In general, Eq. (3.2) has a complete set of $q+t+1$ independent solutions. For example, we may specify $h_n = h_n^{(k)}$, $k = 1,2,\ldots,q+t+1$, by the set of independent initial values

$$h^{(k)}_{M_k - j} = 0, \quad j = 1,2,\ldots,q+t\,, \tag{3.3}$$

with some positive integers $M_k \neq M_n$ for $k\neq n$, and write the general solution of (3.2) as a $(q+t)$-parametric linear superposition

$$h_n = \sum_{k=1}^{q+t+1} \alpha_k h_n^{(k)} \tag{3.4}$$

with an arbitrary normalization. An explicit determinantal representation of the separate components $h_n^{(k)}$ may be derived in full analogy with Eq. (2.9).

According to I, the natural $q+t$ boundary conditions at the origin,

$$h_{-1} = 0, \quad h_{-2} = 0,\ldots,h_{-q-t} = 0\,, \tag{3.5}$$

simply reflect the regularity of $\psi(r)$ at $r\to 0$. They must be complemented by the physical normalization requirement or boundary condition of the asymptotic HD form[1]

$$h_{N+1} = 0, \quad N\to\infty\,. \tag{3.6}$$

### B. The reverse Hill-determinant method

There is a symmetry between (3.5) and (3.6)—we may restrict (3.4) by $q+t$ conditions (3.5) or by the single requirement (3.6), and obtain a unique "regular" solution or $q+t$ "Jost" solutions, respectively. In the former case, an addition of the asymptotical requirement (3.6) and an availability of the determinantal definition (2.9) lead immediately to the Hill-determinant algorithm. Also, let us study the latter combination of "Jost" solutions with the boundary conditions (3.5).

In the first step, we recall I and replace the complete initializations (3.3) and set of solutions $h_n^{(k)}$ by the other complete set of solutions $h_n^{[x,k]}$ specified by the $q+t+1$ independent $n\gg 1$ asymptotics

$$\begin{aligned}h_n^{[d,k]} &\approx |(a/n)^{n/(q+1)}|\exp[2\pi kni/(q+1)]\,,\\ &\qquad n\gg 1, \quad k = 1,2,\ldots,q+1\,,\end{aligned} \tag{3.7a}$$

and

$$\begin{aligned}h_n^{[s,k]} &\approx |(b/n)^{n/t}|\exp(2\pi kni/t)\,,\\ &\qquad b = D^{(t)}/4a, \quad n\gg 1, \quad k = 1,2,\ldots,t\,.\end{aligned} \tag{3.7b}$$

The superscripts $d$ and $s$ mean "dominant" and "subdominant" since

$$\begin{aligned}&(a/n)^{n/(q+1)}/(b/n)^{n/t} = \exp n(\gamma\ln n + \delta)\,,\\ &\gamma = (q+1-t)/t(q+1) > 0,\\ &\delta = (\ln a)/(q+1) - (\ln b)/t\,,\end{aligned} \tag{3.8}$$

increases very quickly with the increasing $n$. An explicit derivation of the asymptotics (3.7) is easy: we insert the leading-order form of coefficients

$$\begin{aligned}&B_n = 4n^2 + O(n), \quad C_n^{(j)} = 4\beta_{j+1}n + O(1),\\ &D^{(k)} = O(1)\end{aligned} \tag{3.9}$$

into (3.2) and arrive at the two possible leading-order forms of this equation,

$$4n^2 h_{n+1} - 4anh_{n-q} + \text{corrections} = 0 \tag{3.10a}$$

and

$$4nah_{n-q} - D^{(t)}h_{n-q-t} + \text{corrections} = 0\,, \quad (3.10b)$$

with the simple solutions (3.7a) and (3.7b), respectively (I).

In the second step, we may incorporate the second-order corrections in (3.10) and find out that the asymptotics of $h_n$ [Eq. (3.4)] are absolutely dominated by a single component $h_n^{[d,q+1]}$. In accordance with the physical boundary condition (3.6), these asymptotics have to change sign at the physical energy $E$ and some large $n = N \gg 1$. Vice versa, the general "Jost" solution $h_N^{(\text{Jost})}$ of (3.2) [compatible with (3.6)] may be characterized simply by an absence of the dominant component $h_N^{[d,q+1]}$,

$$h_N^{(\text{Jost})} = \sum_{k=1}^{q} \rho_k h_N^{[d,k]} + \sum_{k=1}^{t} \rho_{k+q} h_N^{[s,k]}\,. \quad (3.11)$$

In a simplified notation, let us write

$$h_N^{(\text{Jost})} = \sum_{k=1}^{q+t} J_{Nk}\rho_k$$

and define $J_{Nk}$ by Eq. (3.7) in the limit $N \to \infty$. After an incorporation of a few corrections in (3.10), we also may obtain an asymptotic series and define $J_{Nk}$ for $N \geqslant N_0 \gg 1$ with a sufficient precision (see below). Then, we recall Eq. (3.2) with indices $n = N_0 + q + t - 1$, $n = N_0 + q + t - 2,\ldots,n = q + t$, and use it as a recurrent definition of $J_{mk}$ beyond the asymptotic region, at $m = N_0 - 1, N_0 - 2,\ldots,0$, respectively.

Of course, the remaining $q + t$ rows of Eq. (3.2) [together with the boundary conditions (3.5)] represent simply a $(q + t)$-dimensional matrix equation

$$\sum_{m=0}^{q+t-1}\left[\sum_{l=0}^{q+t} Q_{kl}J_{lm} - EJ_{km}\right]\rho_m = 0,$$
$$k = 0,1,\ldots,q+t-1\,. \quad (3.12)$$

Up to an arbitrary normalization, it fixes the physical energies and coefficients $\rho$ in (3.11)—we may put $p_n = h_n^{(\text{Jost})}$ in (1.5).

### C. The general matching conditions

An optimal approximate representation of the wave functions $p_n$ or $\psi(r)$ (1.5) is combined from the $n \gg 1$ asymptotics $p_n = h_n^{(\text{Jost})}$ (3.11) and $n = O(1)$ determinants ($p_n = h_n^{(\text{regular})}$ [Eq. (2.9)], i.e.,

$$h_n^{(\text{regular})} = J_{n0}\rho_0$$

with the single normalization parameter $\rho_0$). In the numerical practice, the particular boundary-condition matchings (3.11) + (3.5) [i.e., (3.12)] or (2.9) + (3.6) [i.e., (1.8)] may be then replaced also by a more flexible condition $h_{n_i}^{(\text{regular})} = h_{n_i}^{(\text{Jost})}$,

$$J_{M+i,0}\rho_0 = \sum_{k=1}^{q+t} J_{M+i,k}\rho_k\,, \quad i = 1,2,\ldots,q+t+1\,, \quad (3.13)$$

of the "determinant = expansion" type with $M > q + t$.

The necessary dimension of determinants or precision of the expansions may be lowered by a transition to the alternative form of (3.13),

$$\left(\sum_{n=m-q-t}^{M} Q_{mn}J_{n0} - EJ_{M0}\delta_{mM}\right)\rho_0$$
$$+ \sum_{k=1}^{q+t}\left[\sum_{n=M+1}^{m+1} Q_{mn}J_{nk} - EJ_{mk}(1-\delta_{mM})\right]\rho_k = 0,$$
$$m = M, M+1,\ldots,M+q+t\,. \quad (3.14)$$

This may be understood as a generalization of (3.12). Now, an analogous generalization has to be found for the Hill-determinant prescription (1.8). We shall interpret the "effective-matrix" Schrödinger equation (2.8) also as a matching condition.

In the first step, let us recall the ECF construction of the corresponding "effective" relation (2.10), and notice that its ECF matrix elements $Q_{Mm}^{\text{eff}}$ must be independent of the parameters $A_0, A_1,\ldots,A_{q+t-1}$. In the present recurrent scheme, we have $J_{n0} = J_{n0}(A_0, A_1,\ldots,A_{n-1})$ and $J_{nk} = J_{nk}(A_{n+q+t}, A_{n+q+t+1},\ldots)$ so that $\rho_k = \rho_k(A_0, A_1,\ldots,A_{q+t-1})$ after a complete matching (3.14), with any $k \geqslant 1$. As a result, the matrix elements of $Q_{Mm}^{\text{eff}}$ with $M > q + t$ must be also independent of our particular choice of the coefficients $\rho_k$ in the "Jost" solution (3.11) and therefore, its insertion in (2.10) should be valid for all the asymptotic components $h_n^{[x,k]}$ separately,

$$Z_M J_{Mk} + \sum_{m=1}^{q} Y_M^{(m)} J_{M-mk} + \sum_{m=1}^{t} X_M^{(m)} J_{M-q-m,k} = 0\,,$$
$$k = 1,2,\ldots,q+t\,. \quad (3.15)$$

This is our main conclusion—we may reinterpret (3.15) as a set of $q + t$ linear homogeneous algebraic equations that define the matrix elements of $Q_{Mm}^{\text{eff}}$ as functions of the known (asymptotic-series) matrix elements $J_{nk}$, $k \geqslant 1$.

## IV. A SIMPLE EXAMPLE

The change of variables in (1.1),

$$\psi \to r^{\text{const}}\psi, \quad r \to r^{1/p}, \quad l + \tfrac{1}{2} \to (l + \tfrac{1}{2})/p,$$
$$p = 1,2,3,4, \quad (4.1)$$

enables us to consider the class of potentials $V = V_p(r)$,

$$\begin{aligned} V_1(r) &= a_1 r^2 + a_2 r^4 + a_3 r^6\,,\\ V_2(r) &= \tfrac{1}{2}a_0/r + 2a_2 r + 4a_3 r^2\,,\\ V_3(r) &= a_0/(3r)^{4/3} + a_1/(3r)^{2/3} + a_3(3r)^{2/3}\,,\\ V_4(r) &= \tfrac{1}{8}a_0/r^{3/2} + \tfrac{1}{4}a_1/r + \tfrac{1}{2}a_2/r^{1/2}\,, \end{aligned} \quad (4.2)$$

as the same eigenvalue problem (1.1) with the energies $E = E_p$,

$$E_1 = -a_0, \quad E_2 = -a_1, \quad E_3 = -a_2, \quad E_4 = -a_3\,, \quad (4.3)$$

respectively. Thus we may restrict our attention to $p = 1$, $q = 1$ and, inserting the ansatz (1.5) in (1.1) ($\beta_1 = \beta$, $\gamma_1 = \gamma$), obtain the matrix equation (1.3) or difference equation (3.2) with the explicit coefficients

$$\begin{aligned} B_n &= (2n+2)(2n+2l+3),\\ A_n &= C_n^{(0)} - a_0 = (4n+2l+3)\beta\,,\\ C_n &= C_n^{(1)} = (4n+2l+1)a + a_1 - \beta^2, \quad n = 0,1,\ldots,\\ D^{(1)} &= D = 2(\gamma-\beta)a, \quad \gamma = a_2/2a > -\beta\,. \end{aligned} \quad (4.4)$$

From this difference Schrödinger equation, we may return to the various physical situations (4.2) via the transformation (1.5) and (4.1).

### A. The asymptotically subdominant solution of the difference Schrödinger equation

Let us complement the explicit coefficients (4.4) by the change of variables

$$h_n = ((\beta-\gamma)/2)^n \varphi_n/n!, \quad n \geqslant 0, \tag{4.5}$$

and use the notation

$$\dot{\varphi}_n = \varphi_{n+1} - \varphi_n, \quad \ddot{\varphi}_n = \dot{\varphi}_{n+1} - \dot{\varphi}_n . \tag{4.6}$$

We obtain the Schrödinger equation (3.2) in the form

$$\begin{aligned} \dot{\varphi}_{n-1} = (1/n)[\alpha_1\varphi_n + \alpha_2\dot{\varphi}_n + \alpha_3\ddot{\varphi}_n \\ + [1/(n+1)](\alpha_4\varphi_{n+1} + \alpha_5\dot{\varphi}_{n+1})], \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} &4\alpha_1 a = \gamma^2 - a_1 - (2l+5)a, \\ &4\alpha_2 a = 2\gamma(\gamma-\beta), \quad 4\alpha_3 a = (\gamma-\beta)^2, \\ &4\alpha_4 a = [(l+\tfrac{3}{2})\gamma + a_0/2](\gamma-\beta), \\ &4\alpha_5 a = (l+\tfrac{3}{2})(\gamma-\beta)^2 . \end{aligned} \tag{4.8}$$

Then, another ansatz

$$\varphi_n = \sum_{k=0}^{\infty} \frac{\Gamma(n+1)}{\Gamma(n+z+k)} c_k \tag{4.9}$$

simplifies the separate differences (4.6)—we may put

$$z = 1 - \alpha_1$$

and get (4.7) in the final form

$$\begin{aligned} &(k+1)c_{k+1} - \mu_k c_k - \nu_k c_{k-1} = 0, \\ &\mu_k = (k-\alpha_1)\alpha_2 - \alpha_4, \\ &\nu_k = (k-\alpha_1-1)(\alpha_1\alpha_3 - k\alpha_3 + \alpha_5), \\ &\quad c_{-1} = 0, \quad k = 0,1,\ldots . \end{aligned} \tag{4.10}$$

Its algebraic determinantal solution is straightforward,

$$c_{k+1} = \frac{c_0}{(k+1)!} \det \begin{pmatrix} \mu_0 & -1 & 0 & \cdots & 0 \\ \nu_1 & \mu_1 & -1 & 0 & \cdots \\ 0 & 2\nu_2 & \mu_2 & -1 & \cdots \\ & \cdots & & & \cdots \\ 0 & \cdots & 0 & k\nu_k & \mu_k \end{pmatrix}. \tag{4.11}$$

For large $k$, we have an estimate

$$c_k \approx (\pm(\gamma-\beta)/2\cdot\sqrt{-k/a})^k \approx k^{k/2}, \quad k \gg 1,$$

which does not contradict the convergence of the ansatz (4.9). Hence we may construct the subdominant solution $h_n^{[s,1]} = h_n^{[s]}$ with an arbitrary precision in principle.

### B. The pair of the asymptotically dominant solutions

Let us put $a = \alpha^2$ and not specify the sign of the new parameter $\alpha$ yet. Then, according to I, the alternative ansatz

$$h_n = \left(\frac{\alpha}{\sqrt{2}}\right)^n \frac{\Gamma(C_{n+1}/8a)}{\Gamma(1+n/2)\Gamma[(n+l+\frac{5}{2})/2]} \varphi_n \tag{4.12}$$

converts the Schrödinger equation (3.2) with $q = p = 1$ into the relation (5.8) of I,

$$\begin{aligned} &[2\sinh\tfrac{1}{2}\nabla + S_n + R_n(\exp\nabla - 1)]\varphi_n = 0, \\ &\nabla = -2\frac{d}{dn}, \quad S_n = T_n A_n C_{n-1}/B_{n-1}, \\ &R_n = DT_n, \\ &T_n = \frac{\sqrt{2}\Gamma[(n+1)/2]\Gamma[(n+l+\frac{3}{2})/2]\Gamma(C_{n-1}/8a)}{(2\alpha)^3\Gamma(n/2)\Gamma[(n+l+\frac{1}{2})/2]\Gamma(C_{n+2}/8a)}, \\ &\quad n = 0,1,\ldots, \end{aligned} \tag{4.13}$$

where the exponentials $\exp(m/2)\nabla$ are infinite Taylor series representing the shift of indices,

$$\varphi_{n+m} = \sum_{k=0}^{\infty} \frac{m^k}{k!}\frac{d^k}{dn^k}\varphi_n = \exp\left(-\frac{m\nabla}{2}\right)\varphi_n \tag{4.14}$$

and an ansatz has to be used again,

$$\begin{aligned} &\varphi_n = \sum_{k=0}^{M} b_k\xi_k + O(\xi_{M+1}), \\ &\xi_k = \exp(G\varepsilon^{-1})\varepsilon^{F+k}, \quad G = (\beta+\gamma)/2\alpha, \\ &F = (\gamma^2-\beta^2)/4a, \\ &\quad \varepsilon = 1/\sqrt{n}, \quad n \gg 1 . \end{aligned} \tag{4.15}$$

Its leading-order coefficients are taken from I, and its general form is dictated by the simple rule for differentiation,

$$\begin{aligned} \nabla^m\xi_k &= \left(\varepsilon^3\frac{d}{d\varepsilon}\right)^m \xi_k \\ &= \sum_{j=0}^{m} (-G)^{m-j} W_{j,k}^{(m)} \xi_{k+m+j}, \\ &W_{0,k}^{(1)} = 1, \quad W_{1,k}^{(1)} = k+F, \\ &W_{j,k}^{(m+1)} = W_{j,k}^{(m)} + (F+k+m+j-1)W_{j-1,k}^{(m)}, \\ &\quad j = 0,1,\ldots,m+1. \end{aligned} \tag{4.16}$$

Obviously, the structure of Eq. (4.13) is much more complicated than the structure of its "subdominantly smooth" counterpart (4.7). Fortunately, the coefficients in (4.13) may be represented by the Barnes formula[13]

$$\begin{aligned} \frac{\Gamma(z+a)}{\Gamma(z+b)} = z^{a-b}\exp\Bigg[\sum_{m=1}^{M} z^{-m}\frac{(-1)^{m+1}}{m(m+1)}\{B_{m+1}(a) \\ - B_{m+1}(b)\} + O(z^{-m-1})\Bigg], \end{aligned} \tag{4.17}$$

where

$$\begin{aligned} &B_2(x) = x^2 - x + \tfrac{1}{6}, \\ &B_3(x) = x^3 - 3x^2/2 + x/2, \\ &\vdots \end{aligned}$$

are the so-called Bernoulli polynomials. As a consequence, an $n \gg 1$ asymptotic form of our Eq. (4.13) may again be interpreted as an equation with the power-series coefficients

$$\begin{aligned} &T_n = t_0\varepsilon + t_1\varepsilon^3 + \cdots, \quad S_n = s_0\varepsilon + \cdots, \\ &t_0 = 1/4\alpha^3, \quad s_0 = G_1\cdots \end{aligned}$$

and schematical structure ($n$ or $\varepsilon$ dependence, $\varepsilon \ll 1$)

$$\Theta_1\varepsilon + \Theta_2\varepsilon^2 + \Theta_3\varepsilon^3 + \cdots = 0, \tag{4.18}$$

where, obviously, each coefficient $\Theta_{M+1}$ must be equal to zero separately.

In general, the algebraic manipulations represented by Eq. (4.18) are combined from the evaluation of the power series (4.14), the action (4.16), and insertions of (4.17). They are complicated but straightforward; at $M=0$, we obtain an identity due to our choice of $G=s_0$, at $M=1$, we obtain an identity as well, due to our specification of the second free parameter $F=GDt_0$ in (4.15). Finally, it is easy to show that the $M\geqslant 2$ items of (4.18) do not contain $b_M$ and may be written in the form

$$(M-1)b_{M-1} = \sum_{l=0}^{M-2} b_l \sum_{k_1k_2k_3} (-s_{k_1}\delta_{k_2,0}\delta_{k_3,0} + \cdots),$$
$$k_1,k_2,k_3 \geqslant 0, \quad 2k_1 + k_2 + k_3 + l = M. \tag{4.19}$$

This defines the sequence of the unknown coefficients in our ansatz (4.15) in a recurrent way similar to the definition (4.10) of the subdominant solutions.

In the present case, our choice of the sign of $\alpha$ in (4.12) is ambiguous. In accord with I, we have to choose both signs when constructing the general solution. For the elimination of the unphysical asymptotic component we used not changing signs, i.e., with $\alpha>0$ in (4.12) (see I for more details) and must guarantee its asymptotic dominance [represented by the exponential $\exp G/\varepsilon$ in (4.15), i.e., obviously, just by the condition (1.7)]. Vice versa, the expansions $h_n^{[d]}$ needed in the Jost solution (3.11) will be given by the present formulas with the negative $\alpha<0$.

### C. The matrix elements of the effective Hamiltonian $Q^{\text{eff}}$

Our basic finite-dimensional and exact form of the Schrödinger equation (2.8) has to be solved numerically, of course. In the present example, it takes the form

$$\begin{pmatrix} (A_0-E) & -B_0 & 0 & \cdots & \cdots & 0 \\ C_1 & (A_1-E) & -B_1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & D & C_{N-1} & (A_{N-1}-E) & -B_{N-1} \\ 0 & \cdots & 0 & D & Y_N & Z_N \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \cdots \\ p_{N-1} \\ p_N \end{pmatrix} = 0. \tag{4.20}$$

The pair of the unknown matrix elements (normalized by the choice of $X_N^{(1)} = D^{(1)} = D$) $Y_N^{(1)}$ ($= Y_N$) and $Z_N$ is defined by Eq. (3.15),

$$Dh_{N-2}^{[d]} + Y_N h_{N-1}^{[d]} + Z_N h_N^{[d]} = 0,$$
$$Dh_{N-2}^{[s]} + Y_N h_{N-1}^{[s]} + Z_N h_N^{[s]} = 0, \tag{4.21}$$

with the asymptotic series representations (4.9) and (4.15) of the Jost components $h_n^{[s]}$ and $h_n^{[d]}$, respectively.

In the $N\gg 1$ asymptotic region, we may interpret (4.21) as an estimate

$$\begin{pmatrix} Y_N \\ Z_N \end{pmatrix} = \begin{pmatrix} O(1/\sqrt{N}) & O(1/N) \\ O(1/N) & O(1/N^2) \end{pmatrix}^{-1} \begin{pmatrix} O(1) \\ O(1) \end{pmatrix}$$
$$= \begin{pmatrix} O(N) \\ O(N^{3/2}) \end{pmatrix}, \tag{4.22}$$

compatible with the ECF results.[16] This estimate enables us to arrange the explicit solution of (4.21) in the form

$$Y_N = \frac{R(s,2) - R(d,2)}{R(s,1) - R(d,1)},$$
$$R(x,i) = \frac{h_{N-i}^{[x]}}{h_N^{[x]}}, \quad x = s,d, \quad i = 1,2,$$
$$Z_N = \frac{R(s,2)R(d,1) - R(d,2)R(s,1)}{R(s,1) - R(d,1)}. \tag{4.23}$$

This is our final formula and has the following features.

(a) After an insertion of the expansions of $h_n^{[s]}$ and $h_n^{[d]}$ as derived in the preceding text, this formula is valid for all $N\geqslant 2$ in principle.

(b) In practice, we shall use the truncated expansions of $h_n^{[x]}$, assuming that the value of the subscript $n$ is sufficiently large, $n\gg 1$. Then, all the second terms in the numerators and denominators in Eq. (4.23) become mere corrections of $O(1/N^{1/2})$. In the numerator of $Y_N$, this correction is even smaller [$O(1/N)$].

(c) The Padé-type structure of Eq. (4.23) is suitable for computations even for the intermediate (not too large) values of the subscript $N$. Vice versa, a power-series rearrangement of Eq. (4.23) could make the $N$ dependence of $Y_N$ and $Z_N$ less complicated in the non-numerical considerations. For this purpose, we also recommend formula (2.5), which defines both $Y_N$ and $Z_N$ as functions of a quantity $F_n$, $n>N$. In this context, we also recall the results of Ref. 16 where an expansion of the asymptotic series type

$$B_{N-1}F_N = \sqrt{N} \sum_{m=0}^{K} \delta_m N^{-m/2} \tag{4.24}$$

has been obtained, with the coefficients

$$\delta_0 = 1/\sqrt{a}, \quad \delta_1 = -(\beta+\gamma)/4a,$$
$$\delta_2 = (1/16a\sqrt{a})[(4l-2)a + (\beta+\gamma)^2 + 4\gamma^2 - 4a_1], \tag{4.25}$$

etc. (generated on the computer in REDUCE).

For illustrative purposes, we may recall once more the tests of Tater,[10] who employed the simplest nontrivial example of Ref. 4 and demonstrated the efficiency of the ECF–FP concept as well as an extension of validity of the eigenvalue condition (1.9) numerically. His conclusions were based on the expansions of the power-series type (4.24), but only insignificant deviations appear after a transition to the Padé-type "effective" matrix elements (4.23) in the corresponding secular equation (2.8) or (2.10) in the simple $q=1$

example. Thus we may omit the tables of energies here, reemphasizing once more that equations of the type (4.23) [and (3.15) in general] seem to represent the only feasible non-numerical definition of $Q^{\text{eff}}$ for the more complicated ($q > 1$) potentials.

## V. SUMMARY

In paper I, the Hill-determinant zeros were shown to coincide with energies in the infinite-dimensional limit only. Now, our main result is the identification of these binding energies with zeros of a finite, $[(M+1)\times(M+1)]$-dimensional effective Hill determinant $\mathcal{G} = \det(Q^{\text{eff}}(E) - E)$.

This paper remains based on the power series ansatz of I, where an algebraic (determinantal) solvability of the resulting difference Schrödinger equation was employed. Here, we solve it iteratively in the $n \gg 1$ asymptotic domain. This has the following advantages.

(i) In the limit $n \to \infty$, our difference equation degenerates into a two-term recurrence. Hence, for any potential, it becomes exactly solvable in terms of the gamma functions. This contrasts with the $(q+t+2)$-term recurrent character of the $n = O(1)$ Hill determinants.

(ii) A transition to large but finite indices $n < \infty$ does not violate the leading-order two-term structure of the equation. A systematic and explicit algebraic evaluation of corrections remains a linear problem to all orders.

(iii) Even in the zeroth-order approximation, we obtain a significant improvement of the Hill-determinant algorithm (cf., e.g., the numerical tests of the particular examples in Refs. 6 and 10). In the present formalism, the reasons are obvious—we have chosen a better matching point for the "regular" and "Jost" solutions.

(iv) The implicit Padé-approximation-like structure of our basic formulas [(3.15) or, in an explicit example, (4.23)] seems promising for an *a priori* control of precision at large $M \ll \infty$.

(v) In contrast to the geometric (fixed-point) analysis of convergence (related to the ECF definition of $Q^{\text{eff}}$ and given in detail in Refs. 6 and 16), our present definitions [containing complex roots (3.7), etc.] seem more likely to clarify the origin and elimination of oscillations as observed in Ref. 14.


[1] M. Znojil, J. Math. Phys. **29**, 1433 (1988).
[2] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982), Part 3.
[3] S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, Phys. Rev. D **4**, 3617 (1971).
[4] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).
[5] C. A. Ginsburg, Phys. Rev. Lett. **48**, 839 (1982).
[6] M. Znojil, J. Math. Phys. **24**, 1136 (1983).
[7] R. N. Chaudhuri, Phys. Rev. D **31**, 2687 (1985).
[8] M. Znojil, Phys. Rev. D **26**, 3750 (1982).
[9] D. Masson, J. Math. Phys. **24**, 2074 (1983).
[10] M. Tater, J. Phys. A: Math. Gen. **20**, 2483 (1987).
[11] A. Hautot, Phys. Rev. D **33**, 437 (1986).
[12] J. Killingbeck, Phys. Lett. A **115**, 301 (1986).
[13] M. Znojil, Phys. Rev. D **34**, 1224 (1986); Phys. Lett. A **116**, 207 (1986).
[14] M. Znojil, Lett. Math. Phys. **5**, 405 (1981).
[15] R. N. Chaudhuri, M. Tater, and M. Znojil, J. Phys. A **20**, 1401 (1987).
[16] M. Znojil, Comm. JINR E5-87-480, 481 (Dubna, 1987).
[17] See the references in M. Znojil, Czech. J. Phys. B **37**, 1072 (1987).

# Local and global spectral shift functions in $\mathbb{R}^2$

D. Bollé and C. Danneels
*Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium*
T. A. Osborn
*Department of Physics, University of Manitoba, Winnipeg R3T 2N2, Canada*

(Received 16 December 1986; accepted for publication 31 August 1988)

The energy-dependent trace-class time-delay operator associated with the transit time of a scattering system through a finite space region $\Sigma \subseteq \mathbb{R}^2$ is used to define a local ($\Sigma$-dependent) version of the Krein spectral shift function. If the region $\Sigma$ is a disk of radius $r$, it is proved that as $r \to \infty$ the local spectral shift function converges, for almost all energies, to the original spectral shift function of Krein. This result continues to be valid for systems exhibiting zero-energy resonance behavior.

## I. INTRODUCTION

Consider the quantum-mechanical scattering in two dimensions of a single spinless particle from a potential $V(\mathbf{x})$. For this system we study representations of the Krein spectral shift function.[1] Let $\{H_0,H\}$ be the pair of free and interacting self-adjoint Hamiltonians that define the scattering system. The free Hamiltonian is the self-adjoint extension in $L^2(\mathbb{R}^2)$ of the negative Laplacian. The spectrum $\sigma(H_0) = [0,\infty)$ associated with $H_0$ is absolutely continuous. Take $R_0(z)$ to be the resolvent operator for $H_0$ defined for complex energies $z$ that lie within the resolvent set $\rho(H_0)$ [$R(z)$, $H$, and $\rho(H)$, respectively]. We recall the following basic result.[1-5]

**Theorem 1.1:** If $\{H,H_0\}$ is such that $[R(z_1) - R_0(z_1)]$ is trace class for some value $z_1 \in \rho(H) \cap \rho(H_0)$, then there exists a real-valued measurable function $\xi$ on $\mathbb{R}$ unique almost everywhere up to an additive constant satisfying

(a) $\xi(\lambda)(1+\lambda^2)^{-1} \in L^1(\mathbb{R})$; (1.1)

(b) $$\mathrm{Tr}[R(z) - R_0(z)] = \int_{-\infty}^{\infty} \xi(\lambda) d\left(\frac{1}{\lambda - z}\right), \quad z \in \rho(H) \cap \rho(H_0); \tag{1.2}$$

(c) if $s(\lambda)$ denotes an element of the $H_0$ direct product representation of the $S$ matrix, then for almost all $\lambda \in \sigma(H_0)$

$$\det s(\lambda) = e^{-2\pi i \xi(\lambda)}. \tag{1.3}$$

Clearly the spectral shift function $\xi(\lambda)$ of Krein is a form of generalized phase shift that remains meaningful even when the scattering potential $V(\mathbf{x})$ is not rotationally invariant. Many of the main features of a scattering system are determined by the knowledge of $\xi(\lambda)$. For example, the function $\xi$ plays a key role in the sum rules (Levinson's theorems) that relate the bound states of $H$ to the scattering continuum.

A second universal quantity in scattering theory is the time delay[6,7] associated with a given finite coordinate space region $\Sigma \subseteq \mathbb{R}^2$. The transit-time difference between the exact and free evolving systems through the region $\Sigma$ leads to the definition[8] of a family of trace-class operators $\tau_\Sigma(\lambda)$ acting on $L^2(\Omega)$, $\lambda \in \sigma(H_0)$, where $\Omega$ is the unit circle in $\mathbb{R}^2$. Denote by tr the trace on $L^2(\Omega)$. It has been previously suggested[9] that one could define a local spectral shift function $\xi_\Sigma(\lambda)$ by the formula

$$\xi_\Sigma(\lambda) = \frac{-\Theta(\lambda)}{2\pi} \int_0^\lambda d\lambda' \,\mathrm{tr}\, \tau_\Sigma(\lambda') - \sum_{\lambda_i < \lambda} \int_\Sigma dx\, |\psi_i(\mathbf{x})|^2. \tag{1.4}$$

Here the functions $\psi_i$ are the unit-normalized $L^2(\mathbb{R}^2)$ eigenfunctions of $H$ having eigenvalue $\lambda_i$. The symbol $\Theta(\lambda)$ is the right continuous form of the Heaviside step function, equal to 1 for non-negative arguments and zero otherwise. It is appropriate to call $\xi_\Sigma(\lambda)$ the local spectral shift function, since it satisfies[9,10] (for all $\Sigma$ with finite Lebesgue measure) a resolvent trace identity analogous to (1.2), i.e.,

$$\mathrm{Tr}\, P_\Sigma [R(z) - R_0(z)] P_\Sigma = \int_{-\infty}^{\infty} \xi_\Sigma(\lambda) d\left(\frac{1}{\lambda - z}\right), \quad z \in \rho(H) \cap \rho(H_0), \tag{1.5}$$

where $P_\Sigma$ is the projector on $L^2(\mathbb{R}^2)$ that is defined by multiplication with the characteristic function $\chi_\Sigma(\mathbf{x})$.

The basic goal of this paper is to characterize in a rigorous manner the way in which the local Krein function $\xi_\Sigma(\lambda)$ converges to the global Krein function $\xi(\lambda)$ as $\Sigma \to \mathbb{R}^2$. This problem has a connection with understanding the manner in which $\tau_\Sigma(\lambda)$ converges to the global time-delay operator defined by

$$\tau(\lambda) = -is^\dagger(k^2) \frac{d}{dk^2} s(k^2), \quad k^2 = \lambda > 0, \tag{1.6}$$

as $\Sigma \to \mathbb{R}^2$. In Eq. (1.6) the symbol $^\dagger$ denotes the adjoint. Results for this latter limit have been established[8-15] either in a distributional sense or in a weak sense for certain smooth functions on $L^2$. For the limiting problem involving the local and global Krein functions, simpler results emerge. Suppose $\Sigma$ is a disk of radius $r$; then we shall prove for a class of potentials having algebraic decay for large argument that $\xi_\Sigma$ converges to $\xi$ pointwise almost everywhere on $(-\infty,\infty)$ as $r \to \infty$.

In Sec. II we state the known results for two-dimensional scattering in a form suitable for the investigation of the local and global Krein functions. Section III provides a representation of $\xi(\lambda)$ in terms of the real-axis limiting value of $\mathrm{Tr}[R(z) - R_0(z)]$. The large-distance behavior of the scat-

 0022-2488/89/020420-13$02.50 © 1989 American Institute of Physics 

tering wave function[16] is used in Sec. IV to obtain an asymptotic relation between $\tau(\lambda)$ and $\tau_\Sigma(\lambda)$. Finally, Sec. V establishes the convergence of $\xi_\Sigma(\lambda)$ to $\xi(\lambda)$.

## II. SCATTERING IN TWO DIMENSIONS

We briefly summarize basic facts about scattering theory that are required in what is to follow. First, let us introduce some useful notation. Throughout the paper we employ the convention that $\hbar^2 = 2m = 1$. The symbols $\mathscr{B}$, $\mathscr{B}_2$, and $\mathscr{B}_1$ will represent, respectively, the Banach spaces of bounded, Hilbert–Schmidt, and trace-class operators on the $L^2$ spaces of interest in this problem—either $L^2(\mathbb{R}^2)$ or $L^2(\Omega)$. The inner product $(\cdot,\cdot)$ on $L^2(\mathbb{R}^2)$ is taken to be antilinear in the left argument. We denote by $\mathscr{S}(\mathbb{R}^2)$ the Schwartz space of rapidly decreasing functions in $\mathbb{R}^2$; $\mathbb{R}^+$ will be the set of positive reals. The complex energy variable takes values in the canonically cut plane $\Pi = \mathbb{C}\backslash\mathbb{R}^+$. Finally, $\Pi_+$ denotes the closure of $\Pi$ that maintains the distinction between the two possible boundary values along the positive real axis and omits the origin.

The potential $V$: $\mathbb{R}^2 \to \mathbb{R}$ is assumed to be a measurable function that is a member of the following class. Let the parameters be $g > 0$ and $\epsilon \geqslant 0$.

*Definition:* The function $V$ is said to be in the class $\mathscr{A}(g,\epsilon)$ if (i) $V \in L^2_{\rm loc}(\mathbb{R}^2) \cap L^{2+\epsilon}_{\rm loc}(\mathbb{R}^2)$, and (ii) for each $V$ there exists some $0 < x_0 < \infty$ such that

$$\sup_{|\mathbf{x}|>x_0} (1 + |\mathbf{x}|^g)|V(\mathbf{x})| = C(x_0,g) < \infty.$$

Condition (i) restricts the severity of the local singularities $V$ may have within the ball $S(x_0) = \{\mathbf{x}\in\mathbb{R}^2;\ |\mathbf{x}| \leqslant x_0\}$, whereas (ii) characterizes the decay of the potential for large $|\mathbf{x}|$. The potential class has the following ordering property: if $g_2 \geqslant g_1$ and $\epsilon_2 \geqslant \epsilon_1$, then $\mathscr{A}(g_2,\epsilon_2) \subseteq \mathscr{A}(g_1,\epsilon_1)$.

The class $\mathscr{A}$ is easily related to standard families of potentials for which scattering theory is fully developed. For example, $g > 1$ implies $\mathscr{A}(g,\epsilon) \subseteq L^2(\mathbb{R}^2) \cap L^{2+\epsilon}(\mathbb{R}^2)$. Recall that if the potential $V \in L^2(\mathbb{R}^2)$, then the associated operator $V$ [defined by maximal multiplicaiton with $V(\mathbf{x})$] is $H_0$ bounded. Thereby the perturbed Laplacian $(-\Delta + V)$ defined on $C_0^\infty(\mathbb{R}^2)$ has a unique self-adjoint extension[17] $H$. In addition, if $g > 1$, then $V$ is in the class of short-ranged potentials studied by Agmon,[18] who shows that for these potentials $H$ has no singular continuous spectrum and that $H$ has a finite number of negative eigenvalues, each with finite multiplicity. Increasing the value of $g$ we see that $g > 2$ leads to $\mathscr{A}(g,0) \subseteq L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$, and $g > n + 2$, $n \geqslant 0$, implies that $(1 + |\mathbf{x}|^n)V(\mathbf{x})$ is $L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$. For example, if $g > 2$ [thereby ensuring that $(1 + |\mathbf{x}|)V(\mathbf{x}) \in L^2(\mathbb{R}^2)$], then it is known[19] that $H$ has no embedded eigenvalues. Our most detailed results for the large $|\mathbf{x}|$ asymptotic expansions of the scattering wave functions will require $g > 8$. If $V$ is replaced by $\lambda_0 V$, where $\lambda_0$ is a real coupling constant, then, for certain values of $\lambda_0$, $H$ may have bound states and/or resonances at zero energy.[20] In this paper this possibility is taken explicitly into account.

Setting $z = k^2$, $\operatorname{Im} k > 0$, $k \neq 0$, the free resolvent

$$R_0(k^2) = (H_0 - k^2)^{-1} \tag{2.1}$$

has a Carleman kernel[21]

$$R_0(k^2;\mathbf{x},\mathbf{y}) = (i/4)H_0^{(1)}(k|\mathbf{x}-\mathbf{y}|), \quad \mathbf{x}\neq\mathbf{y}, \tag{2.2}$$

where $H_0^{(1)}(z)$ is the Hankel function of the first kind and order zero.

Introducing the factorization scheme

$$v(\mathbf{x}) = |V(\mathbf{x})|^{1/2}, \quad u(\mathbf{x}) = v(\mathbf{x})\operatorname{sgn} V(\mathbf{x}) \tag{2.3a}$$

allows one to define a symmetrized free resolvent

$$A(z) = uR_0(z)v, \quad z\in\Pi_+. \tag{2.3b}$$

Specifically, $A(z)$ is the integral operator determined by the kernel $u(\mathbf{x})R_0(z;\mathbf{x},\mathbf{y})v(\mathbf{y})$. If $g > 2$, the operator is Hilbert–Schmidt for all $z\in\Pi_+$. The exceptional set associated with $A$ is defined as

$$\mathscr{E} = \{k^2 > 0:\ A(k^2 + i0)\psi = -\psi, \text{ for some } \psi\in L^2(\mathbb{R}^2),\ k > 0\}.$$

The set $\mathscr{E}$ is a closed subset of $[0,\infty)$ with Lebesgue measure zero containing the singular continuous spectrum and the positive spectrum of $H$. Let $R_0'(z)$ denote the derivative of the free resolvent with respect to $z$. If $g > 3$, then

$$A'(z) = uR_0'(z)v \tag{2.3c}$$

is $\mathscr{B}_2$ valued for all $z\in\Pi_+$. These observations are an immediate consequence of the following estimates, valid for all $z = k^2$, $\operatorname{Im} k \geqslant 0$, $k \neq 0$:

$$\|A(k^2)\|_2^2 \leqslant c|k|^{-1}\|V\|_{4/3}^2, \tag{2.4}$$

$$\|A'(k^2)\|_2^2 \leqslant |k|^{-4}\left[c'\|V\|_1^2 + c''|k|\left(\int d\mathbf{x}(1 + |\mathbf{x}|)|V(\mathbf{x})|\right)^2\right], \tag{2.5}$$

where $c$, $c'$, and $c''$ are constants independent of $k$.

The utility of the operators $A(k^2)$ and $A'(k^2)$ is that they have well-defined $\mathscr{B}_2$ boundary values as $\operatorname{Im} k \to 0$. For this reason the limiting absorption principle is particularly easy to apply in this circumstance. For $g > 2$ we can define the family of transition operators by

$$T(k^2) = (1 + A(k^2))^{-1}, \quad \operatorname{Im} k \geqslant 0, \quad k^2 \notin \mathscr{E}\cup\sigma_p(H)\cup\{0\}. \tag{2.6}$$

The full resolvent is

$$R(k^2) = (H - k^2)^{-1}, \quad \operatorname{Im} k > 0, \quad k^2\notin\sigma_p(H), \tag{2.7}$$

and satisfies the second resolvent equation

$$R(k^2) = R_0(k^2) - R_0(k^2)vT(k^2)uR_0(k^2), \quad \operatorname{Im} k > 0, \quad k^2\notin\sigma_p(H). \tag{2.8}$$

Note that for $V\in L^1(\mathbb{R}^2)$, then $R_0(k^2)v$ and $uR_0(k^2)$ are Hilbert–Schmidt for all $\operatorname{Im} k > 0$.

In a similar fashion one defines a symmetrized pair of wave functions $\Phi^\pm(k\omega)\in L^2(\mathbb{R}^2)$, $k > 0$, $\omega\in\Omega$, as the solution of the inhomogeneous equation

$$\Phi^+(k\omega) = \Phi_0^+(k\omega) - A(k^2 + i0)^\dagger\Phi^+(k\omega), \tag{2.9a}$$

$$\Phi^-(k\omega) = \Phi_0^-(k\omega) - A(k^2 + i0)\Phi^-(k\omega), \tag{2.9b}$$

for all $k \neq 0$, $k^2 \notin \sigma_p(H)$, where the $\Phi_0^{\pm}(k\omega,\cdot)$ are defined by

$$\Phi_0^+(k\omega,\mathbf{x}) = v(\mathbf{x})e^{ik\omega\cdot x}, \quad \Phi_0^-(k\omega,\mathbf{x}) = u(\mathbf{x})e^{ik\omega\cdot\mathbf{x}}. \tag{2.10a}$$

In terms of the transition operators, one obtains $\Phi^{\pm}$ from $\Phi_0^{\pm}$ by

$$\begin{aligned}\Phi^+(k\omega) &= T(k^2+i0)\Phi_0^+(k\omega),\\ \Phi^-(k\omega) &= T(k^2+i0)\Phi_0^-(k\omega).\end{aligned} \tag{2.10b}$$

The functions $\Phi^{\pm}(k\omega)$ are closely related to the incoming and outgoing solutions $\Psi^{\pm}(k\omega,\mathbf{x})$ of the Lippmann–Schwinger equation

$$\Psi^{\pm}(k\omega,\mathbf{x}) = \Psi_0(k\omega,\mathbf{x}) - \int d\mathbf{y}\, R_0(k^2 \pm i0;\mathbf{x},\mathbf{y}) \times V(\mathbf{y})\Psi^{\pm}(k\omega,\mathbf{y}), \tag{2.11a}$$

where $\Psi_0(k\omega,\mathbf{x}) = e^{ik\omega\cdot\mathbf{x}}$ and $k>0$. The solutions $\Psi^{\pm}(k\omega,\mathbf{x})$ of (2.11a) may be recovered from $\Phi^{\pm}(k\omega)$ by

$$\Psi^+(k\omega,\mathbf{x}) = u(\mathbf{x})^{-1}\Phi^-(k\omega,\mathbf{x}), \quad \mathbf{x}\in\mathrm{supp}(u), \tag{2.11b}$$

etc. For $V\in\mathscr{A}(g,0)$, $g>1$, the functions $\Psi^{\pm}$ are the generalized eigenfunctions of $H$, which behave like plane waves as $|\mathbf{x}|\to\infty$. Specifically, for $V\in\mathscr{A}(g,0)$, $g>1$, the $\Psi^{\pm}(k\omega,\mathbf{x})$ are continuous functions of $\mathbf{x}$ with distributional derivatives $D_{\mathbf{x}}^{\alpha}\Psi^{\pm}(k\omega,\mathbf{x})$ that are $L^2_{\mathrm{loc}}(\mathbb{R}^2)$ for multi-indices $|\alpha|\leqslant 2$. For fixed $\mathbf{x}$ and $k>0$, the $\Psi^{\pm}(k\omega,\mathbf{x})$ belong to $L^2(\Omega)$ and are continuous [in the $L^2(\Omega)$ norm] as $(\mathbf{x},k)$ varies throughout $\mathbb{R}^2\times(0,\infty)$ [cf. Ref. 18, Theorem 5.1, part (ii)].

Scattering theory for such a system can be developed along standard lines. We require the following basic results.

**Theorem 2.1:** Let $V\in\mathscr{A}(g,0)$, $g>2$. Then the wave operators $\Omega_{\pm}$ associated with the Hamiltonian pair $\{H_0,H\}$ exist and are complete. The scattering operator $S$ is unitary and commutes with $H_0$, and in the $H_0$ direct integral representation the corresponding on-shell operator $s(k)$ is given by

$$(s(k)\phi)(\omega) = \phi(\omega) + (2\pi)^{-1/2}e^{i\pi/4}k^{1/2} \times \int_{\Omega} d\omega'\, f(k;\omega,\omega')\phi(\omega'), \tag{2.12}$$

for $k>0$, $\phi\in L^2(\Omega)$. The function $f\colon \mathbb{R}_+\times\Omega\times\Omega\to\mathbb{C}$ is the on-shell scattering amplitude determined by the following equivalent expressions ($k>0$):

$$\begin{aligned} f(k;\omega,\omega') &= -e^{i\pi/4}(8\pi k)^{-1/2}(\Phi_0^+(k\omega),\Phi^-(k\omega')) && (2.13a)\\ &= -e^{i\pi/4}(8\pi k)^{-1/2}(\Phi^+(k\omega),\Phi_0^-(k\omega')) && (2.13b)\\ &= e^{-3i\pi/4}(8\pi k)^{-1/2}(\Phi_0^+(k\omega),T(k^2-i0)\Phi_0^-(k\omega')). && (2.13c)\end{aligned}$$

Finally, the operator $[s(k)-1]$ is trace class on $L^2(\Omega)$ for $k>0$.

*Proof:* If $g>1$, then the existence and completeness of the wave operators, together with the unitarity of the scattering operator, are proved by Agmon in Ref. 18, Secs. 5 and 7. The formulas (2.13a)–(2.13c) for the scattering amplitude and $S$ matrix $s(k)$ are standard and follow if $g>2$ (Ref. 18 and Ref. 22, Chap. XI.8). Finally, the trace-class character of $s(k)-1$ follows from formulas (2.12) and (2.13c). The detailed verification that $s(k)-1$ is in $\mathscr{B}_1$ is based on a simple modification of the argument found in Ref. 23, Lemma 1. □

**Theorem 2.2:** Let $V\in\mathscr{A}(g,0)$, $g>3$. For $k>0$, then $(1+|\mathbf{x}|)^{1/2}\Phi^{\pm}(k\omega,\mathbf{x})$ is strongly continuous in $L^2(\mathbb{R}^2_{\mathbf{x}})$ with respect to $k$ for all $\omega\in\Omega$, and furthermore $f(k,\omega,\omega')$ is continuously differentiable with respect to $k$ for all $\omega,\omega'\in\Omega\times\Omega$. In particular, $s(\cdot)\colon \mathbb{R}_+\to\mathscr{B}_1(L^2(\Omega))$ is continuously differentiable in the trace norm. In addition, for $k>0$, $s'(k)$ is trace class on $L^2(\Omega)$ and $s'(k)\colon L^2(\Omega)\to L^2(\Omega)$ is an integral operator with a Hilbert–Schmidt kernal given by

$$(s'(k)\phi)(\omega) = \int_{\Omega} d\omega' \frac{d}{dk}[(2\pi)^{-1/2}e^{i\pi/4}k^{1/2}f(k,\omega,\omega')]\phi(\omega'), \tag{2.14}$$

for $\phi\in L^2(\Omega)$.

*Proof:* A statement parallel to Theorem 2.2 for scattering in $\mathbb{R}^3$ is found in Ref. 23, Lemma 1. Adjusting this derivation of Ref. 23 from $\mathbb{R}^3$ to $\mathbb{R}^2$ gives the result above. □

Finally we recount the finite-region time-delay formalism for our system. For a class of smooth but dense states $\phi\in L^2(\mathbb{R}^2)$ and each measurable set $\Sigma\subseteq\mathbb{R}^2$ having finite measure, the difference in transit times of the exact and free system through region $\Sigma$ is defined as[8,10,13]

$$\Delta T_{\Sigma}^{\pm}(\phi) = \int_{-\infty}^{\infty} dt\, [\|P_{\Sigma}e^{-iHt}\Omega_{\pm}\phi\|^2 - \|P_{\Sigma}e^{-iH_0t}\phi\|^2]. \tag{2.15}$$

The time-dependent functions $e^{-iHt}\Omega_{\pm}\phi$ and $e^{-iH_0t}\phi$ are the exact and free evolving states that converge in the $L^2$ norm to each other as $t\to\pm\infty$. The value of $\Delta T_{\Sigma}^{\pm}(\phi)$ is independent of the choice $\pm$.

The well-known result of Ref. 8 (Theorem 2) constructs the $H_0$ direct product representation associated with $\Delta T_{\Sigma}^{\pm}(\phi)$. The representations (2.16) and (2.17) below are proved in Refs. 10 and 14.

**Theorem 2.3:** Let $\Sigma$ be a measurable subset of $\mathbb{R}^2$ with finite Lebesgue measure. Then for the scattering system $\{H_0,H\}$ defined by $V\in\mathscr{A}(3,0)$ there exists a unique measurable family $\tau_{\Sigma}^{\pm}(k)$ of trace-class operators in $L^2(\Omega)$, i.e., $\tau_{\Sigma}^{\pm}(\cdot)\colon (0,\infty)\to\mathscr{B}_1(L^2(\Omega))$, interpreted as the energy-shell time-delay operator, such that $\Delta T_{\Sigma}^{\pm}(\phi)$ can be written

$$\Delta T_{\Sigma}^{\pm}(\phi) = \int_0^{\infty} dk \int_{\Omega} d\omega (U_0\phi)(k,\omega)^*(\tau_{\Sigma}^{\pm}(k)U_0\phi)(k,\omega), \quad \phi\in\mathscr{D}_0, \tag{2.16}$$

where $\mathscr{D}_0 = \{\phi\in\mathscr{S}(\mathbb{R}^2)\colon \mathrm{supp}(\tilde{\phi}) \text{ is compact}\}$ with $\tilde{\phi}$ the Fourier transform of $\phi$, and where $U_0$ is the following isomorphism, defining the direct product decomposition of $H_0$:

$$U_0\colon \begin{cases} L^2(\mathbb{R}^2)\to L^2((0,\infty);L^2(\Omega)),\\ g(x)\to(U_0g)(k\omega) = k\tilde{g}(k\omega), \quad k>0, \quad \omega\in\Omega. \end{cases}$$

Furthermore, the kernel of $\tau_\Sigma^\pm(k)$ is given by

$$\tau_\Sigma^\pm(k,\omega,\omega') = \int_\Sigma d\mathbf{x}[\Psi^\pm(k\omega,\mathbf{x})^*\Psi^\pm(k\omega',\mathbf{x}) - \Psi_0(k\omega,\mathbf{x})^*\Psi_0(k\omega',\mathbf{x})]. \quad (2.17)$$

## III. REPRESENTATIONS OF THE SPECTRAL SHIFT FUNCTION

In this section we develop two representations of the spectral shift function $\xi(\lambda)$ in terms of the positive energy $z=\lambda\pm i0$, $\lambda>0$, boundary values of $\mathrm{Tr}[R(z)-R_0(z)]$. The first representation is a consequence of Cauchy's residue theorem for meromorphic functions applied to $\mathrm{Tr}[R(z)-R_0(z)]$. The second is the result of combining the first representation with the known form of Levinson's theorem[24–26] in $\mathbb{R}^2$.

It is useful to introduce a family of operators related to $T(z)$. If the decay parameter in $\mathscr{A}$ is $g>3$, then define $\tilde{T}$: $\Pi_+\to\mathscr{B}_2$ by

$$\tilde{T}(z)\equiv A(z)T(z). \quad (3.1)$$

Since $(1+A(z))^{-1}$ is a bounded operator, (3.1) and (2.6) give us

$$T(z)=1-A(z)T(z)=1-\tilde{T}(z), \quad (3.2)$$

$$T(z)=1-\tilde{T}(z). \quad (3.3)$$

The next two lemmas characterize the analytic behavior in $z$ of our system and its boundary-value behavior as $z\to\lambda\pm i0$, $\lambda>0$.

*Lemma 3.1:* Let $V\in\mathscr{A}(g,0)$, $g>3$; then the following statements hold true.

(a) $A(\cdot)\colon \Pi_+\to\mathscr{B}_2(L^2(\mathbb{R}^2))$ is $\|\cdot\|_2$ continuous.

(b) For $z\in\Pi_+$ there exist finite constants $c_1$, $c_2$ independent of $\mathbf{x}\in\mathbb{R}^2$ such that

$$\|\mathbb{R}_0(z;\mathbf{x},\cdot)u(\cdot)\|_2^2 < c_1\|V\|_1 + (c_2/|z|^{1/2})\|V\|_2. \quad (3.4)$$

(c) $\tilde{T}(\cdot)\colon \Pi_+\setminus\sigma_p(H)\to\mathscr{B}_2(L^2(\mathbb{R}^2))$ is $\|\cdot\|_2$ continuous.

(d) $[R(\cdot)-R_0(\cdot)]\colon \Pi\setminus\sigma_p(H)\to\mathscr{B}_1(L^2(\mathbb{R}^2))$ defines a $\mathscr{B}_1$ meromorphic function on $\Pi$, with simple poles at $z\in\sigma_p(H)$. Furthermore,

$$R(z)-R_0(z) = -[R_0(z)v][uR_0(z)] + R_0(z)v\tilde{T}(z)uR_0(z), \quad z\in\Pi\setminus\sigma_p(H). \quad (3.5)$$

(e) For all $z\in\Pi$,

$$\mathrm{Tr}[R_0(z)v][uR_0(z)] = \frac{-1}{4\pi z}\int d\mathbf{x}\, V(\mathbf{x}). \quad (3.6)$$

*Proof:* (a) This follows from the Hilbert–Schmidt norm formula,

$$\|A(z_1)-A(z_2)\|_2^2 = \int\int d\mathbf{x}\,d\mathbf{y}\,|V(\mathbf{x})|\,|V(\mathbf{y})|\,|R_0(z_1;\mathbf{x},\mathbf{y}) - R_0(z_2;\mathbf{x},\mathbf{y})|^2, \quad (3.7)$$

the uniform estimate (Ref. 27, pp. 962–963) for the Hankel function,

$$|H_0^{(1)}(\alpha)| < c|\alpha|^{-1/2}e^{-\mathrm{Im}\,\alpha} \quad (3.8)$$

for all $\alpha\in\mathbb{C}\setminus\{0\}$ with $\mathrm{Im}\,\alpha\geqslant 0$, and the resolvent formula (2.2).

(b) This assertion claims that $R_0(z;\mathbf{x},\cdot)u(\cdot)\colon \mathbb{R}^2\to\mathbb{C}$ is an $L^2(\mathbb{R}^2)$ function. The definition of the $L^2$ norm and the free resolvent kernel expression (2.2) gives us

$$\|R_0(z;\mathbf{x},\cdot)u(\cdot)\|_2^2 = \frac{1}{16}\int d\mathbf{y}\,|V(\mathbf{y})|\,|H_0^{(1)}(\sqrt{z}|\mathbf{x}-\mathbf{y}|)|^2. \quad (3.9)$$

Bound (3.4) results as a consequence of dividing the integral (3.9) into two regions where $\sqrt{z}|\mathbf{x}-\mathbf{y}|\gtrless 1$. For the region with $\sqrt{z}|\mathbf{x}-\mathbf{y}|>1$ use bound (3.8) for $H_0^{(1)}$. In the region $\sqrt{z}|\mathbf{x}-\mathbf{y}|<1$ we use an alternative bound, namely, $|H_0^{(1)}(\sqrt{z}|\mathbf{x}-\mathbf{y}|)| < c\,|\ln|\sqrt{z}|\mathbf{x}-\mathbf{y}||$ [Ref. 27, Eq. 8.444(1)].

(c) This statement is a consequence of the $\mathscr{B}_2$ continuity of $A(z)$, formula (3.1), and the identity

$$T(z_1)-T(z_2) = T(z_2)\{1-[A(z_2)-A(z_1)]T(z_2)\}^{-1}\times[A(z_2)-A(z_1)]T(z_2). \quad (3.10)$$

On the right side of (3.10) the first, second, and fourth operators are bounded as $z_1\to z_2$, while the factor $A(z_1)-A(z_2)$ vanishes in the $\|\cdot\|_2$ norm by virtue of (a).

(d) Both $R(\cdot)$ and $R_0(\cdot)$ are $\mathscr{B}$ holomorphic on $\rho(H)$ and $\rho(H_0)$, respectively. Since $R_0(z)v$, $uR_0(z)$ are in $\mathscr{B}_2$ while $T(z)$ is bounded for $z\in\Pi\setminus\sigma_p(H)$, Eq. (2.8) shows that the resolvent difference is trace class. In addition, Eq. (3.5) results from (2.8) and (3.3).

(e) Since $R_0(z)v$ and $uR_0(z)$ are in $\mathscr{B}_2$ with $L^2(\mathbb{R}^2\times\mathbb{R}^2)$ kernels, $[R_0(z)v][uR_0(z)]$ is trace class, and the trace may be evaluated by the usual diagonal integral formula (Ref. 17, p. 524) and well-known properties of $H_0^{(1)}$. □

*Lemma 3.2:* Let $V\in\mathscr{A}(g,0)$, $g>3$; then the following assertions are true.

(a) $uR_0'(\cdot)v\colon \Pi_+\to\mathscr{B}_2(L^2(\mathbb{R}^2))$ is $\|\cdot\|_2$ continuous.

(b) $\mathrm{Tr}[R(\cdot)-R_0(\cdot)]\colon \Pi_+\setminus\sigma_p(H)\to\mathbb{C}$ is continuous. Furthermore, the boundary value of $\mathrm{Tr}[R(\cdot)-R_0(\cdot)]$ on $\mathbb{R}_+$ can be represented by ($\lambda>0$)

$$B^\pm(\lambda) = \mathrm{Tr}\,\tilde{T}(\lambda\pm i0)[uR_0'(\lambda\pm i0)v] + \frac{1}{4\pi\lambda}\int_{\mathbb{R}^2} d\mathbf{x}\,V(\mathbf{x}). \quad (3.11)$$

*Proof:* (a) Follows from the application of the Hilbert–Schmidt norm formula for operators with $L^2(\mathbb{R}^2\times\mathbb{R}^2)$ kernels. In formula (3.7) replace $R_0(\cdot;\mathbf{x},\mathbf{y})$ with $R_0'(\cdot;\mathbf{x},\mathbf{y})$. The $\mathscr{B}_2$ continuity of $uR_0'(z)v$ is a consequence of the pointwise continuity of the kernel $R_0'(z;\mathbf{x},\mathbf{y})$ and an application of the dominated convergence theorem.

(b) Let $z\in\Pi\setminus\sigma_p(H)$ and take the trace of the operators in Eq. (3.5). After employing the cyclic property of the trace one has

$$\mathrm{Tr}[R(z)-R_0(z)] = -\mathrm{Tr}[R_0(z)v][uR_0(z)] + \mathrm{Tr}\,\tilde{T}(z)uR_0'(z)v. \quad (3.12)$$

In (3.12) we have used the identity $R_0'(z) = R_0^2(z)$. Statement (e) of Lemma 3.1 shows the first term on the right of (3.12) to be continuous for $z \in \Pi_+$. The second term on the right of (3.12) is the trace of the product of two $\mathscr{B}_2$-continuous operators. □

The following two lemmas describe, respectively, the $|z| \to \infty$ and the $|z| \to 0$ behavior of the system $\{H_0, H\}$.

*Lemma 3.3:* Let $V \in \mathscr{A}(g,0)$, $g > 3$; then the following remarks hold true.

(a) There are constants $\Lambda_0 > 0$, $c < \infty$, such that

$$\|B^{\pm}(\lambda)\| < c\lambda^{-1}, \quad \lambda \in [\Lambda_0, \infty). \tag{3.13}$$

(b) For positive integers $n$,

$$\operatorname{Tr} A(z)^n u R_0'(z) v = \frac{1}{n+1} \frac{d}{dz} \operatorname{Tr} A(z)^{n+1}, \quad z \in \Pi_+. \tag{3.14}$$

(c) Suppose $C(\Gamma)$ denotes the circular contour in $\Pi_+$ with center $z = 0$ and radius $\Gamma > 0$, which starts at $\Gamma + i0$ and ends at $\Gamma - i0$. Then

$$\lim_{\Gamma \to \infty} \int_{C(\Gamma)} dz \operatorname{Tr} \tilde{T}(z) u R_0'(z) v = 0. \tag{3.15}$$

*Proof:* (a) The large $\lambda$ decay estimate (3.13) follows from expression (3.11) for $B^{\pm}(\lambda)$ plus the fact that $\|\tilde{T}(\lambda \pm i0)\|_2$ decays like $\lambda^{-1/4}$ and that $\|uR_0'(\lambda \pm i0)v\|_2$ decays like $\lambda^{-3/4}$ [cf. Eq. (2.5)].

(b) The differential identity (3.14) results from the fact that $A(\cdot): \Pi_+ \to \mathscr{B}_2(L^2(\mathbb{R}^2))$ is continuously differentiable in the $\|\cdot\|_2$ norm with derivative $uR_0'(z)v \in \mathscr{B}_2(L^2(\mathbb{R}^2))$. Applying the chain rule for derivatives to $A^{n+1}(z)$ $(n \geqslant 1)$ gives (3.14).

(c) Consider (3.15). The operator identity (3.3) for $T(z)$ and the definition (3.1) for $\tilde{T}(z)$ implies that

$$\tilde{T}(z) = A(z) - A(z)^2 T(z). \tag{3.16}$$

Thus the integrand of (3.15) may be written

$$\operatorname{Tr} \tilde{T}(z) u R_0'(z) v = \operatorname{Tr} A(z) u R_0'(z) v - \operatorname{Tr} A(z)^2 T(z) u R_0'(z) v. \tag{3.17}$$

Note that $T(z)$ is uniformly bounded in $\Pi_+$ for large $z$ since estimate (2.4) says that $\|A(z)\| < 1$ for all $|z| > c^2 \|V\|_{4/3}^4$. We may therefore conclude that the rightmost term in (3.17) is $O(|z|^{-5/4})$ and so gives a vanishing contribution to the $C(\Gamma)$ contour integration as $\Gamma \to \infty$. In view of (3.14), the remaining term from (3.17) can be written as an exact differential. The integral of this term is proportional to $[\operatorname{Tr} A(\Gamma + i0)^2 - \operatorname{Tr} A(\Gamma - i0)^2]$. As $\Gamma \to \infty$, the preceding function in the square bracket vanishes as $\Gamma^{-1/2}$. □

The following lemma gives us an analytic characterization of the small-energy behavior of our system, taking into account the possible occurrence of zero-energy states of the Hamiltonian. Under suitable restrictions on the potential one can describe these states by introducing the function $\phi_0$ as the solution of [20,25,26]

$$\lambda_0 Q M_{00} Q \phi_0 = -\phi_0, \quad \phi_0 \in L^2(\mathbb{R}^2), \quad Q = 1 - P, \quad P = (v,u)^{-1}(v,\cdot)u, \tag{3.18}$$

where $\lambda_0$ is the coupling constant, and $M_{00}$ is a $\mathscr{B}_2(L^2(\mathbb{R}^2))$ operator for all $\operatorname{Im} k \geqslant 0$ obtained from $uR_0(z)v$ by first subtracting out the logarithmic singularity at $z = 0$ and then taking $z = 0$. Its kernel reads

$$M_{00}(\mathbf{x},\mathbf{y}) = (-1/2\pi) u(\mathbf{x}) v(\mathbf{y}) \ln|\mathbf{x} - \mathbf{y}|, \quad \mathbf{x} \neq \mathbf{y}.$$

The function $\phi_0$ can be related to the solution [25,26] $\Psi_0$ of the zero-energy Schrödinger equation, viz.,

$$\psi_0(\mathbf{x}) = -(v,u)^{-1}\lambda_0(v, M_{00}\phi 0) - (2\pi)^{-1}\lambda_0 \int d\mathbf{y} [\ln|\mathbf{x} - \mathbf{y}|] v(\mathbf{y}) \phi_0(\mathbf{y}), \tag{3.19a}$$

$$u(\mathbf{x})\psi_0(\mathbf{x}) = -\phi_0(\mathbf{x}). \tag{3.19b}$$

Looking essentially at the large $|\mathbf{x}|$ behavior of the wave functions $\psi_0(\mathbf{x})$ in (3.19a) one then finds the following possibilities.

Case I: Equation (3.18) has no solutions or, equivalently, there exists no zero-energy state $\Psi_0$.

Case II: Equation (3.18) has $N \leqslant 3$ solutions $\phi_0$ satisfying $\phi_0 \in L^2(\mathbb{R}^2)$, $\Psi_0 \notin L^2(\mathbb{R}^2)$, so that they are all zero-energy resonances. A further distinction has to be made according to exactly what terms survive in the large $|\mathbf{x}|$ limit of (3.19a). Roughly speaking, the following characterization emerges:

(a) $s$-wave type zero-energy resonance: $N = 1$;

(b) $p$-wave type zero-energy resonances: $N \leqslant 2$;

(c) combination of (a) and (b): $2 \leqslant N \leqslant 3$.

Case III: Equation (3.18) has $N$ solutions $\phi_0$, $N$ finite, satisfying $\phi_0 \in L^2(\mathbb{R}^2)$, $\Psi_0 \in L^2(\mathbb{R}^2)$, so that they are all zero-energy bound states.

Case IV: Admixtures of cases II and III.

Then one can show that the following lemma is true (for more details refer to Ref. 26).

*Lemma 3.4:* Let $V \in \mathscr{A}(g,0)$, $g > 4$. If the zero-energy behavior is not of case I or II(a), assume $g > 10$. Then for all $|z| < \eta$ for some $\eta > 0$ and $z \in \Pi_+$,

$$|\operatorname{Tr}[R(z) - R_0(z)] + D/z| < c/(|z|\,|\ln z|), \tag{3.20}$$

$$|\operatorname{Im} \operatorname{Tr}[R(\lambda + i\epsilon) - R_0(\lambda + i\epsilon)]| < c'[\epsilon/(\lambda^2 + \epsilon^2)], \quad \epsilon > 0, \tag{3.21}$$

where $c$ and $c'$ are finite constants and $D$ is given by

$$D^{\mathrm{I}} = 0, \quad D^{\mathrm{II(a)}} = 0, \quad D^{\mathrm{II(b)}} = N, \quad D^{\mathrm{II(c)}} = (N-1), \quad D^{\mathrm{III}} = N, \tag{3.22}$$

$D^{\mathrm{IV}}$ is a combination of $D^{\mathrm{II}}$ and $D^{\mathrm{III}}$.

*Proposition 3.5:* Let $V$ satisfy the hypothesis of Lemma 3.4. Suppose $\lambda_i$ is a negative eigenvalue of $H$ with multiplicity $n_i \geqslant 1$. The eigenvalues will be indexed by magnitude, $\lambda_{i+1} < \lambda_i$. Set $\lambda_0 = 0$ and $n_0 = D$ (of Lemma 3.4). Then the Krein function is (almost everywhere)

$$\xi(\lambda) = -\sum_{i>0} n_i \theta(\lambda - \lambda_i) - \frac{\theta(\lambda)}{\pi} \int_0^{\lambda} d\lambda' \operatorname{Im} B^{+}(\lambda') + \text{const.} \tag{3.23}$$

*Proof:* Define three integration paths in $\Pi$. Choose $\Gamma > |\inf \sigma_p(H)|$ and $|\lambda_1| > \epsilon > \delta > 0$. Paths $C_1(\epsilon,\delta)$ and $C_2(\Gamma,\delta)$ are segments of circular paths about $z=0$. The path $C_1$ has radius $\epsilon$ and $C_2$ has radius $\Gamma$. For both $C_1$ and $C_2$ the portions of the circle in the right half-plane with $|\mathrm{Im}\, z| < \delta$ are absent. Path $C_3(\epsilon,\Gamma,\delta)$ consists of two straight lines parallel to the real axis with imaginary part $\pm i\delta$ such that $C_1 + C_2 + C_3$ are joined into a closed contour. All the poles of $\mathrm{Tr}[R(z) - R_0(z)]$ fall inside the interior of the region defined by the contour $C_1 + C_2 + C_3$. The residue theorem asserts

$$\frac{1}{2\pi i}\int_{C_1+C_2+C_3} dz' \frac{\mathrm{Tr}[R(z') - R_0(z')]}{z' - z} = \mathrm{Tr}[R(z) - R_0(z)] + \sum_{i\geq 1} \mathrm{Res} \frac{\mathrm{Tr}[T(z') - R_0(z')]}{z' - z}\Big|_{z'=\lambda_i} = \mathrm{Tr}[R(z) - R_0(z)] - \sum_{i\geq 1} \frac{n_i}{\lambda_i - z}, \tag{3.24}$$

for $z \neq \lambda_i$.

The right side of (3.24) is independent of $\delta$, $\epsilon$, and $\Gamma$. In Eq. (3.24) we shall take the limits $\delta \to 0$, then $\epsilon \to 0$, and finally $\Gamma \to \infty$. Because $\mathrm{Tr}[R(z') - R_0(z')]$ is uniformly continuous on the compact sets $[\frac{1}{2}\epsilon,\Gamma] \times [0i, \pm i]$, it follows that the $C_3$ integral can be evaluated on the real interval $[\epsilon,\Gamma]$. Specifically using (3.6), (3.11), (3.12), the Schwartz reflection property for $\mathrm{Tr}[R(z') - R_0(z')]$, and the bound (3.21) one finds

$$\lim_{\epsilon\to 0}\lim_{\delta\to 0} \frac{1}{2\pi i}\int_{C_3(\epsilon,\Gamma,\delta)} dz' \frac{\mathrm{Tr}[R(z') - R_0(z')]}{z' - z} = \frac{1}{\pi}\int_0^\Gamma d\lambda' \frac{\mathrm{Im}\, B^+(\lambda')}{\lambda' - z}. \tag{3.25}$$

Next observe that (3.12), (3.6), estimate (2.5), and the fact that $\|\tilde{T}(z)\|_2 = O(|z|^{-1/4})$ imply that for large $|z|$ one has $\mathrm{Tr}[R(z') - R_0(z')] = O(|z'|^{-1})$. Thus as $\Gamma \to \infty$ the $C_2(\Gamma,0)$ contribution on the left of (3.24) vanishes. Also, since $\mathrm{Im}\, B^+(\lambda') = O(\lambda'^{-1})$, the upper limit in (3.25) is absolutely convergent as $\Gamma \to \infty$. Finally, take the $\epsilon \to 0$ limit of the $C_1(\epsilon,0)$ term in (3.24). Using the bound (3.20) one finds without difficulty

$$\lim_{\epsilon\to 0}\int_{C_1(\epsilon,0)} dz' \frac{\mathrm{Tr}[R(z') - R_0(z')]}{z' - z} = -\frac{D}{z} - \frac{n_0}{z}. \tag{3.26}$$

Altogether we have shown

$$\frac{1}{\pi}\int_0^\infty d\lambda' \frac{\mathrm{Im}\, B^+(\lambda')}{\lambda' - z} = \mathrm{Tr}[R(z) - R_0(z)] - \sum_{i\geq 0} \frac{n_i}{\lambda_i - z}. \tag{3.27}$$

The $i = 0$ term in the sum accounts for any zero-energy resonant behavior. In the subsequent theorem on Levinson's sum rule it is shown that the $\Gamma \to \infty$ limit of $\int_1^\Gamma d\lambda'\, \mathrm{Im}\, B^+(\lambda')$ exists and is finite. Thus an integration by parts of the left-hand side of (3.27) gives

$$\frac{1}{\pi}\int_0^\infty d\lambda \frac{\mathrm{Im}\, B^+(\lambda)}{\lambda - z} = -\frac{1}{\pi}\int_0^\infty \left[\int_0^\lambda d\lambda'\, \mathrm{Im}\, B^+(\lambda')\right] d\left(\frac{1}{\lambda - z}\right). \tag{3.28}$$

Continue by writing the summation in (3.27) as

$$-\sum_{i\geq 0} \frac{n_i}{\lambda_i - z} = \int_{-\infty}^\infty \sum_{i\geq 0} n_i \theta(\lambda - \lambda_i) d\left(\frac{1}{\lambda - z}\right). \tag{3.29}$$

Putting together the last three equations gives us

$$\mathrm{Tr}[R(z) - R_0(z)] = \int_{-\infty}^\infty \left[-\sum_{i\geq 0} n_i \theta(\lambda - \lambda_i) - \frac{\theta(\lambda)}{\pi} \times \int_0^\lambda d\lambda'\, \mathrm{Im}\, B^+(\lambda')\right] d\left(\frac{1}{\lambda - z}\right). \tag{3.30}$$

Representation (3.30) has the same integral form for $\mathrm{Tr}[R(z) - R_0(z)]$ as does the Krein function result (1.2). Using the uniqueness property (up to a constant) of the Krein function we conclude that (3.23) is valid. □

In the remainder of this section, we shall employ the 2-D Levinson's theorem in order to obtain from (3.23) a second representation of the Krein function $\xi$. This second representation is essential in the study of the convergence problem that relates the local and global Krein functions to each other.

**Theorem 3.6:** Let $V$ satisfy the hypothesis of Lemma 3.4. Let $n_T = \Sigma_{i\geq 1} n_i$ be the total number of independent negative-energy eigenfunctions of $H$. Then $\mathrm{Im}\, B^+(\cdot)$: $(0,\infty) \to \mathbb{R}$ has an improper Riemann integral on the half axis $(0,\infty)$, which obeys

$$\int_0^{\infty -} d\lambda\, 2\, \mathrm{Im}\, B^+(\lambda) = -2\pi n_T - 2\pi D - \frac{1}{2}\int_{\mathbb{R}^2} d\mathbf{x}\, V(\mathbf{x}). \tag{3.31}$$

The value of the constant $D$ is that found in Lemma 3.4.

*Proof:* The proof of (3.31) is well known[24,26] and so we shall just make a few remarks. The proof is based on the meromorphy of $\mathrm{Tr}[R(z) - R_0(z)]$ in the cut plane $\Pi$ together with knowledge of the small and large $z$ behavior of this function. One starts the derivation with the function

$$F(z) \equiv \mathrm{Tr}[R(z) - R_0(z)] + \frac{D}{z} - \frac{1}{4\pi z}\int d\mathbf{x}\, V(\mathbf{x}), \tag{3.32}$$

and, by applying Cauchy's theorem on the same closed contour $C_1(\epsilon,\delta) + C_2(\Gamma,\delta) + C_3(\epsilon,\Gamma,\delta)$ utilized in Proposition 3.5, maintains the order in which the limits $\delta \to 0$, $\epsilon \to 0$, $\Gamma \to \infty$ are taken. The uniform continuity of $\mathrm{Tr}[R(z)$

$-R_0(z)]$ on $[\frac{1}{2}\epsilon,\Gamma]\times[0,\pm i]$ is used to evaluate the $\delta\to 0$ limit of $C_3$. The $\epsilon\to 0$ limit of $C_1+C_3$ is, by virtue of Lemma 3.4, found to be

$$\lim_{\epsilon\to 0^+}\frac{1}{2\pi i}\int_{C_1(\epsilon,0)+C_3(\epsilon,\Gamma,0)} dz\,F(z)$$
$$=\frac{1}{\pi}\int_0^{\Gamma} d\lambda\,\mathrm{Im}\,B^+(\lambda)+\frac{1}{4\pi}\int d\mathbf{x}\,V(\mathbf{x}). \qquad (3.33a)$$

The large contour term $C_2$ is evaluated by using (3.12) and (3.15). It is found that

$$\lim_{\Gamma\to\infty}\frac{1}{2\pi i}\int_{C_2(\Gamma,0)} dz\,F(z)=D. \qquad (3.33b)$$

Results (3.33) suffice to show that $\mathrm{Im}\,B^+(\cdot)$ has an improper integral on $(0,\infty)$ and that identity (3.31) is obeyed. Note that the asymptotic estimate (3.13) is not strong enough to show that $\mathrm{Im}\,B^+(\cdot)$ is absolutely integrable at $\infty$.

We remark that as long as one guarantees the absence of positive eigenvalues and of the singular continuous spectrum of $H$, Levinson's theorem (3.31) can be derived under the weaker conditions $V\in L^{4/3}(\mathbb{R}^2)$ and $(1+|\mathbf{x}|)V(\mathbf{x})\in L^1(\mathbb{R}^2)$ [resp. $(1+|\mathbf{x}|^{2+\delta})V(\mathbf{x})\in L^1(\mathbb{R}^2)$] in case I [resp. case II(a)]. In the other cases one needs in addition the stronger decay requirement $(1+|x|^{8+\delta})V(x)\in L^1(\mathbb{R}^2)$, $\delta>0$. This latter restriction on $V$ is obeyed if $g>10$. For more details refer to Ref. 26. □

**Theorem 3.7:** Let $V$ be as in Lemma 3.4. Then for almost all $\lambda>0$,

$$\xi(\lambda)=\frac{1}{\pi}\int_{\lambda}^{\infty} d\lambda'\,\mathrm{Im}\,B^+(\lambda')$$
$$+\frac{1}{4\pi}\int_{\mathbb{R}^2} d\mathbf{x}\,V(\mathbf{x})+\text{const.} \qquad (3.34)$$

*Proof:* Equation (3.34) results from adding identity (3.31) to representation (3.23). □

*Corollary 3.8:* Let $V$ be as in Lemma 3.4. Then for all $\lambda>0$,

$$2\,\mathrm{Im}\,B^+(\lambda)=-i\frac{d}{d\lambda}\,\mathrm{tr}\ln s(\lambda). \qquad (3.35)$$

*Proof:* Representation (3.34) asserts that $\xi(\lambda)$ is continuously differentiable for $\lambda>0$ and that

$$\frac{d\xi(\lambda)}{d\lambda}=-\frac{1}{\pi}\,\mathrm{Im}\,B^+(\lambda). \qquad (3.36)$$

If one takes the logarithm of (1.3) and then the $\lambda$ derivative of the resultant identity, it is found that

$$-2\pi i\frac{d\xi(\lambda)}{d\lambda}=\frac{d}{d\lambda}\ln\det s(\lambda)=\frac{d}{d\lambda}\,\mathrm{tr}\ln s(\lambda). \qquad (3.37)$$

Taken together these two equations give (3.35). Since both $\mathrm{Im}\,B^+(\lambda)$ and $(d/d\lambda)\,\mathrm{tr}\ln s(\lambda)$ are continuous functions, identity (3.35) holds for all $\lambda>0$. □

## IV. TIME-DELAY ASYMPTOTICS

In this section we discuss the limit as $\Sigma\to\mathbb{R}^2$ of the trace of the energy-shell time-delay operator $\mathrm{tr}\,\tau_\Sigma^+(k)$, defined by Theorem 2.3. In particular, taking $\Sigma$ to be a disk of radius $r=|\mathbf{x}|$, we will use representation (2.17) for the kernel of the time-delay operator to obtain this limit. So we first have to study the asymptotic behavior as $r\to\infty$ of the wave functions $\Psi_0(k\omega,\mathbf{x})$, $\Psi^+(k\omega,\mathbf{x})$. For fixed values of $k$ and $\mathbf{x}$ it is helpful to interpret $\Psi_0(k\cdot,\mathbf{x})$ and $\Psi^+(k\cdot,\mathbf{x})$ as distributions. Let $\mathscr{D}'$ be the space of distributions dual to $C^\infty(\Omega)$. The continuous linear functional defined by an element $d\in\mathscr{D}'$ we denote as $\langle d,F\rangle$, $F\in C^\infty(\Omega)$. Both $\Psi_0(k\cdot,\mathbf{x})$ and $\Psi^+(k\cdot,\mathbf{x})$ are in $\mathscr{D}'$. We note that our treatment is a generalization of Buslaev's methods[16] to two dimensions and to the more general class of potentials satisfying $\mathscr{A}(g,\epsilon)$.

To determine the wave function asymptotics in two dimensions we prove the following results.

*Lemma 4.1:* The following asymptotic expansion for $R_0(k^2+i0;\mathbf{x},\mathbf{y})$ is valid as $|\mathbf{x}|\to\infty$, $|\mathbf{y}|/|\mathbf{x}|<1$, $k>0$:

$$R_0(k^2+i0;\mathbf{x},\mathbf{y})$$
$$=\frac{i}{(4\pi^{1/2})}e^{ik|\mathbf{x}|}\sum_{n=0}^{N}\frac{1}{n!}\frac{1}{(2ik|\mathbf{x}|)^{n+1/2}}$$
$$\times B_n(\Delta_{\omega'})e^{-ik\omega'\cdot\mathbf{y}}+R^G_{N+1}(k^2;\mathbf{x},\mathbf{y}), \qquad (4.1)$$

where

$$|R^G_{N+1}(k^2;\mathbf{x},\mathbf{y})|$$
$$<C(N)k^{-1/2}|\mathbf{x}|^{-1/2}(1+k|\mathbf{y}|)^{N+1}(|\mathbf{y}|^{N+1}/|\mathbf{x}|), \qquad (4.2)$$

with $C(N)$ a constant independent of $\mathbf{x},\mathbf{y}$. The term $B_n(\Delta_{\omega'})$ denotes the differential operator

$$B_{n+1}(\Delta_{\omega'})=(\Delta_{\omega'}+(n+\tfrac{1}{2})^2)B_n(\Delta_{\omega'}),$$

$$B_0(\Delta_{\omega'})=1,\quad \omega'=\mathbf{x}/|\mathbf{x}|,$$
$$\Delta_{\omega'}=\frac{1}{|\mathbf{x}|^2}\frac{\partial^2}{\partial\Theta^2},\quad \Theta\in[0,2\pi]. \qquad (4.3)$$

*Proof:* Starting from Eq. (2.2) for the free resolvent kernel and using known expansion properties of the Hankel function (Ref. 27, p. 962) we have

$$R_0(k^2+i0;\mathbf{x},\mathbf{y})$$
$$=\left(\frac{i}{8\pi}\right)^{1/2}\frac{e^{ik|\mathbf{x}|}}{k^{1/2}|\mathbf{x}|^{1/2}}\frac{e^{ik|\mathbf{x}|[(|\mathbf{x}-\mathbf{y}|-|\mathbf{x}|)/|\mathbf{x}|]}}{[|\mathbf{x}-\mathbf{y}|/|\mathbf{x}|]^{1/2}}$$
$$\times\left[\sum_{n=0}^{N}\frac{(-1)^n}{(2ik|\mathbf{x}|)^n}\frac{\Gamma(n+\frac{1}{2})}{n!\,\Gamma(-n+\frac{1}{2})}\left(\frac{|\mathbf{x}|}{|\mathbf{x}-\mathbf{y}|}\right)^n\right.$$
$$+\alpha_1\frac{(-1)^{N+1}}{(2ik|\mathbf{x}|)^{N+1}}\frac{\Gamma(N+\frac{3}{2})}{(N+1)!\,\Gamma(-N-\frac{1}{2})}$$
$$\left.\times\left(\frac{|\mathbf{x}|}{|\mathbf{x}-\mathbf{y}|}\right)^{N+1}\right],\quad \alpha_1<1,\quad N>0. \qquad (4.4)$$

Expanding all $|\mathbf{x}-\mathbf{y}|/|\mathbf{x}|$ factors for $|\mathbf{x}|\to\infty$, $|\mathbf{y}|/|\mathbf{x}|<1$, up to order $N$, we arrive at

$$R_0(k^2+i0;\mathbf{x},\mathbf{y})$$

$$=\left(\frac{i}{8\pi}\right)^{1/2}\frac{e^{ik|\mathbf{x}|}}{k^{1/2}|\mathbf{x}|^{1/2}}\sum_{n=0}^{N}\frac{1}{n!}\frac{a_n(k^2,\omega',\mathbf{y})}{(2ik|\mathbf{x}|)^n}+R^G_{N+1}(k^2;\mathbf{x},\mathbf{y}), \tag{4.5}$$

where $R^G_{N+1}(k^2;\mathbf{x},\mathbf{y})$ satisfies the bound (4.2) and the $a_n(k^2;\omega',\mathbf{y})$ are some $n$-dependent coefficients. The latter can be calculated by using the fact that $R_0(k^2+i0;\mathbf{x},\mathbf{y})$ satisfies the Schrödinger equation in the distributional sense, viz.,

$$(-\Delta_{\mathbf{x}}-k^2)R_0(k^2+i0;\mathbf{x},\mathbf{y})=\delta(\mathbf{x}-\mathbf{y}). \tag{4.6}$$

Substituting the expansion (4.5) into Eq. (4.6) and equating common powers of $|\mathbf{x}|$, we find the recursion relation

$$a_{n+1}(k^2;\omega',\mathbf{y})=(\Delta_{\omega'}+(n+\tfrac{1}{2})^2)a_n(k^2;\omega',\mathbf{y}),$$
$$a_0(k^2;\omega',\mathbf{y})=e^{-ik\omega'\cdot\mathbf{y}}. \tag{4.7}$$

This immediately leads to the result (4.1) with $B_{n+1}(\Delta_{\omega'})$ defined by (4.3). □

Next we want the asymptotic behavior of $\Psi_0(k\omega,\mathbf{x})$ as $|\mathbf{x}|\to\infty$. For that purpose we recall the following result from the theory of asymptotic expansion of oscillatory integrals (Ref. 28, Theorem 2.1).

**Theorem 4.2:** Let $g\in C^{2(N+1)}(\Omega)$, $N\geqslant 0$, and let $h\in C^\infty(\Omega)$ be real valued with a finite number of critical points $t_1,\ldots,t_J\in\Omega$ [i.e., $h'(t_j)=0$, $h''(t_j)\neq 0$, $h'(t)\neq 0$, $t\neq t_j, j=1-J$]. The oscillatory integral

$$I(\lambda)=\int_\Omega e^{i\lambda h(t)}g(t)dt \tag{4.8}$$

has the $N$th -order $\lambda>\infty$ asymptotic expansion

$$I(\lambda)=\sum_{j=1}^{J}I_j(\lambda)+R_{N+1}(\lambda), \tag{4.9a}$$

$$I_j(\lambda)=e^{i\lambda h(t_j)}\sum_{n=0}^{N}a_{nj}(g,h)\lambda^{-n-1/2}, \tag{4.9b}$$

where the expansion coefficients are given by

$$\nu_j=\operatorname{sgn}h''(t_j),$$
$$\tilde{h}_j(t)=2|h(t)-h(t_j)|^{1/2}(t-t_j)^{-1},$$
$$a_{nj}(g,h)=\frac{\Gamma(n+1/2)}{(2n)!}e^{i\pi\nu_j(2n+1)/4}\left(\frac{d}{dt}\right)^{2n}\times[\tilde{h}_j(t)^{-2n-1}g(t)]\,|_{t=t_j}. \tag{4.9c}$$

If $\lambda>1$, there is a constant $C(N+1)<\infty$ such that the remainder term has the estimate

$$|R_{N+1}(\lambda)|<C(N+1)\lambda^{-N-3/2}\|g\|_{C^{2(N+1)}(\Omega)}. \tag{4.10}$$

*Lemma 4.3:* For each integer $N\geqslant 0$ and fixed $k,\omega$, $\Psi_0(k\omega,|\mathbf{x}|\cdot)$: $\Omega\to\mathbb{C}$ has the $k|\mathbf{x}|\to\infty$, $k>0$, $\mathscr{D}'^{2(N+1)}$-valued asymptotic expansion

$$\Psi_0(k\omega,|\mathbf{x}|\omega')$$
$$=2\pi^{1/2}\sum_{n=0}^{N}\frac{1}{n!}\frac{1}{(2ik|\mathbf{x}|)^{n+1/2}}\times[e^{ik|\mathbf{x}|}B_n(\Delta_{\omega'})\delta(\omega'-\omega)+i(-1)^n e^{-ik|\mathbf{x}|}B_n(\Delta_{\omega'})\delta(\omega'+\omega)]+R^F_{N+1}(k,|\mathbf{x}|,\omega,\omega'). \tag{4.11a}$$

The term $B_n(\Delta_{\omega'})$ is the derivative operator defined by (4.3), and the remainder term $R^F_{N+1}$ is a distribution in $\mathscr{D}'^{2(N+1)}$ with respect to the variable $\omega'\in\Omega$, which satisfies the estimate

$$\langle R^F_{N+1}(k,|\mathbf{x}|,\omega,\cdot),F\rangle<C(N+1)(k|\mathbf{x}|)^{-N-3/2}\|F\|_{C^{2(N+1)}(\Omega)}, \tag{4.11b}$$

for some finite constant $C(N+1)$.

*Proof:* We know that

$$\Psi_0(k\omega,|\mathbf{x}|\omega')=e^{ik|\mathbf{x}|\omega\cdot\omega'}. \tag{4.12}$$

For $F\in C^{2(N+1)}(\Omega)$ define the functional

$$\langle\Psi_0(k\omega,|\mathbf{x}|),F\rangle=\int_\Omega e^{ik|\mathbf{x}|\cos\theta}F(\theta)d\theta=I(k|\mathbf{x}|). \tag{4.13}$$

We apply Theorem 4.2 to calculate the behavior of $I(k|\mathbf{x}|)$ as $k|\mathbf{x}|\to\infty$. Putting $h=\cos\theta, g=F$, it is evident that there are two isolated stationary points, namely, $\theta_1=0$ and $\theta_2=\pi$. All the conditions of Theorem 4.2 are satisfied, and we obtain from Eqs. (4.9)

$$I(k|\mathbf{x}|)=\sum_{j=1}^{2}\sum_{n=0}^{N}\exp(ik|\mathbf{x}|\cos\theta_j)(k|\mathbf{x}|)^{-n-1/2}\times a_{nj}(F(\theta_j),\cos\theta_j)+R_{N+1}(k|\mathbf{x}|), \tag{4.14}$$

where $R_{N+1}(k|\mathbf{x}|)$ satisfies estimate (4.10) with $\lambda=k|\mathbf{x}|$. The precise form of the coefficients $a_{nj}$ can be calculated from (4.9c). However, we prefer to use the more convenient method of Lemma 4.1 based on the fact that $\Psi_0(k\omega,\mathbf{x})$ also is a solution of the free Schrödinger equation

$$(-\Delta_{\mathbf{x}}-k^2)\Psi_0(k\omega,\mathbf{x})=0. \tag{4.15}$$

It is clear from (4.9c) that in our case the expressions for the coefficients $a_{nj}$ will only involve the function $F$ and its derivatives at the points $\theta_1=0$ ($\omega=\omega'$) [resp. $\theta_2=\pi$ ($\omega=-\omega'$)]. This allows us to extract from (4.14) the following $\mathscr{D}'^{2(N+1)}$-valued asymptotic expansion for $\Psi_0(k\omega,|\mathbf{x}|\cdot)$:

$$\Psi_0(k\omega,|\mathbf{x}|\omega')$$
$$=2(\pi)^{1/2}\sum_{n=0}^{N}\frac{1}{n!\,(2ik|\mathbf{x}|)^{n+1/2}}\times[e^{ik|\mathbf{x}|}a_n(\omega')\delta(\omega'-\omega)+ie^{-ik|\mathbf{x}|}b_n(\omega')\delta(\omega'+\omega)]+R^F_{N+1}(k,|\mathbf{x}|,\omega,\omega'), \tag{4.16}$$

where we have made explicit some numerical factors [in analogy with (4.1)] and where the derivative operators

$a_n(\omega'), b_n(\omega')$ still have to be determined. Substituting this expansion (4.16) in (4.15) and equating the common powers of $|\mathbf{x}|$ we find that $a_n(\omega') = b_n(\omega') = B_n(\Delta_{\omega'})$. Estimate (4.10) becomes the bound (4.11b).

The structure of expansion (4.11) may be somewhat unexpected. Although $\Psi_0(k\omega,|\mathbf{x}|\cdot) \in C^\infty(\Omega)$ the coefficients and remainder terms on the right side of (4.11) are all distributions. However, this is the price we must pay if we seek an expansion of the plane wave $\Psi_0$ in powers of $|\mathbf{x}|^{-1}$. The expansion (4.11) has an obvious symmetry as $\omega \leftrightarrow \omega'$. The differential operators in $B_n(\Delta_{\omega'})$ have only even-order derivatives. For this reason

$$B_n(\Delta_{\omega'})\delta(\omega' - \omega) = B_n(\Delta_{\omega})\delta(\omega' - \omega). \tag{4.17}$$

Identity (4.17) means that expansion (4.11a) can be restated as distribution valued in the variable $\omega$ for a fixed parameter $\omega'$.

In order to obtain a complete characterization of the large $|\mathbf{x}|$ behavior of the wave function $\Psi^+(k\omega,\mathbf{x})$ we examine the large $|\mathbf{x}|$ form of the scattering contribution to $\Psi^+(k\omega,\mathbf{x})$. Define $\Psi_s^+$ by

$$\Psi^+(k\omega,\mathbf{x}) = \Psi_0(k\omega,\mathbf{x}) + \Psi_s^+(k\omega,\mathbf{x}). \tag{4.18}$$

*Lemma 4.4:* Let $V \in \mathscr{A}(g,0)$, $g > 6$. The function $\Psi_s^+(\mathbf{k},\mathbf{x}) \in C(\mathbb{R}^3 \setminus \{O\} \times \mathbb{R}^3)$ has the $|\mathbf{x}| \to \infty$ asymptotic expansion

$$\Psi_s^+(k\omega,\mathbf{x}) = \sum_{n=0}^{1} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|^{1/2}(2ik|\mathbf{x}|)^n} B_n(\Delta_{\omega'}) f(k;\omega,\omega') + R_2^s(k\omega,\mathbf{x}), \tag{4.19a}$$

where $C(k) < \infty$ is an $(\omega,\mathbf{x})$-independent constant such that

$$|R_2^s(k\omega,\mathbf{x})| < C(k)|\mathbf{x}|^{-5/2}. \tag{4.19b}$$

Let $m \geqslant 0$ be an integer. If $g > \max(6, m+4)$, then (for each $k,\mathbf{x}$) the functions $\Psi_s^+(k\omega,\mathbf{x})$, $B_n(\Delta_{\omega'})f(k;\omega,\omega')$, and $R_2^s(k\omega,\mathbf{x})$ are $C^m(\Omega_\omega)$.

*Proof:* The point of departure is the integral representation of $\Psi_s^+$ that follows from (2.11a) and (4.18), $k > 0$:

$$\Psi_s^+(k\omega,\mathbf{x}) = -\int_{\mathbb{R}^2} d\mathbf{y}\, R_0(k^2 + i0;\mathbf{x},\mathbf{y}) \times V(\mathbf{y})\Psi^+(k\omega,\mathbf{y}). \tag{4.20a}$$

Expansion (4.19a) has its origin in the free resolvent expansion (4.1). In order to use (4.1) it is necessary to divide the $\Psi_s^+$ integral into two parts. Let $0 < \gamma < 1$ and define

$$\Psi_s^+(k\omega,\mathbf{x}) = \Psi_<(k\omega,\mathbf{x}) + \Psi_>(k\omega,\mathbf{x}), \tag{4.20b}$$

where

$$\Psi_<(k\omega,\mathbf{x}) = -\int_{|\mathbf{y}|<\gamma|\mathbf{x}|} d\mathbf{y}\, R_0(k^2 + i0;\mathbf{x},\mathbf{y}) V(\mathbf{y})\Psi^+(k\omega,\mathbf{y}). \tag{4.20c}$$

The term $\Psi_>$ involves the integration over region $|\mathbf{y}| \geqslant \gamma|\mathbf{x}|$. The complementary restriction $|\mathbf{y}| < \gamma|\mathbf{x}|$ means that the expansion (4.1) is valid. Inserting (4.1) into the $\Psi_<$ integral gives

$$\Psi_<(k\omega,\mathbf{x}) = -i\frac{e^{ik|\mathbf{x}|}}{(4\pi)^{1/2}} \sum_{n=0}^{1} \frac{1}{(2ik|\mathbf{x}|)^{n+1/2}} \times \int_{|\mathbf{y}|<\gamma|\mathbf{x}|} d\mathbf{y}\, B_n(\Delta_{\omega'}) e^{-ik\omega'\cdot\mathbf{y}} V(\mathbf{y})\Psi^+(k\omega,\mathbf{y}) - \int_{|\mathbf{y}|<\gamma|\mathbf{x}|} d\mathbf{y}\, R_2^G(k^2;\mathbf{x},\mathbf{y}) V(\mathbf{y})\Psi^+(k\omega,\mathbf{y}). \tag{4.21}$$

In order to proceed with the analysis of (4.21) it is helpful to note that $\Psi^+(k\omega,\mathbf{x})$ has an $\mathbf{x}$-independent bound. From (2.11a) we have

$$|\Psi^+(k\omega,\mathbf{x})| \leqslant |\Psi_0(k\omega,\mathbf{x})| + \left|\int_{\mathbb{R}^2} d\mathbf{y}\, R_0(k^2 + i0;\mathbf{x},\mathbf{y}) V(\mathbf{y})\Psi^+(k\omega,\mathbf{y})\right| \tag{4.22a}$$

$$\leqslant 1 + \|R_0(k^2 + i0;\mathbf{x},\cdot)u(\cdot)\|_2 \times \|v(\cdot)\Psi^+(k\omega,(\cdot))\|_2 \tag{4.22b}$$

$$< C(\Psi^+). \tag{4.22c}$$

The estimate (4.22b) follows by interpreting the integral in (4.22a) as an inner product and then using Lemma 3.1(b) and (2.11b) to bound the norm of the right element. The left element of the inner product is bounded by (3.4), which is $\mathbf{x}$ independent.

First we bound the integral in (4.21) that involves the remainder term $R_2^G$. Inserting the estimate (4.2) and (4.22c) shows that this term has the bound

$$\frac{C(\Psi^+)C(2)\gamma^2}{k^{1/2}|\mathbf{x}|^{5/2}} \int_{|\mathbf{y}|<\gamma|\mathbf{x}|} d\mathbf{y}\, |\mathbf{y}|^2(1 + k|\mathbf{y}|)^2 |V(\mathbf{y})|. \tag{4.23a}$$

Now enlarge the domain of integration in (4.23a) by extending it to $\mathbb{R}^2$. If $g > 6$, this integral is finite. This shows that the $R_2^G$ integral term is $O(|\mathbf{x}|^{-5/2})$.

Next examine the two leading integrals in (4.21). These two integrals would construct $f(k;\omega,\omega')$ and $B_1(\Delta_{\omega'}) \times f(k;\omega,\omega')$ if the domain of integration were $\mathbb{R}^2$ [cf. (2.11b) and (2.13a)]. Modify these two integrals by extending the integration domain from $|\mathbf{y}| < \gamma|\mathbf{x}|$ to $\mathbb{R}^2$. Since $B_n(\Delta_{\omega'})e^{-ik\omega'\cdot\mathbf{y}} = O((1 + k|\mathbf{y}|)^n)$, it is easy to see that the contribution from the extended portion of the integral with $|\mathbf{y}| \geqslant \gamma|\mathbf{x}|$ is $O(|\mathbf{x}|^{-5/2})$.

Finally, consider $\Psi_>(k\omega,\mathbf{x})$. This integral must be shown to be $O(|\mathbf{x}|^{-5/2})$. Again subdivide the domain into two regions with

$$D_1 = \{\mathbf{y} \in \mathbb{R}^2 : |\mathbf{y}| \geqslant \gamma|\mathbf{x}|,\ k|\mathbf{x} - \mathbf{y}| > 1\}$$

and

$$D_2 = \{\mathbf{y} \in \mathbb{R}^2 : |\mathbf{y}| \geqslant \gamma|\mathbf{x}|,\ k|\mathbf{x} - \mathbf{y}| \leqslant 1\}.$$

In region $D_1$ the estimate (3.8) shows that $|R_0(k^2 + i0;\mathbf{x},\mathbf{y})| < c$. Thus

$$|\Psi_{>,D_1}| < cC(\Psi^+)\int_{|\mathbf{y}|>\gamma|\mathbf{x}|} d\mathbf{y}\,|V(\mathbf{y})|, \tag{4.23b}$$

$$|\Psi_{>,D_1}| < \frac{cC(\Psi^+)}{\gamma^{5/2}|\mathbf{x}|^{5/2}}\int_{|\mathbf{y}|>\gamma|\mathbf{x}|} d\mathbf{y}\,|\mathbf{y}|^{5/2}|V(\mathbf{y})|,$$

where in the second inequality we have used $|\mathbf{y}|\gamma^{-1} > |\mathbf{x}|$. The integral in (4.23b), after extension to $\mathbb{R}^2$, is finite if $g > 6$. In the case of the integral over region $D_2$ use the resolvent bound (3.8), the estimate (4.22c), and the Schwarz inequality to show

$$|\Psi_{>,D_2}| < cC(\Psi^+)\left(\int_{k|\mathbf{x}-\mathbf{y}|<1} d\mathbf{y}\,\frac{1}{k|\mathbf{x}-\mathbf{y}|}\right)^{1/2} \times\left(\int_{|\mathbf{y}|>\gamma|\mathbf{x}|} d\mathbf{y}\,|V(\mathbf{y})|^2\right)^{1/2}. \tag{4.23c}$$

In the rightmost integral replace $|V(\mathbf{y})|^2$ by $(|\mathbf{y}|/\gamma|\mathbf{x}|)^5 \times |V(\mathbf{y})|^2$ and proceed as in the previous estimate of $\Psi_{>,D_1}$. Taken together the estimates (4.23a), (4.23b), and (4.23c), in combination with (4.20b) and (4.20c), establish (4.19a).

Now consider the $\omega$-smoothness properties of functions entering Eqs. (4.19). To get an idea of the origin of the relationship between $\omega$ smoothness and the large $|\mathbf{x}|$ decay of $V(\mathbf{x})$ we first examine a simple $L^2$ argument. Note that $\Phi_0^-(k\omega)$ is $m$ times continuously differentiable in $L^2(\mathbb{R}^2)$ if $g > 2m+2$. The bounded operator $T(k^2+i0)$ maps $\Phi_0^-(k\omega)$ into $\Phi^-(k\omega)$ via (2.10b). Since $T(k^2+i0)$ is independent of $\omega$, we have that $\Phi^-(k\omega)$ is also $m$ times strongly continuously differentiable. Using Eqs. (2.11) write $\Psi_s^+(k\omega,\mathbf{x})$ as the inner product

$$\Psi_s^+(k\omega,\mathbf{x}) = -(R_0(k^2+i0;\mathbf{x},\cdot)v(\cdot),\Phi^+(k\omega,\cdot)). \tag{4.24}$$

This implies that if $g > 2m+2$, then $\Psi_s^+(k\omega,\mathbf{x}) \in C^m(\Omega_\omega)$. The result quoted in Lemma 4.4 is better than this for $m > 2$. The improved statement of the lemma results if the symmetric operator $A(k^2)$ is replaced by an asymmetric one defined by the kernel

$$|V(\mathbf{x})|^{1/q}R_0(k^2+i0;\mathbf{x},\mathbf{y})|V(\mathbf{y})|^{1/s},$$

where $q^{-1}+s^{-1}=1$. By choosing $s = m+2$ and following an argument similar to the one involving (4.24) we obtain $\Psi_s^+(k\omega,\mathbf{x}) \in C^m(\Omega_\omega)$ if $g > \max(6, m+2)$. The scattering amplitude function $B_1(\Delta_{\omega'})f(k,\omega,\omega')$ has two additional $\omega'$ derivatives and so requires $g > \max(6, m+4)$ in order to be $C^m(\Omega_\omega)$. The $\omega$ smoothness of the error term is a consequence of equality (4.19a) and the fact that both $\Psi_s^+(k,\omega,\mathbf{x})$ and $B_n(\Delta_{\omega'})f(k,\omega,\omega')$ $(n=0,1)$ are $C^m(\Omega_\omega)$. □

By combining the results of Lemma 4.3 and Lemma 4.4 one obtains a distributional-valued large $|\mathbf{x}|$ asymptotic expansion of $\Psi^+(k\omega,\mathbf{x})$ in terms of the scattering matrix.

*Proposition 4.5:* Let $V \in \mathscr{A}(g,0)$, $g > 6$. For each $k > 0$, $\omega \in \Omega$, the wave function $\Psi^+(k\omega,|\mathbf{x}|\cdot)$: $\Omega_{\omega'} \to \mathbb{C}$ has the $k|\mathbf{x}| \to \infty$ $\mathscr{D}'^4$-valued asymptotic expansion

$$\Psi^+(k\omega,|\mathbf{x}|\omega') = 2\pi^{1/2}\sum_{n=0}^{1}\frac{e^{ik|\mathbf{x}|}}{(2ik|\mathbf{x}|)^{n+1/2}}B_n(\Delta_{\omega'})s(k,\omega,\omega') + 2\pi^{1/2}\sum_{n=0}^{1}\frac{i(-1)^n e^{-ik|\mathbf{x}|}}{(2ik|\mathbf{x}|)^{n+1/2}}B_n(\Delta_{\omega'})\delta(\omega+\omega') + R^+(k\omega,\mathbf{x}), \tag{4.25}$$

where the remainder $R^+(k\omega,|\mathbf{x}|\cdot) \in \mathscr{D}'^4$. If $F(\omega')$ is any test function in $C^4(\Omega_{\omega'})$, then there is an $F,|\mathbf{x}|,\omega$-independent constant $C^+ < \infty$ such that

$$\langle R^+(k\omega,|\mathbf{x}|\cdot),F\rangle \leqslant C^+(k|\mathbf{x}|)^{-5/2}\|F\|_{C^4(\Omega)}. \tag{4.26}$$

The results of Lemmas 4.3 and 4.4 also allow us to study the limit of $\operatorname{tr}\tau_\Sigma^+(k)$ as $\Sigma \to \mathbb{R}^2$. We start from Eq. (2.17) on the diagonal $\omega = \omega'$, viz.,

$$\tau_\Sigma^+(k,\omega,\omega) = \int_\Sigma d\mathbf{x}[|\Psi^+(k\omega,\mathbf{x})|^2 - |\Psi_0(k\omega,\mathbf{x})|^2]. \tag{4.27}$$

We will omit the subscript $+$ in the following. Using the Schrödinger equation we obtain

$$\begin{aligned}\int_\Sigma d\mathbf{x}|\Psi(k\omega,\mathbf{x})|^2 &= -\int_\Sigma d\mathbf{x}\left[\Psi^*(k\omega,\mathbf{x})\Delta_x\frac{\partial}{\partial k^2}\Psi(k\omega,\mathbf{x}) - \left(\frac{\partial}{\partial k^2}\Psi(k\omega,\mathbf{x})\right)\Delta_x\Psi^*(k\omega,\mathbf{x})\right]\\ &= -\int_{S_\Sigma} d\vec{S}_\Sigma\cdot\left[\Psi^*(k\omega,\mathbf{x})\vec{\nabla}_x\frac{\partial}{\partial k^2}\Psi(k\omega,\mathbf{x}) - (\vec{\nabla}_x\Psi^*(k\omega,\mathbf{x}))\frac{\partial}{\partial k^2}\Psi(k\omega,\mathbf{x})\right]\end{aligned} \tag{4.28}$$

by Green's theorem, where $S_\Sigma$ is the surface of the region $\Sigma$. Take this region to be a disk with radius $r$. Then (4.28) becomes

$$\begin{aligned}\int_{|\mathbf{x}|<r} d\mathbf{x}\,|\Psi(k\omega,\mathbf{x})|^2 &= -r\int_\Omega d\omega'\left[\Psi^*(k\omega,r\omega')\frac{\partial}{\partial r}\frac{\partial}{\partial k^2}\Psi(k\omega,r\omega') - \left(\frac{\partial}{\partial r}\Psi^*(k\omega,r\omega')\right)\left(\frac{\partial}{\partial k^2}\Psi(k\omega,r\omega')\right)\right]\end{aligned} \tag{4.29}$$

such that [cf. (4.27)]

$$\begin{aligned}\operatorname{tr}\tau_r(k) = \frac{r}{2k}\int_{\Omega\times\Omega} d\omega\,d\omega'\Big\{\Big[\Big(\frac{\partial}{\partial r}\Psi^*(k\omega,r\omega')\Big)&\times\frac{\partial}{\partial k}\Psi(k\omega,r\omega') - \Psi^*(k\omega,r\omega')\\ &\times\frac{\partial}{\partial r}\frac{\partial}{\partial k}\Psi(k\omega,r\omega')\Big] - [\Psi\leftrightarrow\Psi_0]\Big\},\end{aligned} \tag{4.30}$$

where the notation $[\Psi\leftrightarrow\Psi_0]$ is obvious. This expression is not in an optimal form for the purposes of our subsequent calculations in the sense that, if we want to insert the asymptotic expansions for $\Psi_0$ (Lemma 4.3) and $\Psi$ (Proposition 4.5) obtained before, we would have to multiply two distributions.

In order to avoid this difficulty we rewrite (4.30) and replace $\Psi^+[\Psi]$ by $\Psi_s^+[\Psi_s]$ by using Eq. (4.18). With this modification (4.30) becomes

$$\begin{aligned}\operatorname{tr}\tau_r(k) = \frac{r}{2k}\int_{\Omega\times\Omega} d\omega\, d\omega' \Big\{ &\Big(\frac{\partial}{\partial r}\Psi_s^*(k\omega,r\omega')\Big) \\ &\times \frac{\partial}{\partial k}\Psi_0(k\omega,r\omega') \\ &- \Psi_s^*(k\omega,r\omega')\frac{\partial^2}{\partial r\,\partial k}\Psi_0(k\omega,r\omega') \\ &+ \Big(\frac{\partial}{\partial r}\Psi_0^*(k\omega,r\omega')\Big)\frac{\partial}{\partial k}\Psi_s^*(k\omega,r\omega') \\ &- \Psi_0^*(k\omega,r\omega')\frac{\partial^2}{\partial r\,\partial k}\Psi_s(k\omega,r\omega') \\ &+ \Big(\frac{\partial}{\partial r}\Psi^*(k\omega,r\omega')\Big)\frac{\partial}{\partial k}\Psi_s(k\omega,r\omega') \\ &- \Psi_s^*(k\omega,r\omega')\frac{\partial^2}{\partial r\,\partial k}\Psi_s(k\omega,r\omega')\Big\}. \end{aligned} \tag{4.31}$$

Formula (4.31) is an appropriate representation for the investigation of the $r\to\infty$ asymptotics of $\operatorname{tr}\tau_r(k)$. We may employ both asymptotic expansions (4.11a) and (4.19a) in the study of the right side of (4.31). This is because only one of these two asymptotic expansions is distributional valued. In order to have appropriate smoothness and controlling estimates for the calculation of (4.31) we need to supplement Lemma 4.4 by showing that $\partial/\partial|\mathbf{x}|$ and $\partial/\partial k$ derivatives of (4.19a) are also valid asymptotic expansions. For integers $n_1$ and $n_2$ define the partial derivative

$$D(n_1,n_2) = \Big(\frac{\partial}{\partial k}\Big)^{n_1}\Big(\frac{\partial}{\partial|\mathbf{x}|}\Big)^{n_2}. \tag{4.32}$$

*Lemma 4.6:* Let $V\in\mathscr{A}(g,\epsilon)$, $g>6$, $\epsilon>0$. Suppose $0\leqslant n_1\leqslant 1$, $0\leqslant n_2\leqslant 1$, and $0<\gamma<1$. If $|\mathbf{x}|>\gamma^{-1}(x_0+1)$, $k>0$, then $D(n_1,n_2)\Psi_s^+(k\omega,|\mathbf{x}|\omega')\in C(\Omega_\omega\times\Omega_{\omega'})$ has the $k|\mathbf{x}|\to\infty$ asymptotic expansion

$$\begin{aligned} D(n_1,n_2)\Psi_s^+(k\omega,\mathbf{x}) &= \sum_{n=0}^{1} D(n_1,n_2)\Big[\frac{2\pi^{1/2}e^{ik|\mathbf{x}|}}{|\mathbf{x}|^{1/2}(2ik|\mathbf{x}|)^n}B_n(\Delta_{\omega'})f(k;\omega,\omega') \\ &\quad + D(n_1,n_2)R_2^s(k\omega,\mathbf{x})\Big], \end{aligned} \tag{4.33}$$

where, for some $C(n_1,n_2,k)<\infty$,

$$|D(n_1,n_2)R_2^s(k\omega,\mathbf{x})| < C(n_1,n_2,k)|\mathbf{x}|^{-5/2+n_1}. \tag{4.34}$$

For the integer $m\geqslant 0$ let $g>\max(6,m+4)$; then for each fixed $k>0$, $\mathbf{x}$, the function $D(n_1,n_2)\Psi_s^+(k\omega,\mathbf{x})$, and the associated remainder are $C^m(\Omega_\omega)$. The coefficient functions for the terms in (4.33) that are multipliers of $|\mathbf{x}|^{-i/2}$ for $i>\frac{5}{2}-n_1$ are also $C^m(\Omega_\omega)$.

*Proof (sketch):* The pattern of the proof follows that of Lemma 4.4 for each of the three cases $(n_1,n_2)=(0,1)$, $(1,0)$, and $(1,1)$. A new feature not present in the Lemma 4.4 analysis causes one to require $\epsilon>0$. If we take the $\partial/\partial|\mathbf{x}|$ derivative of (4.20a), then the validity of the formal identity

$$\begin{aligned}\frac{\partial}{\partial|\mathbf{x}|}\int_{\mathbf{R}^2} d\mathbf{y}\, R_0(k^2+i0;\mathbf{x},\mathbf{y})V(\mathbf{y})\Psi^+(k\omega,\mathbf{y}) \\ = \int_{\mathbf{R}^2} d\mathbf{y}\Big[\frac{\partial}{\partial|\mathbf{x}|}R_0(k^2+i0;\mathbf{x},\mathbf{y})\Big]V(\mathbf{y})\Psi^+(k\omega,\mathbf{y})\end{aligned} \tag{4.35}$$

must be demonstrated. Here interchanging the derivative and integral order requires care in view of the $\mathbf{x}=\mathbf{y}$ singularity of the resolvent kernel. Titchmarsh (Ref. 29, Theorem 22.11) shows (4.35) to be correct if $V(\mathbf{y})\Psi^+(k\omega,\mathbf{y})\in L^p_{\mathrm{loc}}(\mathbb{R}^2)$, $p>2$. Given the representation of $(\partial/\partial|\mathbf{x}|)\Psi_s^+(k\omega,\mathbf{x})$ in terms of the integral on the right of (4.35), the $|\mathbf{x}|\to\infty$ expansion is derived by the method of Lemma 4.4. □

The $r\to\infty$ behavior of the local time delay $\tau_r(k^2)$ is characterized by the following theorem.

**Theorem 4.7:** Let $V\in\mathscr{A}(g,\epsilon)$, $g>8$, $\epsilon>0$. Then the following asymptotic expansion is valid for $\tau_r(k)$ as $r\to\infty$, $k>0$, $r>r_0=\gamma^{-1}(x_0+1)$:

$$\begin{aligned}\operatorname{tr}\tau_r(k^2) = &-i\operatorname{tr}s^\dagger(k^2)\frac{d}{dk^2}s(k^2) \\ &+ \frac{e^{i\pi/4}}{4(2\pi)^{1/2}}\frac{1}{k^{3/2}}\int_\Omega d\omega\,[e^{-2ikr}f^*(k,\omega,-\omega) \\ &- e^{2ikr}f(k,\omega,-\omega)] + R^T(k,r). \end{aligned} \tag{4.36a}$$

The error term $R^T$ is $C([0,\infty)\times(r_0,\infty))$ and has the bound

$$|R^T(k,r)| < cr^{-1}k^{-3/2}, \tag{4.36b}$$

where $c$ is a $k,r$-independent constant.

*Proof:* The expansion (4.36a) is a consequence of inserting expansions (4.11a), (4.19a), and (4.33) into formula (4.31). The expansion (4.11a) is interpreted as a distribution in the variable $\omega$. The condition $g>8$ ensures that the expressions in (4.19a) and (4.33) (coefficient functions and remainder terms) are $C^4(\Omega_\omega)$. These expressions form the required test functions for the $\mathscr{D}'^4$-valued expansion (4.11a) with $N=1$. Multiplying the six products in (4.31) gives terms proportional to $r$, $r^0$, and $O(r^{-1})$. The terms linear in $r$ add up to zero because of the unitarity property of $s(k^2)$ expressed in terms of $f(k;\omega,\omega')$. The second term on the right of (4.36a) involving backward scattering contributions arises from the $n=0$ contribution $\delta(\omega'+\omega)$ of expansion (4.11a). The error term estimate (4.36b) is a consequence of (4.11b). □

The study of the $\xi_\Sigma(\lambda)$ convergence problem in the next section requires the following estimate for the scattering amplitude term in (4.36a).

*Lemma 4.8:* Let $V\in\mathscr{A}(g,0)$, $g>6$. Then there is a $k_0>0$ such that for $k>k_0$,

$$\Big|\frac{1}{k^{3/2}}\int_\Omega d\omega\,[e^{-2ikr}f^*(k,\omega,-\omega) - e^{2ikr}f(k,\omega,-\omega)]\Big| < \frac{c}{k^{5/2}}, \tag{4.37}$$

with the constant $c$ independent of $r$.

*Proof:* Let $I(k)$ denote the quantity on the left of (4.37).

Write $f$ in terms of representation (2.13c) and express $T(k^2)$ [via (3.2)] as $1 - A(k^2)T(k^2)$. The estimate (2.4) permits the choice of a $k_0 > 0$ such that for all $k > k_0$, $\|A(k^2)\|_2 < 1/2$. Representation (2.6) of $T(k^2)$ implies that $\|T(k^2)\|_2 < 2$, $k > k_0$. Thus $\|A(k^2)T(k^2)\|_2 < C(V)k^{-1/2}$, $k > k_0$. In this way one obtains

$$I(k) < \frac{1}{k^2}\left|\int_\Omega d\omega\, e^{2ikr}\int d\mathbf{x}\, V(\mathbf{x})e^{-2ik\omega\cdot\mathbf{x}}\right| + C(V)\|V\|_1 k^{-5/2}. \tag{4.38}$$

The remaining $\mathbf{x}$ integral is the Born term contribution to $f(k;\omega,\omega')$. Interchanging the order of integration in the double integral one thereby finds

$$I(k) < \frac{1}{2\pi k^2}\left|\int d\mathbf{x}\, V(\mathbf{x})J_0(2k|\mathbf{x}|)\right| + C(V)\|V\|_1 k^{-5/2}. \tag{4.39}$$

Using the detailed decay bounds (Ref. 27, pp. 961–963) for the Bessel function, which assert that $J_0(2k|\mathbf{x}|)$ will decrease as $(k|\mathbf{x}|)^{-1/2}$ for large argument, allows one to verify estimate (4.37). □

## V. CONVERGENCE OF LOCAL AND GLOBAL SPECTRAL SHIFT FUNCTIONS

In order to facilitate the study of the convergence problem for the Krein functions $\xi_\Sigma(\lambda)$ and $\xi(\lambda)$ it is helpful to obtain a representation of $\xi_\Sigma(\lambda)$ that is similar in form to that of Theorem 3.7 for $\xi(\lambda)$. First, recall the sum rule[30] that $\operatorname{tr}\tau_\Sigma(\lambda)$ satisfies. For scattering systems with $V \in \mathscr{A}(3,0)$ and for every measurable subset $\Sigma \subseteq \mathbb{R}^2$ having finite Lebesgue measure, $|\Sigma| < \infty$, one has the identity

$$\int_0^{\infty -} d\lambda \operatorname{tr}\tau_\Sigma(\lambda) = -2\pi\sum_{i\geq 1}\int_\Sigma d\mathbf{x}\,|\psi_i(\mathbf{x})|^2 - \frac{1}{2}\int_\Sigma d\mathbf{x}\, V(\mathbf{x}), \tag{5.1}$$

where the sum over $i$ takes into account the multiplicity of the eigenvalues. Proceed by adding this equation to the definition (1.4) of $\xi_\Sigma(\lambda)$. One finds immediately the following theorem.

**Theorem 5.1:** Let $V \in \mathscr{A}(3,0)$; then for each $\Sigma \subseteq \mathbb{R}^2$ with $|\Sigma| < \infty$,

$$\xi_\Sigma(\lambda) = \frac{1}{2\pi}\int_\lambda^{\infty -} d\lambda' \operatorname{tr}\tau_\Sigma(\lambda') + \frac{1}{4\pi}\int_\Sigma d\mathbf{x}\, V(\mathbf{x}), \quad \lambda > 0. \tag{5.2}$$

Equations (3.34) and (5.2) allow a simple comparison. Set the constant in (3.34) equal to zero. Utilizing (3.35), subtract (3.34) and (5.2) to give [cf. (1.6)]

$$\xi(\lambda) - \xi_\Sigma(\lambda) = \frac{1}{2\pi}\int_\lambda^{\infty -} d\lambda'[\operatorname{tr}\tau(\lambda') - \operatorname{tr}\tau_\Sigma(\lambda')] + \frac{1}{4\pi}\int_{\mathbb{R}^2\setminus\Sigma} d\mathbf{x}\, V(\mathbf{x}). \tag{5.3}$$

Now we let $\Sigma$ be the closed disk of radius $r$ and center $\mathbf{x} = 0$, and furthermore we write $\xi_\Sigma$ as $\xi_r$. Since $V \in L^1(\mathbb{R}^2)$, the rightmost term in (5.3) vanishes as $r \to \infty$. So it remains to evaluate the $d\lambda'$ integral in (5.3). Inserting the asymptotic identity (4.36a) gives us

$$\int_\lambda^{\infty -} d\lambda'[\operatorname{tr}\tau(\lambda') - \operatorname{tr}\tau_r(\lambda')] = -\frac{e^{i\pi/4}}{4(2\pi)^{1/2}}\int_{\sqrt{\lambda}}^\infty 2k\,dk\left\{\frac{1}{k^{3/2}}\int_\Omega d\omega \times[e^{-2ikr}f^*(k,\omega,-\omega) - e^{2ikr}f(k,\omega,-\omega)] + R^T(k,r)\right\}. \tag{5.4}$$

The estimate (4.36b) for $R^T(k,r)$ shows that this function is $L^1$ on the interval $(\sqrt{\lambda},\infty)$ with respect to the measure $k\,dk$. Thus, in view of the $r^{-1}$ dependence of the estimate (4.36b), the remainder term contribution $R^T(k,r)$ to (5.4) vanishes as $r \to \infty$. Finally, consider the term involving the scattering amplitudes $f$ and $f^*$. We know from Lemma 4.8 that the function of $k$ in the $dk$ integral is $L^1(\sqrt{\lambda},\infty)$. An application of the Riemann–Lebesgue lemma implies that this term vanishes as $r \to \infty$.

For $\lambda < 0$, both $\xi_r(\lambda)$ and $\xi(\lambda)$ are piecewise constant with jumps at $\lambda = \lambda_i$. Since $|\psi_i(\mathbf{x})|^2$ is $L^1(\mathbb{R}^2)$, it trivially follows that $\xi_r(\lambda) \to \xi(\lambda)$ as $r \to \infty$, $\lambda \neq \lambda_i$. The value of $\xi(\lambda)$ at the jump, however, is not uniquely determined.

Summarizing the conclusions above, we have demonstrated the following theorem.

**Theorem 5.2:** Let $V \in \mathscr{A}(g,\epsilon)$, $g > 8$, $\epsilon > 0$, and let $\{H_0, H\}$ have the zero-energy behavior described in Lemma 3.4. Let $\xi(\lambda)$ be the spectral shift function in Theorem 3.7 with const = 0; then for $\lambda \notin \sigma_p(H)$ and all $\lambda > 0$,

$$\lim_{r\to\infty}\xi_r(\lambda) = \xi(\lambda). \tag{5.5}$$

It is worthwhile to make several remarks about the basic result stated in (5.5).

(1) Our detailed derivation has been carried out for the case of two-dimensional scattering. However, the method we have used is quite general in nature and will work in other dimensions than two.

(2) The pointwise convergence of $\xi_r(\lambda)$ to $\xi(\lambda)$ requires that the free constant in the specification of $\xi(\lambda)$ [cf. (3.34), (3.23)] be set equal to zero. From (3.23) we see that this means that $\xi(\lambda) = 0$ if $\lambda < \inf\sigma_p(H)$. This is a commonly chosen normalization convention for $\xi(\lambda)$.

(3) In a certain sense $\xi(\lambda)$ is closely related to the difference of the spectral projetors $[E_\lambda - E_\lambda^0]$ for the self-adjoint operators $H$ and $H_0$, respectively. However, $[E_\lambda - E_\lambda^0]$ is in general not a trace-class operator. Nevertheless, for potentials in class $\mathscr{A}(3,0)$ it is easy to show that $[P_r(E_\lambda - E_\lambda^0)P_r]$ is trace class for all $\lambda < \infty$ and $r < \infty$. Here $P_r$ is again the projection operator given by multiplication in $L^2(\mathbb{R}^2)$ by the characteristic function of the disk with radius $r$. Further, it is known[9] that the local spectral shift function obeys the spectral property

$$\xi_r(\lambda) = -\operatorname{Tr} P_r(E_\lambda - E_\lambda^0)P_r. \tag{5.6}$$

In terms of this language Theorem 5.2 asserts that

$$\xi(\lambda) = -\lim_{r\to\infty} \operatorname{Tr} P_r(E_\lambda - E_\lambda^0)P_r. \tag{5.7}$$

This means that the spectral shift function $\xi(\lambda)$ is a type of improper trace of $[E_\lambda - E_\lambda^0]$.

(4) The pointwise convergence result (5.5) for the spectral shift functions $\xi_r(\lambda)$ and $\xi(\lambda)$ is not shared by the local and global time-delay functions tr $\tau_r(k)$ and tr $\tau(k)$. In the identity (4.36a) relating tr $\tau_r(k)$ and tr $\tau(k)$, the remainder term $R^{\tau}(k,r)$ does vanish as $r\to\infty$, for all $k\geqslant 0$; however, the term with $f^*$ and $f$ oscillates as $r\to\infty$ and so has no pointwise limit. The improved limiting behavior of $\xi_r(\lambda)$ and $\xi(\lambda)$ is a consequence of the $d\lambda'$ integration linking tr $\tau_r(\lambda)$ to $\xi_r(\lambda)$ [cf., e.g., (5.2)].

(5) In the circumstances where zero-energy resonances are present there is a marked contrast between the behavior of $\xi_\Sigma(\lambda)$ and $\xi(\lambda)$ in the neighborhood of $\lambda = 0$. Suppose the system exhibits the zero-energy resonant behavior of case II with $D > 0$. In this case bound states of the system all have negative energy. Let $\lambda > 0$ be smaller than the absolute value of the smallest eigenvalue of $H$. Proposition (3.5) implies

$$\xi(\lambda) - \xi(-\lambda) = -D - \frac{1}{\pi}\int_0^\lambda d\lambda' \operatorname{Im} B^+(\lambda'). \tag{5.8}$$

Thus as $\lambda\to 0$ one finds that

$$\xi(0+) - \xi(0-) = -D. \tag{5.9}$$

In contrast, consider $\xi_\Sigma(\lambda)$. From the definition (1.4) we get

$$\xi_\Sigma(\lambda) - \xi_\Sigma(-\lambda) = \frac{-1}{2\pi}\int_0^\lambda d\lambda' \operatorname{tr} \tau_\Sigma(\lambda'). \tag{5.10}$$

For every $\Sigma$ of finite measure the function tr $\tau_\Sigma(\cdot)\colon \mathbb{R}\to\mathbb{R}$ is Lebesgue integrable[8,30] with respect to $d\lambda$ on the finite interval $[0,1]$. Taking the limit $\lambda\to 0$ gives

$$\xi_\Sigma(0+) - \xi_\Sigma(0-) = 0. \tag{5.11}$$

So the local spectral shift function $\xi_\Sigma(\lambda)$ never displays the jump in value at $\lambda = 0$ that is present in the global spectral shift function $\xi(\lambda)$ when $D \neq 0$.


## ACKNOWLEDGMENTS

One of us (D. B.) is supported by the Nationaal Fonds voor Wetenschappelijk Onderzoek (N. F. W. O.), Belgium, as an Onderzoeksdirecteur. In addition, this work has been supported by a grant from the Natural Sciences and Engineering Research Council, Canada.



[1] M. G. Krein, Mat. Sb. **75**, 597 (1953); Sov. Math. Dokl. **3**, 707 (1962).
[2] M. S. Birman and M. G. Krein, Sov. Math. Dokl. **3**, 740 (1962).
[3] K. B. Sinha, "On the theorem of M. G. Krein," University of Geneva preprint, 1975.
[4] M. G. Krein, in *Topics in Differential and Integral Equations and Operator Theory*, edited by I. Gohberg (Birkhäuser, Basel, 1983), p. 107.
[5] H. Baumgärtel and M. Wollenberg, *Mathematical Scattering Theory* (Akademie, Berlin, 1983).
[6] F. T. Smith, Phys. Rev. **118**, 349 (1960).
[7] J. M. Jauch and J.-P. Marchand, Helv. Phys. Acta **40**, 217 (1967).
[8] J. M. Jauch, K. B. Sinha, and B. Misra, Helv. Phys. Acta **45**, 398 (1972).
[9] L. W. MacMillan and T. A. Osborn, Ann. Phys. (NY) **126**, 1 (1980).
[10] Ph. A. Martin, Acta Phys. Austriaca Suppl. **23**, 157 (1981).
[11] Ph. A. Martin, Commun. Math. Phys. **47**, 221 (1976).
[12] R. Lavine, in *Scattering in Mathematical Physics*, edited by J. A. LaVita and J. P. Marchand (Reidel, Dordrecht, 1974); in *Atomic Scattering Theory*, edited by J. Nuttall (University of Western Ontario, London, Canada, 1978), p. 45.
[13] A. Jensen, Commun. Math. Phys. **82**, 435 (1981); Lett. Math. Phys. **7**, 137 (1983).
[14] D. Bollé, F. Gesztesy, and H. Grosse, J. Math. Phys. **24**, 1529 (1983).
[15] H. Narnhofer, J. Math. Phys. **25**, 987 (1984).
[16] V. S. Buslaev, in *Spectral Theory and Waves Processes*, edited by M. Sh. Birman (Consultants Bureau, London, 1967), Vol. 1, p. 69.
[17] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976).
[18] S. Agmon, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **2**, 151 (1975).
[19] R. Froese, I. Herbst, M. Hoffmann-Ostenhof, and T. Hoffmann-Ostenhof, J. Anal. Math. **41**, 272 (1982).
[20] M. Klaus and B. Simon, Ann. Phys. (NY) **130**, 251 (1975).
[21] P. R. Halmos and V. S. Sunder, *Bounded Integral Operators on $L^2$ Spaces* (Springer, Berlin, 1978).
[22] M. Reed and B. Simon, *Scattering Theory*, Vol. 3 of *Methods of Modern Mathematical Physics* (Academic, New York, 1978).
[23] D. Bollé and F. Gesztesy, Helv. Phys. Acta **56**, 1064 (1983).
[24] M. Cheney, J. Math. Phys. **25**, 1449 (1984).
[25] D. Bollé, F. Gesztesy, C. Danneels, and S. F. J. Wilk, Phys. Rev. Lett. **56**, 900 (1986).
[26] D. Bollé, F. Gesztesy, and C. Danneels, Ann. Inst. H. Poincaré **48**, 175 (1988).
[27] I. S. Gradshteyn and I. M. Ryzik, *Tables of Integrals Series and Products* (Academic, New York, 1965).
[28] M. V. Fedoryuk, Usp. Mat. Nauk **26**, 3, 67 (1971) [Russian Math. Surveys **26**, 3, 65 (1971)].
[29] E. C. Titchmarsh, *Eigenfunction Expansions, Part 2* (Oxford U.P., London, 1958).
[30] T. A. Osborn, K. B. Sinha, D. Bollé, and C. Danneels, J. Math. Phys. **26**, 2796 (1985).

# Shear-free normal cosmological models

Andrzej Krasiński
*N. Copernicus Astronomical Center, Polish Academy of Sciences, Bartycka 18, 00716 Warszawa, Poland*

(Received 19 May 1988; accepted for publication 31 August 1988)

Shear-free normal cosmological models are the perfect fluid solutions of Einstein's equations in which rotation and shear vanish, and which are not static [they were all found by A. Barnes, Gen. Relativ. Gravit. **4**, 105 (1973)]. They are either spherically, plane, or hyperbolically symmetric. Their symmetries are discussed in various coordinate systems and related to the conformal group of the three-dimensional flat space. A coordinate representation is introduced which unites all three cases into a single two-parameter family. The limiting transitions to the Friedman–Lemaitre–Robertson–Walker (FLRW) models and to the Schwarzschild–de Sitter-like solutions are presented.

## I. WHAT ARE THE SHEAR-FREE NORMAL COSMOLOGICAL MODELS AND WHY ARE THEY INTERESTING?

All the perfect fluid solutions of Einstein's equations in which rotation and shear vanish were found by Barnes.[1] Some of them are static and thus of no interest in cosmology. One of the nonstatic solutions is conformally flat; it was found earlier by Stephani[2] and studied by this author elsewhere.[3,4] The remaining solutions of Barnes, which are of Petrov type D and nonstatic, qualify as inhomogeneous models of the Universe and will be called here "the Barnes cosmological models."

There are three classes of them: spherically, plane, and hyperbolically symmetric [the first class was in fact discovered by Kustaanheimo and Qvist[5] (KQ) in 1948 and reobtained a few times more,[6–9] also the plane symmetric model was later rediscovered[9]]. Because they have three-dimensional symmetry groups acting on two-dimensional spacelike orbits, all structures in them have one spatial dimension. This is too simple to describe the three-dimensional structures observed in the real Universe, but may be the first step in the right direction. As argued by this author,[10–12] in the plane symmetric Barnes model it seems possible to set up initial conditions in such a way that the matter density is a periodic function of the (invariant) spatial distance. It would then be an example of a model with a discrete symmetry group,[13] combining large scale homogeneity with small scale inhomogeneity.[14] This is all the more attractive because, as will be shown further in this paper, all the classical Friedman–Lemaitre–Robertson–Walker (FLRW) cosmological models are contained in the Barnes models in the limiting case of spatially homogeneous matter distribution. Hence the Barnes models represent an inhomogeneous perturbation (within the exact theory) superimposed on the FLRW background, and are capable of reproducing the classical cosmological results in the limit.

Since they are too simple for cosmological purposes, further generalizations are needed. A study of geometrical properties of the Barnes models may thus be useful. In this paper, a convenient representation of the three Barnes's classes (Sec. III) is used in which each class is generated by the same differential equation. Symmetry groups of the three classes (Sec. V) are investigated, and special subcases possessing four-dimensional groups are revealed (Sec. VI, all but one of them are vacuum solutions with $\Lambda$). Since the symmetries of the models are closely related to the conformal group of the three-dimensional Euclidean space, an account of properties of this group is given (Sec. IV). A new coordinate representation is introduced (Sec. VII) which unites the three classes into a single two-parameter family in which any one model can be continuously deformed into any other. Symmetries of the three models in these coordinates are presented in Sec. VIII. Finally, it is shown in Sec. IX how the FLRW models result as limiting cases of the Barnes models.

In order to make this paper self-contained, an account of Barnes's original results is given in Sec. II.

The readers should be aware that apart from the Barnes models, other generalizations of the FLRW models are also found in the literature. Most important are the geodesic and shearing perfect fluid models of Szafron[15] which generalize those of Szekeres[16] and Lemaitre,[17–20] and the Petrov type N perfect fluid models of Oleson.[21] Many more papers were published, but most of them deal with special cases of those mentioned here and sometimes they duplicate each other. A detailed survey displaying the interdependences between the various models is being prepared by this author. Readers wishing to contribute to the list are welcome to do so.

## II. THE MODELS AS OBTAINED BY BARNES

Of the several solutions found by Barnes we shall consider here only those which have the expansion scalar and the Weyl tensor both nonzero. In the table of Ref. 1 they are contained in the lines IBE and IIE, but those from IBE are special cases of the latter and need not be considered separately. The metric in those solutions [changed to signature $(+---)$] is

$$ds^2 = Y^{-2}(t,r)\left[\left(\frac{Y_{,t}}{3\Theta}\right)^2 dt^2 - \frac{dr^2}{r^2} - d\theta^2 - f^2(\theta)d\phi^2\right], \tag{2.1}$$

where $\Theta(t)$ (the expansion scalar) is an arbitrary function, $Y(t,r)$ is given by the equation

$$r^2 Y_{,rr} + rY_{,r} - KY = b(r)Y^2, \tag{2.2}$$

 0022-2488/89/020433-09$02.50 © 1989 American Institute of Physics 

$b(r)$ is another arbitrary function, $K = +1, 0$, or $-1$, and

$$f(\theta) = \begin{cases} \sin\theta, & \text{for } K = +1, \\ \theta, & \text{for } K = 0, \\ \sinh\theta, & \text{for } K = -1. \end{cases} \tag{2.3}$$

The model with $K = +1$ is spherically symmetric, the one with $K = 0$ is plane symmetric, and the one with $K = -1$ is hyperbolically symmetric. The constant $K$ should not be confused with the FLRW curvature index $k$; the former is the sign of curvature of the two-dimensional orbits of the symmetry groups mentioned above.

The source is in each case a perfect fluid, and the coordinates in (2.1) are comoving; thus the velocity field is

$$u^\alpha = (3\Theta Y/Y_{,t})\delta^\alpha_0. \tag{2.4}$$

From the form (2.1) it is not easy to reobtain the FLRW models, so we shall change to another parametrization.

## III. COORDINATES BETTER ADAPTED TO THE FLRW LIMIT

Let $R(t)$ be a function (which will coincide with the scale factor in the FLRW limit), and let

$$F(t) = -1/(3\Theta). \tag{3.1}$$

In the spherically symmetric case we then have

$$ds^2 = D^2\,dt^2 - [R(t)/V(t,x,y,z)]^2(dx^2 + dy^2 + dz^2), \tag{3.2}$$

where

$$D = F(R/V)(V/R_{,t}), \tag{3.3}$$

$$u \overset{\text{def}}{=} r^2, \tag{3.4}$$

$$(x,y,z) = r(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{3.5}$$

$$Y(t,r) = V(t,u)/[rR(t)], \tag{3.6}$$

and the function $V(t,u)$ is determined by the equation

$$R(t)w_{,uu}/w^2 = f(u), \tag{3.7}$$

where $f(u)$ is an arbitrary function, and, in the present case,

$$w = V(t,u), \tag{3.8}$$

$$b(u) = f(u)u^{5/2}. \tag{3.9}$$

Equation (3.7) is the Kustaanheimo–Qvist equation.[5]

The FLRW models (all of them) are obtained from here when $f = 0$ and $V_{,t} = 0$, thus $V = 1 + (1/4)kr^2$, where $k = \text{const}$. Without the assumption $V_{,t} = 0$ the spherically symmetric subcase of the Stephani universe[2–4] results where $k(t)$ is an arbitrary function.

The matter density and pressure are here equal to

$$\kappa\rho = \frac{3}{F^2} + \frac{8ufV^3}{R^3} + \frac{12VV_{,u}}{R^2} - \frac{12uV_{,u}^2}{R^2}, \tag{3.10}$$

$$\kappa p = -\frac{3}{F^2} + \frac{4(uV_{,u}^2 - VV_{,u})}{R^2} - \frac{2F_{,t}}{(F^2D)} + 4\left[\frac{F}{(R^2D)}\right]\left(1 - \frac{2uV_{,u}}{V}\right)(VV_{,tu} - V_{,t}V_{,u}) \tag{3.11}$$

($\kappa = 8\pi G/c^4$). In the FLRW limit, Eqs. (3.10) and (3.11) reproduce the Friedman equations if the $t$ coordinate is chosen so that $F = -R/R_{,t}$.

In the plane symmetric case the metric form is again (3.2) and (3.3) where this time

$$u = z, \tag{3.12}$$

$$(x,y,z) = (\theta\cos\phi, \theta\sin\phi, \ln r), \tag{3.13}$$

$$Y(t,r) = V(t,z)/R(t), \tag{3.14}$$

and again $V$ is determined by (3.7) and (3.8) where

$$b(u) = f(u). \tag{3.15}$$

When $f = 0$ and $V = 1$, the flat FLRW model results. However, with $f = 0$ and $V = a + bz$; $a,b = \text{const}$, the open ($k < 0$) FLRW model is obtained in a nontrivial disguise, as will be shown in Sec. IX. The matter density and pressure are in this case[22–25]

$$\kappa\rho = \frac{3}{F^2} + \frac{2fV^3}{R^3} - \frac{3V_{,z}^2}{R^2}, \tag{3.16}$$

$$\kappa p = -\frac{3}{F^2} + \frac{V_{,z}^2}{R^2} - \frac{2F_{,t}}{(F^2D)} - 2\left[\frac{FV_{,z}}{(R^2DV)}\right](VV_{,tz} - V_{,t}V_{,z}). \tag{3.17}$$

Finally, in the hyperbolically symmetric case, the metric form can be transformed into (3.2) and (3.3) where

$$u = x/y, \tag{3.18}$$

$$V(t,x,y) = yw(t,u), \tag{3.19}$$

and $w(t,u)$ obeys (3.7). The transformations corresponding to (3.5) and (3.6) and also (3.13) and (3.14) are carried out in two steps. We first observe[26,27] that the two-dimensional metric $(d\theta^2 + \sinh^2\theta\,d\phi^2)$ may be transformed into $(d\tau^2 + e^{2\tau}\,dz^2)$ (see Appendix A), so (2.1) can be written as

$$ds^2 = F^2(t)(Y_{,t}/Y)^2\,dt^2 - r^{-2}Y^{-2}(t,r)(dr^2 + r^2\,d\tau^2 + r^2e^{2\tau}\,dz^2). \tag{3.20}$$

Now we define

$$r = \exp\{\arcsin[(x/y)^2 + 1]^{-1/2}\}, \tag{3.21}$$

$$\tau = -\tfrac{1}{2}\ln(x^2 + y^2), \tag{3.22}$$

$$Y = [(x/y)^2 + 1]^{-1/2}w(t,u)/R(t), \tag{3.23}$$

and change (3.20) into (3.2) with (3.3), (3.19), and (3.18). Moreover, if variables are changed in (2.2) according to (3.21), (3.18), and (3.19), then $w$ will obey (3.7) where this time

$$b(u) = f(u)(u^2 + 1)^{5/2}. \tag{3.24}$$

With $f = 0$ and $w_{,t} = 0$, this Barnes model can reproduce the flat and the open FLRW model, but the limit is not achieved trivially; see Sec. IX. The matter density and pressure are here[22]

$$\kappa\rho = \frac{3}{F^2} + \frac{2(u^2 + 1)fw^3}{R^3} + \frac{6uww_{,u}}{R^2} - \frac{3(u^2 + 1)w_{,u}^2}{R^2} - \frac{3w^2}{R^2}, \tag{3.25}$$

$$\kappa p = -\frac{3}{F^2} + \frac{(u^2+1)w_{,u}^2}{R^2} - \frac{2uww_{,u}}{R^2} - \frac{2F_{,t}}{(F^2D)} + \frac{w^2}{R^2} + 2\left[\frac{F}{(R^2D)}\right] \times \left[u - \frac{(u^2+1)w_{,u}}{w}\right](ww_{,tu} - w_{,t}w_{,u}). \quad (3.26)$$

The transformations changing (2.2) into (3.7) fulfill in each case the Barnes equations (6.12), (6.13), and those following them.

In the previous papers of this author,[10–12] the Barnes models were unknowingly reobtained and some were named differently. The hyperbolically symmetric model was called there "line homogeneous." The "axially symmetric universe" of Ref. 11 is the unified representation from Sec. VII of this paper, and the other cases of Ref. 11 lead to one or another of the three Barnes models.

Several special solutions of Eq. (3.7) were found by different authors[5,6,8,28–37] (the list of references is not guaranteed to be complete), and the existence of solutions consisting of elementary functions was systematically investigated by Stephani.[38] The result was that Eq. (3.7) can be integrated to a first-order equation when $f(u) = u^n$ or $f(u) = e^u$ or $f(u) = (u+\alpha)^n (u+\beta)^{-n-5}$ ($\alpha,\beta,n$ are constants), and it can be completely solved in elementary functions when $f(u) = u^{-15/7}$ or $f(u) = (au^2 + 2bu + c)^{-5/2}$ ($a,b,c$ are constants). For this last case Wyman[30] provided a general formal solution (not necessarily elementary). In the coordinates used in this section, all these results translate immediately into the corresponding plane and hyperbolically symmetric cases even though the papers quoted were concerned with spherically symmetric space-times. However, as argued in Refs. 10 and 11, it may be more important from the physical point of view to solve Eq. (3.7) in the variable $l(u,t)|_{t=t_0}$, where $l$ is the affine parameter on the geodesics orthogonal to the fluid flow and to the group orbits, since $l$ has a direct geometrical meaning while $u$ has not. The transformation $u \to l(u)$ may not be elementary.

The transformations that change (2.1) into (3.2) preserve in each case the comoving character of the coordinate system, and so

$$u^\alpha \underset{*}{=} D^{-1}\delta_0^\alpha. \quad (3.27)$$

The Weyl tensor[22] is in each case proportional to the arbitrary function $f(u)$, and so will vanish whenever $f(u) = 0$. Then, each of the Barnes models becomes a subcase of the Stephani universe.[2–4] With $f \neq 0$, the Barnes models are of Petrov type D. Note that the Stephani universe has in general no symmetry,[3] so only its special cases are contained as limits $f \to 0$ in the Barnes models.

Since the space metric in (3.2) is manifestly conformally flat, its symmetries will be closely related to conformal symmetries of the flat space. These are shortly described in the next section.

## IV. THE CONFORMAL GROUP OF A FLAT SPACE

In this section we shall consider a flat Riemannian space of arbitrary signature and arbitrary dimension $n$. Let $x^A$, $A = 1,...,n$, be the Cartesian coordinates so that the metric form is

$$ds_x^2 = \epsilon_1(dx^1)^2 + \cdots + \epsilon_n(dx^n)^2, \quad (4.1)$$

where each $\epsilon_i$ equals either $+1$ or $-1$. A transformation of coordinates $x \to y^A(x)$ is called a conformal symmetry of (4.1) if it changes the metric form (4.1) to

$$ds_y^2 = \Phi(y) ds_{x\to y}^2, \quad (4.2)$$

where $\Phi$ is a function and $ds_{x\to y}^2$ is obtained from (4.1) by replacing all $x^A$ by $y^A$. The conformal group (i.e., the group of conformal symmetries) of an $n$-dimensional flat space has $[\frac{1}{2}n(n+1)+1]$ parameters; $\frac{1}{2}n(n-1)$ of them belong to the symmetry group (for which $\Phi = 1$) and 1 belongs to the dilatation transformation, $x^A/y^A = l = \text{const}$ (where $\Phi = l^{-2}$). The remaining $n$ parameters are connected with the following transformations:

$$x^A = (y^A + C^A y_R y^R)/T, \qquad T \overset{\text{def}}{=} 1 + 2C_S y^S + C_S C^S y_P y^P, \quad (4.3)$$

where $C^A$, $A = 1,...,n$, are the group parameters. After Plebański[39] we shall call (4.3) the Haantjes transformations, although Haantjes derived in fact only the special cases of (4.3) where (1) $C_S C^S = 0$ (Ref. 40) and (2) just one $C^A$ was nonzero.[41] The group (4.3) is Abelian. Composing two such transformations with the sets of parameters $C^A$ and $D^A$ results in a single transformation (4.3) with the set of parameters $(C^A + D^A)$. Consequently, the inverse transformation to (4.3) is obtained by interchanging $x$'s with $y$'s and replacing all $C^A$ by $(-C^A)$. The generators of (4.3) are

$$J_B = y_R y^R \frac{\partial}{\partial y^B} - 2y_B y^R \frac{\partial}{\partial y^R}. \quad (4.4)$$

The following properties of (4.3) are useful in calculations:

$$x_A x^A = y_A y^A / T, \quad (4.5)$$

$$dx_A\, dx^A = dy_A\, dy^A / T^2. \quad (4.6)$$

Equation (4.3) can be interpreted as the succession of the following three transformations.[39]

(1) Inversion in the (pseudo-) sphere of radius $L$ centered at $x^A = 0$, $x^A = L^2 u^A / u_S u^S$.

(2) Translation by the vector $L^2 C^A$, $u^A = w^A + L^2 C^A$.

(3) Inversion in an identical pseudosphere centered at $w^A = 0$, $w^A = L^2 y^A / y^S y_S$.

Since $L$ cancels in the end, it may be assumed that $L = 1$ without loss of generality. In the following we shall often denote (4.3) by

$$x^A = H(C^1,...,C^n) y^A; \quad (4.7)$$

its inverse is then $y^A = H(-C^1,...,-C^n) x^A$.

## V. SYMMETRY GROUPS OF THE BARNES MODELS

For the spherically symmetric and the plane symmetric model, the symmetries are well known. The generators are $J_i = \epsilon_{ijk} x^j (\partial/\partial x^k)$ in the former case and $J_1 = \partial/\partial x$, $J_2 = \partial/\partial y$, $J_3 = x(\partial/\partial y) - y(\partial/\partial x)$ in the latter. For the hyperbolically symmetric model in the form given by (3.2), (3.3), (3.7), (3.18), and (3.19) the generators are found from the Killing equations to be

$$J_1 = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}, \tag{5.1}$$

$$J_2 = \frac{\partial}{\partial z}, \tag{5.2}$$

$$J_3 = -2xz\frac{\partial}{\partial x} - 2yz\frac{\partial}{\partial y} + (x^2 + y^2 - z^2)\frac{\partial}{\partial z}. \tag{5.3}$$

The corresponding symmetry transformations are, respectively,

$$x'^i = lx^i, \quad i = 1,2,3, \quad l = \text{const}, \tag{5.4}$$

$$z' = z + a, \quad a = \text{const}, \quad (x',y') = (x,y), \tag{5.5}$$

and the Haantjes transformation,

$$(x',y',z') = H(0,0,C)(x,y,z), \quad C = \text{const}, \tag{5.6}$$

which will be for once written out explicitly:

$$x' = \frac{x}{T_H}, \quad y' = \frac{y}{T_H}, \quad z' = \frac{[z + C(x^2 + y^2 + z^2)]}{T_H},$$
$$T_H = 1 + 2Cz + C^2(x^2 + y^2 + z^2). \tag{5.7}$$

It may be verified that the algebras of the symmetry groups of the three models are of Bianchi types IX, $\text{VII}_0$, and VIII, respectively. These are all the Bianchi types possible with two-dimensional orbits.

As usual, while solving the Killing equations several alternatives are encountered, of the form: either a certain differential expression vanishes and a constant in the Killing vector survives, or else the constant vanishes and a possible one-parameter symmetry group is absent. In this way, special subcases of a class of metrics are revealed that have higher symmetry. Such special cases of the Barnes models will be presented in the next section. However, a few more cases show up in several places which will be ignored for the reasons explained below.

(1) If $V_{,t} = 0$, then Eq. (3.7) implies that either (a) $R_{,t} = 0$, and Eq. (3.3) becomes invalid, or (b) $w_{,uu} = 0$. Equation (3.3) results from the Einstein equations[22] $G_{0i} = 0$, $i = 1,2,3$, for the metric (3.2) if $(V/R)_{,t} \neq 0$. Otherwise, $G_{0i} \equiv 0$ and $D$ remains arbitrary. However, then matter density does not depend on time and $\Theta = 0$ [see Eq. (7.2) in Ref. 14]. Thus case 1(a) is of no interest for cosmology and will be ignored here. In case 1(b) a FLRW model results (see Sec. IX) whose symmetries are well known.

(2) If $V$ separates, $V = g(t)v(x,y,z)$, then Eq. (3.7) implies again that either $(g/R)_{,t} = 0$, which is equivalent to case 1(a), or else case 1(b) occurs.

(3) If $f(u) = 0$, then a subcase of the Stephani universe results (see Sec. IX) that is beyond the scope of this paper.

These subcases will be called "trivial."

## VI. SUBCASES OF HIGHER SYMMETRY

### A. The spherically symmetric model

A detailed analysis of the Killing equations shows that only one subcase of higher symmetry exists here which is in fact also trivial, but different from those mentioned above. The subcase results when the arbitrary function $f(u)$ from (3.7) is $f = B/u^{5/2}$, where $B = \text{const}$ and $w(t,u) = V(t,u)$ is given by

$$\left(\frac{V}{R}\right)_{,u}^2 - \frac{1}{u}\frac{V}{R}\left(\frac{V}{R}\right)_{,u} - \frac{2}{3}\frac{B}{u^{5/2}}\left(\frac{V}{R}\right)^3 - \frac{1}{4u}\left(\frac{1}{F^2} - \frac{\Lambda}{3}\right) = 0, \tag{6.1}$$

where $\Lambda$ is (the cosmological) constant. Equation (6.1) is a first integral of (3.7) and is equivalent to Eq. (6.7) from Ref. 42 (the correspondence is $V/R = \Phi$, $u = r^2$, $B = f$, $1/F = g$). Therefore (6.1) defines the Schwarzschild–de Sitter solution which in the more familiar "standard" spherical coordinates has the form

$$ds^2 = P\,dt^2 - P^{-1}\,dr^2 - r^2\,d\theta^2 - (r^2/K)\sin^2(K^{1/2}\theta)d\phi^2, \tag{6.2}$$

where $P \overset{\text{def}}{=} K - 2m/r - \frac{1}{3}\Lambda r^2$, $m = -\frac{4}{3}B$, and $K = +1$ (for the proof see Ref. 42). The reason for the peculiar notation will become clear further on. The additional symmetry is time independence in the coordinates of (6.2).

### B. The plane symmetric model

Two subcases of higher symmetry exist here, one of which is again trivial in the same sense as (6.1). In the trivial case the function $f(u)$ from (3.7) is $f = C = \text{const}$ and $w = V$ is given by the following first integral of (3.7):

$$\left(\frac{V}{R}\right)_{,z}^2 - \frac{2}{3}C\left(\frac{V}{R}\right)^3 - \frac{1}{F^2} + \frac{1}{3}\Lambda = 0. \tag{6.3}$$

As indicated in Appendix B, this case can be transformed into (6.2) with $K = 0$, i.e., is the plane symmetric analog of the Schwarzschild–de Sitter solution. The additional symmetry is again time independence in the form (6.2).

The other subcase of higher symmetry has $F = \text{const}$ and

$$V = R(t)azv(azt), \tag{6.4}$$

where $a$ is an arbitrary constant and $v(X)$ is defined by

$$X^2 v_{,XX} + 2Xv_{,X} = Bv^2, \tag{6.5}$$

where $X = azt$, $B$ is another arbitrary constant, and the function $f(u)$ from (3.7) is

$$f = B/(az^3). \tag{6.6}$$

The metric is

$$ds^2 = (azv_{,X}/v)^2\,dt^2 - (azv)^{-2}(dx^2 + dy^2 + dz^2). \tag{6.7}$$

With (6.6), Eqs. (3.7) and (6.5) are consistent. The additional symmetry is

$$t = t'/l, \quad (x,y,z) = l(x',y',z'), \tag{6.8}$$

and its generator

$$J_4 = -t\frac{\partial}{\partial t} + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z} \tag{6.9}$$

has the following commutators with the other three generators, $J_1 = \partial/\partial x$, $J_2 = \partial/\partial y$, and $J_3 = x(\partial/\partial y) - y(\partial/\partial x)$:

$$[J_i, J_4] = J_i, \quad i = 1,2; \quad [J_3, J_4] = 0. \tag{6.10}$$

Note that the group generated by $J_1, J_2, J_4$ is of Binachi type V and has three-dimensional orbits. The orbits are not, however, orthogonal to the flow lines of the fluid and have indefinite geometry ($J_4$ may be spacelike in one place and timelike in another). Thus (6.7) is a tilted Bianchi type V space-time, with an additional symmetry generated by $J_3$. Up to inessential reparametrizations, the solution coincides with the one investigated in detail by Collins and Wainwright[43,44] and so has a barotropic equation of state.

### C. The hyperbolically symmetric model

As in the spherically symmetric case, the analysis of the Killing equations shows that only one subcase of higher symmetry exists here, in which $w(t,u)$ is given by the following first integral of (3.7):

$$\left(\frac{w}{R}\right)_{,u} = \frac{u}{1+u^2}\frac{w}{R} \pm \left\{\frac{2}{3}\frac{B}{(1+u^2)^{5/2}}\left(\frac{w}{R}\right)^3 - \left[\frac{w}{R(1+u^2)}\right]^2 + \frac{1}{1+u^2}\left(\frac{1}{F^2} - \frac{\Lambda}{3}\right)\right\}^{1/2}, \tag{6.11}$$

where $B$ is a constant given by $f(u) = B/(1+u^2)^{5/2}$, $f$ is the function from (3.7) while $\Lambda$ is the cosmological constant. By coordinate transformations the solution defined by Eq. (6.11) can be reduced to (6.2) with $K = -1$ [then $\sin(K^{1/2}\theta) = \sin(i\theta) = i\sinh\theta$] and so it is the hyperbolically symmetric counterpart of the Schwarzschild–de Sitter solution. The additional symmetry is time independence in the coordinates of (6.2). The transformation from (6.11) to (6.2) is sketched in Appendix C.

The metrics represented by (6.2) belong to the type D metrics investigated by Kinnersley,[45] and are a subcase of those given by Eq. (25.74) in Ref. 26, but in a different coordinate system.

## VII. COORDINATES COVERING ALL THREE MODELS SIMULTANEOUSLY

The following metric represents all the three models of Barnes simultaneously,

$$ds^2 = F^2(t)\left(\frac{R}{S}\right)^2\left(\frac{S}{R}\right)^2_{,t} dt^2 - \frac{R^2(t)}{(z+b)^2 S^2(t,Z)}(dx^2 + dy^2 + dz^2), \tag{7.1}$$

where $F(t)$ and $R(t)$ are arbitrary functions, $b$ is an arbitrary constant, the variable $Z$ is defined by

$$Z = \frac{a - x^2 - y^2 - z^2}{2(z+b)}, \tag{7.2}$$

$a$ is another arbitrary constant, and the function $S(t,Z)$ is defined by the KQ equation,

$$R(t)S_{,ZZ}/S^2 = f_G(Z), \tag{7.3}$$

$f_G(Z)$ being an arbitrary function. With any given $S(t,Z)$, Eqs. (7.1) and (7.2) represent a two-parameter family of metrics. However, we shall show that (7.1)–(7.3) are always equivalent to one of the Barnes models, namely, (I) with $a < b^2$ to the spherically symmetric one, (II) with $a = b^2$ to the plane symmetric one, (III) with $a > b^2$ to the hyperbolically symmetric one.

Thus in the coordinates of (7.1)–(7.3), continuous deformations of one model into another are possible.

### A. The spherically symmetric model

We introduce the constant $\alpha$ by

$$a = b^2 - \alpha^2. \tag{7.4}$$

The cases when $a \geqslant 0$ and $a \leqslant 0$ will have to be considered separately. When $a \geqslant 0$, we define the constants $A$ and $l$ by

$$a = A^2(1+2Al)^{-2}, \quad \alpha = [2l(1+Al)]^{-1}, \tag{7.5}$$

which implies

$$b = (1 + 2Al + 2A^2l^2)/[2l(1+Al)(1+2Al)]. \tag{7.6}$$

The equations are solvable for $A$ and $l$ without further conditions on $a$ and $\alpha$. Then we perform on (7.1)–(7.3) the sequence of three transformations

$$(1)\ z = z' + A/(1+2Al), \tag{7.7}$$

(2) the Haantjes transformation,

$$(x,y,z') = H(0,0,-l)(x',y',z''), \tag{7.8}$$

$$(3)\ z'' = z''' - A. \tag{7.9}$$

The final result is transforming $Z$ given by (7.2) into

$$Z = \frac{l(1+Al)}{1+2Al}\frac{u - A^2}{l^2u - (1+Al)^2}, \tag{7.10}$$

where $u = x'^2 + y'^2 + z'''^2$, and the metric (7.1) into

$$ds^2 = D^2\,dt^2 - (R/V)^2(dx'^2 + dy'^2 + dz'''^2), \tag{7.11}$$

where $D = F(t)(R/V)(V/R)_{,t}$ and

$$V = [l^2u - (1+Al)^2]S/[2l(1+Al)]. \tag{7.12}$$

This suggests that the substitutions (7.10) and (7.12) change (7.3) into (3.7) and (3.8). This is indeed the case, and $f(u)$ from (3.7) is here

$$f(u) = 2[l(1+Al)]^3[l^2u - (1+Al)^2]^{-5} f_G(Z(u)). \tag{7.13}$$

Note that all the operations make sense also for $a = 0$ (then $A = 0$).

When $a \leqslant 0$, we define $A$ and $l$ by

$$a = -A^2(1 + 2Al + 2A^2l^2)^{-2}, \quad \alpha = [2l(1+Al)]^{-1}, \tag{7.14}$$

which implies

$$b = (1+2Al)/[2l(1+Al)(1+2Al+2A^2l^2)].$$

Again, (7.14) are solvable for $A$ and $l$ without further conditions on $a$ and $\alpha$. In the sequence of transformations step (1) is to be replaced by

(1′) $z = z' + A(1 + 2Al)/(1 + 2Al + 2A^2l^2)$, (7.15)

steps (2) and (3) remaining unchanged. The final result for $Z$ is

$$Z = \frac{l(1+Al)}{1+2Al+2A^2l^2}\,\frac{u+A^2}{l^2u-(1+Al)^2}, \tag{7.16}$$

and for the metric it is again (7.12) and (7.13). Just as before, the substitutions (7.16) and (7.12) change (7.3) into (3.7) and (3.8) with $f(u)$ given again by (7.13). The limit $a \to 0$ (thus $A \to 0$) of (7.14) and the subsequent operations is the same as of (7.5)–(7.13). In either case, each step of the transformation was invertible, so we proved that (7.1)–(7.3) with $a < b^2$ is *equivalent* to the spherically symmetric model for all possible values of $a$ and $b$.

### B. The plane symmetric model

The cases $a = b^2 \neq 0$ and $a = b = 0$ have to be treated separately. In the first case, we perform the sequence of two transformations

(4) $z = z' + b$, (7.17)

(5) the Haantjes transformation,

$(x,y,z') = H(0,0,-1/(2b))(x',y',z'')$. (7.18)

The result on $Z$ in (7.2) is

$Z = bz''/(z'' - 2b)$, (7.19)

and the metric (7.1) is transformed into (7.11) with

$V = (z'' - 2b)S$. (7.20)

Just as should be expected, the substitutions (7.19) and (7.20) change (7.3) into (3.7) and (3.8) where this time $u = z''$ and

$f(z'') = 4b^4 f_G(Z(z''))/(z'' - 2b)^5$. (7.21)

When $a = b = 0$, the sequence (4) and (5) should be changed to

(4′) $z = z' + 1$, (7.22)

(5′) $(x,y,z') = H(0,0,-1)(x',y',z'')$, (7.23)

resulting in

$Z = \frac{1}{2}(z'' - 1)^{-1}$ (7.24)

and

$V(t,z'') = (z'' - 1)S(t,Z(z''))$. (7.25)

The function $f(z'')$ from (3.7) is here

$f(z'') = \frac{1}{4} f_G(Z(z''))/(z'' - 1)^5$. (7.26)

Each step of the transformations is invertible, so (7.1)–(7.3) with $a = b^2$ is *equivalent* to the plane symmetric model for each possible value of $a$.

### C. The hyperbolically symmetric model

We introduce $\alpha$ by

$a = b^2 + \alpha^2$. (7.27)

The cases $b > 0$, $b < 0$, and $b = 0$ have to be considered separately. When $b > 0$ we define $l_1$ and $l_2$ by

$b^{-1} = 2l_1^2$, (7.28)

$\alpha^{-1} = 2l_1 l_2$, (7.29)

and then perform the following sequence of three transformations:

(6) $x = (l_1^2 + l_2^2)^{1/2} z'/(l_1 l_2)$, $y = x'/l_2 + y'/l_1$, $z = -x'/l_1 + y'/l_2$, (7.30)

(7) $x' = x'' + 1/(2l_1)$, (7.31)

(8) $(x'',y',z') = H(-l_1,-l_2,0)(x''',y'',z'')$. (7.32)

The final result for $Z$ given by (7.2) is

$Z = au/(bu - \alpha)$, (7.33)

where

$u = x'''/y''$, (7.34)

and for the metric (7.1) it is (7.11), where

$V = y''(bu - \alpha)S(t,Z(u))/a^{1/2}$. (7.35)

This suggests that the substitution (7.33) together with

$S = a^{1/2} w(t,Z(u))/(bu - \alpha)$ (7.36)

will change (7.3) into (3.7) with (3.18). This is indeed the case. Thus the metric (7.1)–(7.3) with $b > 0$ and $a > b^2$ can be transformed into the hyperbolically symmetric Barnes model; the function $f(u)$ is then

$f(u) = \alpha^2 a^{5/2} f_G(Z(u))/(bu - \alpha)^5$. (7.37)

When $b < 0$, we change (7.28) to

$b^{-1} = -2l_1^2$, (7.38)

and the transformation of $z$ in (7.30) to

(6′) $z = x'/l_1 - y'/l_2$, (7.39)

the other steps in the sequence of transformations remaining unchanged. Instead of (7.33) we then obtain

$Z = au/(bu + \alpha)$, (7.40)

which corresponds to replacing $b$ by $-b$ and $Z$ by $-Z$ in (7.33). Hence (7.33) also covers, in fact, the case $b < 0$.

The sequence of transformations is different for $b = 0$. Then we define

$a = \alpha^2 \overset{\text{def}}{=} 1/(4l^2)$, (7.41)

and perform the transformations

(9) $x = x' + 1/(2l)$, (7.42)

(10) $(x',y,z) = H(-l,0,0)(x'',y',z')$, (7.43)

which result in

$Z = -\alpha u$ (7.44)

with

$u = x''/y'$. (7.45)

Equation (7.44) is in fact the limit of (7.33) when $b = 0$. However, the transformations (6)–(8) do not have a meaningful limit $b \to 0$, so the effect of (9) and (10) on the metric (7.1) has to be calculated separately. It turns out that (7.1) changes to (7.11) where

$$V = y'S(t,Z(u)), \tag{7.46}$$

which shows that the resulting metric is the hyperbolically symmetric Barnes model. The function $f$ from (3.7) is here

$$f(u) = \alpha^2 f_G(Z(u)) \tag{7.47}$$

[this is again the limit $b \to 0$ of (7.37)]. Thus (7.1)–(7.3) can be transformed into the hyperbolically symmetric model when $a > b^2$ for each possible value of $b$. Since each transformation in the sequences (6)–(8) and (9) and (10) is invertible, (7.1)–(7.3) with $a > b^2$ is in fact *equivalent* to the hyperbolically symmetric model.

## VIII. SYMMETRIES OF THE BARNES MODELS IN THE COORDINATES OF SEC. VII

Since in the form (7.1)–(7.3) the symmetries are rather difficult to recognize, we shall present them explicitly. Apart from special cases discussed in Sec. VI, the generators of symmetries are

$$J_1 = -2xy\frac{\partial}{\partial x} + (x^2 - y^2 + z^2 + 2bz + a)\frac{\partial}{\partial y} - 2y(z+b)\frac{\partial}{\partial z}, \tag{8.1}$$

$$J_2 = (-x^2 + y^2 + z^2 + 2bz + a)\frac{\partial}{\partial x} - 2xy\frac{\partial}{\partial y} - 2x(z+b)\frac{\partial}{\partial z}, \tag{8.2}$$

$$J_3 = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}. \tag{8.3}$$

The commutation relations are

$$[J_1,J_2] = 4(b^2 - a)J_3, \quad [J_2,J_3] = J_1, \quad [J_3,J_1] = J_2. \tag{8.4}$$

Just as the results of Sec. VII suggest, when $a < b^2$, $a = b^2$, or $a > b^2$, the Bianchi type of the algebra (8.4) is IX, $\mathrm{VII}_0$, or VIII, and it corresponds to the symmetry, spherical, plane, or hyperbolic, respectively.

The transformations generated by $J_3$ are evidently rotations. Those generated by $J_1$ are generalizations of the Haantjes transformations,

$$\begin{aligned} &x' = x/W, \quad z' = (z+b)/W - b, \\ &y' = [y\cosh(2\beta\tau) + (U/2\beta)\sinh(2\beta\tau)]/W, \end{aligned} \tag{8.5}$$

where $\tau$ is the group parameter, $\beta \overset{\text{def}}{=} (a - b^2)^{1/2}$,

$$\begin{aligned} &U \overset{\text{def}}{=} x^2 + y^2 + z^2 + 2bz + a, \\ &W \overset{\text{def}}{=} [a - x^2 - y^2 - z^2 - 2bz - 2b^2 \\ &\qquad + U\cosh(2\beta\tau)]/(2\beta^2) + (y/\beta)\sinh(2\beta\tau). \end{aligned} \tag{8.6}$$

Equations (8.5) and (8.6) cover all three cases given after Eq. (8.4): when $a < b^2$, $\beta$ becomes imaginary, so $\beta = iB$ and $\cosh(2\beta\tau) = \cos(2\beta\tau)$, $\beta^{-1}\sinh(2\beta\tau) = B^{-1}\sin(2B\tau)$. With $a = b^2$ we have

$$\lim_{\beta\to 0} W = 1 + 2\tau y + \tau^2(x^2 + y^2 + z^2 + 2bz + b^2),$$

in that case (8.5) and (8.6) is a composition of $z'' = z + b$ and the Haantjes transformation $(x',y',z') = H(0,0,\tau)(x,y,z'')$.

The transformation generated by $J_2$ is obtained from (8.5) and (8.6) by just interchanging $x$ with $y$ and $x'$ with $y'$.

## IX. THE FLRW LIMIT OF THE BARNES MODELS

Let us change the $t$ coordinate in (3.2) and (3.3) to $T(t)$ defined by

$$\frac{dT}{dt} = -\frac{FR_{,t}}{R}. \tag{9.1}$$

Then, in the new coordinates

$$D = -R^2(V/R)_{,T}/(VR_{,T}) = -V_{,T}R/(VR_{,T}) + 1. \tag{9.2}$$

Thus the $T$ coordinate was chosen so as to make $F(T) = -R(T)/R_{,T}$. In what follows, we shall use just this coordinate.

In order to obtain the FLRW models from the Barnes models, we must first make the latter conformally flat. As was stated after (3.27), this happens when the function $f(u)$ from (3.7) vanishes. Solving (3.7) in that case we obtain

$$w(t,u) = a(t) + b(t)u. \tag{9.3}$$

With arbitrary $a$ and $b$ the metric corresponding to (9.3) is in each of the three cases a subcase of the Stephani universe[2–4] [in order to verify this, one has to occasionally use transformations like (9.7), (9.9), or (9.10) below].

Now the three models have to be considered separately. In the spherically symmetric case, Eq. (9.3) implies that $V$ in (3.2) will be

$$V = a(T) + b(T)(x^2 + y^2 + z^2). \tag{9.4}$$

If $a \neq 0$, then it can be scaled to 1 by redefining $b(T)$ and $R(T)$. Let us then assume $a = 0$ first. Then obviously $b \neq 0$, and so $b$ can be scaled to 1 by redefining $R(T)$. From (3.10) and (3.11) we then see that $\rho = \rho(t)$ and $p = p(t)$, i.e., (9.4) with $a = 0$, $b = 1$ should be a FLRW model. Indeed, the standard form of the FLRW metric,

$$ds^2 = dT^2 - [R(T)/V]^2(dx^2 + dy^2 + dz^2) \tag{9.5}$$

with

$$V = 1 + \tfrac{1}{4}k(x^2 + y^2 + z^2) \tag{9.6}$$

is obtained in this case by the sequence of two transformations,

$$\begin{aligned} &(1)\ z = z' - 1, \\ &(2)\ (x,y,z') = H(0,0,1)(x',y',z''). \end{aligned} \tag{9.7}$$

The resulting $k = 0$.

With $a \neq 0$, we scale $a$ to $+1$ and find from (3.11) that $p_{,u} = 0$ implies either $b = \text{const}$ or $\rho = -p = \text{const}$. The latter case is the de Sitter solution, the former covers all the FLRW metrics ($k = 0, \pm 1$). Thus (9.4) leads to (any) FLRW model when $a$ and $b$ are both constants.

In the plane symmetric case (9.3) implies for (3.2)

$$V = a(T) + b(T)z. \tag{9.8}$$

If $b=0$, then this is evidently the flat FLRW model. Any $b\neq 0$ can be scaled to 1 by redefining $R(T)$. In that case we find from (3.17) that $p_{,z}=0$ implies either $a=$ const or $\rho=-p=$ const. The latter case is again the de Sitter solution, while the former is a FLRW model. The transformations to (9.5) and (9.6) are the following. When $a\neq 0$,

$$\begin{aligned}&(3)\ (x,y,z)=H(0,0,-1/(2a))(x',y',z'),\\&(4)\ (x',y',z')=a(x'',y'',z'');\end{aligned}\tag{9.9}$$

and, when $a=0$,

$$\begin{aligned}&(5)\ (x,y,z)=H(0,0,\tfrac{1}{4})(x',y',z'),\\&(6)\ z'=z''-2.\end{aligned}\tag{9.10}$$

In both cases the resulting $k$ is necessarily $-1$. Hence the plane symmetric Barnes model can reproduce only the flat ($k=0$) and the open ($k=-1$) FLRW models.

Finally, in the hyperbolically symmetric case, (9.3) implies for (3.2),

$$V=a(T)y+b(T)x.\tag{9.11}$$

If either $a=0$ or $b=0$, then (9.11) is equivalent to (9.8) with $a=0$. We shall thus consider the case $ab\neq 0$. Then $a$ can be scaled to 1 redefining $R(T)$, and from (3.26) we conclude that $p_{,u}=0$ implies either $b=$ const or $\rho=-p=$ const. The latter is once more the de Sitter solution, while the former is a FLRW metric. It can be transformed to the standard form (9.5) and (9.6) by the following sequence of transformations:

$$\begin{aligned}&(7)\ (x,y,z)=H(-\tfrac{1}{2},b,0)(x',y',z'),\\&(8)\ x'=x''-1,\quad y'=y''-1/b,\\&(9)\ (x'',y'',z')=[(b^2+1)^{1/2}/(2b)](x''',y''',z''').\end{aligned}\tag{9.12}$$

The resulting $k$ is necessarily $-1$. Hence the hyperbolically symmetric Barnes model can reproduce the open ($k<0$) FLRW model. It can also reproduce the flat ($k=0$) FLRW model if the following trick is applied to (9.11) (with $a$ and $b$ being constant). We first transform $x=x'=+B/b$, and then let $b\to 0$. In this way (9.11) becomes $V=ay+B$, and in the limit $a\to 0$ this is the flat FLRW metric.


## ACKNOWLEDGMENT

I am grateful to H. Knutsen for providing me with his large bibliography on nonstatic spherically symmetric solutions.


## APPENDIX A: TRANSFORMING THE HYPERBOLIC TWO-METRIC

We shall transform the two-dimensional metric

$$ds_2^2=d\theta^2+\sinh^2\theta\,d\phi^2\tag{A1}$$

into

$$ds_2^2=d\tau^2+e^{2\tau}dz^2\tag{A2}$$

(compare Refs. 26 and 27). The following formulas alternately present the transformation and its result on the previous metric form:

$$\theta=\ln(1+2p)-\ln(1-2p),\tag{A3}$$

$$ds_2^2=16(1-4p^2)^{-2}(dp^2+p^2\,d\phi^2),\tag{A4}$$

$$u=p\sin\phi-\tfrac{1}{2},\quad v=p\cos\phi,\tag{A5}$$

$$ds_2^2=(u+u^2+v^2)^{-2}(du^2+dv^2),\tag{A6}$$

$$u=[w-(w^2+z^2)]/T,\quad v=z/T,$$

$$T\overset{\text{def}}{=}1-2w+w^2+z^2,\tag{A7}$$

$$ds_2^2=(dw^2+dz^2)/w^2\tag{A8}$$

[if the previous step looks miraculous, then consult Sec. IV, it is the Haantjes transformation in two dimensions, $(u,v)=H(-1,0)(w,z)$],

$$w=e^{-\tau},\tag{A9}$$

and the resulting metric is (A2). ■

## APPENDIX B: THE PLANE SYMMETRIC ANALOG OF THE SCHWARZSCHILD–de SITTER SOLUTION

In order to bring out the analogy of (6.2) with Eq. (6.7) in Ref. 42, let us change the variable $z$ and the function $w$ according to

$$z=\ln\zeta,\quad w=R\Phi(t,\zeta)/\zeta.\tag{B1}$$

Then (6.3) will change to

$$\Phi_{,\zeta}=(\Phi/\zeta)\pm[\tfrac{2}{3}C(\Phi/\zeta)^3+F^{-2}(t)-\Lambda/3]^{1/2},\tag{B2}$$

where $\zeta$ plays the role of $r$ from Ref. 42, and the metric becomes

$$ds^2=F^2(t)\Phi^{-2}\Phi_{,t}^2\,dt^2-\Phi^{-2}[d\zeta^2+\zeta^2(d\theta^2+d\phi^2)],\tag{B3}$$

where $x=\theta\cos\phi$, $y=\theta\sin\phi$. By the methods of Ref. 42, it can now be shown that a coordinate transformation of the form $t(t',\zeta')$, $\zeta(t',\zeta')$ exists that preserves the algebraic form of (B3), but changes $\Phi$ so that the term $F^{-2}$ in (B2) disappears, thus making the new $\Phi$ time independent. The new $g_{00}$ becomes $(\zeta\Phi_{,\zeta}-\Phi)^2$ (primes dropped). The further coordinate transformation $r=\zeta\Phi^{-1}$ reduces then the metric to (6.2) with $K\to 0$ and $m=-C/3$. The generator of the additional symmetry, which in the original coordinates $(t,x,y,z)$ was

$$J_4=-\frac{1}{2F}\left[\frac{(w/R)_{,z}}{(w/R)_{,t}}\right]\frac{\partial}{\partial t}+\frac{1}{2F}\frac{\partial}{\partial z},\tag{B4}$$

is then transformed into $\partial/\partial t$.

## APPENDIX C: THE HYPERBOLICALLY SYMMETRIC ANALOG OF THE SCHWARZSCHILD–de SITTER SOLUTION

Let us perform the change of variables inverse to (3.21)–(3.23) in the metric (3.2) with (3.18) and (3.19) and $w$ given by (6.11), i.e.,

$$x=e^{-\tau}\cos(\ln\zeta),\quad y=e^{-\tau}\sin(\ln\zeta),\tag{C1}$$

which implies

$$u=x/y=1/\tan(\ln\zeta),\tag{C2}$$

where $\zeta$ plays the role of $r$ from (3.21). The metric then becomes

$$ds^2 = F^2(R/w)^2(w/R)_{,t}^2\,dt^2 - (1+u^2)[R/w(t,u)]^2(d\zeta^2/\zeta^2 + d\tau^2 + e^{2\tau}dz^2). \tag{C3}$$

Let us next introduce the new function $Y(t,u)$ in (6.11) by

$$w = (1+u^2)^{1/2}RY, \tag{C4}$$

then change the variable according to (C2), and finally introduce the new function $\Phi(t,\zeta)$ by

$$Y = \Phi/\zeta. \tag{C5}$$

The resulting equation for $\Phi(t,\zeta)$ will be

$$\Phi_{,\zeta} = \Phi/\zeta \pm [-(\Phi/\zeta)^2 + \tfrac{2}{3}B(\Phi/\zeta)^3 + F^{-2}(t) - \Lambda/3]^{1/2}. \tag{C6}$$

This is analogous to Eq. (6.7) in Ref. 42 [the sign before $(\Phi/\zeta)^2$ is opposite here]. The methods of Ref. 42 work here exactly as described in Appendix B, so a transformation to (6.2) with $K = -1$ exists, where $m = -B/3$ and $r = \zeta'/\Phi$, $\zeta'$ being that variable in which $1/F^2$ drops out from (C6). The generator of additional symmetry which in the coordinates of (6.11) is

$$J_4 = \left[(1+u^2)\frac{(w/R)_{,u}}{(w/R)_{,t}} - u\frac{w/R}{(w/R)_{,t}}\right] \times \frac{1}{F}\frac{\partial}{\partial t} - \frac{y}{F}\frac{\partial}{\partial x} + \frac{x}{F}\frac{\partial}{\partial y} \tag{C7}$$

reduces in the coordinates of (6.2) to $J_4 = \partial/\partial t$.


[1]A. Barnes, Gen. Relativ. Gravit. **4**, 105 (1973).
[2]H. Stephani, Commun. Math. Phys. **4**, 137 (1967).
[3]A. Krasiński, Gen. Relativ. Gravit. **13**, 1021 (1981).
[4]A. Krasiński, Gen. Relativ. Gravit. **15**, 673 (1983).
[5]P. Kustaanheimo and B. Qvist, Soc. Sci. Fenni. Comment. Phys. Math. **XIII** (16), 1 (1948).
[6]A. H. Taub, Ann. Inst. H. Poincaré A **9**, 153 (1968).
[7]M. E. Cahill and G. C. McVittie, J. Math. Phys. **11**, 1382 (1970).
[8]E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[9]A. Banerjee and N. Chakravarty, Acta Phys. Pol. B **10**, 3 (1979).
[10]A. Krasiński, in *Proceedings of the 4th Marcel Grossman Meeting on General Relativity*, edited by R. Ruffini (Elsevier, Amsterdam, 1986), p. 989.
[11]A. Krasiński, in *Gravitational Collapse and Relativity*, Proceedings of Yamada Conference XIV, edited by H. Sato and T. Nakamura (World Scientific, Singapore, 1986), p. 500.
[12]References 10 and 11 discussed the Barnes models, in particular the plane symmetric one, even though no mention of Barnes's paper was made there.
[13]G. F. R. Ellis, in *General Relativity and Gravitation*, edited by B. Bertotti, F. de Felice, and A. Pascolini (Reidel, Dordrecht, 1984), p. 215.
[14]A. Krasiński, in *The Big Bang and Georges Lemaitre*, edited by A. Berger (Reidel, Dordrecht, 1984), p. 63.
[15]D. A. Szafron, J. Math. Phys. **18**, 1673 (1977).
[16]P. Szekeres, Commun. Math. Phys. **41**, 55 (1975).
[17]G. Lemaitre, Ann. Soc. Sci. Bruxelles A **53**, 51 (1933).
[18]For mysterious reasons, the solution obtained in Ref. 17 is now commonly called "the Tolman–Bondi model" or "the Tolman model" although Tolman (Ref. 19) quoted Lemaitre (Ref. 17) and Bondi (Ref. 20) quoted Tolman (Ref. 19), none of them claiming priority.
[19]R. C. Tolman, Proc. Natl. Acad. Sci. USA **20**, 169 (1934).
[20]H. Bondi, Mon. Not. R. Astron. Soc. **107**, 410 (1947).
[21]M. Oleson, J. Math. Phys. **12**, 666 (1971).
[22]These formulas were calculated by the algebraic program ORTOCARTAN (see also Refs. 23–25).
[23]A. Krasiński, SIGSAM Bulletin **17**, 12 (1983).
[24]A. Krasiński, in *International Conference on Computer Algebra and its Applications in Theoretical Physics*, edited by N. N. Govorun (Joint Institute for Nuclear Research, Dubna, 1986), p. 50.
[25]A. Krasiński and M. Perkowski, *The System ORTOCARTAN—User's Manual* (University of Cologne, Cologne, 1983), 3rd ed. (available as a magnetic tape recording or a printout).
[26]D. Kramer, H. Stephani, E. Herlt, and M. A. H. MacCallum, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980), pp. 103 and 104.
[27]H. P. Robertson and T. W. Noonan, *Relativity and Cosmology*. (Saunders, Philadelphia, 1968), pp. 190–192.
[28]P. Kustaanheimo, Soc. Sci. Fenn. Comment. Phys Math. **XIII** (12), 1 (1947).
[29]M. Wyman, Phys. Rev. **70**, 396 (1946).
[30]M. Wyman, Aust. J. Phys. **31**, 111 (1978).
[31]G. C. McVittie, Mon. Not. R. Astron. Soc. **93**, 325 (1933).
[32]G. C. McVittie, Ann. Inst. H. Poincaré A **6**, 1 (1967).
[33]N. Chakravarty, S. B. Dutta-Choudhury, and A. Banerjee, Aust. J. Phys. **29**, 113 (1976).
[34]E. N. Glass and B. Mashhoon, Astrophys. J. **205**, 570 (1976).
[35]A. Banerjee and S. Banerji, Acta Phys. Pol. B **7**, 389 (1976).
[36]H. Nariai. Prog. Theor. Phys. **38**, 92 (1967).
[37]M. Wyman, Can. Math. Bull. **19**, 343 (1976).
[38]H. Stephani, J. Phys. A **16**, 3529 (1983).
[39]J. Plebański, "On conformally equivalent Riemannian spaces," Centro de Investigacion y de Estudios Avanzados del I.P.N., Mexico, 1967, p. 326.
[40]J. Haantjes, Koninkl. Ned. Akad. Wet. Proc. **40**, 700 (1937).
[41]J. A. Schouten and J. Haantjes, Koninkl. Ned. Akad. Wet. Proc. **43**, 1288 (1940).
[42]A. Krasiński and J. Plebański, Rep. Math. Phys. **17**, 217 (1980).
[43]C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[44]C. B. Collins, J. Math. Phys. **26**, 2009 (1985).
[45]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).

# A family of inhomogeneous cosmological Einstein–Rosen metrics

Gabriel Oliver
*Departament de Tecnologia de Computadors, Facultat d'Informàtica de Barcelona (U.P.C.), Pau Gargallo 5, 08028 Barcelona, Spain*

Enric Verdaguer
*Departament de Física Teòrica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

(Received 24 June 1987; accepted for publication 28 September 1988)

Some generalized soliton solutions of the cosmological Einstein–Rosen type defined in the space-time region $t^2 \geqslant z^2$ in terms of canonical coordinates are considered. Vacuum solutions are studied and interpreted as cosmological models. Fluid solutions are also considered and are seen to represent inhomogeneous cosmological models that become homogeneous at $t \to \infty$. A subset of these evolve toward isotropic Friedmann–Robertson–Walker metrics.

## I. INTRODUCTION

Many known cosmological solutions of the Einstein–Rosen form can be deduced as generalized soliton solutions.[1,2] Soliton solutions are obtained by application of the Belinskii and Zakharov soliton transformation[3] and can be easily generalized by taking advantage of the linearity of Einstein's equations for the Einstein–Rosen metric.

The main ingredients in the construction of a soliton solution are the seed metric, which is the starting solution to be transformed, and the so-called pole trajectories, a set of well defined functions that may be real or complex. Although solutions with real pole trajectories may be obtained as the limit of solutions with complex poles when some of the parameters are null, they also form a class on their own.

In terms of canonical coordinates $(t,z)$[3] (for instance, Einstein–Rosen coordinates), the generalized soliton solutions with real poles are defined either in the space-time region $z^2 \geqslant t^2$ or in the region $t^2 \geqslant z^2$. These solutions may be completed by matching them to the seed solutions in the region $t^2 \geqslant z^2$ or in $z^2 \geqslant t^2$, respectively.[4,5] The completed solutions, however, have discontinuous first derivatives along the matching light cones $z^2 = t^2$. These light cone discontinuities disappear if we take complex poles: the metrics are then defined in the whole $(t,z)$ coordinate range and the two regions are smoothly matched.

All soliton solutions can be understood in terms of one-pole and two-pole solutions.[6] Real one-pole generalized soliton solutions valid in $z^2 \geqslant t^2$ have been seen[1,2] to include well-known solutions such as the spatially homogeneous Ellis and MacCallum[7] metrics and their Wainwright, Ince, and Marshman[8] inhomogeneous generalizations; the first have the cosmological singularity only, at $t = 0$, but the second are also singular at $|z| \to \infty$ and $z^2 = t^2$. The Wainwright *et al.* solutions completed with the seed solutions in $t^2 \geqslant z^2$ are better interpreted as the limits of complex pole solutions[9] or as composite universes.[4,10] Real two-pole generalized soliton solutions, on the other hand, have been seen to include the Carmeli and Charach[11,12] pulse wave solutions, which are not singular at $|z| \to \infty$. In all of these solutions the spatially homogeneous Kasner metric has been taken as the seed metric.

In this paper we consider the generalized soliton solutions with real poles defined in $t^2 \geqslant z^2$. To our knowledge these solutions have not been studied previously. In some sense they are complementary to the solutions just mentioned and their possible relevance as cosmological models should be emphasized.

We consider vacuum one-pole and two-pole solutions. For the one-pole solutions we see that, unlike the metrics in $z^2 \geqslant t^2$, they do not include homogeneous metrics. They have only the cosmological singularity at $t = 0$ and the light cone singularity at $z^2 = t^2$. This second singularity, however, disappears when one takes complex poles. Therefore, all of these solutions are potentially interesting as limits of perfectly regular inhomogeneous cosmological models, but are much simpler and easier to study. On the other hand, two-pole solutions give, as in the previous case, pulse wave type metrics. All of the generalized soliton solutions evolve in time to the seed solution (the Kasner metric in our case) and are classified as Petrov type I metrics.

We also consider solutions representing the coupling with a massless scalar field. Such solutions are easily obtained from the vacuum metrics. The new solutions admit a fluid interpretation[8,13,14] according to the space-time properties of the scalar field; in some regions it is a perfect fluid with a stiff equation of state whereas it is an anisotropic fluid in others. The scalar field also admits a generalized soliton solution and we show that the final metric can be seen as a generalization of a Tabensky and Taub fluid plane symmetric metric.[13] The most interesting aspect of this solution is that it approaches spatially homogeneous metrics for $t \to \infty$ and for some values of the parameters it approaches the isotropic Friedmann–Robertson–Walker (FRW) metric. Thus this is an example of cosmological isotropization of initially inhomogeneous metrics.

The vacuum solutions with one and two poles are studied in Sec. II by means of the curvature tensor. The fluid solutions, which are easier to interpret because of the existence of a coordinate system attached to the fluid, are considered in Sec. III.

## II. VACUUM SOLUTIONS

In this section we consider the generalized soliton solutions with real poles that are defined in the space-time region $t^2 \geqslant z^2$. We also consider briefly the solutions defined in $z^2 \geqslant t^2$.

The Einstein–Rosen metrics can be written as

 0022-2488/89/020442-04$02.50 © 1989 American Institute of Physics 

$$ds^2 = f(dz^2 - dt^2) + t(e^{\Phi}\,dx^2 + e^{-\Phi}\,dy^2), \tag{1}$$

where $f$ and $\Phi$ are functions of $t$ and $z$ only.

According to Einstein's equations the potential function $\Phi(z,t)$ verifies a linear wave equation. Using the linearity of this equation, soliton solutions of $\Phi$ may be generalized easily.[1,2,5] In fact, if we take the spatially homogeneous Kasner metric as the seed metric,

$$\Phi_0 = d \ln t, \quad f_0 = (d^2 - 1)/2 \ln t, \tag{2}$$

where $d$ is an arbitrary real parameter, the soliton solutions with $n$ simple poles may be written as[6]

$$\Phi \equiv \Phi_0 + \Phi_s = d \ln t + \sum_{i=1}^{n} \ln\left(\frac{\mu_i}{t}\right), \tag{3}$$

$$f = f_0 t^{n(4-n)/2}\left[\prod_{k=1}^{n}\left(\frac{\mu_k}{t}\right)\right]^{2+d-n} \times \prod_{\substack{k,l=1 \\ k>l}}^{n} (\mu_k - \mu_l)^2 \prod_{k=1}^{n} (\mu_k^2 - t^2)^{-1}, \tag{4}$$

where

$$\mu_i^{\pm} = z_i \pm (z_i^2 - t^2)^{1/2}, \quad z_i \equiv z_i^0 - z, \tag{5}$$

are the pole trajectories that are real if the parameters $z_i^0$ are real or complex otherwise; they verify that $\mu_i^+/t = t/\mu_i^-$.

The analysis of these solutions is usually performed by considering one and two poles only.[6]

By taking $z_i^0$ real we see that $\Phi_s$, for $\mu_i^+$, is a linear superposition of terms of type $\cosh^{-1}(z_1/t)$. Therefore, the one-pole solution may be generalized as

$$\Phi_s = h \cosh^{-1}(z_1/t), \quad |z_1| \geqslant t, \tag{6a}$$

where $h$ is a real parameter. The corresponding $f$ coefficient is easily found from (4) by taking appropriate limits:

$$f = t^{(d^2 + h^2 - 1)/2}(z_1^2 - t^2)^{-h^2/2} \exp[hd \cosh^{-1}(z_1/t)]. \tag{6b}$$

Solution (6), which is defined in the space-time region $|z_1| \geqslant t$, is the Wainwright *et al.* solution.[8] Generally it has singularities at $t = 0$, $|z| \to \infty$, and $t^2 = z_1^2$. When $h^2 = d^2 + 3$ it has the cosmological singularity only ($t = 0$) and is the Ellis and MacCallum[7] spatially homogeneous anisotropic solution: Bianchi V if $d = 0$, Bianchi III if $d^2 = 1$, and Bianchi $\mathrm{VI}_h$ otherwise. In this case it is better to use coordinates $(T,Z)$ adapted to spatial homogeneity[12] $t = \exp(-2aZ)\sinh(2aT)$, $z = \exp(-2aZ)\cosh(2aT)$, where $a$ is a positive constant.

When this solution is completed by matching it to the Kasner solution in the space–time region $t^2 \geqslant z_1^2$, it may be seen as the limit of the one complex pole solution with no light cone singularities[9] and it may be interpreted as a composite universe.[4,10]

To complete this short review of soliton solutions defined in $z_1^2 \geqslant t^2$ we now consider the two-pole solutions that are necessary in order to give an overview of all the soliton solutions. These are the solutions obtained with $\mu_1^+$ and $\mu_2^-$,

$$\Phi_s = (h/2)[\cosh^{-1}(z_1/t) - \cosh^{-1}(z_2/t)], \quad \min(|z_1|,|z_2|) \geqslant t, \tag{7a}$$

and the $f$ coefficient is

$$f = t^{(d^2-1)/2}(\mu_2 - \mu_1)^{h^2/2}(\mu_1/\mu_2)^{h(2d-h)/4} \times ((z_1^2 - t^2)(z_2^2 - t^2))^{-n^2/8}. \tag{7b}$$

Metrics (7), which are not singular at $|z| \to \infty$, are the Carmeli and Charach[11,12] pulse wave solutions. They may be seen as the limit of the corresponding complex pole solutions that describe gravisolitons propagating on a Kasner background.[16]

The solutions we wish to consider here are the family of real pole generalized soliton solutions defined in the space-time region $t \geqslant |z_1|$. In some sense they may be considered as complementary to the solutions just mentioned. For one pole such a family may be obtained in a way similar to (6a) but changing $h$ to $ih$ (Refs. 5 and 17):

$$\Phi_s = h \cos^{-1}(z_1/t), \quad |z_1| \leqslant t. \tag{8a}$$

The $f$ coefficient may be obtained also from (4) by taking appropriate limits:

$$f = t^{(d^2 - h^2 - 1)/2}(t^2 - z_1^2)^{h^2/2} \exp[dh \cos^{-1}(z_1/t)]. \tag{8b}$$

This solution may be matched to the Kasner metric in the region $|z_1| \geqslant t$.

We may now study the intrinsic properties of solution (8). By taking the null tetrad

$$n = (2f)^{-1/2}(\partial_t + \partial_z), \quad l = (2f)^{-1/2}(\partial_t - \partial_z),$$
$$m = (2g_{xx})^{-1/2}\partial_x + i(2g_{yy})^{-1/2}\partial_y,$$

and the complex conjugate of $m$, $m^*$, the Riemann tensor has three non-null components only.[15] For the metric (8) these are

$$\begin{aligned}
\Psi_2 &= -(8f)^{-1}[(1 + h^2 - d^2)t^{-2} \\
&\quad - 2hz_1\,dt^{-2}(t^2 - z_1^2)^{-1/2}], \\
\Psi_0 &= -(2f)^{-1}X^+, \quad \Psi_4 = -2(f)^{-1}X^-, \\
X^{\pm} &= (t^2 - z^2)^{-1/2}(hzt^{-2}(3d^2 - h^2 - 1) \\
&\quad \pm ht^{-1}(3d^2 - h^2 - 3))/4 \\
&\quad + (t^2 - z^2)^{-1}h^2 d(2 + z^2 t^{-2} \pm 3zt^{-1})/2 \\
&\quad + (t^2 - z^2)^{-3/2}h(z + z^3 \pm t(2 + h^2 \\
&\quad + h^2 z^2 t^{-2})/2) + dt^{-2}(d^2 - h^2 - 1)/4.
\end{aligned} \tag{9}$$

The algebraic classification of this metric is easily done by following the d'Inverno and Russell-Clark algorithm.[15,18] For $h \neq 0$ the metrics are of Petrov type I. Of course for $h = 0$ the metric reduces to the Kasner seed, which is of Petrov type D for $d = 0$, flat space for $d^2 = 1$, and Petrov type I otherwise.

The metric has only the cosmological singularity at $t = 0$ and the light cone singularity at $z_1^2 = t^2$; but this last one may be avoided with complex poles. Unlike the solution (6), there are no values of the parameters for which the metric is spatially homogeneous. Therefore for a cosmological interpretation we match metric (8) with the Kasner metric in $|z_1| \geqslant t$ and the new solution may be seen as the limit of inhomogeneous complex pole solutions that have the cosmological singularity only. Such complex pole solutions are acceptable cosmological models, and may be considered as

composite universes,[4,10]; however, they are not so easily deduced and analyzed.

Metric (8) evolves to the spatially homogeneous Kasner metric when $t \to \infty$.

We may now consider the two-pole solutions. They are obtained, similarly to (8), by changing $h$ to $ih$ in (7a):

$$\Phi_s = (h/2)[\cos^{-1}(z_1/t) - \cos^{-1}(z_2/t)], \quad t \geqslant \max(|z_1|,|z_2|), \tag{10a}$$

and the metric coefficient $f$ is found to be

$$f = t^{(d^2-h^2-1)/2}[(t^2-z_1^2)(t^2-z_2^2)]^{h^2/8} \times (z_2\sqrt{t^2-z_1^2} - z_1\sqrt{t^2-z_2^2})^{h^2/4} \times (\sqrt{t^2-z_2^2} + \sqrt{t^2-z_1^2})^{-h^2/4} \times \exp[(dh/2)(\cos^{-1}(z_1/t) - \cos^{-1}(z_2/t))]. \tag{10b}$$

This metric is complementary to the Carmeli and Charach[11] pulse wave solutions (7). It may be understood as the (destructive) superposition of the two solutions (8). In (7) this superposition was essential in order to avoid the singularity at $|z| \to \infty$ of the one-pole solutions (6). Here this is not necessary because the solution (8) is not singular at $|z| \to \infty$ (it is not defined there). The interpretation of the solution, however, is similar. The space-time may be divided by the light cones $|z_1| = t$ and $|z_2| = t$. In the intersection region we have solution (10); it is matched to the one-pole solution at the "inner" light cones and to the seed solution at the "outer" cones. At $t \to \infty$ the completed metric becomes the Kasner metric and thus it may be interpreted as pulse waves propagating on a Kasner background. It is the limit of the corresponding complex pole solution; in such a solution an observer sitting at some fixed $z$ will start in an homogeneous model and will end up in the same model after having gone through inhomogeneous regions of type (8).

## III. SOLUTIONS WITH FLUIDS

We now consider the coupling of Einstein's equations with a massless scalar field $\sigma$. These equations read[13]

$$R_{\mu\nu} = \sigma_{,\mu}\sigma_{,\nu}, \tag{11a}$$

$$\sigma_{,\mu}{}^{,\mu} = 0. \tag{11b}$$

It is well known that a solution of this system may have a fluid interpretation.[8,13,14] Given $\sigma$ this is done in the following way. If $\sigma_{,\mu}$ is a timelike vector, $\sigma_{,\mu}\sigma^{,\mu} < 0$, $\sigma$ may be considered as the potential of a perfect fluid with a stiff equation of state $p = \rho$ ($p$ = pressure, $\rho$ = energy density). This is achieved by defining the density, pressure, and four-velocity of the fluid as

$$\rho = p = -\tfrac{1}{2}\sigma_{,\mu}\sigma^{,\mu}, \quad u_\mu = (-\sigma_{,\mu}\sigma^{,\mu})^{-1/2}\sigma_{,\mu}. \tag{12}$$

The energy-momentum tensor of the fluid is identified from the rhs of (11a) as $T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T$, where

$$T_{\mu\nu} = 2\rho u_\mu u_\nu + \rho g_{\mu\nu}, \tag{13}$$

that is, a perfect fluid with a stiff equation of state.

If $\sigma_{,\mu}$ is a spacelike vector, $\sigma_{,\mu}\sigma^{,\mu} > 0$, the above identification is still formally valid but now $u_\mu$ is a spacelike vector and the perfect fluid interpretation does not hold. Following Tabensky and Taub[13] we can see that the rhs of (11a) may be identified with an anisotropic fluid. For this we define an orthonormal tetrad $(\hat{\tau}_\mu, \hat{\sigma}_{,\mu}, \hat{x}_\mu, \hat{y}_\mu)$, where $\hat{\tau}_\mu$ is a timelike vector, $\hat{\sigma}_\mu \equiv u_\mu$, and $\hat{x}_\mu$, $\hat{y}_\mu$ are spacelike vectors. Now $g_{\mu\nu} = -\hat{\tau}_\mu\hat{\tau}_\nu + \hat{\sigma}_{,\mu}\hat{\sigma}_{,\nu} + \hat{x}_\mu\hat{x}_\nu + \hat{y}_\mu\hat{y}_\nu$ and the rhs of (11a) can be written as $T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T$ with

$$T_{\mu\nu} = \tfrac{1}{2}\sigma_{,\lambda}\sigma^{\lambda}(\tau_\mu\tau_\nu + \sigma_{,\mu}\sigma_{,\nu} - x_\mu x_\nu - y_\mu y_\nu), \tag{14}$$

which corresponds to the energy-momentum tensor of an anisotropic fluid with energy density $\rho = \frac{1}{2}\sigma_{,\lambda}\sigma^{\lambda}$ and vanishing heat-flow vector. The weak and strong energy conditions[19] are satisfied and the fluid interpretation is reasonable.[13]

For the Einstein–Rosen metrics (1) the solutions of (11) are easily found. In such a case the scalar field $\sigma(t,z)$ verifies the same linear wave equation, (11b), that the potential field $\Phi(t,z)$ verifies. Only the metric coefficient $f(t,z)$ is modified by the presence of such a scalar field and it is simply found as the product of two functions, each one of them determined, respectively, from $\Phi(t,z)$ and $\sigma(t,z)$ by similar equations.[13,20]

Therefore we may take generalized soliton solutions for $\sigma$. For the real one-pole case we take, as in (8a),

$$\sigma = a\ln t + b\cos^{-1}(z_1/t), \quad t \geqslant |z_1|, \tag{15}$$

where $a$ and $b$ are arbitrary parameters. This is the solution used by Tabensky and Taub[13] in their study of plane symmetric metrics that evolve to FRW models.

Now the $f$ coefficient is easily obtained by making use of (4) and (8b) as

$$f = t^{(d^2-h^2+2a^2-2b^2-1)/2}(t^2-z_1)^{(h^2+2b^2)/2} \times \exp[(dh+2ab)\cos^{-1}(z_1/t)]. \tag{16a}$$

The potential $\Phi$ has not been modified, i.e., $\Phi = \Phi_0 + \Phi_s$ with $\Phi_s$ being (8a),

$$\Phi = d\ln t + h\cos^{-1}(z_1/t). \tag{16b}$$

Metric (16) gives a solution to Einstein's equations (11a) with the coupling of the massless scalar field (15). It reduces to the Tabensky and Taub plane symmetric metric when $d = h = 0$. For $t$ this metric approaches a spatially homogeneous metric.

The space-time regions where $\sigma_{,\mu}$ is, respectively, timelike and spacelike are divided by the straight line,

$$t = -(a^2+b^2)(a^2-b^2)^{-1}z_1. \tag{17}$$

According to the previous discussion we have a perfect fluid in the space-time region between the straight line (17) and $t = z_1 > 0$ and an anisotropic fluid in the complementary region.

The presence of a fluid makes this metric easier to study and interpret because we may adapt the coordinate system to the fluid. Following Refs. 13 and 21 we shall introduce comoving coordinates. In the region where $\sigma_{,\mu}$ is timelike we may use $\sigma(t,z)$ as the time coordinate and define a space coordinate $Z(t,z)$ by

$$dZ = a^{-1}t(\sigma_{,z}\,dt + \sigma_{,t}\,dz); \tag{18}$$

this ensures that $Z_{,\mu}\sigma^{,\mu} = 0$ and that $Z_{,\mu}$ is spacelike.

Equation (18) is easily integrated as

$$Z = z_1 - a^{-1}b(t^2-z_1^2)^{1/2}. \tag{19a}$$

The fluid lines are the hyperbolas defined by $Z = \text{const}$; they approach straight lines for $t \to \infty$. The time coordinate may be defined as

$$T = \exp[a^{-1}\sigma(t,z) - a^{-1}b\cos^{-1}(b(a^2+b^2)^{-1/2})], \tag{19b}$$

where the constant parameters have been introduced for convenience and $\sigma(t,z)$ is given in (15).

In the region where $\sigma_{,\mu}$ is spacelike, $T$ and $Z$ are space and time coordinates, respectively, and the fluid lines are defined by $\sigma(t,z) = \text{const}$.

The coordinate change defined by (19) is not explicitly invertible. However, for large $t$, which is the region we are interested in, it is

$$t \simeq T + b(a^2+b^2)^{-1/2}Z, \quad z \simeq Z + b(a^2+b^2)^{-1/2}T,$$

and the metric (16) can be written in comoving coordinates as

$$\begin{aligned} ds^2 = {} & T^{(d^2+h^2+2a^2+2b^2-1)}\{1 + (Z/T)(a^2+b^2)^{-1/2} \\ & \times[(b/2)(d^2-h^2-2a^2-2b^2-1) - ahd]\} \\ & \times\{[1 - 2(Z/T)b(a^2+b^2)^{-1/2}]dZ^2 - dT^2\} \\ & + T[1 + (Z/T)b(a^2+b^2)^{-1/2} \\ & \times(T^d A\,dx^2 + T^{-d}A^{-1}\,dy^2)], \end{aligned} \tag{20}$$

where

$$\begin{aligned} A = {} & 1 + (Z/T)(a^2+b^2)^{-1/2}(db - ah) \\ & \times\exp\{h\cos^{-1}[b(a^2+b^2)^{-1/2}]\}. \end{aligned}$$

For $d = h = 0$, i.e., the Tabensky and Taub plane symmetric solution, and $2(a^2+b^2) = 3$ the metric approaches at $T \to \infty$ the flat FRW metric with a stiff perfect fluid

$$ds^2 = T(dZ^2 + dx^2 + dy^2 - dT^2). \tag{21}$$

Metric (20) also approaches the isotropic flat FRW model, (21), when $d = 0$ and $2(a^2+b^2) = 3 - h^2$. For all other values of the parameters the metric approaches a spatially homogeneous but anisotropic model.

To finite values of time the metric is spatially inhomogeneous and may be interpreted as representing inhomogeneous finite perturbations on homogeneous metrics. Therefore the solutions (20) may be considered as an example of inhomogeneous cosmologies that become spatially homogeneous and, for some values of the parameters, isotropic, as a result of cosmological evolution.

Solution (16) may be matched to the space-time region $|z_1| \geqslant t$ with the seed soution obtained by setting $h = b = 0$. There we again have discontinuities in the first derivatives for the metric and the fluid potential. Rather than interpreting it as a solution with shock waves, it is better interpreted as the limit of the corresponding complex one-pole solutions that have no fluid discontinuities in the pressure or the density. The asymptotic behavior at $t \to \infty$ of such a solution is that of the real pole solution described above, i.e., it evolves to spatially homogeneous metrics.

We could also describe fluid solutions with two or more poles; the interpretations, however, are now rather obvious.

## ACKNOWLEDGMENTS

EV is grateful to A. Feinstein for correspondence regarding these solutions.

We acknowledge partial financial support from research project CAICYT.

[1] D. W. Kitchingham, Class. Quantum Gravit. **1**, 677 (1984).
[2] E. Verdaguer, *Observational and Theoretical Aspects of Relativistic Astrophysics and Cosmology*, edited by J. L. Sanz and L. J. Giocoechea (World Scientific, Singapore, 1985), p. 311.
[3] V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **49**, 985 (1979).
[4] A. Feinstein and Ch. Charach, Class. Quantum Gravit. **3**, L5 (1985).
[5] J. Céspedes and E. Verdaguer, Class. Quantum Gravit. **4**, L7 (1987).
[6] B. J. Carr and E. Verdaguer, Phys. Rev. D **28**, 2995 (1983).
[7] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[8] J. Wainwright, W. C. W. Ince, and B. J. Marshman, Gen. Relativ. Gravit. **10**, 259 (1979).
[9] E. Verdaguer, Gen. Relativ. Gravit. **18**, 1045 (1986).
[10] A. Feinstein and Ch. Charach, "Composite universes," Ben-Gurion University of the Negev preprint, Israel, 1986.
[11] M. Carmeli and Ch. Charach, Phys. Lett. A **75**, 333 (1980).
[12] M. Carmeli, Ch. Charach, and S. Malin, Phys. Rep. **76**, 79 (1981).
[13] R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[14] P. Letelier, J. Math. Phys. **20**, 2078 (1979).
[15] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein Field Equations* (Cambridge U.P., Cambridge, 1980).
[16] J. Ibañez and E. Verdaguer, Phys. Rev. Lett. **51**, 1313 (1983).
[17] D. W. Kitchingham, Ph.D. thesis, Queen Mary College, 1986.
[18] R. A. d'Inverno and R. A. Russell-Clarck, J. Math. Phys. **12**, 1258 (1971).
[19] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973).
[20] V. A. Belinskii, Sov. Phys. JETP **50**, 623 (1979).
[21] R. Tabensky and P. S. Letelier, J. Math. Phys. **15**, 594 (1974).

# Rarita–Schwinger fields in algebraically special vacuum space-times

G. F. Torres del Castillo
*Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, 72000 Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Apartado Postal 14-740, 07000 México, D. F., Mexico*

(Received 26 January 1988; accepted for publication 31 August 1988)

It is shown that previous results concerning test massless fields on algebraically special vacuum backgrounds can be extended to the case of massless spin-$\frac{3}{2}$ Rarita–Schwinger fields. A decoupled equation is derived from the Rarita–Schwinger equation on an algebraically special vacuum space-time and it is shown that all the components of the field can be obtained from a scalar potential that obeys a wavelike equation. In the case of type D metrics, identities of the Teukolsky–Starobinsky-type are obtained. Some relations induced by Killing spinors are also included.

## I. INTRODUCTION

In recent years there has been some interest in massless spin-$\frac{3}{2}$ (Rarita–Schwinger) fields, mainly in connection with supergravity.[1–3] According to the supergravity field equations the Rarita–Schwinger field acts as a source of torsion and curvature; when the Rarita–Schwinger field vanishes the supergravity field equations reduce to the Einstein vacuum field equations. Therefore, the first-order perturbations of a solution of the supergravity field equations with a vanishing spin-$\frac{3}{2}$ field are determined by the Rarita–Schwinger equation on a background space-time that is a solution of the Einstein vacuum field equations.

By considering the specific case of a type D vacuum background, making use of a supersymmetry transformation to eliminate some components of the field, Güven[3] found that there exist two decoupled equations derivable from the Rarita–Schwinger equation; he also showed that, in the case of the Kerr metric, these two decoupled equations can be solved by separation of variables and it turns out that something similar occurs for all the type D vacuum metrics.[4,5] All the components of the field are, essentially, determined by these two decoupled components provided that the conformal curvature be different from zero.[3] Güven's decoupled equations correspond to make $s = \pm\frac{3}{2}$ in the master equations found by Teukolsky,[6] which govern certain components of other perturbing massless fields (Weyl neutrino fields, electromagnetic fields, and gravitational perturbations).

Actually, Teukolsky's procedure applies to all the algebraically special vacuum space-times: the component of a Weyl neutrino field, an electromagnetic field, or the perturbed Weyl spinor, obtained by fully contracting the field with a multiple principal spinor of the conformal curvature (Weyl spinor), satisfies a decoupled equation. On the other hand, in these cases, all the components of the fields are obtainable from a scalar potential that satisfies a wavelike equation.[7–12]

In this paper we show that in an algebraically special vacuum space-time, for each multiple principal spinor of the conformal curvature there exists a decoupled equation that follows from the Rarita–Schwinger equation, thus extending Güven's result[3] to all the algebraically special types, including the flat case, and that all the components of the field can be obtained by differentiation from a single scalar potential. We also show that in the case of a vacuum type D space-time the solutions of the two existing decoupled equations are differentiably related and that the existence of a two-index Killing spinor leads to a relation between decoupled components and potentials.

In the derivation of the expression for the Rarita–Schwinger field in terms of potentials we make use of the form of the metric of an algebraically special vacuum space-time given in Refs. 13 and 14. Although this involves the complex extension of space-time, the final expressions apply directly to any real algebraically special vacuum space-time. In general, a complex space-time may not possess real slices; however, the derivation followed here amounts to complexifying the background space-time, going back at the end to the original real slice. A similar procedure has been applied previously in Refs. 9–12 to various other fields. In most of this paper we employ the spinor formalism following the notation and conventions of Plebański *et al.* (see, e.g., Ref. 15), but the final expressions are also given in the Newman–Penrose notation (see, e.g., Ref. 16). The spinor indices are raised and lowered according to $\psi_A = \epsilon_{AB}\psi^B$ and $\psi^B = \psi_A\epsilon^{AB}$.

## II. PRELIMINARIES

The Rarita–Schwinger equation on a curved background space-time can be written as

$$\nabla_{A\dot{D}}\psi^A{}_{B\dot{C}} = \nabla_{B\dot{C}}\psi^A{}_{A\dot{D}}, \tag{1}$$

or, equivalently, in the form,

$$H_{ABC} = H_{(ABC)}, \tag{2a}$$

$$H_{AB\dot{C}} = 0, \tag{2b}$$

where

$$H^A{}_{BC} \equiv \nabla_{(B}{}^{\dot{D}}\psi^A{}_{C)\dot{D}}, \quad H^A{}_{B\dot{C}} \equiv \nabla^{\dot{D}}{}_{(\dot{B}}\psi^A{}_{|D|\dot{C})}. \tag{3}$$

(The parentheses denote symmetrization on the indices enclosed and the indices between bars are excluded from the

 0022-2488/89/020446-06$02.50 © 1989 American Institute of Physics 

symmetrization.) The definitions (3) are similar to those of the electromagnetic spinors in terms of the vector potential (see, e.g., Ref. 16) and Eq. (2b) is analogous to the definition of an (anti-) self-dual field. There exist transformations, analogous to the gauge transformations for the electromagnetic field, that map a solution of the Rarita–Schwinger equation into another solution provided the Ricci tensor vanishes. These transformations are given by

$$\psi_{AB\dot{C}} \to \psi_{AB\dot{C}} + \nabla_{B\dot{C}}\alpha_A, \tag{4}$$

where $\alpha_A$ is an arbitrary spinor field. From Eq. (3) and the Ricci identities it follows that, under the transformation (4), the fields $H_{ABC}$ and $H_{AB\dot{C}}$ are transformed into

$$H^A{}_{BC} + \nabla_{(B}{}^{\dot{D}}\nabla_{C)\dot{D}}\alpha^A = H^A{}_{BC} - 2C^A{}_{BCD}\alpha^D - (R/6)\delta^A_{(B}\alpha_{C)} \tag{5a}$$

and

$$H^A{}_{B\dot{C}} + \nabla^D{}_{(\dot{B}}\nabla_{|D|\dot{C})}\alpha^A = H^A{}_{B\dot{C}} + 2C^A{}_{DB\dot{C}}\alpha^D, \tag{5b}$$

respectively, where $C_{ABCD}$ are the spinor components of the conformal curvature, $R$ is the scalar curvature, and $C_{AB\dot{C}\dot{D}}$ are the spinor components of the traceless part of the Ricci tensor. From Eqs. (2) and (5) it is clear that if $\alpha_A$ is arbitrary and $\psi_{AB\dot{C}}$ is a solution of Eq. (1), then the transformed field (4) will also be a solution of (1) if and only if the Ricci tensor vanishes. In contrast with the electromagnetic case, the field $H_{ABC}$ is not invariant under the transformation (4), unless the conformal curvature is also equal to zero.

In the supergravity theory the metric of the space-time is coupled to a massless spin-$\frac{3}{2}$ Rarita–Schwinger field. The supergravity field equations are invariant under supersymmetry transformations which, on the spin-$\frac{3}{2}$ field, are given by Eq. (4). When the Rarita–Schwinger field vanishes, the supergravity field equations reduce to Einstein's vacuum equations; therefore, since the torsion and the curvature produced by the spin-$\frac{3}{2}$ field depend quadratically on this field, in the linear approximation about a solution with $\psi_{AB\dot{C}} = 0$ the supergravity field equations reduce to Eq. (1) together with the Einstein vacuum field equations.[2,3]

In flat space-time the field $H_{ABC}$ satisfies the massless free-field equations. If the Ricci tensor vanishes, from Eqs. (1) and (3) and the Ricci identities one finds that

$$\begin{aligned}\nabla^{A\dot{R}}H_{ABC} &= -\nabla_A{}^{\dot{R}}\nabla_{(B}{}^{\dot{D}}\psi^A{}_{C)\dot{D}} \\ &= -\nabla_{(B}{}^{\dot{D}}\nabla_{|A|}{}^{\dot{R}}\psi^A{}_{C)\dot{D}} + 2C_{ABC}{}^D\psi^A{}_D{}^{\dot{R}} \\ &= -\nabla_{(B}{}^{\dot{D}}\nabla_{C)\dot{D}}\psi^A{}_A{}^{\dot{R}} + 2C_{ABC}{}^D\psi^A{}_D{}^{\dot{R}} \\ &= 2C_{ABC}{}^D\psi^A{}_D{}^{\dot{R}}.\end{aligned} \tag{6}$$

These equations are not restricted by algebraic consistency conditions. In fact, from Eq. (2a) it follows that

$$\nabla^B{}_{\dot{R}}\nabla^{A\dot{R}}H_{ABC} = \nabla^{(B}{}_{\dot{R}}\nabla^{A)\dot{R}}H_{ABC} = -2C^{ABD}{}_CH_{ABD},$$

and applying $\nabla^B{}_{\dot{R}}$ to the right-hand side of Eq. (6), using the Bianchi identities and Eq. (3), one obtains

$$\begin{aligned}\nabla^B{}_{\dot{R}}(2C_{ABC}{}^D\psi^A{}_D{}^{\dot{R}}) &= 2C_{ABC}{}^D\nabla^B{}_{\dot{R}}\psi^A{}_D{}^{\dot{R}} \\ &= -2C_{ABC}{}^DH^{AB}{}_D,\end{aligned}$$

which coincides with the foregoing expression.

When the Ricci tensor is different from zero, apart from the fact that Eq. (1) is not invariant under the transformations (4), there exist integrability conditions on Eq. (1) involving the Ricci tensor, the field $\psi_{AB\dot{C}}$, and its derivatives, which are very restrictive. For these reasons we will restrict ourselves to the case where the Ricci tensor is equal to zero. Furthermore, in what follows we shall impose the condition that the conformal curvature be algebraically special.

## III. DECOUPLED EQUATIONS AND THE TEUKOLSKY–STAROBINSKY IDENTITIES

In this section we shall show, following the procedure used in Ref. 6, that in an algebraically special vacuum space-time the contraction of $H_{ABC}$, defined in (3), with a multiple principal spinor of the conformal curvature satisfies a decoupled equation, thus extending Güven's result[3] to a wider class of space-times.

The algebraic degeneracy of the conformal curvature means that there exists a spinor field $l_A$ that is a multiple principal spinor of $C_{ABCD}$ (the Weyl spinor), i.e.,

$$l^Al^Bl^CC_{ABCD} = 0. \tag{7}$$

Then, from Eqs. (5a) and (7) it follows that $l^Al^Bl^CH_{ABC}$ is invariant under the transformations (4). This "gauge-invariant" component satisfies a decoupled equation. In order to obtain this equation we shall write down the equivalent of part of Eq. (6) in the Newman–Penrose notation, allowing the spinor indices to take the values 0 and 1, instead of 1 and 2 as in Refs. 9, 10, 14, and 15. In a frame such that $\Psi_0 = 0 = \Psi_1$, according to the Sachs–Goldberg theorem, $\kappa = 0 = \sigma$ and one has

$$\begin{aligned}(D - \epsilon - 3\rho)H_{001} - (\bar{\delta} - 3\alpha + \pi)H_{000} &= \Psi_2\psi_{000'}, \\ (\delta - \beta - 3\tau)H_{001} - (\Delta - 3\gamma + \mu)H_{000} &= \Psi_2\psi_{001'},\end{aligned} \tag{8}$$

where our convention for the definition of $H_{ABC}$ is such that, e.g.,

$$H_{000} = (\delta - 2\beta - \bar{\alpha} + \bar{\pi})\psi_{000'} - (D - 2\epsilon + \bar{\epsilon} - \bar{\rho})\psi_{001'}. \tag{9}$$

Then by applying $(\delta - 2\beta - \bar{\alpha} - 3\tau + \bar{\pi})$ to the first equation in (8) and $(D - 2\epsilon + \bar{\epsilon} - 3\rho - \bar{\rho})$ to the second one and subtracting, we obtain

$$\begin{aligned}&[(D - 2\epsilon + \bar{\epsilon} - 3\rho - \bar{\rho})(\Delta - 3\gamma + \mu) \\ &\quad - (\delta - 2\beta - \bar{\alpha} - 3\tau + \bar{\pi})(\bar{\delta} - 3\alpha + \pi) - \Psi_2]H_{000} \\ &\quad = 0,\end{aligned} \tag{10}$$

where we have used Eq. (9) and the relations

$$(D - 3\rho)\Psi_2 = 0, \quad (\delta - 3\tau)\Psi_2 = 0, \tag{11}$$

which follow from the Bianchi identities (cf. Ref. 6).

In the case of a vacuum type D space-time there are two linearly independent solutions of Eq. (7) and correspondingly there are two decoupled equations that follow from Eq. (6). In a frame such that $\Psi_2$ is the only nonvanishing component of the Weyl spinor, the components $H_{000}$ and $H_{111}$ satisfy Eq. (10) and

$$[(\Delta+2\gamma-\bar{\gamma}+3\mu+\bar{\mu})(D+3\epsilon-\rho)$$
$$-(\bar{\delta}+2\alpha+\bar{\beta}+3\pi-\bar{\tau})(\delta+3\beta-\tau)-\Psi_2]H_{111}$$
$$=0, \tag{12}$$

respectively. In all the type D vacuum metrics, Eqs. (10) and (12) can be solved by separation of variables.[4,5]

The solutions of Eqs. (10) and (12) satisfy certain differential relations, namely,

$$(D+\epsilon-2\bar{\epsilon}-5\rho)(D+2\epsilon-\bar{\epsilon}-3\rho)(D+3\epsilon-\rho)H_{111}$$
$$=(\bar{\delta}-\alpha-2\bar{\beta}+5\pi)(\bar{\delta}-2\alpha-\bar{\beta}+3\pi)$$
$$\times(\bar{\delta}-3\alpha+\pi)H_{000} \tag{13}$$

and

$$(\delta+\beta+2\bar{\alpha}-5\tau)(\delta+2\beta+\bar{\alpha}-3\tau)(\delta+3\beta-\tau)H_{111}$$
$$=(\Delta-\gamma+2\bar{\gamma}+5\mu)(\Delta-2\gamma+\bar{\gamma}+3\mu)$$
$$\times(\Delta-3\gamma+\mu)H_{000}, \tag{14}$$

which are similar to the Teukolsky–Starobinsky identities found in the case of the electromagnetic field (see, e.g., Ref. 17). The validity of Eq. (13) can be verified by a straightforward but rather lengthy computation using Eqs. (1), (6), and (11) and the relations

$$(D-\epsilon-\bar{\epsilon}-\rho)\rho=0, \quad (\bar{\delta}+\alpha-\bar{\beta}+\pi)\pi=0, \tag{15}$$

which follow from the Ricci identities with the Ricci tensor being equal to zero, and

$$(\Delta+3\mu)\Psi_2=0, \quad (\bar{\delta}+3\pi)\Psi_2=0. \tag{16}$$

The computation can be simplified by employing the Geroch–Held–Penrose formalism.[18]

A shorter proof of (13) can be given in the case where $\Psi_2\neq 0$. From Eq. (5a) it is clear that by means of a transformation (4) one can make $H_{001}=0=H_{011}$; then from Eq. (6) one has

$$(D+3\epsilon-\rho)H_{111}=\Psi_2\psi_{110'},$$
$$\psi_{100'}+\psi_{010'}=0,$$
$$(\bar{\delta}-3\alpha+\pi)H_{000}=-\Psi_2\psi_{000'}.$$

On the other hand, from Eq. (1) it follows that

$$(D-\bar{\epsilon}-\rho)\psi_{010'}-\rho\psi_{100'}=(\bar{\delta}-2\alpha-\bar{\beta}+\pi)\psi_{000'},$$
$$(\bar{\delta}-\bar{\beta}+\pi)\psi_{100'}+\pi\psi_{010'}=(D+2\epsilon-\bar{\epsilon}-\rho)\psi_{110'}.$$

Using these equations together with Eqs. (11), (15), and (16), Eq. (13) can be readily verified. Equation (14) is just the "primed version" of Eq. (13) in the sense defined in Ref. 18; hence it does not require a separate proof.

Equations (13) and (14) can be written in other equivalent forms, making use of Eq. (15), e.g.,

$$(D+\epsilon-2\bar{\epsilon}-4\rho)(D+2\epsilon-\bar{\epsilon}-2\rho)(D+3\epsilon-3\rho)H_{111}$$
$$=(\bar{\delta}-\alpha-2\bar{\beta}+4\pi)(\bar{\delta}-2\alpha-\bar{\beta}+2\pi)$$
$$\times(\bar{\delta}-3\alpha+3\pi)H_{000}. \tag{17}$$

## IV. SOLUTION OF THE RARITA–SCHWINGER EQUATION IN TERMS OF POTENTIALS

In this section we shall show that in a vacuum algebraically special space-time the solution of the Rarita–Schwinger equation can be expressed in terms of a scalar potential that satisfies a single second-order partial differential equation.

As shown in Refs. 13 and 14, the metric of a vacuum algebraically special space-time can be written in the form

$$ds^2=2\phi^{-2}\,dq^{\dot{A}}\underset{s}{\otimes}(dp_{\dot{A}}+Q_{\dot{A}\dot{B}}\,dq^{\dot{B}}), \tag{18}$$

where $q^{\dot{A}}$ and $p^{\dot{A}}$ are complex coordinates, $Q_{\dot{A}\dot{B}}=Q_{\dot{B}\dot{A}}$, and $\phi$ is a function that satisfies

$$l^C\nabla_{A\dot{B}}l_C=l_Al^C\partial_{C\dot{B}}\ln\phi, \tag{19}$$

where $l_A$ is a multiple principal spinor of the conformal curvature $C_{ABCD}$ [see Eq. (7)]. Then the vector fields

$$\partial_{1\dot{A}}=\sqrt{2}\,\frac{\partial}{\partial p^{\dot{A}}}\equiv\sqrt{2}\partial_{\dot{A}},$$
$$\partial_{2\dot{A}}=\sqrt{2}\,\phi^2\left(\frac{\partial}{\partial q^{\dot{A}}}+Q_{\dot{A}\dot{B}}\,\partial^{\dot{B}}\right)\equiv\sqrt{2}\,\phi^2D_{\dot{A}}, \tag{20}$$

form a null tetrad satisfying $\partial_{A\dot{B}}\cdot\partial_{C\dot{D}}=-2\epsilon_{AC}\epsilon_{\dot{B}\dot{D}}$.

We shall consider the complex conjugate of Eq. (1) written in the form

$$H_{\dot{A}\dot{B}\dot{C}}=H_{(\dot{A}\dot{B}\dot{C})}, \tag{21a}$$
$$H_{\dot{A}BC}=0, \tag{21b}$$

where

$$H^{\dot{A}}{}_{\dot{B}\dot{C}}\equiv\nabla^R{}_{(\dot{B}}\psi^{\dot{A}}{}_{\dot{C})R}, \quad H^{\dot{A}}{}_{BC}\equiv\nabla_{(B}{}^{\dot{R}}\psi^{\dot{A}}{}_{|\dot{R}|C)}. \tag{22}$$

By using the components of the connection for the tetrad (20) given in Ref. 14 one finds that, with respect to this tetrad, the equation $H^{\dot{A}}{}_{11}=0$ amounts to

$$\partial_{\dot{R}}(\phi^{-1/2}\psi^{\dot{A}\dot{R}}{}_1)+\phi^{-1}J^{\dot{A}}\phi^{-1/2}\psi_{\dot{R}}{}^{\dot{R}}{}_1=0, \tag{23}$$

where $J_{\dot{A}}\equiv\partial_{\dot{A}}\phi$. Contracting Eq. (23) with $J_{\dot{A}}$, it follows that $\partial_{\dot{R}}(J_{\dot{A}}\phi^{-1/2}\psi^{\dot{A}\dot{R}}{}_1)=0$, since $\partial_{\dot{R}}\cdot J_{\dot{A}}=\partial_{\dot{R}}\partial_{\dot{A}}\phi=-\phi C_{11\dot{R}\dot{A}}=0$. Therefore there exists (locally) a function $M$ such that $J_{\dot{A}}\phi^{-1/2}\psi^{\dot{A}\dot{R}}{}_1=\partial^{\dot{R}}M$; hence $\phi^{-1/2}\psi^{\dot{A}\dot{R}}{}_1=K^{\dot{A}}\partial^{\dot{R}}M+J^{\dot{A}}B^{\dot{R}}$, where $K^{\dot{A}}$ is independent of $p^{\dot{B}}$ and satisfies $K^{\dot{A}}J_{\dot{A}}=1$ and $B^{\dot{R}}$ is some pair of functions. Substituting into Eq. (23) one obtains $\partial_{\dot{R}}(\phi B^{\dot{R}}-MK^{\dot{R}})=0$, which implies that $\phi B^{\dot{R}}-MK^{\dot{R}}=\partial^{\dot{R}}N$, where $N$ is some function. Thus

$$\phi^{-1/2}\psi^{\dot{A}\dot{R}}{}_1=K^{\dot{A}}\partial^{\dot{R}}M+\phi^{-1}J^{\dot{A}}(MK^{\dot{R}}+\partial^{\dot{R}}N)$$
$$=\sqrt{2}\{\partial^{\dot{R}}(\phi^{-1/2}\xi^{\dot{A}})+\phi^{-1}J^{\dot{A}}(\phi^{-1/2}\xi^{\dot{R}})\},$$

where $\phi^{-1/2}\xi^{\dot{A}}\equiv(MK^{\dot{A}}+N\phi^{-1}J^{\dot{A}})/\sqrt{2}$. Therefore, $H^{\dot{A}}{}_{11}=0$ implies that locally

$$\phi^{-1/2}\psi^{\dot{A}\dot{R}}{}_1=\sqrt{2}\{\partial^{\dot{R}}(\phi^{-1/2}\xi^{\dot{A}})+\phi^{-1}J^{\dot{A}}(\phi^{-1/2}\xi^{\dot{R}})\} \tag{24}$$

for some $\xi^{\dot{R}}$. Even though the preceding derivation requires $J_{\dot{A}}\neq 0$, it is easy to see directly from Eq. (23) that in the case $J_{\dot{A}}=0$, Eq. (24) also applies. Equation (24) can be written in the form

$$\psi^{\dot{A}\dot{R}}{}_1=\nabla_1{}^{\dot{R}}\xi^{\dot{A}}, \tag{25}$$

as it can be readily verified by using the connection coefficients for the tetrad (20).

From Eqs. (22) and (25) one gets

$$\begin{aligned} 2H^{\dot{A}}{}_{12} &= -\nabla_{1\dot{R}}\psi^{\dot{A}\dot{R}}{}_{2} - \nabla_{2\dot{R}}\psi^{\dot{A}\dot{R}}{}_{1} \\ &= -\nabla_{1\dot{R}}\psi^{\dot{A}\dot{R}}{}_{2} - \nabla_{2\dot{R}}\nabla_{1}{}^{\dot{R}}\xi^{\dot{A}} \\ &= -\nabla_{1\dot{R}}[\psi^{\dot{A}\dot{R}}{}_{2} - \nabla_{2}{}^{\dot{R}}\xi^{\dot{A}}] + \nabla_{(1}{}^{\dot{R}}\nabla_{2)\dot{R}}\xi^{\dot{A}}. \end{aligned}$$

The last term is equal to zero by virtue of the Ricci identities since we are assuming that the Ricci tensor vanishes, and evaluating the first term explicitly we obtain that $H^{\dot{A}}{}_{12} = 0$ amounts to

$$\partial_{\dot{R}}[\phi^{-5/2}(\psi^{\dot{A}\dot{R}}{}_{2} - \nabla_{2}{}^{\dot{R}}\xi^{\dot{A}})] + \phi^{-1}J^{\dot{A}}\phi^{-5/2}(\psi_{\dot{R}}{}^{\dot{R}}{}_{2} - \nabla_{2}{}^{\dot{R}}\xi_{\dot{R}}) = 0.$$

Hence by comparing with Eqs. (23) and (24) we see that

$$\phi^{-5/2}\psi^{\dot{A}\dot{R}}{}_{2} = \phi^{-5/2}\nabla_{2}{}^{\dot{R}}\xi^{\dot{A}} + \sqrt{2}(\partial^{\dot{R}}\zeta^{\dot{A}} + \phi^{-1}J^{\dot{A}}\zeta^{\dot{R}}), \quad (26)$$

for some $\zeta^{\dot{R}}$.

In order to write Eq. (26) in a covariant form we introduce a spinor field $\chi^{\dot{A}}{}_{B\dot{C}} = \chi^{\dot{A}}{}_{C\dot{B}}$ such that, in the frame (20), it has components $\chi^{\dot{A}}{}_{11} = 0 = \chi^{\dot{A}}{}_{12}$, $\chi^{\dot{A}}{}_{22} = \phi^{5/2}\zeta^{\dot{A}}$. Then one finds that Eq. (26) is equivalent to

$$\psi^{\dot{A}\dot{R}}{}_{2} = \nabla_{2}{}^{\dot{R}}\xi^{\dot{A}} + \phi^{-2}\nabla^{B\dot{R}}(\phi^{2}\chi^{\dot{A}}{}_{B2}). \quad (27)$$

Moreover, since $\phi^{-2}\nabla^{B\dot{R}}(\phi^{2}\chi^{\dot{A}}{}_{B1}) = 0$, we can write Eqs. (25) and (27) in the form

$$\psi^{\dot{A}\dot{R}}{}_{C} = \nabla_{C}{}^{\dot{R}}\xi^{\dot{A}} + \phi^{-2}\nabla^{B\dot{R}}(\phi^{2}\chi^{\dot{A}}{}_{BC}). \quad (28)$$

Substituting now this last expression into the first of Eqs. (22) and using the Ricci identities with the scalar curvature equal to zero we get

$$H^{\dot{A}}{}_{B\dot{C}} = -2C^{\dot{A}}{}_{\dot{B}\dot{C}D}\xi^{\dot{D}} + \nabla^{S}{}_{(\dot{B}}[\phi^{-2}\nabla^{\dot{R}}{}_{\dot{C})}(\phi^{2}\chi^{\dot{A}}{}_{RS})]. \quad (29)$$

Evaluating the second term and expressing the result in terms of $\zeta^{\dot{A}}$ we find

$$\begin{aligned} H^{\dot{A}}{}_{B\dot{C}} = &-2C^{\dot{A}}{}_{\dot{B}\dot{C}D}\xi^{\dot{D}} \\ &+ 2\phi^{5/2}[\partial_{\dot{B}}\partial_{C}\zeta^{\dot{A}} + 2\phi^{-1}J^{\dot{A}}\partial_{(\dot{B}}\zeta_{\dot{C})}]. \end{aligned} \quad (30)$$

The symmetry condition (21a) is equivalent to $H^{\dot{A}}{}_{A\dot{C}} = 0$; hence $\zeta^{\dot{A}}$ is restricted by the condition

$$\partial_{\dot{C}}\partial_{A}\zeta^{\dot{A}} + \phi^{-1}J^{\dot{A}}\partial_{A}\zeta_{\dot{C}} + \phi^{-1}J^{\dot{A}}\partial_{\dot{C}}\zeta_{A} = 0,$$

which can be rewritten as

$$\phi^{-1}\partial_{C}[\phi^{3}\partial_{A}(\phi^{-2}\zeta^{\dot{A}})] = 0. \quad (31)$$

Assuming $J_{\dot{A}} \neq 0$ one immediately obtains that the most general solution of Eq. (31) is given by

$$\phi^{-2}\zeta^{\dot{A}} = \sqrt{2}[b\phi^{-2}K^{\dot{A}} + \partial^{\dot{A}}(\phi^{-2}\tilde{H})], \quad (32)$$

where $b$ is a function independent of $p^{\dot{A}}$ and $\tilde{H}$ is some function (the factor $\phi^{-2}$ multiplying $\tilde{H}$ is inserted for later convenience). [On the other hand, the condition $J_{\dot{A}} = 0$ is very restrictive; it implies that the spinor $l_{A}$ is at least a triple principal spinor of $C_{ABCD}$, thus excluding types II and D, and that the shear-free congruence of null geodesics defined by $l_{A}$ has expansion and twist equal to zero.[13,14] In what follows only the case $J_{\dot{A}} \neq 0$ will be considered.] Substituting Eq. (32) into (26) one gets

$$\begin{aligned} \phi^{-5/2}\psi^{\dot{A}\dot{R}}{}_{2} = &\phi^{-5/2}\nabla_{2}{}^{\dot{R}}\xi^{\dot{A}} \\ &+ 2(\partial^{\dot{R}}\partial^{\dot{A}}\tilde{H} - \phi^{-1}J^{\dot{A}}\partial^{\dot{R}}\tilde{H} \\ &+ b\phi^{-1}J^{\dot{A}}K^{\dot{R}}) \\ = &\phi^{-5/2}\nabla_{2}{}^{\dot{R}}\xi^{\dot{A}} \\ &+ 2(\partial^{\dot{R}}\partial^{\dot{A}}H - \phi^{-1}J^{\dot{A}}\partial^{\dot{R}}H), \end{aligned} \quad (33)$$

where

$$H \equiv \tilde{H} - bK_{\dot{B}}p^{\dot{B}}. \quad (34)$$

Equations (25) and (33) can now be written in the covariant form

$$\psi^{\dot{A}\dot{R}}{}_{C} = \nabla_{C}{}^{\dot{R}}\xi^{\dot{A}} + \phi^{-2}\nabla^{B\dot{R}}[\phi\nabla^{S\dot{A}}(\phi\chi_{CBS})] \quad (35)$$

[cf. Eq. (28)], where $\chi_{ABC}$ is totally symmetric and in the frame (20) has components $\chi_{222} = \phi^{5/2}H$ and all others equal to zero. In a covariant way, this means that $\chi_{ABC}$ is proportional to $l_{A}l_{B}l_{C}$.

The only restriction on the potential $\chi_{ABC}$ comes from the equations $H^{\dot{A}}{}_{22} = 0$, which amount to

$$\nabla_{2\dot{R}}\{\phi^{-2}\nabla^{B\dot{R}}[\phi\nabla^{S\dot{A}}(\phi\chi_{2BS})]\} = 0.$$

After some rearrangement these conditions can be written as

$$\phi\partial^{\dot{A}}\{\phi^{-1}[D_{\dot{R}}\partial^{\dot{R}}H - \tfrac{1}{2}(\partial^{\dot{B}}Q_{\dot{B}\dot{R}})\partial^{\dot{R}}H - \tfrac{1}{2}\phi^{-2}C^{(3)}H]\} = 0,$$

where $C^{(3)} \equiv 2C_{1122}$ and we have used the fact that

$$\tfrac{1}{2}R\phi^{-2} = -\partial_{\dot{A}}\partial_{\dot{B}}Q^{\dot{A}\dot{B}} - 6\phi\,\partial_{\dot{A}}(\phi^{-2}D^{\dot{A}}\phi) = 0,$$

$$\phi^{-2}C_{12\dot{A}\dot{B}} = \tfrac{1}{2}\partial^{C}\partial_{(\dot{A}}Q_{\dot{B})\dot{C}} - \phi^{-1}\partial_{(\dot{A}}D_{\dot{B})}\phi = 0,$$

and that

$$\phi^{-3}C^{(3)} = -\tfrac{1}{3}\phi^{-1}\partial_{\dot{A}}\partial_{\dot{B}}Q^{\dot{A}\dot{B}}$$

is independent of $p^{\dot{A}}$ (see Ref. 14), which implies that the expression between braces is a function independent of $p^{\dot{A}}$. On the other hand, from Eq. (33) one finds that the field is unchanged if $H$ is replaced by $H + a(q^{\dot{R}})\phi^{2} + c(q^{\dot{R}})$. Therefore, by using this freedom, one can make the potential $H$ to satisfy the wavelike equation

$$D_{\dot{R}}\partial^{\dot{R}}H - \tfrac{1}{2}(\partial^{\dot{R}}Q_{\dot{B}\dot{R}})\partial^{\dot{R}}H - \tfrac{1}{2}\phi^{-2}C^{(3)}H = 0. \quad (36)$$

Expressed in covariant form, in terms of $\chi_{ABC}$, this equation is equivalent to

$$\nabla_{\dot{R}(S}\phi^{-3}\nabla^{C\dot{R}}\phi^{3}\chi_{AB)C} - 2C_{(SA}{}^{CD}\chi_{B)CD} = 0, \quad (37)$$

and according to Eqs. (22) and (35) the field $H_{A\dot{B}\dot{C}}$ is given by

$$H_{A\dot{B}\dot{C}} = -2C_{A\dot{B}\dot{C}D}\xi^{\dot{D}} + \nabla^{S}{}_{(\dot{B}}\phi^{-2}\nabla^{R}{}_{\dot{C})}\phi\nabla^{D}{}_{\dot{A}}\phi\chi_{SRD}. \quad (38)$$

The last term can also be expressed in the form

$$\phi\nabla^{S}{}_{(\dot{A}}\phi^{-2}\nabla^{R}{}_{\dot{B}}\phi^{-2}\nabla^{D}{}_{\dot{C})}\phi^{3}\chi_{SRD} \quad (39)$$

[see Eq. (3.16) of Ref. 9].

In terms of the Newman–Penrose notation, in a frame such that $\kappa = 0 = \sigma$ (and hence $\Psi_{0} = 0 = \Psi_{1}$), Eq. (37) takes the form

$$[(\Delta + 2\gamma - \bar{\gamma} + \bar{\mu})(D + 3\epsilon + 2\rho) - (\bar{\delta} + 2\alpha + \bar{\beta} - \bar{\tau})(\delta + 3\beta + 2\tau) - \Psi_{2}]\chi = 0, \quad (40)$$

where $\chi$ is the only nonvanishing component of $\chi_{ABC}$ and we have made use of Eq. (19), which in this case gives $\kappa = 0 = \sigma$, $\rho = D \ln \phi$, $\tau = \delta \ln \phi$. According to Eq. (35) the components of the field are given by

$$\begin{aligned}
&\psi_{0'0'0} = (D - \bar{\epsilon})\xi_{0'}, \quad \psi_{0'1'0} = (\delta - \bar{\alpha})\xi_{0'} + \bar{\rho}\xi_{1'},\\
&\psi_{1'0'0} = (D + \bar{\epsilon})\xi_{1'} - \bar{\pi}\xi_{0'}, \quad \psi_{1'1'0} = (\delta + \bar{\alpha})\xi_{1'} - \bar{\lambda}\xi_{0'},\\
&\psi_{0'0'1} = (\bar{\delta} - \bar{\beta})\xi_{0'} + (D + 2\epsilon - \bar{\epsilon} + \rho)(D + 3\epsilon)\chi,\\
&\psi_{0'1'1} = (\Delta - \bar{\gamma})\xi_{0'} + \bar{\tau}\xi_{1'}\\
&\qquad + (\delta + 2\beta - \bar{\alpha} + \tau)(D + 3\epsilon)\chi + \bar{\rho}(\delta + 3\beta)\chi,\\
&\psi_{1'0'1} = (\bar{\delta} + \bar{\beta})\xi_{1'} - \bar{\mu}\xi_{0'}\\
&\qquad + (D + 2\epsilon + \bar{\epsilon} + \rho)(\delta + 3\beta)\chi - \bar{\pi}(D + 3\epsilon)\chi,\\
&\psi_{1'1'1} = (\Delta + \bar{\gamma})\xi_{1'} - \bar{\nu}\xi_{0'}\\
&\qquad + (\delta + 2\beta + \bar{\alpha} + \tau)(\delta + 3\beta)\chi - \bar{\lambda}(D + 3\epsilon)\chi.
\end{aligned} \tag{41}$$

By comparing Eq. (35) with (the complex conjugate of) Eq. (4) we see that the effect of a "gauge transformation" (4) on $\psi_{\dot{A}\dot{B}\dot{C}}$ can be taken into account by simply shifting the spinor field $\xi^{\dot{A}}$ (which is not restricted by the field equations) and leaving the potential $\chi_{ABC}$ unchanged.

In Refs. 19 and 20 it was shown that the existence of a two-index Killing spinor establishes a relation between decoupled components and potentials for massless fields of spins $\frac{1}{2}$, 1, and 2 (see also Ref. 8). A similar relation holds for Rarita–Schwinger fields. A two-index Killing spinor $L_{AB}$ satisfies

$$\nabla_{(A}{}^{\dot{R}} L_{BC)} = 0 \tag{42}$$

and its existence requires that the conformal curvature be of type D or N. All the vacuum type D space-times admit a two-index Killing spinor. In a frame such that $\Psi_2$ is the only nonvanishing component of the Weyl spinor the solution of (42) is given by $L_{00} = 0 = L_{11}$, and $L_{01} = L_{10} = \text{const}(\Psi_2)^{-1/3}$ satisfies

$$\begin{aligned}
&(D + \rho)L_{01} = 0, \quad (\Delta - \mu)L_{01} = 0,\\
&(\delta + \tau)L_{01} = 0, \quad (\bar{\delta} - \pi)L_{01} = 0,
\end{aligned} \tag{43}$$

where we are using 0 and 1 to label the spinor components in order to get a closer agreement with the Newman–Penrose conventions. By using Eq. (43) it is easy to see that if $H_{111}$ is a solution of Eq. (12), then

$$\chi = (L_{01})^3 H_{111} \tag{44}$$

is a solution of Eq. (40), and conversely. This means that the component $H_{111}$ acts as a potential for all the components of a field that, in general, is different from the original one. Since in all the type D vacuum metrics Eq. (12) is known to admit separable solutions,[4,5] so does Eq. (40) (see also Ref. 21).

A restricted class of type N metrics admits a solution of Eq. (42). In this case the solution of Eq. (42) must be of the form $L_{AB} = L_A L_B$ and in a frame such that $\Psi_4$ is the only nonvanishing component of the Weyl spinor, $L_0 = 0$, and $L_1$ must satisfy

$$\begin{aligned}
&(D + \epsilon + \rho)L_1 = 0, \quad (\Delta + \gamma)L_1 = 0,\\
&(\delta + \beta + \tau)L_1 = 0, \quad (\bar{\delta} + \alpha)L_1 = 0.
\end{aligned} \tag{45}$$

By commuting the derivatives acting on $H_{000}$ one can rewrite Eq. (10) as

$$\begin{aligned}
&[(\Delta - 4\gamma - \bar{\gamma} + \bar{\mu})(D - 3\epsilon - 4\rho)\\
&\quad - (\bar{\delta} - 4\alpha + \bar{\beta} - \bar{\tau})(\delta - 3\beta - 4\tau) - 10\Psi_2]H_{000}\\
&\quad = 0,
\end{aligned} \tag{46}$$

and using Eq. (45) one can easily verify that if $H_{000}$ satisfies Eq. (10), then

$$\chi = (L_{11})^3 H_{000} = (L_1)^6 H_{000} \tag{47}$$

satisfies Eq. (40), and conversely. Furthermore, if $\chi$ is a solution of Eq. (40) then, using Eq. (45), one finds that $(L_1)^{-3}\chi$ satisfies the massless field equation for spin-0 and that $(L_1)^{-2}\chi$, $(L_1)^{-1}\chi$, and $L_1\chi$ are potentials for a Weyl neutrino field, an electromagnetic field, and a gravitational perturbation, respectively (see Ref. 20). This relationship between potentials yields a relationship between fields. For instance, using Eq. (38) and following the same steps as in Ref. 20, one finds that, in terms of the components $\phi_{\dot{A}\dot{B}}$ of the electromagnetic field generated by $(L_1)^{-1}\chi$,

$$\begin{aligned}
H_{\dot{A}\dot{B}\dot{C}} = &-\{2C_{\dot{A}\dot{B}\dot{C}\dot{D}}\xi^{\dot{D}} + L^{S}\nabla_{S\dot{A}}\phi_{\dot{B}\dot{C}}\\
&- \tfrac{3}{2}\phi_{(\dot{A}\dot{B}}\nabla^{S}{}_{\dot{C})}L_S\}.
\end{aligned}$$

On the other hand, using the fact that $L_A$ is a quadruple principal spinor of the conformal curvature, from Eq. (6) it follows that $\phi_{AB} = H_{ABC}L^{C}$ satisfies the source-free Maxwell equations. This shows that the spin-raising and spin-lowering operations, valid in the case of conformally flat space-time (see §6.4 of Ref. 22), also apply in the present case.

## V. CONCLUSIONS

The results presented here show that despite the differences between Eq. (1) and the usual massless free-field equations and the equations for gravitational perturbations, in an algebraically special space-time all of them share several properties such as the existence of decoupled equations and the possibility of expressing all the components of the field in terms of a single scalar potential. Moreover, the decoupled equations and the equations for the potentials can be summarized by certain general expressions applicable to the cases $s = 0, \frac{1}{2}, 1, \frac{3}{2}$, and 2.

In the cases $s = \frac{1}{2}$, 1, and 2, it is known that the equation for the potential is the adjoint, in the sense of Ref. 8, of the corresponding decoupled equation, which is related to the fact that the respective field equations are self-adjoint. Similarly, Eq. (1) is self-adjoint and Eq. (40) is precisely the adjoint of the decoupled equation (10).


## ACKNOWLEDGMENT

The author acknowledges support from the Sistema Nacional de Investigadores.



[1] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976).

[2] P. C. Aichelburg and H. K. Urbantke, Gen. Relativ. Gravit. **13**, 817 (1981).

[3] R. Güven, Phys. Rev. D **22**, 2327 (1980).

[4]N. Kamran, J. Math. Phys. **26**, 1740 (1985); N. Kamran and R. G. McLenaghan, in *Gravitation and Geometry: A Volume in Honor of I. Robinson*, edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987).
[5]G. F. Torres del Castillo, J. Math. Phys. **29**, 2078 (1988).
[6]S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[7]L. S. Kegeles and J. M. Cohen, Phys. Rev. D **19**, 1641 (1979).
[8]R. M. Wald, Phys. Rev. Lett. **41**, 203 (1978).
[9]G. F. Torres del Castillo, J. Math. Phys. **25**, 342 (1984).
[10]G. F. Torres del Castillo, J. Math. Phys. **27**, 1586 (1986).
[11]M. S. Hickman and C. B. G. McIntosh, Gen. Relativ. Gravit. **18**, 1275 (1986).
[12]B. P. Jeffryes, preprint, Max-Planck-Institut, MPA 263, 1986.
[13]J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976); in *Asymptotic Structure of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).
[14]G. F. Torres del Castillo, J. Math. Phys. **24**, 590 (1983).
[15]J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[16]R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U.P., Cambridge, England, 1984), Vol. 1.
[17]G. F. Torres del Castillo, J. Math. Phys. **29**, 971 (1988).
[18]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[19]G. F. Torres del Castillo, Proc. R. Soc. London Ser. A **400**, 119 (1985).
[20]G. F. Torres del Castillo, J. Math. Phys. **27**, 1583 (1986).
[21]N. Kamran, C. R. Acad. Sci. **304**, 299 (1987).
[22]R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U.P., Cambridge, England, 1986), Vol. 2.

# Simplical minisuperspace. III. Integration contours in a five-simplex model

James B. Hartle
*Department of Physics, University of California, Santa Barbara, California 93106*

(Received 9 August 1988; accepted for publication 21 September 1988)

The no boundary proposal for the wave function of the universe is investigated in a minisuperspace model of pure gravity with cosmological constant. The model's four geometries consist of five four-simplices joined together to make the surface of a five-simplex from which one four-simplex face has been removed. The model is further simplified by symmetrically choosing all the interior edges of equal length and all the edges of the four-simplex boundary of equal length. The wave function is thus a function of a single boundary squared edge length and is specified by an integral over the single interior edge length. The analytic properties of the action in the space of complex edge lengths are exhibited, its classical extrema are calculated, and the possible contours of integration defining the wave function of the universe are discussed. A descending contour of constant imaginary action is proposed along which the integral defining the wave function is convergent and which predicts classical space-time in the late universe. This contour is the analog for the model of the conformally rotated contour appropriate to Euclidean sums over asymptotically flat space-times. The wave function is evaluated numerically for this contour both directly and by semiclassical methods.

## I. INTRODUCTION

The "no boundary" proposal[1] for the initial conditions of our universe prescribes, among other amplitudes, the wave function of a closed universe on a connected spacelike surface as a Euclidean sum over histories of the form

$$\Psi_0(h,\phi,\partial M) = \sum_M \nu(M) \int_C \delta g\, \delta\Phi \exp(-I[g,\Phi,M]). \tag{1.1}$$

The arguments of the wave function, $h$ and $\phi$, denote the three metric and matter field configurations, respectively, on the three-manifold $\partial M$; $I$ is the Euclidean gravitational action for the metric $g$ and matter field configurations $\Phi$ on a four-manifold $M$. The sum over manifolds is over a class of four-manifolds $M$ that have a boundary $\partial M$ *and no other boundary*. The functional integral is over the four-metrics $g$ and matter field configurations $\Phi$ that induce $h$ and $\phi$, respectively, on the boundary $\partial M$. Other amplitudes are prescribed by this proposal. For example, there are the amplitudes associated with a surface that has disconnected parts,[2,3] important for the value of the cosmological constant,[4] or the "multisurface" amplitudes important for the recovery of a notion of time.[5] These have an analogous construction to that of the wave function (1.1).

To make a construction such as (1.1) definite, the class of manifolds, the measure for the functional integrals, and the contour $C$ over which these integrals are to be taken must all be specified. Various possibilities have been discussed for the class of manifolds[2,6] and for the measure.[7] In this paper we shall discuss some possibilities for the contour $C$ in the context of a simplical minisuperspace model.

For several reasons the contour of integration defining the wave function of the universe may be expected to run over complex metrics. First, were the action such as to make an integral over real metrics convergent, the wave function defined by (1.1) with a real contour of integration would contradict one of the immediate facts of our experience—the classical space-time of the late universe. Classical space-time is a prediction of an oscillatory wave function in those regions of configuration space (the classically allowed regions) where it is well approximated semiclassically.[2,8] The integral of $\exp(-I)$ over real Euclidean geometries can never oscillate. A complex contour is therefore necessary. Second, were the action such as to make an integral over real metrics convergent, it seems unlikely that (1.1) would yield a wave function satisfying the constraints required by diffeomorphism invariance and, in particular, the Wheeler–DeWitt equation.[9] A complex contour of appropriate range could well give a construction by which the constraints are satisfied.

Fortunately, the Euclidean Einstein action—the low energy limit of any quantum theory of gravity—does not permit a real contour of integration with the unacceptable properties described above. It can assume arbitrarily negative values when evaluated on certain real metrics[10] and (1.1) integrated over all real metrics would diverge. Merely from finiteness, one is naturally led to a complex contour.

In simple, familiar, flat space quantum field theories, there is no issue of the choice of contour for the Euclidean sum over histories defining the ground state. The sum is typically over real Euclidean field configurations with appropriate asymptotic behavior. Why then should the contour be an issue, or even a possibility for choice, for the sum over histories defining the analog of the ground state in the quantum mechanics of closed cosmologies? It is perhaps appropriate to briefly review the reasons.

Einstein gravity and gauge field theories are examples of theories that are most straightforwardly formulated in terms of redundant variables. Physical properties such as the positivity of the energy necessary for a stable ground state are features of the theory expressed in terms of these physical degrees of freedom. Sums over histories defining quantum amplitudes are sums over these physical degrees of freedom. Indeed, the machinery of gauge fixed functional integrals is a

 0022-2488/89/020452-09$02.50 © 1989 American Institute of Physics 

formalism for carrying out just such sums without explicitly isolating the physical degrees of freedom. Redundant gauge degrees of freedom can be "fixed" in such constructions. However, if there are *gauge invariant* redundant degrees of freedom, there can be many contours of integration that correspond to summing the physical degrees of freedom over a real, physically appropriate range, but which differ in the contour assigned to the redundant variables. Such contours are physically equivalent when they give convergent results. It is the existence of gauge invariant redundant variables that makes the choice of contour an issue in (1.1).

Linearized general relatively is an example of a theory of this type for which the physical and redundant degrees can be explicitly identified. The action can be arbitrarily negative for some linearized metrics, but is positive when restricted to the physical degrees of freedom. The energy expressed in terms of the physical degrees of freedom is positive; therefore, there is a stable ground state. Indeed, it could not be otherwise since linearized gravity is just the theory of a free spin-2 field in flat space-time. The possible contours of integration for constructing the ground state wave function by the analog of (1.1) can be explicitly discussed.[11,12] There is no real contour giving the ground state wave function since the resulting integral would be divergent. However, there are contours in which the redundant variables are integrated over complex values that correctly yield the ground state wave function of the Hamiltonian theory. Indeed, this can be demonstrated to all orders in perturbation theory.[12]

In the full theory of general relativity we have no explicit decomposition into physical and redundant variables. However, the positive energy theorem of classical general relativity[13] suggests the existence of a stable ground state when the theory is restricted to asymptotically flat space-times. The work of Gibbons, Hawking, and Perry,[10] coupled with the positive action theorem of Schoen and Yau[14] has shown how sensible convergent results can be obtained for Euclidean sums over asymptotically flat space-times, with contours along which the conformal degree of freedom takes complex values. However, in the case of closed cosmologies we have, as yet, no explicit prescription for the complex contour that defines the "no boundary" wave functions. Here we have neither an explicit decomposition into physical and redundant variables nor a notion of total energy to guide us. Therefore, it seems appropriate to search *generally* for suitable contours. In particular, we can ask whether there are complex contours of integration that (i) are convergent, (ii) lead to a wave function that predicts classical space-time in the late ($t \gtrsim 10^{-43}$ sec) universe, and (iii) lead to a wave function that satisfies the constraints implementing diffeomorphism invariance. In this paper we discuss this question in the simplest simplical minisuperspace model.

Minisuperspace models have a history of utility in the exploration of quantum gravity and quantum cosmology.[15,16] In a minisuperspace model the parameters needed to describe a family of space-time histories is truncated to a manageable number leading to a tractable quantum mechanical model of general relativity. In some circumstances these models may give rise to approximations to quantities of physical interest.[17,18] A useful class of systematically improvable models can be obtained by using the methods of the Regge calculus,[19] to restrict the geometries contributing to (1.1) to the possible simplicial geometries built on a fixed simplicial manifold. Such models were generally discussed in the first two papers of this series.[18,20]

The question of the possible contours of integration obeying criteria (i)–(iii) can be usefully explored in minisuperspace models for which the possible contours can be explicitly displayed. Simplicial minisuperspace models are particularly useful in this way. The possible contours of integration are the contours in the space of complex squared edge lengths of the simplicial geometry. Since the Regge action is an *algebraic* function of the squared edge lengths, its analytic properties as a function of these many complex variables are straightforwardly displayed and the consequences of a particular choice of contour analyzed in an elementary way. In this paper we carry out such an analysis for the simplest minisuperspace model. The model is specified in Sec. II. In Sec. III the semiclassical approximation to the sum over geometries is discussed. In Sec. IV it is shown that the steepest descents contour through the extrema that give classical space-time in the late universe is a contour satisfying the applicable criteria above. Some brief conclusions are drawn in Sec. V.

## II. THE MODEL

The surface of a tetrahedron (a three-simplex) consists of four triangles that together form a two-dimensional simplicial geometry without boundary. If one of these triangles is removed, the result is a two-dimensional simplicial geometry with a single one-dimensional boundary consisting of three edges (see Fig. 1). Two dimensions up, a similar procedure can be used to construct a four-dimensional simplicial geometry with a single $S^3$ boundary. A five-simplex consists of six points in five dimensions, with every pair defining an edge. The surface of a five-simplex consists of six four-simplices that together form a four-dimensional geometry without boundary. If one of these four-simplices is removed, the remaining five four-simplices form a simplicial four-geometry with a single three-sphere boundary. There are five vertices in the boundary and a single interior vertex. There are thus ten boundary edges and five interior ones that join the

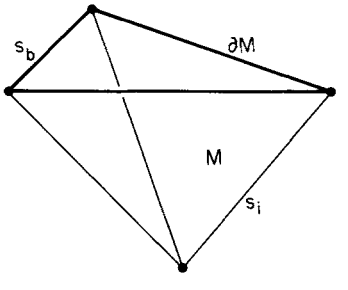



FIG. 1. A two-dimensional simplicial geometry. Remove one triangle from the surface of a tetrahedron and one obtains the two-dimensional simplicial geometry shown. The geometry consists of three triangles meeting in the single interior vertex. There is a single closed boundary consisting of the three edges (heavy lines) of the removed triangle. The simplical geometry used in constructing the minisuperspace model of this paper is the four-dimensional analog of that pictured here.

interior vertex to each of the five boundary vertices (see Fig. 1). The simplicial manifold is clearly invariant under permutations of the five boundary vertices. If all boundary edges are chosen equal and all interior edges separately equal, one obtains a family of symmetric simplicial geometries. Each is characterized by just two numbers: the squared boundary edge length $s_b$ and the squared interior edge length $s_i$. This symmetric family of simplicial four-geometries defines our minisuperspace model. We include no other geometrical degrees of freedom, no other manifolds, and no matter degrees of freedom. The wave function is thus a function only of the boundary edge length $s_b$ and given by the transcription of (1.1):

$$\Psi_0(s_b) = \int_C d\mu(s_i)\exp[-I(s_b,s_i)]. \tag{2.1}$$

The only integration is over the interior edge $s_i$ over a contour $C$ with a measure $\mu$. To complete the model it remains to specify $C$, $\mu$, and the action $I$.

For the action we take the Regge action for Euclidean Einstein gravity with cosmological constant, that is, we take the simplicial analog of

$$l^2 I = -2\int_{\partial M} d^3x\sqrt{h}K - \int_M d^4x\sqrt{g}R + 2\Lambda\int_M d^4x\sqrt{g}. \tag{2.2}$$

Here, $R$ is the scalar curvature, $\Lambda$ is the cosmological constant, $K$ is the extrinsic curvature scalar of the boundary, and $l = (16\pi G)^{1/2}$ is the Planck length in the units with $\hbar = c = 1$ used throughout. The first term is an integral over the boundary of the manifold and the second over its interior. The simplicial analog of (2.2) is[19,21]

$$l^2 I = -2\sum_{\sigma\in\partial M} A(\sigma)\psi(\sigma) - 2\sum_{\sigma\in\mathrm{int}(M)} A(\sigma)\theta(\sigma) + 2\Lambda\sum_{\tau\in\mathrm{int}(M)} V_4(\tau). \tag{2.3}$$

The sums are, respectively, over triangles $\sigma$ in the boundary $\partial M$, over triangles $\sigma$ in the interior of $M$, and over interior four-simplices $\tau$. Here, $\theta(\sigma)$ is the deficit angle of triangle $\sigma$ and $\psi(\sigma)$ is the angle between the normals to the boundary tetrahedra meeting in triangle $\sigma$. The area of triangle $\sigma$ is $A(\sigma)$ and $V_4(\tau)$ is the four-volume of the four-simplex $\tau$. Further details of definition, as well as practical prescriptions for expressing these quantities in terms of squared edge lengths, are reviewed in Paper I.

The analytic properties of the action as a function of the complex squared edge lengths will be important for an analysis of possible complex contours of integration: Although they can be explicitly exhibited for the model, they are also generally read off easily from Eq. (2.3) and the relations (Paper I) defining volumes, areas, and angles in terms of squared edge lengths. In particular, let $e_1, e_2, \ldots, e_n$ be vectors lying along the edges of an $n$-simplex emanating from one chosen vertex 0. The volume $n$-form for the simplex is $\omega_n = e_1 \wedge \ldots \wedge e_n$ and the squared volume is given in terms of it by [see Paper I, Eq. (3.6)]

$$V_n^2 = \omega_n\cdot\omega_n = [1/(n!)^2]\det(e_\alpha\cdot e_\beta). \tag{2.4}$$

Since $e_\alpha\cdot e_\beta = (s_{0\alpha} + s_{0\beta} - s_{\alpha\beta})/2$, where $s_{\alpha\beta}$ is the squared edge length between vertices $\alpha$ and $\beta$, we conclude that $V_4^2$ is a polynomial in the squared edge lengths. The deficit angle $\theta(\sigma)$ is $2\pi$ minus the sum of the "dihedral angles" between the three-simplices meeting at $\sigma$. Similarly, $\psi(\sigma)$ is $\pi$ minus the sum of the dihedral angles between the interior three-simplices meeting at $\sigma$. The dihedral angle $\phi$ between two three-simplices with the volume forms $\omega_3$ and $\omega_3'$ is [see Paper I, Eq. (3.9)]

$$\phi = \cos^{-1}((\omega_3\cdot\omega_3')/V_3V_3') \tag{2.5}$$

and $\omega_3\cdot\omega_3' = (3!)^{-2}\det(e_\alpha\cdot e_\beta')$. Using this formula, Eq. (2.4), the relation

$$\cos^{-1}(z) = -i\log(z + \sqrt{z^2-1}), \tag{2.6}$$

and the law of cosines to express the vector scalar products in terms of squared edge lengths, the analytic properties of the angles entering the action may be explicitly exhibited.

The analytic properties of the action in the complex squared edge lengths may be summarized as follows: There are logarithmic infinities on those surfaces that correspond to the vanishing of the polynomial which gives a three-simplex squared volume. The action is not single valued. Evidently, there are branch surfaces where the volume of a triangle, three-simplex, or four-simplex vanishes and, also, there are branch surfaces on which the squared cosine of any dihedral angle equals unity. However, a degenerate triangle or three-simplex implies the degeneracy of the four-simplex that contains it. Further, the identity [see Paper I, Eq. (3.12)]

$$\sin\phi = \tfrac{4}{3}AV_4/V_3V_3' \tag{2.7}$$

gives a relation between the dihedral angle between two three-simplices meeting in a triangle, their volumes $V_3$ and $V_3'$, the volume $V_4$ of the four-simplex they span, and the area $A$ of the triangle in which they meet. This shows that $\cos\phi = \pm 1$ only when either $A$ or $V_4$ vanishes. The branch surfaces of the action are therefore entirely contained in those surfaces on which the volume of some four-simplex vanishes. Except for these branch surfaces and logarithmic singularities the action is an analytic function of the squared edge lengths.

To carry out the integral (2.1) for the present model the action must be expressed in terms of the two edge lengths $s_b$ and $s_i$. The quantities occurring in (2.3) are straightforwardly calculated by use of (2.4) and (2.5). The results are most conveniently expressed in terms of the dimensionless ratios

$$\xi = s_i/s_b, \quad S = H^2 s_b/l^2, \tag{2.8}$$

where $H^2 = l^2\Lambda/3$. The volume of each four-simplex is

$$V_4 = (1/(24\sqrt{2}))s_b^2\,(\xi - \tfrac{3}{8})^{1/2}. \tag{2.9}$$

The area of each interior triangle is

$$A_i = \tfrac{1}{2}s_b(\xi - \tfrac{1}{4})^{1/2} \tag{2.10}$$

and the associated deficit angle is

$$\theta = 2\pi - 3\cos^{-1}\{\tfrac{1}{2}[(2\xi - 1)/(3\xi - 1)]\}. \tag{2.11}$$

The area of a boundary triangle is

$$A_b = (\sqrt{3}/4)s_b \tag{2.12}$$

and the angle $\psi$ for each is

$$\psi = \pi - 2\cos^{-1}\{(1/(2\sqrt{2}))[1/(3\xi-1)^{1/2}]\}. \quad (2.13)$$

The action (2.3) is then

$$I = [-S\mathscr{F}(\xi) + S^2\mathscr{G}(\xi)]/H^2 \equiv \mathscr{I}(S,\xi)/H^2, \quad (2.14)$$

where

$$\mathscr{F}(\xi) = 5\{\sqrt{3}\,[\pi - 2\cos^{-1}(z_1)] + (4\xi-1)^{1/2}\,[2\pi - 3\cos^{-1}(z_2)]\} \quad (2.15)$$

and

$$\mathscr{G}(\xi) = (5\sqrt{2}/8)(\xi - \tfrac{3}{8})^{1/2}, \quad (2.16)$$

with

$$z_1 = \frac{1}{2\sqrt{2}}\frac{1}{(3\xi-1)^{1/2}}, \quad z_2 = \frac{1}{2}\left(\frac{2\xi-1}{3\xi-1}\right). \quad (2.17)$$

In familiar quantum theories the choice of measure for a sum-over-histories formulation is dictated by requiring correspondence with the Hamiltonian version of the theory.[6] However, there may be no natural Hamiltonian formulation of the quantum mechanics of closed cosmologies from which to draw this information.[22] The corresponding constraints in the more general formulations of quantum mechanics have not yet been fully explored; they certainly have not been for the Regge calculus. Fortunately, the main results of this investigation for the contour $C$ do not seem very sensitive to the choice of measure among those in a "reasonable" class, e.g., measures that are polynomials in the squared edge lengths. Several choices have been suggested as natural in one way or another. For illustrative purposes we shall choose the simplest possibility and write

$$d\mu(s_i) = ds_i/(2\pi i l^2). \quad (2.18)$$

The factor $2\pi i l^2$ is a convenient normalization. Thus we can write (2.1) as

$$\Psi_0(S) = \frac{S}{2\pi i H^2}\int_C d\xi \exp\left[-\frac{\mathscr{I}(S,\xi)}{H^2}\right]. \quad (2.19)$$

The analytic and asymptotic properties of the action $\mathscr{I}(S,\xi)$ as a function of the complex variable $\xi$ are easily deduced from the general discussion of the analytic properties of the action or from the explicit expressions (2.9)–(2.13) and the definition of $\cos^{-1}(z)$. There is a square root branch point of $\mathscr{I}(S,\xi)$ at $\xi = \frac{3}{8}$ where the four-simplices become degenerate; there is another square root branch point at $\xi = \frac{1}{4}$ where the interior triangles become degenerate; and, finally, at $\xi = \frac{1}{3}$ there is a square root branch point and a logarithmic branch point near which $\mathscr{I}(S,\xi)$ behaves as

$$\mathscr{I}(S,\xi) \sim 10\sqrt{3}\log(3\xi-1)S. \quad (2.20)$$

There is also a branch point of the logarithms at infinity.

Choosing phases so that $\cos^{-1}(z)$ is real for $-1 < z < 1$, real values of the *squared* edge length $\xi$ correspond to real geometries with real metrics. Indeed, the metric inside each four-simplex is easily displayed in the basis whose defining basis vectors $e_i$ lie along the edges from the single interior vertex 0 to the five vertices $i$ of the boundary four-simplex; it is

$$g_{ij} = e_i \cdot e_j = \tfrac{1}{2}(s_{0i} + s_{0j} - s_{ij}), \quad (2.21)$$

where $s_{\alpha\beta}$ is the squared edge length joining vertices $\alpha$ and $\beta$. The metric is real for real values of $s_{\alpha\beta}$. For the symmetric choice of edges $s_{0i} = s_b\xi$ and $s_{ij} = s_b$, the eigenvalues of (2.21) are $\lambda = \frac{1}{2}$, $\lambda = \frac{1}{2}$, and the two values

$$\lambda = 1 + \tfrac{1}{2}\xi \pm [(1+\tfrac{1}{2}\xi)^2 - 2(\xi - \tfrac{3}{8})]^{1/2}. \quad (2.22)$$

Thus we have the signature $(+,+,+,+)$ for $\xi > \frac{3}{8}$ and $(-,+,+,+)$ for $\xi < \frac{3}{8}$. The real axis for $\xi > \frac{3}{8}$ is the regime of real Euclidean geometries; the regime for $\xi < \frac{3}{8}$ is the regime of real Lorentzian geometries.

The phases of the complex functions are chosen so that on the real axis for $\xi > \frac{3}{8}$ one has real volumes (2.9), real areas (2.10), real deficit angles (2.11), and real Euclidean action (2.14). It is thus convenient to define a first sheet for the action function cut from $\frac{3}{8}$ to $-\infty$. The reality of $\mathscr{I}$ above $\xi = \frac{3}{8}$ establishes that the action is real analytic:

$$\overline{\mathscr{I}(S,\bar{\xi})} = \mathscr{I}(S,\xi). \quad (2.23)$$

The Euclidean action for the Lorentzian geometries in the range $\xi < \frac{1}{4}$ is pure imaginary, taking opposite signs above and below the cut. In this range, about the cut,

$$\mathscr{F}(\xi) = i5\left\{-2\sqrt{3}\sinh^{-1}\left(\frac{1}{2\sqrt{2}(1-3\xi)^{1/2}}\right) + (1-4\xi)^{1/2} \times\left[2\pi - 3\cos^{-1}\left(\frac{1}{2}\frac{2\xi-1}{3\xi-1}\right)\right]\right\}, \quad (2.24a)$$

$$\mathscr{G}(\xi) = i(5\sqrt{2}/8)\,(\tfrac{3}{8} - \xi)^{1/2}. \quad (2.24b)$$

If the function $\mathscr{I}(S,\xi)$ is continued in $\xi$ once around all the branch points at $\xi = \frac{1}{4}$, $\frac{1}{3}$, and $\frac{3}{8}$, that is, through the cut along $\xi < \frac{1}{4}$, we reach its second sheet. The value of $\mathscr{I}(S,\xi)$ on the second sheet is the negative of its value on the first sheet. This is easily seen as follows: The function $\cos^{-1}(z)$ defined by (2.6) has branch points at $z = -1$, $+1$, and $\infty$ and may be discussed in the plane cut from $-\infty$ to $-1$ and 1 to $\infty$. The corresponding branch points of $\cos^{-1}(z_2)$ are at $\xi = \frac{3}{8}$, $\frac{1}{4}$, and $\frac{1}{3}$, respectively, so that the cuts defining its first sheet run from $\frac{1}{4}$ to $\frac{1}{3}$ and $\frac{1}{3}$ to $\frac{3}{8}$. Thus $\cos^{-1}(z_2)$ does not change when continued around $\xi = \frac{1}{4}$, $\frac{1}{3}$, and $\frac{3}{8}$. The branch points of $\cos^{-1}(z_1)$ corresponding to $z = -1$, 1, and $\infty$ lie at $\xi = \frac{1}{4}$, $\frac{3}{8}$, and $\frac{1}{3}$, respectively, so again the cuts defining the first sheet lie between these points. Thus, from (2.17), when continued once around, $\cos^{-1}(z_1) \to \cos^{-1}(-z_1) = \pi - \cos^{-1}(z_1)$. These results, together with the changes in sign of the roots in (2.15) and (2.16), are sufficient to establish that $\mathscr{I}(S,\xi)$ on the second sheet is the negative of its value on the first sheet. In particular, if one continues twice around $\xi = \frac{1}{4}$, $\frac{1}{3}$, and $\frac{3}{8}$, the function $\mathscr{I}(S,\xi)$ does not change.

The asymptotic behavior of the action for large $\xi$ is important for establishing the convergence of any proposed contour $C$. This is dominated by the scalar curvature and volume terms in (2.14). On the first sheet we have, for large $|\xi|$,

$$\mathscr{F}(\xi) \sim (5\sqrt{2}/8)S_{\text{crit}}(\xi - \tfrac{1}{4})^{1/2}, \quad (2.25a)$$

$$\mathcal{G}(\xi) \sim (5\sqrt{2}/8)\sqrt{2}(\xi - \tfrac{3}{8})^{1/2}, \tag{2.25b}$$

where

$$S_{\text{crit}} = (16/\sqrt{2})\,[2\pi - 3\cos^{-1}(\tfrac{1}{3})] = 29.306. \tag{2.26}$$

Thus

$$\mathcal{I}(S,\xi) \sim (5\sqrt{2}/8)\sqrt{2}S(S - S_{\text{crit}})\xi^{1/2}. \tag{2.27}$$

The asymptotic behavior on the first sheet thus depends crucially on whether $S$ is greater or less than the critical value $S_{\text{crit}}$. On other sheets the asymptotic behavior will change as the branch points at $\xi = \frac{1}{4}$ and $\frac{3}{8}$ are circled, resulting in changes of signs of the factors in (2.25).

## III. SEMICLASSICAL APPROXIMATIONS

The wave function defined by (2.19) will predict classical space-time for those values of $S$ and $H$ for which the semiclassical approximation appropriate to Lorentz signatured classical geometries is valid. In this section we explore the classical geometries predicted by the action (2.14) and the semiclassical approximations which can be built upon them.

Classical simplicial geometries are the extrema of the Regge action, here determined by the single algebraic equation

$$I'(S,\xi) = 0, \tag{3.1}$$

where a prime denotes the derivative with respect to $\xi$. For a given boundary edge length, Eq. (3.1) is to be solved for the value of $\xi$ that extremizes the action. This extremum determines the interior edges through (2.8) and thus a complete simplicial geometry. From (2.14), condition (3.1) is equivalent to

$$S = \mathcal{F}'(\xi)/\mathcal{G}'(\xi), \tag{3.2}$$

a relation which permits an easy graphical analysis of the extrema. The left-hand side of (3.2) is real. The right-hand side of (3.2) is real for real $\xi > \frac{3}{8}$ because both $\mathcal{F}$ and $\mathcal{G}$ are real and real for real $\xi < \frac{1}{4}$ because they are purely imaginary. The same right-hand side is reached whether one continues to above or below the cut $\xi < \frac{1}{4}$. Figure 2 shows a plot of the right-hand side of (3.2). From (2.25) asymptotically for large $\xi$ one has

$$\mathcal{F}'(\xi)/\mathcal{G}'(\xi) \sim S_{\text{crit}}, \tag{3.3}$$

where $S_{\text{crit}}$ is given by (2.26). Thus for every $0 < S < S_{\text{crit}}$ there is a Euclidean geometry with $\xi > \frac{3}{8}$, which is a solution. For every $S > S_{\text{crit}}$ there is a Lorentzian geometry with $\xi < \frac{1}{4}$, which is a solution.

The real solutions of (3.2) correspond to *pairs* of extrema of the action. For $S > S_{\text{crit}}$, $\xi < \frac{1}{4}$ the two extrema are reached from $\xi > \frac{3}{8}$ by continuing to either above the cut along $\xi < \frac{1}{4}$ or to below it. The action of both these extrema is purely imaginary [cf. (2.24)], but of opposite sign. For $S < S_{\text{crit}}$ there is an extremum on the first sheet with $\xi > \frac{3}{8}$ and a real Euclidean action. The second member of the pair is at the same location on the second sheet, reached by continuing around all the branch points at $\xi = \frac{1}{4}$, $\frac{1}{3}$, and $\frac{3}{8}$. It also has real action, but of opposite sign. As $S$ is varied smoothly from below $S_{\text{crit}}$ to above it, this pair of extrema migrate to ever larger values of real $\xi$ and reappear as the two extrema with $S > S_{\text{crit}}$ at large negative $\xi$.

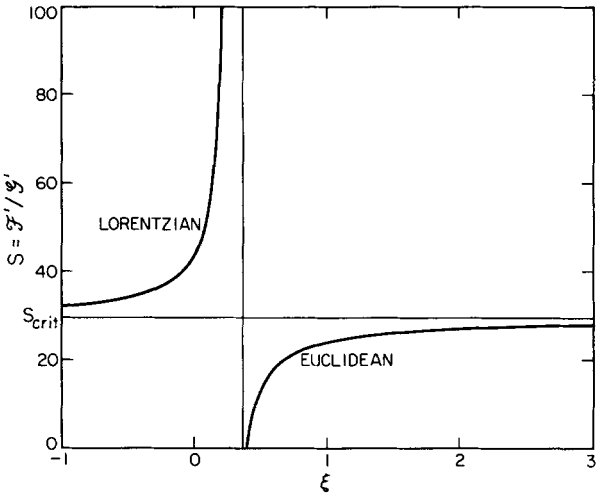



FIG. 2. Classical solutions. The solution of the algebraic equation that determines the extrema of the Euclidean Regge action can be obtained from the curve plotted here. It is a plot of the function $\mathcal{F}'(\xi)/\mathcal{G}'(\xi)$, which is equal to the scaled squared boundary edge length $S$ at an extremum. A value of $S$ thus determines $\xi$ and a unique squared interior edge length through $s_i = H^2 l^2 \xi / S$. For $S < S_{\text{crit}}$ the solutions have $\xi > \frac{3}{8}$, real action, and Euclidean signature. For $S > S_{\text{crit}}$ the solutions have $\xi < \frac{1}{4}$, imaginary Euclidean action, and Lorentzian signature.

It is interesting to compare the distribution of extrema obtained here with those of the analogous continuum minisuperspace model in which the geometries are restricted to be homogeneous and isotropic.[2] There the solution of the Euclidean Einstein equation with cosmological constant is a round four-sphere with radius $H^{-1}$. There are real Euclidean solutions for round three-sphere boundaries whose boundary radius is less than the critical value $H^{-1}$. In the continuum model there are *two* solutions for a boundary radius less than $H^{-1}$ corresponding to a four-geometry consisting of greater than a hemisphere of the four-sphere or less than a hemisphere. The action (2.2) is negative for both. However, because of the $\sqrt{g}$ in (2.2), it can be continued in the metric so as to reverse its overall sign. We should, therefore, count *four* extrema of the action in the space of complex metrics—two with positive action and two with negative.

Analogously to the continuum case the simplicial model has real extrema when the boundary is sufficiently small, $S < S_{\text{crit}}$. Unlike the continuum case there are two extrema of opposite sign of the action rather than four. A model with a single interior vertex is incapable of approximating both more than a hemisphere of a four-sphere and less.

For boundary radii greater than $H^{-1}$ the continuum model exhibits two pairs of extrema. Each pair has complex conjugate values of the action. The pairs differ in the sign of the real part of the action. The imaginary parts are the action of Lorentzian de Sitter space normalized to vanish at the minimum radius of contraction. For $S > S_{\text{crit}}$ the simplicial model displays two extrema with purely imaginary complex conjugate actions.

The wave function (2.19) will predict the correlations of classical space-time where it is well approximated by the semiclassical approximation associated with one or both of the Lorentzian extrema. This will be the case when the con-

tour $C$ can be distorted to pass as a steepest descents contour through these points and the range of integration in their neighborhoods gives the dominant contribution to the integral. For this to be the case $S$ and $H$ must be such that, locally, the integrand is sharply peaked about the extrema. Globally the contour must be such that no greater contributions arise from other extrema on the steepest descents contour or its end points.

Semiclassical approximations for $\Psi_0(S)$ that predict classical geometry are therefore linear combinations of the steepest descents approximation to the integral (2.19) arising from the Lorentzian extrema. That is, defining $\mathscr{I} = i\mathscr{S}$ so that $\mathscr{S}$ is real at a Lorentzian extremum, such semiclassical approximations are linear combinations of

$$\Psi_0(S) \sim \left[\frac{S^2}{2\pi H^2 \mathscr{S}''_{\rm ext}(S)}\right]^{1/2} \times \exp\left\{\pm i\left[\frac{\mathscr{S}_{\rm ext}(S)}{H^2} - \frac{\pi}{4}\right]\right\}, \tag{3.4}$$

where $\mathscr{S}_{\rm ext}$ and $\mathscr{S}''_{\rm ext}$ are evaluated at the extremum value of $\xi$. For the *CPT* symmetric wave function of the no boundary proposal we expect the real combination of these two exponentials.

If the local and global properties of the contour are appropriate as described above, we expect the steepest descents approximation to be valid when the argument of the exponentials in (3.4) is large. This will be the case for the large $s_b$ of the late universe, where

$$\mathscr{S}_{\rm ext}(S)/H^2 \sim \tfrac{5}{16}(S/H)^2 = \tfrac{5}{16}(\Lambda s_b^2/l^2); \tag{3.5}$$

it will also be the case over the whole range of $S$ (except for turning points) when $H^2 = l^2\Lambda/3$ is sufficiently small, as it certainly is in our late universe.

## IV. STEEPEST DESCENTS CONTOUR FOR THE NO BOUNDARY WAVE FUNCTION

The descending contour of constant imaginary action passing through both complex conjugate extrema for $S > S_{\rm crit}$ yields a convergent integral defining a real $\Psi_0$ that predicts classical space-time when $H^2$ is small. Therefore, it is a natural candidate for the contour defining the no boundary wave function of the universe. In this section we shall demonstrate these results and discuss the continuation of the resulting wave function to values of $S < S_{\rm crit}$. For the one complex variable of this model, a descending contour of constant imaginary action is a steepest descents contour.

Figure 3 shows the steepest descents contour that passes through both Lorentzian extrema when $S > S_{\rm crit}$. The contour consists of two complex conjugate sections, each passing through one extremum. Along with the real analyticity of the action, this ensures that the wave function resulting from (2.19) is real. Each section is a curve of constant $\mathrm{Im}(\mathscr{I})$ equal to its value $\mathscr{S}_{\rm ext}$ at the extremum through which it passes. Descending most steeply from the extremum one could generally end either at infinity, a singular point of the function $\mathscr{I}$, or at another extremum with the same value of $\mathrm{Im}(\mathscr{I})$. The only singular point is at $\xi = \frac{1}{3}$, at which $\mathrm{Im}(\mathscr{I})$ diverges; therefore no steepest descents contour can end there. The two exhibited extrema have opposite

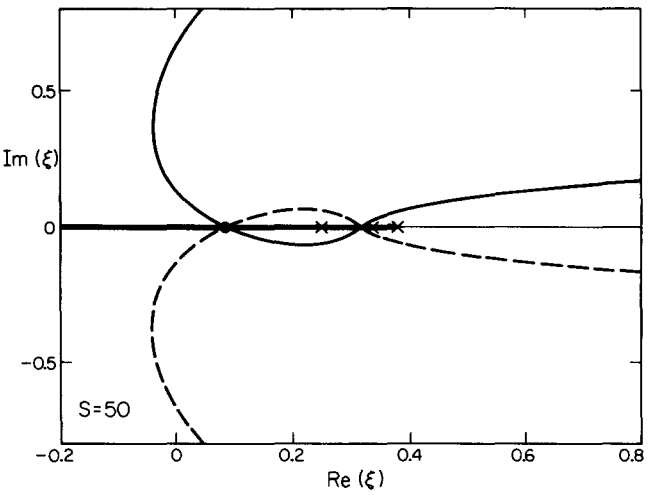



FIG. 3. The steepest descents contour for $S > S_{\rm crit}$. Shown here is the complex $\xi$ plane with branch points ($\times$) at $\xi = \frac{1}{4}$, $\frac{1}{3}$, and $\frac{3}{8}$ cut from $-\infty$ to $\frac{3}{8}$. The steepest descents contour for $S = 50$ is plotted. It consists of two complex conjugate sections plotted as the solid and dotted lines. Each passes (on different sheets) through an extremum of the action (●) at $\mathrm{Re}\,\xi = -0.0836$, $\mathrm{Im}\,\xi = 0$ and each has infinite end points.

signs of $\mathrm{Im}(\mathscr{I})$ so that a steepest descents contour cannot connect them. In the absence of any other extrema the steepest descents contour must pass from infinity to infinity; the numerical integration of Fig. 3 bears this out. Proceeding upward from the extremum, Eq. (2.27) shows the contour is asymptotic on the first sheet to the parabola

$$(5\sqrt{2}/8)S(S - S_{\rm crit})\,\mathrm{Im}(\xi^{1/2}) = \mathscr{S}_{\rm ext}. \tag{4.1}$$

Along this curve the asymptotic behavior of $\mathrm{Re}(\mathscr{I})$ is

$$\mathrm{Re}\,\mathscr{I}(S,\xi) \sim (5\sqrt{2}/8)(S - S_{\rm crit})|\xi|^{1/2}, \tag{4.2}$$

so that the defining integral (2.19) converges with any polynomial measure. Proceeding downward from the extremum the contour enters the second sheet. It cannot proceed directly to infinity since the action is asymptotically negative on the second sheet. Rather, for sufficiently large $S$ the contour passes through the cut between $\xi = \frac{1}{3}$ and $\frac{3}{8}$ reaching a third sheet, in effect changing the asymptotic sign of $\mathscr{G}$, but not of $\mathscr{F}$. On this third sheet the contour proceeds to infinity along the parabola

$$(5\sqrt{2}/8)(S + S_{\rm crit})\,\mathrm{Im}(\xi^{1/2}) = \mathscr{S}_{\rm ext}. \tag{4.3}$$

$\mathrm{Re}\,\mathscr{I}$ behaves as

$$\mathrm{Re}\,\mathscr{I}(S,\xi) \sim (5\sqrt{2}/8)(S + S_{\rm crit})|\xi|^{1/2} \tag{4.4}$$

and the integral from infinity to infinity is therefore convergent. For smaller values of $S$ the behavior near the branch points is slightly more complicated, but the contour is still infinite in extent. Since each section of the contour has no finite end points or other extrema along it, as $\mathscr{S}$ becomes large or $H^2$ small the behavior of integral (2.19) along each section is given increasingly accurately by the steepest descents approximation based on the Lorentzian extremum. Thus taking both sections of the contour together, the semiclassical approximation for $S > S_{\rm crit}$ becomes

$$\Psi_0(S) \sim \left[\frac{S^2}{2\pi H^2 \mathscr{S}''_{\rm ext}(S)}\right]^{1/2} 2\cos\left[\frac{\mathscr{S}_{\rm ext}(S)}{H^2} - \frac{\pi}{4}\right]. \tag{4.5}$$

Thus we recover a real $\Psi_0$ and classical space-time when the universe is large.[23]

The third sheet, reached by following the steepest descents contour downward from the extremum reached from the upper half of the first sheet, is the same as that reached from following the steepest descents contour upward from the extremum reached from the lower half of the first sheet. It is this crucial fact that allows the closure of the contour and a simple analytic continuation of integral (2.19) to $S < S_{\rm crit}$. The fact may be deduced from an elementary, but detailed analysis of the change in phases of the various terms in (2.14) as one proceeds along the contour. A more general argument is as follows: We are considering the two continuations $\mathscr{I}_{\rm I}(\xi)$ and $\mathscr{I}_{\rm II}(\xi)$ along complex conjugate curves. Initially, on the first sheet, as a consequence of real analyticity, the continuations are related by $\overline{\mathscr{I}_{\rm I}(\bar{\xi})} = \mathscr{I}_{\rm II}(\xi)$; this relation will continue to be maintained because the contours are conjugate to each other. Therefore if the continuations cross the real axis in a range where each are separately real, we have

$$\mathscr{I}_{\rm I}(\xi) = \overline{\mathscr{I}_{\rm I}(\bar{\xi})} = \mathscr{I}_{\rm II}(\xi) \tag{4.6}$$

and they will agree. However, no matter how the branch point are circled, $\mathscr{I}(\xi)$ is real for $\xi > \frac{1}{8}$. Circling one of the square root branch points can at most change its sign. Circling the logarithmic branch point changes $\cos^{-1}(z) \to \cos^{-1}(z) \pm \pi$. Therefore the two continuations must reach a common third sheet.

The asymptotic behaviors (4.2) and (4.4) allow the two sections of the steepest descents contour to be joined at infinity on the first and third sheets. The result is a closed contour defining the wave function of the universe. The contour can be distorted smoothly from the third sheet to the second because $\mathscr{I}(\xi)$ is not singular at the branch point $\xi = \frac{1}{8}$ through which it passes. The result is the closed contour of Fig. 4.

Since the contour is closed and finite, there is no obstacle to continuing the wave function defined by (2.19) to values of $S$ less than $S_{\rm crit}$. One can then investigate the semiclassical behavior of $\Psi_0$ in the limit where $H^2$ becomes small. In fact, the contour can be distorted into a steepest descents contour for $S < S_{\rm crit}$. The extrema are located at real values of $\xi$ on the first and second sheets. The asymptotic behavior of (2.19) for small $H^2$ is given by the integral in the neighborhood of the extremum on the first sheet. This is

$$\Psi_0(S) \sim -\left[S^2/2\pi H^2 I''_{\rm ext}(S)\right]^{1/2} \exp\left[-I_{\rm ext}(S)\right], \tag{4.7}$$

where $I_{\rm ext}(S)$ and $I''_{\rm ext}(S)$ are evaluated at the extremum value of $\xi$.

With the contour defined, both the integral (2.19) and its semiclassical approximations (4.5) and (4.7) are easily evaluated. Figure 5 shows a numerical evaluation of Eq. (2.16) for $H^2 = 50$ along the actual steepest descents contour. Figure 6 shows the corresponding semiclassical approximation evaluated from (4.5) and (4.7). For this value

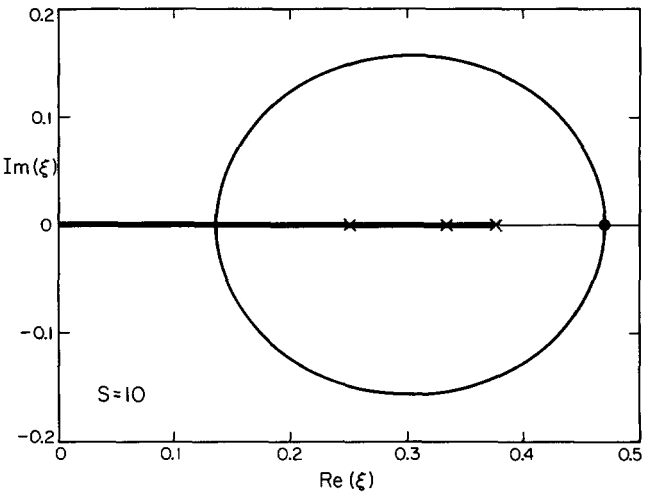



FIG. 4. Continuation to $S < S_{\rm crit}$. The steepest descents contour shown in Fig. 3 can be distorted into the closed contour shown. Starting on the positive real axis the contour winds around all finite branch points, through the cut $-\infty < \xi < \frac{1}{8}$, onto the second sheet, and around all the finite branch points again to close on the first sheet. Thus it winds around the branch points *twice* on the curve shown. Using the closed contour the integral can be continued to $S < S_{\rm crit}$. For these values the integral can be distorted into a steepest descents contour, shown here for $S = 10$. This contour passes through both real extrema (●) at Re $\xi = 0.469$, Im $\xi = 0$, one on the first sheet, one on the second. It is the integral on the first sheet that gives the dominant contribution to the integral defining the wave function in the semiclassical approximation.

of $H^2$, $\mathscr{S}/H^2 \gtrsim 20$ in the classical allowed region $S > S_{\rm crit}$, so that we expect the semiclassical approximation to give a good approximation to the actual integral. It does for large values of $S$. The semiclassical approximation gives a poor approximation for $S$ near $S_{\rm crit}$ and only a moderate approximation below $S_{\rm crit}$.

Both the wave function and its semiclassical approximation show the characteristic features expected from similar minisuperspace models based on symmetry.[2] Here, $S > S_{\rm crit}$ is the classically allowed region in which the wave function

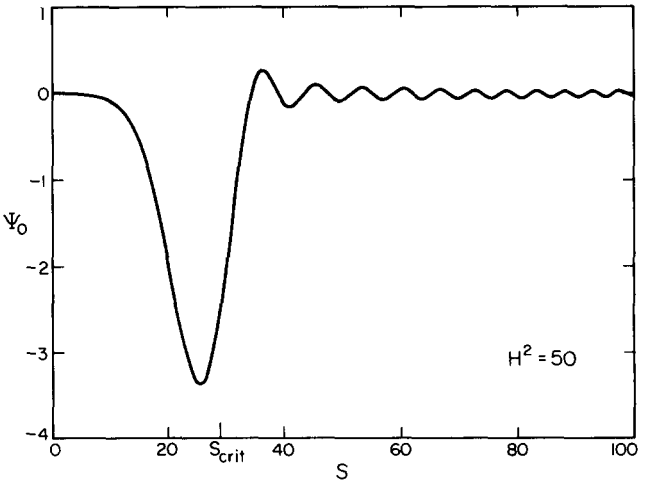



FIG. 5. The wave function. A numerical integration of the wave function defined by Eq. (2.19) and the steepest descents contour of Figs. 3 and 4 is plotted for $H^2 = 50$. The wave function oscillates in the classically allowed range of $S > S_{\rm crit}$. In the classically forbidden range of $S < S_{\rm crit}$ the wave function decays inward from the very large peak in $|\Psi_0|$ at $S_{\rm crit}$.

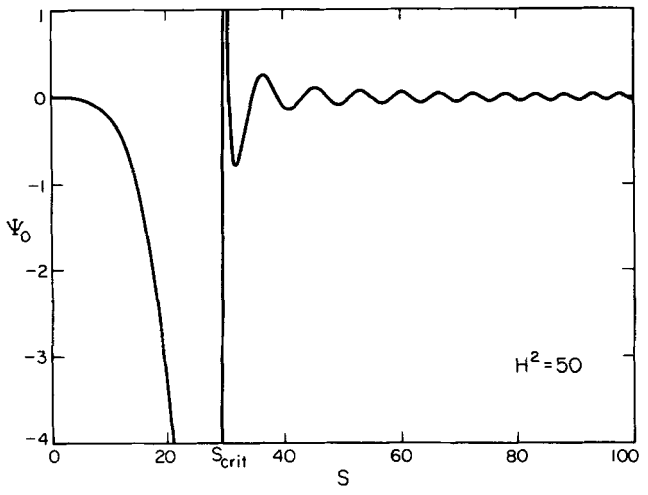



FIG. 6. The semiclassical wave function. The semiclassical approximation to the wave function specified by Eqs. (4.5) and (4.7) is plotted for $H^2 = 50$. The semiclassical approximation is infinite at the "turning point" $S = S_{crit}$ because $I''_{ext}(S)$ vanishes there. The semiclassical wave function becomes an increasingly accurate approximation to the wave function of Fig. 4 for large values of $S$. Where this approximation is valid we may say that the wave function predicts the correlations of classical de Sitter space.

oscillates. The semiclassical approximation here corresponds to classical de Sitter space. The oscillation at arbitrarily large values of $S$ corresponds to the limitless expansion of de Sitter space. The boundary between the classically allowed and classically forbidden regions at $S = S_{crit}$ corresponds to the minimum radius of contraction $H^{-1}$. The wave function is large near the boundary $S_{crit}$, corresponding to the most probable three spheres in de Sitter space. It decays exponentially from these high values for $S < S_{crit}$, reflecting the classically forbidden nature of this region.

## V. CONCLUSIONS

The steepest descents contour for the no boundary wave function of the universe explicitly displayed in this paper manifestly meets two of the criteria set forth in Sec. I. It leads to a convergent integral for $\Psi_0$ and to a wave function that predicts classical space-time when the universe is large. The third criterion, which concerned the constraints implementing diffeomorphism invariance, cannot be an issue for this simple Regge model since a simplicial geometry exhibits no exact nontrivial invariances, but only approximate ones (see, e.g., Ref. 18).

The proposed contour is in no sense a distortion or rotation of the contour of integration over real Euclidean geometries. That contour runs along the real $\xi$ axis from $\xi = \frac{3}{8}$ to $\xi = \infty$. Therefore, it has a finite end point at $\xi = \frac{3}{8}$ that will not be displaced by distortion or rotation. For $S > S_{crit}$ the integral (2.19) over this real contour converges. However, the asymptotic behavior for small $H^2$ is not governed by a Lorentzian extremum, but rather by the end point at $\xi = \frac{3}{8}$. The wave function defined by this contour does not predict classical space-time in the late universe.

The results of this model suggest that it will be of considerable interest to investigate whether a descending contour of constant imaginary action through the classical extrema can provide a general definition of the no boundary wave function consistent with convergence, classical space-time, and invariance.

## ACKNOWLEDGMENT

Thanks are due to D. Page and J. Halliwell for constructive criticism.

This work was supported in part by the National Science Foundation under Grant No. PHY 85-06686.

[1] S. W. Hawking, in *Astrophysical Cosmology: Proceedings of the Study Week on Cosmology and Fundamental Physics*, edited by H. A. Brüch, G. V. Coyne, and M. S. Longair (Pontificae Academiae Scientarium Scripta Varia, Vatican City, 1982); Nucl. Phys. B **264**, 185 (1984).
[2] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
[3] S. W. Hawking, Phys. Scr. T **15**, 151 (1987).
[4] S. Coleman (to be published); S. Giddings and A. Ströminger (to be published).
[5] See, e.g., J. B. Hartle, in *Proceedings of the Osgood Hill conference on conceptual problems of quantum gravity* (to be published) or *Proceedings of the 5th Marcel Grossmann conference on recent developments in general relativity* (to be published).
[6] See, e.g., J. B. Hartle, Class. Quantum Gravit. **2**, 707 (1985).
[7] See, e.g., H. Leutwyler, Phys. Rev. **134**, B1155 (1964); B. S. DeWitt, in *Magic Without Magic: John Archibald Wheeler*, edited by J. Klauder (Freeman, San Francisco, 1972); E. Fradkin and G. Vilkovisky, Phys. Rev. D **8**, 4241 (1973); L. Faddeev and V. Popov, Usp. Fiz. Nauk **111**, 427 (1973); M. Kaku, Phys. Rev. D **15**, 1019 (1977); C. Teitelboim, *ibid.* **28**, 310 (1983); H. Hamber, in *Critical Phenomena, Random Systems and Gauge Theories: Les Houches 1984*, edited by R. Stora and K. Osterwalder (Elsevier Sciences, Amsterdam, 1986); M. Bander, Phys. Rev. Lett. **57**, 1825 (1986).
[8] See, e.g., B. DeWitt, Phys. Rev. **160**, 571 (1967); U. Gerlach, Phys. Rev. **177**, 1929 (1969); V. Lapchinsky and V. Rubakov, Acta Phys. Polonica B **10**, 1041 (1979); T. Banks, Nucl. Phys. B **249**, 332 (1985); J. Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985). There are many later references.
[9] C. Teitelboim, Phys. Rev. Lett. **50**, 705 (1983); J. Halliwell (to be published).
[10] G. Gibbons, S. W. Hawking, and M. Perry, Nucl. Phys. B **138**, 141 (1978).
[11] J. B. Hartle, Phys. Rev. D **29**, 2730 (1984); J. B. Hartle and K. Schleich, in *Quantum Field Theory and Quantum Statistics*, edited by T. A. Batalin, C. J. Isham, and G. A. Vilkovisky (Adam Hilger, Bristol, 1987); H. Arisue, T. Fujiwara, M. Kato, and K. Ogawa, Phys. Rev. D **35**, 2309 (1987); D. Boulware (unpublished).
[12] K. Schleich, Phys. Rev. D. **36**, 2342 (1987).
[13] R. Schoen and S.-T. Yau, Phys. Rev. Lett. **43**, 1457 (1979); E. Witten, Commun. Math. Phys. **80**, 381 (1987).
[14] R. Schoen and S.-T. Yau, Phys. Rev. Lett. **42**, 547 (1979).
[15] For reviews of some of the earlier literature on the application of minisuperspace methods to canonical quantum gravity, see C. W. Misner, in *Magic Without Magic: John Archibald Wheeler*, edited by J. Klauder (Freeman, San Francisco, 1972); M. Ryan, *Hamiltonian Cosmology* (Springer, New York, 1972); M. A. H. MacCullum, in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. Sciama (Clarendon, Oxford, 1975).
[16] For a review of some of the earlier applications of minisuperspace to the no boundary proposal see, e.g., J. B. Hartle, in *High Energy Physics 1985: Proceedings of the Yale Theoretical Advanced Study Institute*, edited by M. J. Bowick and F. Gürsey (World Scientific, Singapore, 1986). References to both later and earlier papers can be found in J. J. Halliwell, *Bibliography of Papers on Quantum Cosmology* (to be published).
[17] M. Ryan and K. Kuchař, in *Gravitational Collapse and Relativity*, edited by H. Sato and T. Nakamura (World Scientific, Singapore, 1986).
[18] J. B. Hartle, J. Math. Phys. **26**, 804 (1985) (referred to as paper I).
[19] T. Regge, Nuovo Cimento **19**, 558 (1961).

[20]J. B. Hartle, J. Math. Phys. **27**, 287 (1986) (referred to as paper II).

[21]For the boundary term see J. B. Hartle and R. Sorkin, Gen. Relat. Gravit. **13**, 541 (1981).

[22]See, e.g., C. Teitelboim, Phys. Rev. D **25**, 3159 (1983); **28**, 297, 310 (1983); J. B. Hartle, in *Gravitation in Astrophysics*, edited by B. Carter and J. B. Hartle (Plenum, New York, 1987); J. B. Hartle, Phys. Rev. D **37**, 2818 (1988); J. B. Hartle (to be published).

[23]By taking only a single section of the contour we would recover a wave function that in the semiclassical limit behaves like $\exp(i\mathcal{S}_{\rm ext}/H^2)$ in the classically allowed region; possessing "outgoing waves," it is analogous to the "tunneling" wave function of A. Vilenkin [Phys. Lett. B **117**, 25 (1983); Phys. Rev. D **27**, 2848 (1983); **30**, 509 (1984); **37**, 2872 (1988)]. However, we have not investigated the continuation of this wave function to $S < S_{\rm crit}$.

# The hidden variable problem in algebraic relativistic quantum field theory

Miklós Rédei[a)]
*Loránd Eötvös University, H-1088 Budapest, Rákóczi út 5, Hungary*

(Received 17 May 1988; accepted for publication 5 October 1988)

Given two quasilocal $C^*$-algebras $A$ and $B$ of relativistic quantum field theory, their state spaces $E(A)$ and $E(B)$, and a positive, unit preserving map $L: B \to A$ respecting the relativistic quasilocal structure of $A$ and $B$, $(B,E(B))$ is said to be a local hidden theory of $(A,E(A))$ via $L$ if for every state $\varphi$ in $E(A)$ the state $L^*\varphi \in E(B)$ can be decomposed in $E(B)$ via a subcentral measure into states with pointwise strictly less dispersion than the dispersion of $\varphi$. It is shown that if there is a unique, locally normal, locally faithful, analytic vacuum state in $E(A)$ then $(A,E(A))$ cannot have a local hidden theory $(B,E(B))$ via $L$. This improves the result obtained in J. Math. Phys. **28**, 833 (1987).

## I. INTRODUCTION

In previous papers[1,2] the problem of hidden variables was redefined in the operator algebraic framework of quantum mechanics as the problem of finding conditions on the unital $C^*$-algebras $A$ and $B$ and on a positive, unit preserving map $L: B \to A$ that implies that $(B,E(B))$ is not a hidden theory of $(A,E(A))$ via $L$ [where $E(A),E(B)$ are the state spaces of $A$ and $B$][3] in the sense of the following definition.

*Definition 1:* Here $(B,E(B))$ is a hidden theory of $(A,E(A))$ via $L$ if for each state $\varphi \in E(A)$ one can find a positive (regular Borel) measure $\mu$ on the state space $E(B)$ with the help of which the composite state $\varphi \circ L = L^*\varphi \in E(B)$ can be obtained in the integral form

$$L^*\varphi(x) = \int \omega(x) d\mu(\omega), \quad x \in B, \tag{1}$$

such that for all $x \in B$ the following conditions hold:

$$\text{if} \quad \sigma_\varphi(Lx) > 0, \quad \text{then} \quad \sigma_\omega(x) < \sigma_\varphi(Lx), \tag{2}$$

$$\text{if} \quad \sigma_\varphi(Lx) = 0, \quad \text{then} \quad \sigma_\omega(x) = 0, \tag{3}$$

for all $\omega \in \operatorname{supp}\mu$, where $\sigma_\varphi(Lx) = \varphi((Lx)^2) - \varphi(Lx)^2$ and $\sigma_\omega(x) = \omega(x^2) - \omega(x)^2$ are the dispersions of the states $\varphi$ and $\omega$, and $\operatorname{supp}\mu$ denotes the support set of $\mu$.

This definition is a natural generalization of earlier hidden variable definitions, especially the one due to von Neumann.[4] In view of this definition, if, besides having positivity and being unit preserving, $L$ has certain additional algebraic properties, a negative result on the existence of a hidden theory of $(A,E(A))$ via $L$ can be interpreted as determining an algebraic structure responsible for a given statistical uncertainty inherent in the description of physical systems by $C^*$-algebras that represent the algebraic structure in question. It can be proved, for instance, that $(B,E(B))$ is not a hidden theory of $(A,E(A))$ via a Jordan homomorphism $L$ if $A$ is a simple $C^*$-algebra, which means that if $(B,E(B))$ is a hidden theory of $(A,E(A))$ via $L$ then the Jordan algebra structures of $A$ and $B$ must be regarded differently.[5]

Based on definition 1, the problem of *local* hidden variables can also be naturally reformulated in terms of quasilocal $C^*$-algebras[6]: if both $A$ and $B$ are quasilocal $C^*$-algebras then $(B,E(B))$ is said to be a local hidden theory of $(A,E(A))$ via $L$ if $L$ preserves the local structure of $A$ and $B$ in some appropriate sense, $(B,E(B))$ is a hidden theory of $(A,E(A))$ in the sense of Definition 1, and the measure $\mu$ in (1) can be chosen subcentral, which is to be interpreted as a natural locality property of $\mu$.[7]

A negative result on the existence of a local hidden theory of $(A,E(A))$ shows what locality properties must be violated if $(A,E(A))$ has a local hidden theory. It was proved in a previous paper that if $L$ maps the local algebras onto local algebras in such a way that disjoint algebras are mapped onto disjoint ones and, in addition, $L$ factorizes on disjoint algebras, then (modulo some technical assumptions) $(A,E(A))$ does not have a local hidden theory $(B,E(B))$ via $L$ provided the local algebras in both $A$ and $B$ are (isomorphic to) type I von Neumann algebras.[8]

However, this result is unsatisfactory in at least two respects: (i) the local algebras in relativistic quantum field theory cannot be type I von Neumann algebras[9] so the previous result does not apply to the relativistic case; and (ii) even more importantly, due to the lack of expression of relativistic covariance on the quasilocal algebra, the locality properties of $L$ are not related to relativistic locality in any way. But it is just the relativistically local (also called Einstein local[10]) hidden variable theories that have been the main subject of interest since Bell's work.[11] To have a result on the nonexistence of a local hidden theory of $(A,E(A))$, where $A$ is a quasilocal $C^*$-algebra of relativistic quantum field theory, is, furthermore, particularly desirable in light of the recent work of Summers and Werner, who have proved that Bell's inequalities are maximally and typically violated in relativistic quantum field theory.[12] The violation of Bell's inequalities in a theory $T$ is commonly interpreted in most papers in the literature on local hidden variables as the impossibility of $T$ being Einstein local. The results of Summers and Werner show that this interpretation is certainly not valid in relativistic quantum field theory, which implies that the usual approach to the local hidden variable problem through Bell's inequalities does not work in relativistic quantum field theory.

In this paper we prove a proposition asserting the nonexistence of a local hidden theory $(B,E(B))$ of $(A,E(A))$ via $L$, where $A$ and $B$ are quasilocal $C^*$-algebras of relativistic

[a)] Temporary address until June 1989: Department of Physics, University of Florida, Gainesville, Florida 32611.

 0022-2488/89/020461-03$02.50 © 1989 American Institute of Physics 

quantum field theory and $L$ is a positive, unit preserving map from $B$ into $A$ that has natural Einstein local properties (see Definition 2 below). Before stating the proposition we briefly recall the notion of a quasilocal algebra in relativistic quantum field theory and define the Einstein local properties of $L$.

## II. EINSTEIN LOCAL HIDDEN THEORIES

Recall that $A$ is a quasilocal $C^*$-algebra of relativistic quantum field theory if $A$ is the uniform closure of a net $\{A(V), V \subset M\}$ of (strictly) local $C^*$-algebras $A(V)$ (with common unit) associated to the open, bounded subsets $V$ of the Minkowski space $M$, where the net has the following properties: (i) if $V_1$ is contained in $V_2$ then $A(V_1)$ is a subalgebra of $A(V_2)$, (ii) if $V_1$ is spacelike separated from $V_2$ then every element of $A(V_1)$ commutes with every element of $A(V_2)$, and (iii) there is a representation $\alpha$ of the identity-connected component $P$ of the Poincaré group by automorphisms on $A$ such that $\alpha_g A(V) = A(gV)$ for all $V$ and $g \in P$.

Part of the axioms of relativistic quantum field theory is also the assumption of existence of at least one physical representation of the quasilocal $C^*$-algebra $A$, which means mathematically that one postulates the existence of an $\alpha$-invariant state $\varphi \in E(A)$ (vacuum) such that the spectrum condition holds in the corresponding cyclic representation $\pi_\varphi$.[13]

In what follows "relativistic quasilocal algebra" will always mean a quasilocal $C^*$-algebra of this type with the further assumption that all local algebras are von Neumann algebras and "relativistic quantum field theory" will mean a pair $(A, E(A))$ with a relativistic quasilocal algebra $A$ and its state space $E(A)$, which is supposed to contain at least one vacuum state.

Let $A$ and $B$ be two relativistic quasilocal algebras and denote by $\beta$ the representation of $P$ on $B$. We wish to define Einstein local properties of the positive, unit preserving map $L: B \to A$, by which we mean properties that express the similarity of the relativistic quasilocal structure of $A$ and $B$. In the following definition (a) is a natural locality demand, the content of (c) is that $L$ does not destroy the relativistic covariance whereas (b) is of technical nature.

*Definition 2:* The positive, unit preserving map $L: B \to A$ between two relativistic quasilocal algebras $A$ and $B$ is called Einstein local if (a) $L$ maps the local algebras into local algebras, (b) the restriction of $L$ to each local algebra is continuous in the ultraweak operator topology, and (c) $L$ commutes with the two representations $\alpha$ and $\beta$ in the sense that

$$\alpha_g \circ L = L \circ \beta_g, \quad g \in P. \tag{4}$$

We sum up with the following definition.

*Definition 3:* Let $A$ and $B$ be two relativistic quasilocal algebras. Here $(B, E(B))$ is a local hidden theory of $(A, E(A))$ if (a) $L$ is Einstein local, (b) $(B, E(B))$ is a hidden theory of $(A, E(A))$ via $L$ in the sense of Definition 1, and (c) $\mu$ in (1) can be chosen subcentral.

Before formulating the proposition let us recall a few definitions and facts that will be used in the proof. A state $\varphi$ on $A$ is called locally normal if the restriction of $\varphi$ to every local von Neumann algebra $A(V)$ is ultraweakly continuous. The state $\varphi$ is said to be locally faithful if the condition $x > 0$ implies $\varphi(x) > 0$ for any strictly local element $x \in A(V)$. Let $T$ be the translation subgroup of $P$. A state $\varphi$ is called a translation clustering state if

$$\lim_{t \to \infty} \varphi(x \alpha_{tg} y) = \varphi(x)\varphi(y)$$

for all spacelike $g \in T$ and for all local $x, y$.

For a positive, unit preserving map $L: B \to A$ the Cauchy–Schwartz inequality $L(x^2) \geqslant L(x)^2$ holds for all self-adjoint $x \in B$.[14] Let $\mu$ be a subcentral measure on $E(B)$ that decomposes a state $\psi \in E(B)$ in the sense of (1). Then if $\psi$ is a factor state, i.e., if the center $Z_\psi = \pi_\psi(B)'' \cap \pi_\psi(B)'$ of the von Neumann algebra $\pi_\psi(B)''$ generated by $\psi$ in the Gel'fand–Naimark–Segal (GNS) representation $\pi_\psi$ consists of the multiples of the identity only, then $\mu$ is the Dirac measure $\delta_\psi$ concentrated at $\psi$.

*Proposition 1:* Let $(A, E(A))$ and $(B, E(B))$ be two relativistic quantum field theories. If there is a non-dispersion-free, locally normal, locally faithful, $\alpha$-invariant, translation clustering state $\varphi$ on $A$ such that the spectrum condition holds in the representation $\pi_{L^*\varphi}$ then $(A, E(A))$ does not have a local hidden theory via $L$ in the sense of Definition 3.

*Proof:* The proof is similar to the proof of the proposition in Ref. 2. One shows first that if there is a state $\psi \in E(A)$ with nonzero dispersion $\sigma_\psi(Lx_0) > 0$ on some self-adjoint $x_0 \in B$, then $(B, E(B))$ cannot be a hidden theory of $(A, E(A))$ via $L$ if the only measure that decomposes $L^*\psi \in E(B)$ is the Dirac measure $\delta_{L^*\psi}$. To show this assume that $(B, E(B))$ is a hidden theory of $(A, E(A))$ via $L$. By integrating (2) with respect to $\mu = \delta_{L^*\psi}$ one obtains

$$\psi((Lx_0)^2 - Lx_0^2) > \psi(Lx_0)^2 - \int \omega(x_0)\, d\mu(\omega) = 0, \tag{5}$$

which is a contradiction since the left-hand side of (5) is not greater than zero by the Cauchy–Schwartz inequality for $L$.

Thus to prove Proposition 1 it is enough to show that there is a non-dispersion-free state $\psi \in E(A)$ such that $L^*\psi \in E(B)$ is a factor state. We prove that $L^*\varphi$ is a factor state over $B$ by showing that the assumption of $L^*\varphi$ not being a factor state contradicts the clustering property of $\varphi$.

Since $\varphi$ is $\alpha$ invariant and $L^*\varphi$ is $\beta$ invariant both $\alpha$ and $\beta$ are implemented by unitary representations $U$ and $W$ in the GNS representations $\pi_\varphi$ and $\pi_{L^*\varphi}$ of $A$ and $B$, respectively. Denote $L_\pi$ as the "representation" of $L$ in $\pi_\varphi$ and $\pi_{L^*\varphi}$, i.e., $L_\pi(\pi_{L^*\varphi}(x)) = \pi_\varphi(Lx)$. Then (c) in Definition 2 takes the form

$$U_g L_\pi(\cdot) U_g^* = L_\pi W_g(\cdot) W_g^*. \tag{6}$$

Assume that $L^*\varphi$ is not a factor state. Then there is a nontrivial projector $\pi_{L^*\varphi}(P)$ in the center:

$$Z = \pi_{L^*\varphi}(B)'' \cap \pi_{L^*\varphi}(B)'$$

$$= \bigcap_{V \subset M} (\pi_{L^*\varphi}(B(V))'' \cap \pi_{L^*\varphi}(B(V))').$$

The last equality follows because $B$ is a quasilocal algebra.[15] By the local normality of $\varphi$ and $L$, $L^*\varphi$ is locally normal, too, and so $\pi_{L^*\varphi}(B(V))'' = \pi_{L^*\varphi}(B(V))$. Thus $\pi_{L^*\varphi}(P)$ is a nontrivial projector contained in (the center of) each local

algebra $\pi_{L*\varphi}(B(V))$. Fix a $V$. Now $L$ maps the local algebras into local ones thus $L_\pi\pi_{L*\varphi}(P)$ is an element in some local algebra $\pi_\varphi A(V')$. We may assume that $L_\pi\pi_{L*\varphi}(P)$ is nonzero for if it were then we could take the orthogonal complement $(\pi_{L*\varphi}(P))^\perp$, which shares all the properties of $\pi_{L*\varphi}(P)$ stated so far and the unit preserving property of $L$ implies that $L_\pi\pi_{L*\varphi}(P)$ and $L_\pi(\pi_{L*\varphi}(P))^\perp$ cannot both be zero. So $L_\pi\pi_{L*\varphi}(P)$ is a nonzero element, which is also positive by positivity of $L$ and this implies that $L_\pi\pi_{L*\varphi}(P)$ has a nontrivial spectral projector $L_\pi\pi_{L*\varphi}(e)$. By a theorem of Araki[16] the elements of $Z$ commute with $W_g$ $(g\in P)$ and so $L_\pi\pi_{L*\varphi}(P)$ commutes with $U_g$ $(g\in P)$ by (6). But then $U_g$ commutes with $L_\pi\pi_{L*\varphi}(e)$, too, and it follows that $L_\pi\pi_{L*\varphi}(e)$ is a $U$-invariant nontrivial projector in $\pi_\varphi A(V')$. Let $\pi_\varphi(R)\in\pi_\varphi A(V')$ be any nonzero local projector orthogonal to $L_\pi\pi_{L*\varphi}(e)$. Since $\varphi$ is locally faithful we have

$$\langle\Omega_\varphi,\pi_\varphi(R)\Omega_\varphi\rangle 0, \quad \langle\Omega_\varphi,L_\pi\pi_{L*\varphi}(e)\Omega_\varphi\rangle 0, \tag{7}$$

where $\Omega_\varphi$ is the cyclic vector representing $\varphi$ in the $\pi_\varphi$ representation. On the other hand, by the clustering property of $\varphi$ and by the $U$ invariance of $L_\pi\pi_{L*\varphi}(e)$,

$$\begin{aligned} 0 &= \langle\Omega_\varphi,\pi_\varphi(R)L_\pi\pi_{L*\varphi}(e)\Omega_\varphi\rangle \\ &= \lim_{t\to\infty}\langle\Omega_\varphi,\pi_\varphi(R)U_{tg}L_\pi\pi_{L*\varphi}(e)U_{tg}^*\Omega_\varphi\rangle \\ &= \langle\Omega_\varphi,\pi_\varphi(R)\Omega_\varphi\rangle\langle\Omega_\varphi,L_\pi\pi_{L*\varphi}(e)\Omega_\varphi\rangle \end{aligned}$$

must hold, too, which contradicts (7).

## III. DISCUSSION

Let $\varphi$ be a vacuum state in $E(A)$. If the vacuum vector $\Omega_\varphi$ is analytic for the generator of time translations (in the representation $\pi_\varphi$) then by the Reeh–Schlieder theorem[17] $\varphi$ is cyclic and separating for the local algebras, which means that in this case $\varphi$ is locally faithful, too. Obviously, the requirement that the vacuum state is not dispersion-free is not a strong one; moreover, if the vacuum state $\varphi$ in $E(A)$ is the unique $\alpha$-invariant state then $\varphi$ is known to have also the translation clustering property (both in massive and in massless theories[18]). In this case, by (c) of Einstein locality of $L$, $L^*\varphi$ is the unique $\beta$-invariant state, therefore, by the assumption that $E(B)$ contains at least one vacuum state, $L^*\varphi$ is the (unique) vacuum state in $E(B)$; in particular, the spectrum condition is fulfilled in $\pi_{L*\varphi}$. So if $(A,E(A))$ is such that the vacuum state $\varphi$ is a unique $\alpha$-invariant, analytic, locally normal state then the assumptions of Proposition 1 are fulfilled and Proposition 1 tells us that such $(A,E(A))$ relativistic quantum field theories are the best possible ones if "best" means "containing the least statistical uncertainty," i.e., the statistical uncertainty inherent in the description of quantum fields by these theories cannot be reduced without violating either at least one of the Einstein local properties (1)–(3) in Definition 2 or the subcentrality of $\mu$.

Note that "statistical uncertainty" has been assumed in this paper to be measured by the dispersion of quantum states. However, the dispersion is not the only conceivable measure of uncertainty of a probability distribution: the entropy also expresses a kind of uncertainty, which is conceptually different from what is expressed by dispersion. Based on the entropic measure of uncertainty the nonlocal hidden variable problem was reformulated in a previous paper[19] and it will be interesting to investigate the local hidden variable problem in relativistic quantum field theory using entropic uncertainty.


## ACKNOWLEDGMENTS

This work was done while the author was holding a Courtesy Appointment in the Department of Physics, University of Florida. I wish to thank Professor D. B. Tanner for his support.



[1] M. Rédei, Found. Phys. **16**, 807 (1986).
[2] M. Rédei, J. Math. Phys. **28**, 833 (1987).
[3] For all the definitions and elementary facts in connection with the operator algebra theory we refer to O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, Berlin, 1979), Vol. I.
[4] See Refs. 1 and 2 for a detailed motivation of this definition and its relation to previous ones.
[5] Corollary 1 in Ref. 1.
[6] The theory of quasilocal $C^*$-algebras is described, e.g., in Chap. 2.6 in Ref. 3.
[7] For the definition of subcentral measure see Ref. 3, p. 363, for a detailed explanation why subcentrality does express physical locality see Ref. 2. It should be noted that the locality (i.e., subcentrality) of the decomposing measure has nothing to do with the locality of $L$ or with the locality of the net of local algebras. The subcentrality can be—and is in fact—defined quite generally as a property of a measure on the state space of a general (unital) $C^*$-algebra.
[8] See the Proposition in Ref. 2.
[9] H. Araki, Prog. Theor. Phys. **32**, 956 (1964); K. Fredenhagen, in *Operator Algebras and Mathematical Physics*, edited by Palle E. T. Jorgensen, and Paul S. Muhly (Am. Math. Soc., Providence, RI, 1987).
[10] Here the informal notion "Einstein local theory " means a theory that complies with relativistic principles such as the principle of "no action at a distance."
[11] J. S. Bell, Physics (NY) **1**, 195 (1961); J. S. Bell, in the *Proceedings of the International School of Physics*, "Enrico Fermi" Course 49 (Academic, New York, 1971).
[12] S. J. Summers and R. Werner, Commun. Math. Phys. **110**, 247 (1987); J. Math. Phys. **28**, 2440,2448 (1987).
[13] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[14] Proposition 3.2.4 in Ref. 3.
[15] See Ref. 3, p. 122.
[16] H. Araki, Prog. Theor. Phys. **32**, 844 (1964).
[17] H. Reeh and S. Schlieder, Nuovo Cimento **22**, 1051 (1961). For an algebraic formulation and proof of the Reeh–Schlieder theorem see H.-J. Borchers, Commun. Math. Phys. **1**, 57 (1965).
[18] H. Araki, K. Heep, and D. Ruelle, Helv. Phys. Acta **35**, 164 (1962).
[19] M. Rédei, Synthese **73**, 371 (1987).

# Classical solutions of a nonrelativistic model exhibiting spontaneous symmetry breakdown

D. Spehler
*Université Louis Pasteur, Strasbourg, France*

G. C. Marques
*Instituto de Fisica, Universidade de São Paulo, C.P. 20516, 01498 São Paulo, São Paulo, Brazil*

(Received 5 June 1987; accepted for publication 17 August 1988)

The inverse scattering method is employed in order to obtain solutions of a nonrelativistic and nonlinear model exhibiting spontaneous breakdown of symmetry. Explicit expressions for multiple-soliton solutions are developed. The one-soliton solution exhibits some interesting properties. In particular, they interpolate between two different vacua of the theory differing by a phase related to the soliton velocity.

## I. INTRODUCTION

This paper deals with the problem of finding solutions to nonlinear field theoretical equations through the use of the inverse scattering technique.[1] The equation that we will study in this paper is

$$i\frac{\partial}{\partial t}q(x,t)+\frac{\partial^2}{\partial x^2}q(x,t)-k|q|^2q(x,t)+k\rho q(x,t)=0\,, \tag{1.1}$$

where $k$ and $\rho$ are positive constants.

One of the motivations for studying (1.1) is that the nonrelativistic model associated to it describes a system exhibiting a manifold of degenerate vacua ($|q|^2=\rho$), thus being a typical model exhibiting spontaneous breakdown of symmetry. Furthermore, model (1.1) describes the effect of a net background charge in a nonrelativistic Bose gas whose interaction potential is a $\delta$ function.[2] In this context $k\rho$ is just the chemical potential.

The classical, finite energy solution to (1.1) should obey nontrivial asymptotic conditions

$$\lim_{x\to\infty}q(x,t)=\rho^{1/2}\,, \tag{1.2a}$$

$$\lim_{x\to-\infty}q(x,t)=\rho^{1/2}e^{i\alpha}\,, \tag{1.2b}$$

that is, the classical solution should tend to one of the field theoretical configurations associated to the degenerate vacua.

The usual repulsive nonlinear Schrödinger (NLS) equation is just a particular case of ours. That is, it is just the particular case in which the background charge is zero. Formally we take the $\rho\to0$ limit of (1.1). The application of the inverse scattering method (ISM) to solving the NLS equation is found in Ref. 3.

As is already well known, the ISM is applicable to equations of the form

$$\frac{\partial\Phi}{\partial t}=\hat{S}[\Phi]\,, \tag{1.3}$$

where $\hat{S}$ is a nonlinear operator (differential in $x$), whenever Eq. (1.3) can be represented in the form

$$\frac{\partial L}{\partial t}=i[L,\mathscr{H}]\,, \tag{1.4}$$

where $L$ and $\mathscr{H}$ are linear differential operators containing the function $\Phi(x,t)$.

In the case of Eq. (1.1), $L$ and $\mathscr{H}$ are given by

$$L=-i\sigma_3\frac{\partial}{\partial x}+\begin{pmatrix}0 & q\\ q^* & 0\end{pmatrix}, \tag{1.5}$$

$$\begin{aligned}\mathscr{H}=&-2\sigma_3\frac{\partial^2}{\partial x^2}-i\begin{pmatrix}0 & q_x\\ q_x^* & 0\end{pmatrix}\\&-2i\begin{pmatrix}0 & q\\ q^* & 0\end{pmatrix}\frac{\partial}{\partial x}+\begin{pmatrix}|q|^2 & 0\\ 0 & -|q|^2\end{pmatrix}-\rho\sigma_3\\=&H-\rho\sigma_3\,,\end{aligned} \tag{1.6}$$

$$\sigma_3=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{1.7}$$

The fact that the model studied here is just an extension of NLS model becomes more transparent when one looks to (1.5)–(1.7). The $H$ in (1.7) is the operator analogous to $\mathscr{H}$ in the context of the NLS model. Thus, the difference between the two, at the level of the operators $\mathscr{H}$ and $L$, resides in the piece $\rho\sigma_3$ in $\mathscr{H}$.

Since we have obtained the operators $L$ and $\mathscr{H}$ and, as we will show later, because of our ability to solve the inverse problem for the operator $L$, Eq. (1.1) can be solved by the ISM.

The order of the paper is the following: In Sec. II we discuss the direct problem, and in Sec. III we obtain the Gel'fand–Levitan–Marchenko equation, thus implementing the inverse program. Explicit applications for $N$-soliton solutions are shown in Sec. IV. In this section, we discuss some interesting properties of the one-soliton solution that we obtained explicitly. The final part, Sec. V, is dedicated to our conclusions.

## II. THE DIRECT PROBLEM

The next step in the ISM consists of the analyses of the eigenvalue problem for the differential operator $L$. That is, we consider the eigenvalue equation

 0022-2488/89/020464-06$02.50 © 1989 American Institute of Physics 

$$LN=\lambda N,\quad N=\begin{pmatrix}n_1\\ n_2\end{pmatrix}. \tag{2.1}$$

Since $\mathscr{H}$ is self-adjoint, it follows from (1.4) that the spectrum of $L$ is independent of time. On the other hand, the eigenvectors $N$ evolve in time in accordance with the equation

$$\frac{dN}{dt}=-i\mathscr{H}N\equiv -i(H+\rho\sigma_3)N, \tag{2.2}$$

where $\mathscr{H}$ in (2.2) is defined in (1.6). The explicit definition of (1.5) leads us to the following equation for the eigenvectors:

$$\frac{d}{dt}\begin{pmatrix}n_1\\ n_2\end{pmatrix}=\begin{pmatrix}-i(2\lambda^2+|q|^2)+i\rho & q_x+2iq\lambda\\ q_x^*-2iq^*\lambda & i(2\lambda^2+|q|^2-\rho)\end{pmatrix}\begin{pmatrix}n_1\\ n_2\end{pmatrix}. \tag{2.3}$$

As explained previously we will be interested in a set of solutions for the eigenvalue equations (2.1) where the unknown "potential" $q(x,t)$ satisfies the nontrivial asymptotic conditions (1.2). For $q(x,t)$ satisfying (1.2a) one can write, considering these asymptotic values for $q(x,t)$, the asymptotic eigenvalue equation

$$\begin{pmatrix}n_1\\ n_2\end{pmatrix}_x=\begin{pmatrix}i\lambda & -i\sqrt{\rho}\\ i\sqrt{\rho} & -i\lambda\end{pmatrix}\begin{pmatrix}n_1\\ n_2\end{pmatrix}. \tag{2.4}$$

If one looks for asymptotic solutions of the form

$$n_i=n_i e^{i\beta x}, \tag{2.5}$$

one gets the following set of values for $\beta$:

$$\beta=\pm\sqrt{\rho-\lambda^2}. \tag{2.6}$$

Thus, for

$$\rho>\lambda^2,$$

$\beta$ is real and the solution is of the form

$$A\exp(\sqrt{\rho-\lambda^2}\,x)+B\exp(-\sqrt{\rho-\lambda^2}\,x), \tag{2.7}$$

thus corresponding to a discrete spectra.

On the other hand, if

$$\rho<\lambda^2,$$

the solution takes the form

$$A'\exp(i\sqrt{\lambda^2-\rho}\,x)+B'\exp(-i\sqrt{\lambda^2-\rho}\,x)\equiv A'e^{iKx}+B'e^{-iKx}, \tag{2.8}$$

and one has a continuous spectrum. This is the one we will be interested in from now on.

We define the functions $\Phi$ and $\psi$ as solutions of Eq. (2.2) whose asymptotic forms are $\varphi$ and $\psi$, that is,

$$\psi(x,\lambda)\underset{x\to+\infty}{\to}\psi(x,\lambda)\equiv\begin{pmatrix}\lambda+\sqrt{\lambda^2-\rho}\\ \sqrt{\rho}\end{pmatrix}\exp(i\sqrt{\lambda^2-\rho}\,x), \tag{2.9a}$$

$$\bar{\psi}(x,\lambda)\underset{x\to+\infty}{\to}\bar{\psi}(x,\lambda)\equiv\begin{pmatrix}\sqrt{\rho}\\ \lambda+\sqrt{\lambda^2-\rho}\end{pmatrix}\exp(-i\sqrt{\lambda^2-\rho}\,x), \tag{2.9b}$$

$$\Phi(x,\lambda)\underset{x\to-\infty}{\to}\varphi(x,\lambda)\equiv\begin{pmatrix}e^{i\alpha}(\lambda+\sqrt{\lambda^2-\rho})\\ \sqrt{\rho}\end{pmatrix}\exp(i\sqrt{\lambda^2-\rho}\,x), \tag{2.9c}$$

$$\bar{\Phi}(x,\lambda)\underset{x\to-\infty}{\to}\bar{\varphi}(x,\lambda)\equiv\begin{pmatrix}\sqrt{\rho}\\ e^{-i\alpha}(\lambda+\sqrt{\lambda^2-\rho})\end{pmatrix}\exp(-i\sqrt{\lambda^2-\rho}\,x). \tag{2.9d}$$

We note that $\bar{\psi}$ and $\bar{\varphi}$ in (2.9) are defined as

$$\bar{\psi}\equiv\sigma_1\psi^*,\quad \bar{\Phi}\equiv\sigma_1\Phi^*. \tag{2.10}$$

In view of the asymptotic conditions (2.9a) and (2.9c) one can write the following integral equations for $\psi$ and $\Phi$:

$$\psi(x,\lambda)=\psi(x,\lambda)-\int_x^\infty K^1(x,x')\psi(x',\lambda)dx', \tag{2.11a}$$

$$\Phi(x,\lambda)=\varphi(x,\lambda)-\int_{-\infty}^x K^2(x,x')\varphi(x',\lambda)dx', \tag{2.11b}$$

where $K^1$ and $K^2$ in (20) are two by two kernel matrices, that is,

$$K^{(i)}(x,x')=\begin{pmatrix}K_{11}^{(i)}(x,x') & K_{12}^{(i)}(x,x')\\ K_{21}^{(i)}(x,x') & K_{22}^{(i)}(x,x')\end{pmatrix}. \tag{2.12}$$

One can further show that the following restrictions hold true:

$$K_{12}^{(i)*}=K_{21}^{(i)},\quad K_{11}^{(i)}=K_{22}^{(i)*}. \tag{2.13}$$

The pair of solutions $\psi$, $\bar{\psi}$ forms a complete set of solutions. In this way one can write the solutions $\Phi$ and $\bar{\Phi}$ as linear combinations of these two:

$$\begin{aligned}\Phi(x,\lambda)&=C_{11}(\lambda,k)\psi(x,\lambda)+C_{12}(\lambda,k)\bar{\psi}(x,\lambda),\\ \bar{\Phi}(x,\lambda)&=C_{11}^*(\lambda,k)\bar{\psi}(x,\lambda)+C_{12}^*(\lambda,k)\psi(x,\lambda),\end{aligned} \tag{2.14}$$

where $k$ in (2.14) is given by

$$k=\sqrt{\lambda^2-\rho}. \tag{2.15}$$

One can invert (2.14) and write, analogously, $\psi$ and $\bar{\psi}$ as linear combinations of $\Phi$ and $\bar{\Phi}$:

$$\begin{aligned}\psi(x,\lambda)&=C_{22}(\lambda,k)\Phi(x,\lambda)+C_{21}(\lambda,k)\bar{\Phi}(x,\lambda),\\ \bar{\psi}(x,\lambda)&=C_{21}^*(x,k)\Phi(x,\lambda)+C_{22}^*(\lambda,k)\bar{\Phi}(x,\lambda).\end{aligned} \tag{2.16}$$

The coefficients $C_{ij}$ in (2.14) and (2.15) can be expressed in terms of Wronskians of the functions $\Phi$ and $\psi$, being the Wronskian of two functions $u$, $v$ defined as

$$W(u,v)=u_1v_2-u_2v_1. \tag{2.17}$$

The coefficient $C_{12}(\lambda)$ defined in (2.14), for instance, can be expressed as

$$C_{12}(\lambda)=W(\Phi,\psi)/W(\Phi,\bar{\Phi}). \tag{2.18}$$

In order to compute the Wronskians we remember that they are $x$ independent and consequently they can be deter-

mined from the asymptotic expressions for the various functions given in Eqs. (2.9a)–(2.9d). One obtains, for instance,

$$W(\Phi,\bar{\Phi}) = -W(\psi,\bar{\psi}) = -2\sqrt{\lambda^2-\rho}(\sqrt{\lambda^2-\rho}+\lambda)$$
$$= -2k(k+\lambda)\,. \qquad (2.19)$$

By computing other Wronskians one gets the following relevant relation among the coefficients:

$$C_{22}^*(\lambda) = -C_{11}(\lambda)\,,$$
$$C_{12}(\lambda) = C_{21}(\lambda)\,, \qquad (2.20)$$
$$C_{11}(\lambda)C_{11}^*(\lambda) - C_{12}(\lambda)C_{12}^*(\lambda) = -1\,.$$

The time dependence of the coefficients $C_{ij}$ can be inferred from the time dependence of the eigenvalues $N(x,t)$. By assuming that this time dependence is of the form

$$N(x,t) = f(t)\psi(x,t), \qquad (2.21)$$

and since $N(x,t)$ evolutes in accordance with (2.2), where $\mathscr{H}$ is given in (1.6), one has

$$C_{12}(\lambda,t) = C_{12}(\lambda,0)\,,$$
$$C_{11}(\lambda,t) = C_{11}(\lambda,0)\exp(-4i\lambda\sqrt{\lambda^2-\rho}\,t)\,. \qquad (2.22)$$

## III. THE INVERSE SCATTERING PROBLEM

We now turn to the problem of reconstructing $q(x,t)$ from the scattering data $[C_{ij}(\lambda,t),\ -\sqrt{\rho}<\lambda<\sqrt{\rho}]$. As a matter of fact, it suffices to reconstruct $\Phi(x)$ (or $\psi$) from $C_{ij}(\lambda,0)$.

In the problem we are studying here the solution $q(x,t)$ can be expressed as

$$q(x,t) = 2iK_{12}^{(1)}(x,x,t) + \sqrt{\rho}\,, \qquad (3.1)$$

where the kernel $K_{ij}^1$ is defined in (2.11a).

In order to implement the inverse program, the basic tool is the Gel'fand–Levitan–Marchenko (GLM) integral equation.[4] This equation allows us to determine the kernel $K^{(i)}$ from our knowledge of the scattering data. With the kernel in hand, the solution $q(x,t)$ is determined from (3.1).

In order to write the GLM equation we consider the scattering function defined as

$$\frac{\Phi(x,\lambda)}{C_{12}(\lambda,k)} = \bar{\psi}(x,\lambda) + \frac{C_{11}(\lambda,k)\psi(x,\lambda)}{C_{12}(\lambda,k)}\,. \qquad (3.2)$$

One can rewrite (3.2) as

$$\left\{\frac{1}{C_{12}(\lambda)}\Phi(x,\lambda) - \bar{\psi}(x,\lambda)\right\}\frac{\exp[i\sqrt{\lambda^2-\rho}\,y]}{\sqrt{\lambda^2-\rho}}$$
$$= \left\{\frac{C_{11}(\lambda)}{C_{12}(\lambda)}\psi(x,\lambda) + \bar{\psi}(x,\lambda) - \bar{\psi}(x,\lambda)\right\}$$
$$\times\frac{\exp[i\sqrt{\lambda^2-\rho}\,y]}{\sqrt{\lambda^2-\rho}}\,. \qquad (3.3)$$

For convenience from now on we will take $y>x$.

The GLM equation is obtained by integrating (3.3) in the complex plane $\lambda$ along the contour shown in Fig. 1. The contour, which we call $C$ in the following, is the dotted one in Fig. 2. We restrict ourselves to $K=\sqrt{\lambda^2-\rho}$ such that $\mathrm{Im}\,K(\lambda)>0$.

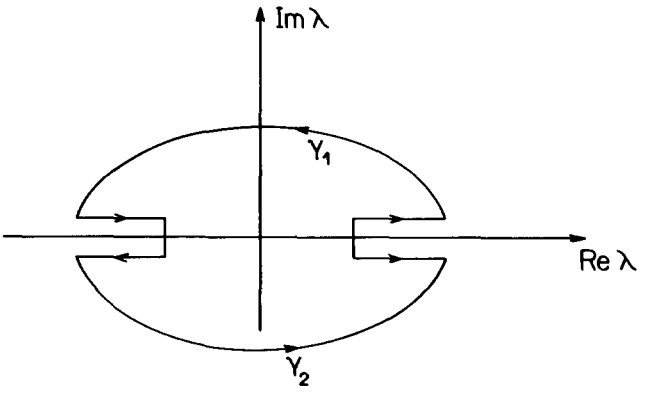



FIG. 1. Integration contour in the complex plane used in order to write the GLM equation.

First of all we will integrate the left-hand side of (3.3) along contour $C$. In order to do this one notes first that $C_{12}(\lambda,k)$, $\Phi(x,\lambda)\exp(i\sqrt{\lambda^2-\rho}\,y)$, and $\bar{\psi}(x,\lambda)\times\exp(i\sqrt{\lambda^2-\rho}\,y)$ have analytic continuation for $\mathrm{Im}(\sqrt{\lambda^2-\rho})>0$. Furthermore, if $C_{12}(\lambda)$ has poles located at $\lambda_j$ (where $\lambda_j\in\,]-\sqrt{\rho},\sqrt{\rho}\,[\,$), then $\sqrt{\lambda_j^2-\rho}\equiv\pm i\nu_j$.

Since, on the other hand, we are restricted to $\mathrm{Im}(\sqrt{\lambda^2-\rho})>0$ one can write

$$\sqrt{\lambda_j^2-\rho} = i\nu_j,\quad \nu_j>0\,. \qquad (3.4)$$

Since the left-hand side of (3.3) is analytic in the Riemann surface in which $\mathrm{Im}\sqrt{\lambda^2-\rho}>0$, one integrates the lhs over the closed contour of Fig. 1.

One can then write

$$\int_C \mathrm{lhs} + \int_{\gamma_1}\mathrm{lhs} + \int_{\gamma_2}\mathrm{lhs} = 2\pi i\sum_j \mathrm{residue}\,. \qquad (3.5)$$

Remembering that we have chosen $y>x$ and $\mathrm{Im}(\sqrt{\lambda^2-\rho})>0$, one obtains

$$\int_{\gamma_i}\left(\frac{\Phi}{C_{12}} - \bar{\psi}\right)\frac{\exp(i\sqrt{\lambda^2-\rho}\,y)}{\sqrt{\lambda^2-\rho}} \underset{\lambda\to\infty}{\sim} 0\,. \qquad (3.6)$$

Consequently one can write, from (3.5) and (3.6),

$$\int_C \mathrm{lhs} = 2\pi i\sum_j \mathrm{residue}$$
$$= 2\pi i\sum_j \frac{C_{11}(\lambda_j)e^{-\nu_j y}\psi(x,\lambda_j)}{V_j(d/d\lambda)C_{12}(\lambda,k)\big|_{\lambda=\lambda_j}}\,, \qquad (3.7)$$

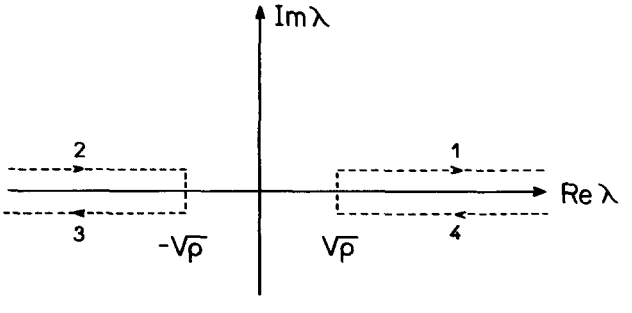



FIG. 2. Integration contour that we have named $C$ in the paper.

where $\nu_j$ in (3.7) is given, by using (3.4), as

$$\nu_j = \sqrt{\rho - \lambda_j^2}\,. \tag{3.8}$$

We integrate now the right-hand side of (3.3) over the contour $C$. We divide the integration contour into four pieces, as shown in Fig. 2. That is,

$$\int_C \text{rhs} = \int_1 + \int_2 + \int_3 + \int_4 .$$

The following observations are relevant in order to carry out the integration.

When $\lambda$ varies from $\sqrt{\rho} + i\epsilon$ until $+\infty + i\epsilon$ (path 1) ($\epsilon > 0$), then $\lambda$ can be written as $\lambda = \lambda_0 + i\epsilon$ with $\lambda_0^2 - \rho > 0$ and consequently one can write

$$K(\lambda) = \sqrt{\lambda^2 - \rho} \simeq \sqrt{\lambda_0^2 - \rho}\,(1 + i\epsilon\lambda_0/\sqrt{\lambda_0^2 - \rho})\,,$$

that is, for path 1, $\operatorname{Im} K(\lambda) > 0$. Analogously one has that $\operatorname{Im} k(\lambda) < 0$ for paths 2 and 4. Whereas, as in path 1, $\operatorname{Im} k(\lambda) > 0$ in path 3. This just means that, since we integrate the rhs for $\operatorname{Im} k(\lambda) > 0$, one has to make in paths 2 and 4 the replacement of $\sqrt{\lambda^2 - \rho}$ by $-\sqrt{\lambda^2 - \rho}$.

Finally we will make the change of variables $\lambda \to k$; using the $k$ defined in (2.15) it follows that $k\,dk = \lambda\,d\lambda$. The integration over $k$ then leads us to

$$\frac{1}{2}\int_C \text{rhs} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} 2\pi K^{(1)}(x,y)\theta(y-x) + \begin{pmatrix} \sqrt{\rho}\,B(x+y) \\ A(x+y) + i[\partial/\partial(x+y)]B(x+y) \end{pmatrix} - \int_x^\infty K^1(x,x') \begin{pmatrix} \sqrt{\rho}\,B(x'+y) \\ A(x'+y) + i[\partial/\partial(x'+y)]B(x'+y) \end{pmatrix} dx' , \tag{3.9}$$

where

$$A(z) = \int_{-\infty}^{\infty} \frac{C(\lambda,k) + C(-\lambda,k)}{2} e^{ikz}\,dk\,,$$
$$B(z) = \int_{-\infty}^{\infty} \frac{C(\lambda,k) - C(-\lambda,k)}{2\lambda} e^{ikz}\,dk\,, \tag{3.10}$$

and $C(\lambda,k)$ in (3.10) is defined as

$$C(\lambda,k) = C_{11}(\lambda,k)/C_{12}(\lambda,k)\,, \tag{3.11}$$

if one defines $C_j$ as

$$C_j \equiv \frac{C_{11}(\lambda,k)}{(d/d\lambda)C_{12}(\lambda,k)}\bigg|_{\substack{\lambda=\lambda_j \\ k=i\nu_j}} . \tag{3.12}$$

The Gel'fand–Levitan–Marchenko equation follows by equalizing equations (3.7) and (3.9). One obtains

$$\begin{aligned} &2\pi K^1(x,y)\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \sqrt{\rho}\,B(x,y) \\ A(x,y) + i[\partial/\partial(x+y)]B(x+y) \end{pmatrix} \\ &\quad - \int_x^{+\infty} dx'\,K^1(x,x') \times \begin{pmatrix} \sqrt{\rho}\,B(x'+y) \\ A(x',y) + i[\partial/\partial(x'+y)]B(x'+y) \end{pmatrix} \\ &\quad = \pi \sum_j C_j \left\{ e^{-\nu_j(x+y)} \begin{pmatrix} \sqrt{\rho}/\nu_j \\ \lambda_j/\nu_j - i \end{pmatrix} - \int_x^\infty dx'\,K^1(x,x') \begin{pmatrix} \sqrt{\rho}/\nu_j \\ \lambda_j/\nu_j - i \end{pmatrix} e^{-\nu_j(x'+y)} \right\} . \end{aligned} \tag{3.13}$$

## IV. MULTISOLITON AND SINGLE-SOLITON SOLUTIONS

Solutions describing $N$ solitons can be obtained by considering reflectionless potentials. In this case $A(x,y) = B(x,y) = 0$ and the GLM equation reduces to

$$\begin{aligned} &2K^1(x,y)\begin{pmatrix} -1 \\ 0 \end{pmatrix} \\ &\quad = \sum_j C_j e^{-\nu_j(x+y)} \begin{pmatrix} \sqrt{\rho}/\nu_j \\ \lambda_j/\nu_j - 1 \end{pmatrix} - \int_x^{+\infty} dx'\,K^1(x,x') \begin{pmatrix} \sqrt{\rho}/\nu_j \\ \lambda_j/\nu_j - 1 \end{pmatrix} e^{-\nu_j(x'+y)} . \end{aligned} \tag{4.1}$$

In order to get multisoliton solutions it follows from (3.1) that the relevant matrix elements are $K_{12}^{(1)}(x,y)$, which satisfies the GLM equation

$$\begin{aligned} 2K_{12}^{(1)}(x,y) &= -\sum_j C_j \frac{(\lambda_j + i\nu_j)}{\nu_j} e^{-\nu_j(x+y)} \\ &\quad + \sum_j C_j \int_x^{+\infty} dx'\,e^{-\nu_j(x'+y)} \\ &\quad \times \left\{ K_{11}^{(1)}(x,x) \frac{(\lambda_j + i\nu_j)}{\nu_j} + K_{12}^{(1)}(x,x') \frac{\sqrt{\rho}}{\nu_j} \right\} . \end{aligned} \tag{4.2}$$

Since, from (2.13), $K_{11}(x,x')$ can be obtained from $K_{22}(x,x')$ as a simple complex conjugation, one can just write that the equation satisfied by $K_{22}^{(1)}(x,x')$ is

$$\begin{aligned} 2K_{22}^{(1)}(x,y) &= -\sum C_j \frac{(\sqrt{\rho})}{\nu_j} e^{-\nu_j(x+y)} \\ &\quad + \sum_j \int_x^{+\infty} C_j\,dx'\,e^{-\nu_j(x'+y)} \\ &\quad \times \left\{ K_{21}^{(1)}(x,x') \frac{(\lambda_j + i\nu_j)}{\nu_j} + K_{22}^{(1)}(x,x') \frac{\sqrt{\rho}}{\nu_j} \right\} . \end{aligned} \tag{4.3}$$

The one-soliton solution is found by assuming that $C_{12}(\lambda,k)$ has a single pole, that is, $C_{12}(\lambda,k) = 0$ for $k = i\nu = i\sqrt{\rho^2 - \lambda^2}$ ($\lambda \in R$, $\nu \in R$). Under these conditions one can solve explicitly the preceding system of equations [(4.1)–(4.3)] and obtain

$$K_{12}^{(1)}(x,x) = -\frac{2(C/\nu)(\lambda + i\nu)e^{-2\nu x}}{4 - 2(C/\nu^2)\sqrt{\rho}\, e^{-2\nu x}}. \tag{4.4}$$

From relation (3.1) it follows that

$$2iK_{12}(x,x) = -\sqrt{\rho} + q(x), \tag{4.5}$$

thus the procured solution reads

$$q(x) = \sqrt{\rho} + \frac{-4i(C/\nu)(\lambda + i\nu)e^{-2\nu x}}{4 - 2(C/\nu^2)\sqrt{\rho}\, e^{-2\nu x}}. \tag{4.6}$$

One gets $q(x,t)$ from expression (3.1) by considering the explicit time evolution of the scattering data. This dependence appears through the parameter $C$. In this way the dependence of $C$ in $t$ can be inferred from (3.12). That is,

$$C(t) = \left.\frac{C_{11}(\lambda,k,t)}{(d/d\lambda)C_{12}(\lambda,k,t)}\right|_{K=i\nu}. \tag{4.7}$$

From (2.22) one gets

$$C(t) = C_0 e^{4\lambda\nu t}, \tag{4.8}$$

where $C_0$ is given as

$$C_0 = \left.\frac{C_{11}(\lambda,0)}{(d/d\lambda)C_{12}(\lambda,0)}\right|_{K=i\nu}. \tag{4.9}$$

Together all this leads to the explicit solution

$$q(x,t) = \frac{(\lambda + i\nu)^2 + \sqrt{\rho}(-2\nu^2/C_0)e^{-4\lambda\nu t}e^{2\nu x}}{\sqrt{\rho} - (2\nu^2/C_0)e^{-4\lambda\nu t}e^{2\nu x}}. \tag{4.10}$$

A more convenient form for the one-soliton solution is[5] [here we change notation $q(x,t) \equiv \varphi_c(x,t,v)$]

$$\varphi_c(x,t,v) = [v - i\gamma \tanh(\gamma m(x - vt))]\sqrt{\rho}\, e^{-i(\lambda\rho t - \vartheta)}, \tag{4.11}$$

where $\vartheta$ is an arbitrary phase, $\gamma = \sqrt{1 - v^2}$, and $v$ is the velocity of the soliton ($-1 < v < 1$). This parameter $v$ is related to the old one as $v = \lambda/\sqrt{\rho}$. We have also taken $C_0\sqrt{\rho} = -2\nu^2$.

Solution (4.11) describes charged solutions moving with velocity $v$. The conserved charge associated to one such solution is given as

$$Q_c = \int dx\, \varphi_c^*(x,t)\varphi_c(x,t),$$

whereas the momentum and energy are

$$P = -i\int dx[\varphi_c^*\, \partial_x \varphi_c],$$

$$E = \int dx\left[\varphi_c^*\left(-\frac{\partial_x^2}{2m}\right)\varphi_c + \frac{K}{2}(\varphi_c\varphi_c^*)^2 - K\rho(\varphi_c\varphi_c^*)\right].$$

The charge and energy of the soliton will be given by

$$Q_s = Q_c - Q_0, \quad E_s = E_c - E_0,$$

where $Q_0$ and $E_0$ are the charge and energy of the condensate (or the vacuum energy) given by

$$Q_0 = L\rho, \quad E_0 = L\lambda\rho^2/2,$$

where $L$ is the size of the system that, at this point, we take finite. For the solitons (4.11) one gets

$$Q_s = -2\rho\gamma/\sqrt{m\lambda\rho}, \quad P = -2\rho v\gamma, \quad E_s = \tfrac{4}{3}\rho\gamma^3.$$

This solution $\varphi_v(x,t)$ describes a pulse that moves in the $x$ direction with velocity $v$ in a medium of density $\rho$ (the condensate density) and is interpreted as a coherent state of the bosonic system.

The large $x$ behavior (for fixed $t$) of $\varphi(x,t)$ gives

$$\lim_{x\to\infty} \varphi(x,t) = \sqrt{\rho}\exp i(\vartheta + \delta - \lambda\rho t),$$

whereas

$$\lim_{x\to+\infty} \varphi(x,t) = \sqrt{\rho}\exp i(\varphi - \delta - \lambda\rho t),$$

where $\delta = \arccos v$.

Therefore these solutions cut the condensate (the manifold of degenerate vacua) in two sectors of phases where the difference in phase is given by $2\delta$. Since this difference is time independent (even when there exist quasiparticles in the system) it constitutes itself in a charge.[5]

Figure 3 exhibits in the complex $\varphi$ plane the "trajectories" of different $\varphi_c$ solutions when $x$ varies from $-\infty$ to $+\infty$. This figure corresponds to a choice of $\vartheta = \delta$ and we have fixed the time as $t = 0$.

## V. CONCLUSIONS

We applied the inverse scattering technique in order to get classical solutions to a nonlinear nonrelativistic model exhibiting a spontaneous breakdown of symmetry.

The model we have studied is just an extension of the usual nonlinear Schrödinger theory (NLS model). Physically the model considered here describes a nonrelativistic Bose gas whose particles interact via a $\delta$-like potential but, and this is the difference from the NLS model, it takes into account the possibility of a net background charge in the system.

As a result of the above-mentioned distinction between the two theories, one expects that the differences show up at the classical and at the quantum level.

At the classical level this is manifest by the fact that the $x$-independent solution (condensate or vacuum configuration) is characterized by an infinite set of solutions. That is, the vacuum is infinitely degenerate. This has an interesting reflection on the set of solitonlike solutions. For instance, the one-soliton solution describes a pulse moving with velocity $v$ in such a way that for each velocity there corresponds a certain phase difference in the manifold of degenerate vacua.

In the process of quantizing model (1.1) the approach developed here is also relevant. This is a result of the fact that the inverse scattering technique is a way to cast the Hamiltonian written in terms of action-angle variables.[6] The quantization of model (1.1) is in progress and will be dealt with in a future publication.

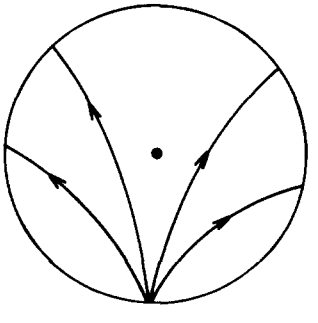


FIG. 3. Solitons with different velocities interpolate different points in the manifold of degenerate vacua. To each velocity there corresponds a phase difference.

We believe that the classical solutions described here can also be obtained from the NLS equation by making the substitution $q \to qe^{-ipkt}$ and at the same time considering nontrivial boundary conditions. This approach has been carried out in Ref. 7. For the quantized version, we believe that work with the spontaneously broken symmetry version is more advantageous since the introduction of a chemical potential (as we do) is a very systematic way of taking into account the presence of a background charge.

## ACKNOWLEDGMENTS

We are particularly indebted to Professor I. Ventura for his interest and many discussions at several stages of this work as well as for calling our attention to this subject. We thank him also for a critical reading of the manuscript.

One of us (DS) would like to thank the Departamento de Física Matemática, Universidade de São Paulo for their kind hospitality.

We would like to thank Fundaçîo de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional do Desenvolvimento Cientifico e Tecnologico for their partial financial support.

[1] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967). For a review article, see A. C. Scott, F. Y. F. Chu Seym, and D. W. McLaughlin, Proc. IEEE **61**, 1443 (1973).

[2] N. N. Bogoliubov, J. Phys. **11**, 23 (1947); F. W. Wiegel, Phys. Rep. C **16**, 57 (1975).

[3] V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972); T. Tsuzuki, J. Low Temp. Phys. **4**, 441 (1971).

[4] I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl. **1**, 253 (1955).

[5] This is the form suggested by I. Ventura, Rev. Bras. Fis. **9**, 375 (1979).

[6] V. E. Zakharov and L. D. Faddeev, Funct. Anal. Appl. **5**, 280 (1971); P. P. Kulish, S. V. Manakov, and L. D. Faddeev, Teor. Mat. Fiz. **28**, 615 (1976).

[7] M. Boiti and F. Pempinelli, "The spectral transforms for the NLS equation with left–right asymmetric boundary conditions," Preprint Lab. de Physique Mathematique, 1981.

# BRST quantization of the spinning particle

P. Thomi
*Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland*

(Received 16 August 1988; accepted for publication 12 October 1988)

Using rigged Hilbert space techniques, the spinning particle is BRST quantized. It is shown that there is a unique admissible Hermitian form on the extended phase space. With respect to it, the BRST operator is anti-Hermitian. The doubling of the BRST cohomology is removed by choosing a subspace on which the Hermitian form is positive definite.

## I. INTRODUCTION

In recent times BRST methods have become increasingly important in the attempts to quantize constrained systems. Despite all these efforts there still exist problems even at a basic level. For instance, the inner product structure of the BRST cohomology has not yet been satisfactorily defined.

We illustrate and solve this problem in the case of the spinning relativistic particle. We use (a slightly modified version of) the model given in Ref. 1. We find (in a sense) all irreducible representations of the operator algebra of this model. What one might suspect, to represent the odd canonical variables by $\gamma$ matrices and the odd ghosts by multiplication with a Grassmann variable, does not work because then the odd canonical variables and the odd ghosts would commute with each other, whereas they must anticommute. So this is not a representation. As we will see, insisting on the usual Grassmann form leads to a reducible representation.

On the space of wave functions one then has to give a Hermitian form that satisfies, among other requirements, that the BRST operator $\hat{\Omega}$ is Hermitian or anti-Hermitian with respect to it. It turns out that there is no (nonzero) Hermitian form on the extended phase space such that $\hat{\Omega}$ is Hermitian w.r.t. it. There do exist, however, two linearly independent Hermitian forms such that $\hat{\Omega}$ is anti-Hermitian w.r.t. them (Theorem 1). One of them leads to a pseudoscalar product on the physical space, however, and thus has to be rejected. Hence there remains a unique (up to normalization) Hermitian form on the extended phase space that fulfills all the requirements. With respect to this form, real even and imaginary odd classical variables are represented by Hermitian operators, and real odd and imaginary even variables by anti-Hermitian ones. This is in conflict with the usual opinion that real variables always have to be represented by Hermitian and imaginary variables by anti-Hermitian operators. (For the above mentioned reducible representation there exists a Hermitian form such that $\hat{\Omega}$ is Hermitian; however, when restricted to the invariant subspaces of the operator algebra this form vanishes.)

Our Hermitian form would be well defined if the wave functions were square integrable. However, as in the bosonic case[2] this is not sufficiently general. We can give the Hermitian form a well-defined meaning even when generalized functions occur using the same rigged Hilbert space techniques that were introduced in Ref. 2 to handle the bosonic case.

The wave functions are eight component spinors. The Hermitian form is such that the first four components are paired with the last four. If we choose the components to lie in the parts of a Gel'fand triple in such a way that in the Hermitian form distributions are always paired with test functions, then the Hermitian form is well defined. To describe in which part of the Gel'fand triple the components of the wave functions lie, we introduce the "characteristic function." There are several characteristic functions such that the respective Hermitian form on the extended phase space is well defined, but only one that leads to a nondegenerate induced Hermitian form on the BRST cohomology (Theorem 2). That the choice of the function spaces in which the components of the wave functions lie is unique here (in contrast to the bosonic systems treated in Ref. 2) stems from the simplicity of the considered system. (Indeed for the minimal sector of the scalar particle the characteristic function is also unique.) It will not hold for general systems. We conjecture that the characteristic function is monotonically increasing in the component corresponding to the commuting ghosts and monotonically decreasing in the component corresponding to the canonical momenta.

For the above unique choice of function spaces we then calculate the BRST cohomology. This is more difficult than in the bosonic case because, e.g., such subtleties as the domain of definition of $\hat{\Omega}$ have to be properly taken into account. The result is given in Theorem 3. We see that there occurs a doubling of the same nature as in the bosonic case. The induced Hermitian form on the BRST cohomology is not positive definite. In order to have a positive definite inner product one has to choose a linear subspace of the BRST cohomology that is left invariant by the BRST observables. Using the special form of the BRST observables (Theorem 4) we prove that such subspaces exist and that they all lead to unitarily equivalent theories (Theorem 5). By choosing the subspace we arrive at a positive definite inner product and, at the same time, remove the (unwanted) doubling.

Our notation has been taken mostly from Refs. 2 and 3. We refer to Ref. 3 for an excellent review on the Hamiltonian BRST formalism.

## II. DEFINITION OF THE MODEL

We start from a slightly modified version of the classical model of the relativistic particle of Galvão and Teitelboim.[1] The classical variables fulfill the Dirac brackets,

$$\begin{aligned} &\{x^\mu, p_\nu\} = \delta^\mu_\nu, \\ &\{\theta^\mu, \theta^\nu\} = i\eta^{\mu\nu}, \\ &\{\theta^5, \theta^5\} = i, \end{aligned} \tag{1}$$

 0022-2488/89/020470-07$02.50 © 1989 American Institute of Physics 

where $\theta^\mu, \theta^5$ are real Grassmann variables and $\eta^{\mu\nu} = \text{diag}(-+++)$. The constraints are

$$\begin{aligned} &\mathscr{H} = \theta(p^0)p^2 + m^2, \\ &\mathscr{S} = \theta(p^0)\theta^\mu p_\mu + m\theta^5. \end{aligned} \tag{2}$$

We have introduced the $\theta(p^0)$ in the constraints to get rid of the negative frequencies from the start, as in Ref. 4. In Ref. 1 this $\theta$ function is not included. Since $\theta^2(p^0) = \theta(p^0)$ the constraint algebra does not change. As in Ref. 1 we have

$$\begin{aligned} &\{\mathscr{S},\mathscr{S}\} = i\mathscr{H}, \\ &\{\mathscr{S},\mathscr{H}\} = 0 = \{\mathscr{H},\mathscr{H}\}. \end{aligned} \tag{3}$$

The action is

$$S = \int_{\tau_1}^{\tau_2} d\tau \left\{ \dot{x}^\mu p_\mu + \frac{i}{2}(\dot{\theta}^\mu\theta_\mu + \dot{\theta}^5\theta^5) - \lambda^a(\tau)G_a \right\} + \frac{i}{2}(\theta^\mu(\tau_1)\theta_\mu(\tau_2) + \theta^5(\tau_1)\theta^5(\tau_2)), \tag{4}$$

where $\lambda^a(\tau)$ are Lagrange multipliers, $a = 1,2$, and $G_a$ are the constraints $G_1 = \mathscr{S}$, $G_2 = \mathscr{H}$. The surface terms are needed in order for the variational problem to be well defined (see Ref. 1). For the sake of simplicity we do not introduce the Lagrange multipliers and their associated momenta as additional canonical variables, i.e., we work in the so-called minimal sector. The ghosts and the ghost momenta satisfy

$$\{\mathscr{P}_a,\eta^b\} = -\delta_a^b. \tag{5}$$

The ghosts and ghost momenta have the opposite Grassmann parity as the constraint to which they belong. The general formula for the BRST charge[3] gives, using (3),

$$\Omega = \eta^1\mathscr{S} + \eta^2\mathscr{H} + (\eta^1)^2(i/2)\mathscr{P}_2. \tag{6}$$

To pass to quantum theory we substitute the Dirac brackets by $(-i)$ times (anti-) commutators. So

$$\begin{aligned} &[\hat{x}^\mu,\hat{p}^\nu] = i\eta^{\mu\nu}, \\ &[\hat{\theta}^\mu,\hat{\theta}^\nu] = -\eta^{\mu\nu}, \\ &[\hat{\theta}^5,\hat{\theta}^5] = -1, \\ &[\hat{\mathscr{P}}_a,\hat{\eta}^b] = -i\delta_a^b. \end{aligned} \tag{7}$$

All the other (anti-) commutators vanish. We have to represent this algebra on some space of wave functions. To this end we note that when we combine the anticommuting ghost $\hat{\eta}^2$ and its momentum $\hat{\mathscr{P}}_2$ as $\hat{g}^+ = \hat{\eta}^2 + i\hat{\mathscr{P}}^2$ and $\hat{g}^- = \hat{\eta}^2 - i\hat{\mathscr{P}}^2$, then $\hat{\theta}^\mu$, $\hat{\theta}^5$, $\hat{g}^+$, and $\hat{g}^-$ are the generators of a Clifford algebra with signature $+---\,-+-$. This algebra has exactly two irreducible ($2^{[7/2]} = 8$ dimensional) representations.[5] Thus we represent the operator algebra (7) on eight-component Dirac spinors $\psi(p,\eta^1)$. The operators are given as

$$\begin{aligned} &\hat{x}^\mu = i\frac{\partial}{\partial p_\mu}, \quad \hat{p}_\mu = p_\mu, \quad \hat{\eta}^1 = \eta^1, \quad \mathscr{P}_1 = -i\frac{\partial}{\partial\eta^1}, \\ &\hat{\theta}^\mu = \frac{\lambda}{\sqrt{2}}\begin{pmatrix}\gamma^5\gamma^\mu & 0\\ 0 & -\gamma^5\gamma^\mu\end{pmatrix}, \quad \hat{\theta}^5 = \frac{i}{\sqrt{2}}\begin{pmatrix}\gamma^5 & 0\\ 0 & -\gamma^5\end{pmatrix}, \\ &\hat{\eta}^2 = \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \quad \mathscr{P}_2 = \frac{1}{i}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \end{aligned} \tag{8}$$

where $\{\gamma^\mu,\gamma^\nu\} = 2\eta^{\mu\nu}$ and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$; $(\gamma^5)^2 = \mathbf{1}$ are the usual $\gamma$ matrices. [The other representation would be obtained by changing the sign of $g^-$. Using this representation would lead to the same final result.] The BRST operator is [since no factor ordering ambiguities arise we may just insert carets in (6)]

$$\hat{\Omega} = \hat{\eta}^1(\theta(\hat{p}^0)\hat{p}_\mu\hat{\theta}^\mu + m\hat{\theta}^5) + \hat{\eta}^2(\theta(\hat{p}^0)\hat{p}^2 + m^2) + (i/2)(\hat{\eta}^1)^2\hat{\mathscr{P}}_2. \tag{9}$$

Introducing the abbreviations

$$D = \theta(p^0)p_\mu\gamma^5\gamma^\mu + mi\gamma^5, \quad G = (1/\sqrt{2})\eta^1\cdot\mathbf{1}, \tag{10}$$

we may write it as

$$\hat{\Omega} = \begin{pmatrix} GD & G^2 \\ -D^2 & -DG\end{pmatrix}. \tag{11}$$

The nilpotence of $\hat{\Omega}$ is easily established.

*Remark:* The "representation"

$$\begin{aligned} &\hat{\theta}^\mu = \frac{1}{\sqrt{2}}\gamma^5\gamma^\mu, \quad \hat{\theta}^5 = \frac{i}{\sqrt{2}}\gamma^5, \\ &\hat{\eta}^2 = \eta^2, \quad \mathscr{P}_2 = i\frac{\partial^l}{\partial\eta^2}, \quad \text{etc.}, \end{aligned} \tag{12}$$

on Dirac spinors $\psi(p,\eta^1,\eta^2) = \psi_0(p,\eta^1) + \psi_1(p,\eta^1)\eta^2$, for which $\hat{\eta}^2$ acts by multiplication with a Grassmann number $\eta^2$, does not work. It is true that all the relations written down in (7) are satisfied; however, for (12) $\hat{\eta}^2$, $\hat{\mathscr{P}}_2$ commute with $\hat{\theta}^\mu$, $\hat{\theta}^5$, whereas they should anticommute.

One can remedy this by introducing a Clifford algebra with six generators,

$$\begin{aligned} &\{\tilde{\gamma}^r,\tilde{\gamma}^s\} = 2\tilde{g}^{rs}, \\ &r,s = 0,1,\ldots,5, \quad \tilde{g}^{rs} = \text{diag}(+-----). \end{aligned}$$

As the space of wave functions one then takes the Dirac spinors $\psi(p,\eta^1,\eta^2)$ belonging to this Clifford algebra. The operators are then represented by

$$\begin{aligned} &\hat{\theta}^\mu = \frac{1}{\sqrt{2}}\tilde{\gamma}^\mu, \quad \hat{\theta}^5 = \frac{1}{\sqrt{2}}\tilde{\gamma}^5, \\ &\hat{\eta}^2 = \tilde{\gamma}^A\eta^2, \quad \hat{\mathscr{P}}_2 = i\tilde{\gamma}^A\frac{\partial^l}{\partial\eta^2}, \quad \text{etc.} \end{aligned} \tag{13}$$

They are $8\times 8$ matrices. The representation is effectively 16-dimensional and it is reducible. To see this choose

$$\begin{aligned} &\tilde{\gamma}^\mu = \begin{pmatrix}\gamma^5\gamma^\mu & 0\\ 0 & -\gamma^5\gamma^\mu\end{pmatrix}, \quad \tilde{\gamma}^5 = \begin{pmatrix}i\gamma^5 & 0\\ 0 & -i\gamma^5\end{pmatrix}, \\ &\tilde{\gamma}^A = i\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \end{aligned}$$

then writing $\psi = \psi_0 + \eta^2\psi_1$ as vector $\begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix}$ we get, for $\hat{\Omega}$,

$$\hat{\Omega}\psi = \begin{pmatrix} GD & 0 & 0 & -iG^2 \\ 0 & -GD & -iG^2 & 0 \\ 0 & -iD^2 & GD & 0 \\ -iD^2 & 0 & 0 & -GD \end{pmatrix}\begin{pmatrix}\bar{\psi}_0\\ \psi_0\\ \bar{\psi}_1\\ \psi_1\end{pmatrix}$$

with $G = (1/\sqrt{2})\eta^1$, $D = \theta(p^0)p_\mu\tilde{\gamma}^\mu + m\tilde{\gamma}^5$. Making the change of basis

$$\begin{pmatrix}\bar{\psi}_0\\ \underline{\psi}_0\\ \bar{\psi}_1\\ \underline{\psi}_1\end{pmatrix}\to\begin{pmatrix}\underline{\bar{\psi}}_1\\ \bar{\psi}_0\\ \underline{\psi}_0\\ \bar{\psi}_1\end{pmatrix}=\begin{pmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{pmatrix}\begin{pmatrix}\bar{\psi}_0\\ \underline{\psi}_0\\ \bar{\psi}_1\\ \underline{\psi}_1\end{pmatrix}=B\psi,$$

one gets, for $\hat{\Omega}'=B\hat{\Omega}B^{-1}$,

$$\hat{\Omega}'=\begin{pmatrix}-GD&-iD^2&0&0\\-iG^2&GD&0&0\\0&0&-GD&-iG^2\\0&0&-iD^2&GD\end{pmatrix},$$

all other operators are of a similar form.

If we used this representation, we would get an additional doubling as we shall see later.

## III. THE HERMITIAN FORM

On the space of wave functions we want to define a Hermitian form such that (a) the BRST observables are Hermitian w.r.t. it, (b) the BRST operator is Hermitian or anti-Hermitian w.r.t. it, (c) it induces a scalar product on the physical space.

Condition (b) deserves some comment. In the literature it is usually stated that the BRST operator has to be Hermitian in order for the formalism to work properly. However, what one really needs is that a state of the form $\hat{\Omega}\psi$ satisfies $(\phi,\hat{\Omega},\psi)=0$ for any physical state $\phi$ (i.e., any $\phi\in\mathrm{Ker}\,\hat{\Omega}$). This is also satisfied when $\hat{\Omega}$ is anti-Hermitian. In fact, there is no (nonzero) Hermitian form such that $\hat{\Omega}$ is Hermitian w.r.t. it.

**Theorem 1:** For the representation (8) there is no nonzero Hermitian form such that $\hat{\Omega}$ is Hermitian w.r.t. it. There are only two linearly independent Hermitian forms such that $\hat{\Omega}$ is anti-Hermitian w.r.t. them, namely,

$$(\psi,\phi)=\int d^4p\,d\eta^1(\psi_0^*\gamma^0\phi_1-\psi_1^*\gamma^0\phi_0) \tag{14}$$

and

$$(\psi,\phi)=-i\int d^4p\,d\eta^1(\psi_0^*\gamma^5\gamma^0\phi_1-\psi_1^*\gamma^5\gamma^0\phi_0), \tag{15}$$

where $\psi_0$, $\psi_1$ denote the first and last four components of $\psi$, respectively.

*Proof:* The most general (formal) Hermitian form on the space of spinors $\psi(p,\eta^1)$ is

$$(\psi,\phi)=\int d^4p\,d\eta^1(\psi_0^*,\psi_1^*)\begin{pmatrix}A&B\\B^*&C\end{pmatrix}\begin{pmatrix}\phi_0\\ \phi_1\end{pmatrix},$$

$$A=A^*,\quad C=C^*, \tag{16}$$

where $\psi_0$, etc., are four component spinors and $A$, $B$, $C$ are $4\times4$ matrices.

The condition for $\hat{\Omega}$ to be Hermitian w.r.t. (16) is

$$\begin{pmatrix}GD&G^2\\-D^2&-GD\end{pmatrix}^*\begin{pmatrix}A&B\\B^*&C\end{pmatrix}=\begin{pmatrix}A&B\\B^*&C\end{pmatrix}\begin{pmatrix}GD&G^2\\-D^2&-GD\end{pmatrix},$$

which leads to the equations

$$\begin{aligned}&GD^*A-D^2B^*=AGD-BD^2,\\&GD^*B-D^2C=AG^2-BGD,\\&G^2A-GD^*B^*=B^*GD-CD^2,\\&G^2B-GD^*C=B^*G^2-CGD.\end{aligned} \tag{17}$$

They have to hold for any $\eta^1$, i.e., for any $G$. At $G=0$ they reduce to ($D^2=D^2\cdot1\neq0$ in general)

$$B^*=B,\quad C=0.$$

Setting this back into (17) leads to ($G=G\cdot1$)

$$\begin{aligned}&AD=D^*A,\\&BD+D^*B=AG,\\&BD-D^*B=AG.\end{aligned}$$

Adding and subtracting the last two equations gives, respectively,

$$\begin{aligned}&2BD=2AG,\\&2D^*B=0,\end{aligned}$$

and this leads to $B=0$ and $A=0$. So together $A=B=C=0$. This proves the first claim. The condition for $\hat{\Omega}$ to be anti-Hermitian is

$$\begin{aligned}&D^2B^*-GD^*A=AGD-BD^2,\\&D^2C-GD^*B=AG^2-BGD,\\&GD^*B^*-G^2A=B^*GD-CD^2,\\&GD^*C-G^2B=B^*G^2-CGD,\end{aligned} \tag{17'}$$

at $G=0$ we get this time $B^*=-B$ and $C=0$. Setting this in (17') gives

$$\begin{aligned}&AD+D^*A=0,\\&-GD^*B=G^2A-GBD.\end{aligned}$$

Comparison of the coefficients of $G$ and $G^2$ in the second equation leads to

$$A=0\quad\text{and}\quad D^*B=BD,$$

so together we have $A=C=0$, $B=-B^*$, $D^*B=BD$. Now because $D=\theta(p^0)p_\mu\gamma^5\gamma^\mu+mi\gamma^5$ we must have

$$B=-B^*, \tag{18}$$

$$-i\gamma^5B=Bi\gamma^5, \tag{19}$$

$$(\gamma^5\gamma^\mu)^*B=B(\gamma^5\gamma^\mu). \tag{20}$$

A basis of the space of $4\times4$ matrices is $1$, $\gamma^\mu$, $\gamma^\mu\gamma^\nu,\gamma^5\gamma^\mu$, $\gamma^5$. In order that (19) is satisfied we must have $B=\alpha\gamma^\mu+\beta\gamma^5\gamma^\mu$. Then (20) requires $B=\alpha\gamma^0+\beta\gamma^5\gamma^0$ and (18) implies $\alpha\in\mathbb{R}$, $i\beta\in\mathbb{R}$. ■

The decision between (14) and (15) is made by the above requirement (c). Equation (15) leads to the Dirac product $\int\bar{\psi}\phi$ on the physical space, as we will see below, whereas (14) would lead to $\int\bar{\psi}\gamma^5\phi$. The latter is not a scalar w.r.t. the Lorentz transformations but a pseudoscalar; we thus have to use (15).

With respect to this form $\hat{x},\hat{p},\hat{\eta}^1$, $\hat{\mathscr{P}}_2$ are Hermitian and $\hat{\eta}^2$, $\hat{\mathscr{P}}_1$, $\hat{\theta}^\mu$, $\hat{\theta}^5$, and $\hat{\Omega}$ are anti-Hermitian. In general, real even and imaginary odd variables are represented by Hermitian operators and real odd and imaginary even ones by anti-Hermitian operators. Since BRST observables are real and even,[3] condition (a) above is thus satisfied.

*Remark:* In the representation (13) we may define a Hermitian form by

$$(\psi,\phi) = \int d^4p\, d\eta^1\, d\eta^2\, \psi^* \tilde{\gamma}^0 \phi$$

$$= \int d^4p\, d\eta^1 (\psi_1^* \tilde{\gamma}^0 \phi_0 - \psi_0^* \tilde{\gamma}^0 \phi_1). \tag{21}$$

With respect to this form, $\hat{\Omega}$ in the representation (13) is Hermitian. However, when restricted to one of the invariant subspaces (21) is zero. Nevertheless one can calculate the BRST cohomology in the same way as in the main text. One finds that it is four times as large as the physical space of the Dirac theory. One doubling is the usual one, the other has its root in the reducibility of the chosen representation. Since, restricted to the BRST observables, this representation is a direct sum of four times the same irreducible representation, there are linear subspaces such that they are left invariant by the BRST observables, the observables act irreducibly, and the Hermitian form restricted to them is definite. These subspaces are all isomorphic to the physical space of the Dirac theory.

The form (15) would be well defined if both components ($\psi_0, \psi_1$, etc.) of the wave function were square integrable. This is not general enough, however. Since we want $\psi$ to have something to do with the solutions of the Dirac equation, which are $D\delta(p^2 + m^2)$, we must allow for distributions. On the other hand we have to require that (15) is well defined.

To achieve this we use a similar method as in Ref. 2, where the bosonic case was treated. In contrast to the bosonic case where we had only one group of variables (the canonical momenta) to integrate over, here we have two, namely the canonical momenta and the commuting ghost. We therefore have to generalize the "characteristic function" introduced in Ref. 2.

*Definition:* The characteristic function $\chi$: $\{0,1\}^2 \to \{0,1\}^2$ is defined to be

$$\chi(\psi_j) \doteq \chi(j) = \begin{matrix}(0,0)\\(1,0)\\(0,1)\\(1,1)\end{matrix} \text{ if } \psi_j \text{ is a } \begin{matrix}\text{test function}\\\text{distribution}\\\text{test function}\\\text{distribution}\end{matrix} \text{ in } p \text{ and } a \begin{matrix}\text{test function}\\\text{test function}\\\text{distribution}\\\text{distribution}\end{matrix} \text{ in } \eta^1.$$

This definition implies the following.

*Lemma:* The Hermitian form (15) is well-defined iff

$$\chi_i(0) + \chi_i(1) \leqslant 1 \quad (i = 1,2),$$

i.e., whenever every distribution is paired with a test function.

As discussed in Ref. 2 one should choose $\chi$ such that in the above inequality the equality sign holds. Then there are four possibilities:

$$\begin{aligned}&\text{(i)} && \chi(0) = (1,0), && \chi(1) = (0,1),\\ &\text{(ii)} && \chi(0) = (0,1), && \chi(1) = (1,0),\\ &\text{(iii)} && \chi(0) = (0,0), && \chi(1) = (1,1),\\ &\text{(iv)} && \chi(0) = (1,1), && \chi(1) = (0,0).\end{aligned} \tag{22}$$

Only (i) is sensible. Possibilities (ii)–(iv) lead to induced [from (15)] forms on Ker $\hat{\Omega}$/Im $\hat{\Omega}$ which are pathological. To prove this claim we first have to determine Ker $\hat{\Omega}$. To this end note that

$$\hat{\Omega}\psi = \begin{pmatrix} GD\psi_0 + G^2\psi_1 \\ -D^2\psi_0 - DG\psi_1 \end{pmatrix} = \begin{pmatrix} G(D\psi_0 + G\psi_1) \\ -D(D\psi_0 + G\psi_1) \end{pmatrix}. \tag{23}$$

So for $\psi \in \text{Ker } \hat{\Omega}$ we must have

$$D\psi_0 + G\psi_1 = \tilde{C}(p,\eta^1), \tag{24}$$

where $\tilde{C}$ fulfills the equations

$$G\tilde{C} = 0, \quad D\tilde{C} = 0. \tag{25}$$

The solution of (25) is

$$\tilde{C}(p,\eta^1) = \delta(\eta^1)C(p) \quad \text{with} \quad DC(p) = 0. \tag{26}$$

The solution of (24) is thus

$$\begin{aligned}\psi_0 &= GA + \delta(\eta^1)B + E,\\ \psi_1 &= -DA + \delta(\eta^1)H - \sqrt{2}\delta'(\eta^1)J,\end{aligned} \tag{27}$$

with $A(p,\eta^1)$ and $H(p)$ arbitrary; $B(p)$ and $J(p)$ such that $DB + J = C$; and $E$ such that $DE = 0$.

Next we calculate $(\psi,\phi)$ for $\psi,\phi \in \text{Ker } \hat{\Omega}$. We obtain, setting (27) into (15)

$$\begin{aligned}(\psi,\phi) = -i \int d^4p\, d\eta^1 (&GA_\psi^* + \delta(\eta^1)B_\psi^*\\ &+ E_\psi^*)\gamma^5\gamma^0(-DA_\phi + \delta(\eta^1)H_\phi - \sqrt{2}\delta'(\eta^1)J_\phi)\\ &-(-(DA_\psi)^* + \delta(\eta^1)H_\psi^*\\ &-\sqrt{2}\delta'(\eta^1)J_\psi^*)\gamma^5\gamma^0(GA_\phi + \delta(\eta^1)B_\phi + E_\phi).\end{aligned} \tag{28}$$

This is only a formal expression, of course; terms containing $\delta^2$, etc., are not well defined. We have to go through the four possibilities listed in (22).

(i) In order to have $\chi(0) = (1,0)$ and $\chi(1) = (0,1)$, we must have $A$ of the form $A = A_0 + \delta(\eta^1)A_\eta + D\delta(\theta(p^0)p^2 + m^2)A_p$ with $A_0$, $A_\eta$, and $A_p$ test functions in both variables. Further we must have $B = J = 0$, $E$ a test function in $\eta^1$, and $H$ a test function $p$. When we set this in (28) what remains is

$$\begin{aligned}(\psi,\phi) = -i \int d^4p\, d\eta^1 (&E_\psi^*(p,\eta^1)\gamma^5\gamma^0\delta(\eta^1)H_\phi(p)\\ &-\delta(\eta^1)H_\psi^*\gamma^5\gamma^0E_\phi).\end{aligned} \tag{29}$$

(ii) The term $A$ must be a test function both in $\eta^1$ and $p$. The term $B$ is a test function in $p$ with $D^2B = 0$; this implies $B = 0$. Further, we must have $E = H = J = 0$ and thus (use $D^*\gamma^5\gamma^0 = \gamma^5\gamma^0 D$)

$$\|\psi\|^2 = 0. \tag{30}$$

(iii) The term $A$ is of the form $A = A_0 + \delta(\eta^1)A_\eta$ with $A_0, A_\eta$ test functions $B = E = 0$ and $J(p)$ such that $DJ = 0$.

This gives

$$(\psi,\phi) = -i\int d^4p\, d\eta^1 \left[\sqrt{2}\delta'(\eta^1) J^*_\psi \gamma^5\gamma^0 G A_{0\phi} - G A^*_{0\psi}\gamma^5\gamma^0\sqrt{2}\delta'(\eta^1) J_\phi\right]$$

$$= -i\int d^4p \left[A^*_{0\psi}|_{\eta^1=0}\gamma^5\gamma^0 J_\phi - J^*_\psi\gamma^5\gamma^0 A_{0\phi}|_{\eta^1=0}\right]. \tag{31}$$

(iv) $A = A_0 + D\delta(\theta(p^0)p^2 + m^2)A_p$ with $A_0$, $A_p$ test functions and $D^2B = 0$, $DE = 0$, $H = J = 0$. This leads to

$$(\psi,\phi) = -i\int d^4p\, d\eta^1 \left[\delta(\eta^1) B^*_\psi\gamma^5\gamma^0(-DA_\phi) + (DA_\psi)^*\gamma^5\gamma^0\delta(\eta^1)B_\phi\right]$$

$$= -i\int d^4p \left[A_{0\psi}|_{\eta^1=0}\gamma^5\gamma^0 DB_\phi - (DB_\psi)^*\gamma^5\gamma^0 A_{0\phi}|_{\eta^1=0}\right]. \tag{32}$$

We are through for case (ii). For cases (iii) and (iv) we have to show that the induced Hermitian form on the BRST cohomology has degenerate directions. Since $\hat\Omega$ is nilpotent and anti-Hermitian we always have $(\psi,\phi) = 0$ for $\psi\in$Ker $\hat\Omega$ and $\phi\in$Im $\hat\Omega$ and thus by adding states of Im $\hat\Omega$ to $\psi$ we do not change the norm. The image of $\hat\Omega$ is given by

$$\hat\Omega\chi = \begin{pmatrix} G(D\chi_0 + G\chi_1) \\ -D(D\chi_0 + G\chi_1)\end{pmatrix}. \tag{33}$$

Case (iii): $\psi + \hat\Omega\chi$ is of the form

$$\psi + \hat\Omega\chi = \begin{pmatrix} G(A_0 + D\chi_0 + G\overset{(0)}{\chi}_1) \\ -D(A_0 - \delta(\eta^1)A_\eta + D\chi_0 + G\overset{(0)}{\chi}_1 - \overset{(2)}{\chi}_1\delta(\eta^1)) + \delta(\eta^1)H - \sqrt{2}\delta'(\eta^1)J\end{pmatrix},$$

where $\chi_1 = \overset{(0)}{\chi}_1 + \overset{(1)}{\chi}_1\delta(\eta^1) + \overset{(2)}{\chi}\delta'(\eta^1)$. We see that it is not possible to completely remove the $H$ term by adding states in Im $\hat\Omega$. On the other hand, $H$ does not appear in (31). So a state of the form $\psi_0 = 0$, $\psi_1 = \delta(\eta^1)H$ represents a degenerate direction in the BRST cohomology.

Case (iv): We have

$$\psi + \hat\Omega\chi = \begin{pmatrix} G(A_0 + D\delta(\theta(p^0)p^2 + m^2)A_p + D\overset{(0)}{\chi}_0 + D\delta(\theta(p^0)p^2 + m^2)\overset{(1)}{\chi}_0 + G\chi_1) + \delta(\eta^1)B + E \\ -D(A_0 + D\overset{(0)}{\chi}_0 + G\chi_1)\end{pmatrix},$$

where $\chi_0 = \overset{(0)}{\chi}_0 + \delta(\theta(p^0)p^2 + m^2)\overset{(1)}{\chi}_0$. This time $E$ cannot be completely removed by adding states in Im $\hat\Omega$. Comparing with (32) we see that states of the form $\psi_0 = E$, $\psi_1 = 0$ are degenerate directions. Together this proves the following theorem.

**Theorem 2:** The choices (22)(ii)–(iv) lead to a degenerate induced Hermitian form on the BRST cohomology.

We are thus left with the choice (22)(i).

*Corollary:* The term $\psi_0$ is a test function in $\eta^1$ (and a distribution in $p$). The term $\psi_1$ is a test function in $p$ (and a distribution in $\eta^1$).

## IV. THE BRST COHOMOLOGY AND THE PHYSICAL SUBSPACE(S)

We have to calculate the BRST cohomology for the choice (22)(i). We have already determined Ker $\hat\Omega$ in (27), it is

$$\text{Ker}\,\hat\Omega = \left\{\psi \,\middle|\, \begin{matrix}\psi_0 = GA_0 + GD\delta(\theta(p^0)p^2 + m^2)A_p + E \\ \psi_1 = -DA_0 - \delta(\eta^1)DA_\eta + \delta(\eta^1)H\end{matrix} \quad \text{with} \quad \begin{matrix}A_0, A_p, A_\eta, H \text{ test functions in both} \\ \text{variables, } E \text{ test function in } \eta^1,\ DE = 0\end{matrix}\right\}. \tag{34}$$

To compute the image of $\hat\Omega$ we first have to specify the domain of definition, i.e., that subset of all (generalized) functions (22)(i) for which also $\hat\Omega\chi$ satisfies (22)(i). These are the states that fulfill [see (22) and (33)]

$$G^2\chi_1 \text{ test function in } \eta^1,\ D^2\chi_0 \text{ test function in } p. \tag{35}$$

These states are given by

$$\begin{matrix}\chi_0 = R_0 + \delta(\theta(p^0)p^2 + m^2)R_1 \\ \chi_1 = S_0 + \delta(\eta^1)S_1 + \delta'(\eta^1)S_2\end{matrix} \quad \text{with} S_j, R_i \text{ test functions.} \tag{36}$$

So for Im $\hat\Omega$ we get [using $x\delta'(x) = -\delta(x)$ and $x^k\delta^{(n)}(x) = 0$ for $k > n$]

$$\text{Im}\,\hat\Omega = \left\{\chi \,\middle|\, \begin{matrix}\chi_0 = GDR_0 + GD\delta(\theta(p^0)p^2 + m^2)R_1 + G^2S_0 \\ \chi_1 = -D^2R_0 - DGS_0 + D(1/\sqrt{2})\delta(\eta^1)S_2\end{matrix} \quad R_i, S_j \text{ test functions}\right\}. \tag{37}$$

In order to compute the cohomology, choose as a first step $R_1 = -A_p$ and $S_2 = \sqrt{2}A_\eta$ to see that in every equivalence class there is a state with $A_p = A_\eta = 0$. Now add another state of Im $\hat{\Omega}$, namely, one satisfying

$$A_0 + DR_0 + D\delta(\theta(p^0)p^2 + m^2)R_1 = 0.$$

This equation always has a solution: with the $R_0$ term one can change $A_0$ arbitrarily outside Ker $D$, with the $R_1$ term in Ker $D$. Together we can thus bring $A_0$ to zero everywhere. Another indication that the $A$ terms should be removable comes from the fact that they do not contribute in the induced form as seen from (29). By adding further $R_1$ terms one can change $E$ at will except on the hypersurface $\eta^1 = 0$. So the only invariant information contained in $E$ is $E(\eta^1 = 0, p)$, which is in fact also the only part contributing in (29). Thus so far we have shown that in every equivalence class there are states of the form

$$\psi_0 = E(p), \quad DE = 0, \quad \psi_1 = \delta(\eta^1)H. \tag{38}$$

Next we show that by adding terms $(1/\sqrt{2})\delta(\eta^1)DS_2$ one can change $H$ at will whenever $D\gamma^5 H \neq 0$.

Choosing $S_2 = D\tilde{S}_2$ it is easily seen that $H$ can be changed everywhere outside $\theta(p^0)p^2 + m^2 = 0$. To see that in fact only solutions of $D\gamma^5 H = 0$ have an invariant meaning is somewhat more difficult. The task is, however, simplified by the fact that this is a Lorentz invariant statement and it therefore suffices to show it in the rest frame of the particle $(p^0 = m, \vec{p} = 0)$. First, we choose an explicit representation for the $\gamma$ matrices

$$\gamma^0 = i\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \vec{\gamma} = i\begin{pmatrix} 0 & \vec{\sigma} \\ -\vec{\sigma} & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{39}$$

Transforming $H$ away means solving the equation $(1/\sqrt{2}) \times DS_2 + H = 0$, i.e.,

$$(1/\sqrt{2})(\theta(p^0)p_\mu\gamma^5\gamma^\mu + im\gamma^5)S_2 + H = 0, \tag{40}$$

or written differently

$$[\theta(p^0)p_\mu(-i\gamma^\mu) + m]S_2 = i\sqrt{2}\gamma^5 H.$$

In the rest frame $p^0 = m, p_0 = -m, \vec{p} = 0$, using (39), this is

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2m & 0 \\ 0 & 0 & 0 & 2m \end{pmatrix}\begin{pmatrix} S^1 \\ S^2 \\ S^3 \\ S^4 \end{pmatrix} = i\sqrt{2}\begin{pmatrix} H^3 \\ H^4 \\ H^1 \\ H^2 \end{pmatrix}. \tag{41}$$

This has solutions iff $H^3 = H^4 = 0$. The equation $D\gamma^5 H = 0$ in the rest frame is (after multiplication with $-i\gamma^5$ from the left)

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2m & 0 \\ 0 & 0 & 0 & 2m \end{pmatrix}\begin{pmatrix} H^3 \\ H^4 \\ H^1 \\ H^2 \end{pmatrix} = 0$$

and its solutions are $H^1 = H^2 = 0, H^3, H^4$ arbitrary. So our claim that the solutions of $D\gamma^5 H = 0$ have exactly an invariant meaning emerges. We thus have the following theorem.

**Theorem 3:** The BRST cohomology for the spinning particle is (isomorphic to)

$$\text{Ker } \hat{\Omega}/\text{Im } \hat{\Omega} \cong \{\psi | \psi_0 = E(p), \ DE = 0, \ \psi_1 = \delta(\eta^1)H, \ D\gamma^5 H = 0\}. \tag{42}$$

Now $DE = 0$ iff $E$ is a solution of the Dirac equation [modified by $\theta(p^0)$], and also for $\gamma^5 H$. So, as for the scalar particle, we have a doubling. The induced form on the BRST cohomology is

$$(\psi,\phi) = -i\int d^4p(E^*_\psi\gamma^5\gamma^0 H_\phi - H^*_\psi\gamma^5\gamma^0 E_\phi),$$

or if we set

$\gamma^5 H \doteq F$,

$$(\psi,\phi) = i\int d^4p(E^*_\psi\gamma^0 F_\phi + F^*_\psi\gamma^0 E_\phi), \tag{43}$$

where $E$ and $F$ are solutions of the Dirac equation [modified by $\theta(p^0)$]. The norm of a state $\psi$ is

$$\|\psi\|^2 = \int d^4p\, E^*(i\gamma^0)F + F^*(i\gamma^0)E. \tag{44}$$

If we expand $E$ and $F$ in plane waves and do the $p^0$ integration we get

$$\|\psi\|^2 = \int d\mu(p) \sum_{\alpha=1}^{2} f^*(\vec{p},\alpha)e(\vec{p},\alpha) + e^*(\vec{p},\alpha)f(\vec{p},\alpha), \tag{45}$$

where $d\mu(p) = d^3p/(2\pi)^3 2p^0$ and $e, f$ are the coefficients in the expansion of $E$ and $F$, respectively. As one sees from (45), the induced Hermitian form (43) is not positive definite.

The situation we have is analogous to the bosonic case. Again we have a doubling and the induced Hermitian form on the BRST cohomology is not positive definite. Note that it is necessary (at least in our rigged Hilbert space approach to the BRST cohomology) to have the doubling in the cohomology, because otherwise, as one sees from Theorem 1, there would be no nonzero Hermitian form on it which is induced from an admissible (see text before Theorem 1) form on the extended phase space. However, the cohomology cannot be the physical space since the induced form on it is not positive definite. So as in Ref. 2 we try to find linear subspaces of it, which are left invariant by the algebra of BRST observables and restricted to which (43) is positive definite.

According to Theorem 3, the BRST cohomology is the space of pairs, $(E,F)$ of solutions of the Dirac equation. The linear subspaces of it with which we will be concerned are

$$\Pi_{\mathscr{K}} = \{(E,F) | F = \mathscr{K}E\}, \tag{46}$$

where $\mathscr{K}$ is a linear map from the space of solutions of the Dirac equation to itself. We look for $\mathscr{K}$ such that $\Pi_{\mathscr{K}}$ satisfies the above two requirements. To this end we first prove a theorem on the form of BRST observables.

**Theorem 4:** Any BRST observable $\hat{\mathbb{A}}$ is of the form $\begin{pmatrix} \hat{A} & 0 \\ 0 & \hat{A} \end{pmatrix}$ with $[\hat{A}, \hat{G}_a] = 0$.

*Proof:* Any $\hat{\mathbb{A}}$ is a matrix $\hat{\mathbb{A}} = \begin{pmatrix} x & y \\ z & u \end{pmatrix}$ acting on the $\begin{pmatrix} \psi_0 \\ \psi_1 \end{pmatrix}$ space. We demand that it maps Ker $\hat{\Omega}$ into itself and is defined on the whole of Ker $\hat{\Omega}$, i.e., using (26),

$$\tilde{\psi} = \hat{\mathbb{A}}\psi = \begin{pmatrix} x(GA_0 + GD\delta(\theta(p^0)p^2 + m^2)A_p + E) + y(-DA_0 - \delta(\eta^1)DA_\eta + \delta(\eta^1)H) \\ z(GA_0 + GD\delta(\theta(p^0)p^2 + m^2)A_p + E) + u(-DA_0 - \delta(\eta^1)DA_\eta + \delta(\eta^1)H) \end{pmatrix}$$
$$\dot{=} \begin{pmatrix} G\tilde{A}_0 + GD\delta(\theta(p^0)p^2 + m^2)\tilde{A}_p + \tilde{E} \\ -D\tilde{A}_0 - \delta(\eta^1)D\tilde{A}_\eta + \delta(\eta^1)\tilde{H} \end{pmatrix}.$$

From this one sees immediately that $y = z = 0$. So, $\hat{\mathbb{A}}$ is of the form $\hat{\mathbb{A}} = \left(\begin{smallmatrix} x & 0 \\ 0 & u \end{smallmatrix}\right)$. The operator $\hat{\mathbb{A}}$ must commute with $\hat{\Omega}$, so with (11) we get

$$0 = [\hat{\Omega},\hat{\mathbb{A}}] = \begin{pmatrix} [GD,x] & G^2u - xG^2 \\ uD^2 - D^2x & -[DG,u] \end{pmatrix}.$$

Since $x$ and $u$ stand in the diagonal of $\hat{\mathbb{A}}$, and $\hat{\mathbb{A}}$ has ghost number zero, so must $x$ and $u$, i.e., they must be of the form

$$\sum_i x_i(\hat{\eta}^1)^i(\hat{\mathscr{P}}_1)^i,$$

where the $x_i$ do not contain any ghosts or odd products of $\hat{\theta}$'s. (Since $\hat{\mathbb{A}}$ is real and even, so must be $x$ and $u$.) Thus $x_i = x_i(\hat{p},\hat{q},\hat{\theta}\hat{\theta})$. The order of $(\hat{\eta}^1)^i(\hat{\mathscr{P}}_1)^i$ is ambiguous; different factor ordering changes the $x_j$ for $j<i$ by additive constants. Assume that all $\hat{\eta}$'s stand to the left of all the $\hat{\mathscr{P}}$'s. Then we have

$$0 = \left[GD, \sum_j x_j(\eta^1)^j(\mathscr{P}_1)^j\right]$$
$$= \frac{1}{\sqrt{2}} \sum_j \{[D,x_j](\eta^1)^{j+1}(\mathscr{P}_1)^j + jx_j(\eta^1)^j(\mathscr{P}_1)^{j-1}D\}$$
$$\curvearrowright [D,x_j] + (j+1)x_{j+1}D = 0$$
$$\curvearrowright x_{j+1} = [1/(j+1)](x_j - Dx_jD^{-1}).$$

Next we use the fact that BRST observables are extensions of Dirac observables, i.e., $\mathbb{A}|_{\eta^a = 0 = \mathscr{P}_b} = \mathbb{A}_0 = \left(\begin{smallmatrix} x_0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ commutes with the constraints, so $[x_0,D] = 0 = [x_0,D^2]$. From the above recursion formula we thus get $x_j = 0$, $\forall j > 0$, so $x = x_0$. From the equation $0 = uD^2 - D^2x = (u-x)D^2$, we get $u = x \dot{=} A$. ■

We note that although the representation (8) is irreducible for the whole algebra of quantum operators, it becomes reducible when restricted to the subalgebra of BRST observables.

Using Theorem 4 we see that the condition for $\Pi_{\mathscr{K}}$ to be invariant under the action of the BRST observables is $\hat{A}\mathscr{K} = \mathscr{K}\hat{A}$ for any $\hat{A}$ that commutes with $D$ and $D^2$.

The most general form of a linear transformation on the space of four-spinors is

$$(\mathscr{K}E)_\alpha(p) = \int \mathscr{K}_{\alpha\beta}(p,p')E_\beta(p')dp',$$

but since $\hat{A} = A(\hat{p})$ is an observable for any $A(\hat{p})$ it is clear that $\mathscr{K}_{\alpha\beta}(p,p')$ has to be of the form $\mathscr{K}_{\alpha\beta}(p)\delta(p-p')$ and we have to discuss only

$$(\mathscr{K}E)_\alpha(p) = \mathscr{K}_{\alpha\beta}(p)E_\beta(p).$$

Further $\mathscr{K}$ should map Ker $D$ into Ker $D$ so that we must have $[\mathscr{K},\hat{D}] = \hat{B}\hat{D}$. Since the Lorentz generators $\hat{J}_{\mu\nu} = \hat{p}_\mu\hat{q}_\nu - \hat{p}_\nu\hat{q}_\mu + (i/2)(\hat{\theta}_\mu\hat{\theta}_\nu - \hat{\theta}_\nu\hat{\theta}_\mu)$ and the $C$, $P$, and $T$ operations commute with $D$, the operator $\mathscr{K}_{\alpha\beta}(p)$ must be a scalar, and thus it is of the form $\mathscr{K}(p^2)\delta_{\alpha\beta}$.

However, not all $\mathscr{K}(p^2)$ are allowed; we also have to fulfill the second requirement, namely that (35) restricted to $\Pi_{\mathscr{K}}$ is positive definite. The situation is as in the bosonic case and, exactly as in Ref. 2, Sec. IV, one proves the following theorem.

**Theorem 5:** On the subspaces $\Pi_{\mathscr{K}}$ of the BRST cohomology defined by $f(\vec{p},\alpha) = \mathscr{K}((\vec{p})^2)e(\vec{p},\alpha)$, with $\mathrm{Re}\,\mathscr{K} > 0$, the induced form (35) is positive definite and all the $(\Pi_{\mathscr{K}}; (\cdot,\cdot)|_{\Pi_{\mathscr{K}}})$ are isomorphic scalar product spaces.

A convenient choice for $\mathscr{K}$ is

$$\mathscr{K} = \tfrac{1}{2}, \quad \text{i.e.,} \quad f = \tfrac{1}{2}e \dot{-} \tfrac{1}{2}b$$

as it leads to

$$(\psi,\phi) = \int \frac{d^3p}{(2\pi)^3 2p^0} \sum_{\alpha=1}^{2} b^*_\psi(\vec{p},\alpha)b_\phi(\vec{p},\alpha)$$

for the induced product on $\Pi_{1/2}$. This is the usual Dirac product with the negative energy states removed. [There are no negative energy oscillators since we have included a $\theta(p^0)$ in the constraints.]

## ACKNOWLEDGMENTS

It is a pleasure to thank J. Abraham, J. Fisch, and especially P. Hajicek and M. Henneaux for helpful and stimulating discussions.

This work was supported by Schweizerischer National-Fonds.

[1] C. Galvão and C. Teitelboim, J. Math. Phys. **21**, 1863 (1980).
[2] P. Thomi, J. Math. Phys. **29**, 1014 (1988).
[3] M. Henneaux, Phys. Rep. **126**, 1 (1985).
[4] S. Monaghan, Phys. Lett. B **178**, 231 (1986).
[5] R. Coquereaux, "Spinors, reflections and Clifford algebras: A review," in *Spinors in Physics and Geometry*, edited by R. Furlan and A. Trautman (World Scientific, Singapore, 1987).

# Covariant factor ordering of gauge systems using ghost variables. I. Constraint rescaling

D. McMullan[a] and J. Paterson[b]
*Department of Physics and Astronomy, The University of Glasgow, Glasgow G12 8QQ, Scotland*

(Received 22 June 1988; accepted for publication 28 September 1988)

A dynamical consequence of local gauge invariance is the existence of first class constraints that are linear in momenta. By studying finite-dimensional systems with this constraint structure one finds that, in order to quantize such theories, it is not enough to find a consistent factor ordering, rather, one must also maintain covariance under rescaling of the constraints, point transformations, and weak changes to the observables. Within the standard Dirac constraint formalism, covariance under these symmetries obstructs a full Hilbert space description. It is shown how this difficulty may be overcome by the use of ghost variables. In the present paper, I, an analysis of the Becchi–Rouet–Stora–Tyutin (BRST) structure of such systems is presented. The constraint rescaling is shown to be implemented by a canonical transformation on the super phase space that can be evaluated to a unitary transformation on a suitably defined state space. In the following paper, II [J. Math. Phys. **30**, 487 (1989)] these techniques are used to solve the constraint factor ordering problem.

## I. INTRODUCTION

Quantization involves the metamorphosis of a classical theory into its quantum version. Since we believe that, at some fundamental level, the quantum description is basic, it cannot then come as too much of a surprise if sometimes the quantization procedure appears to break down. Van Hove[1] was the first to show that, in finite-dimensional systems, one cannot expect to represent all the classical structures in the quantum theory. In particular, the classical invariance under canonical transformations is not compatible with quantum theory. On top of this type of problem, field theory has the additional requirement of renormalizability that can render (at least perturbatively) many classical systems as unviable in the quantum regime.

Systems that have a local gauge, or reparametrization, invariance have the additional complication brought about by the occurrence of (first class) constraints. These imply the existence of unphysical modes that should not affect any physical result in either the classical or quantum descriptions. The most expedient way to deal with this problem is to remove all unphysical objects and deal directly with the true degrees of freedom. However, often one cannot or does not wish to implement this reduction prior to quantization and hence a constraint quantization procedure is needed. Obviously, any such scheme will hit the Van Hove type of problem when the reduction to physical quantum states is implemented. But, one would hope that there would be no additional complications induced by the unphysical modes prior to the reduction.

In the context of canonical quantization, Dirac[2] developed a general constraint quantization scheme whereby the constraints are turned into operators on the extended state space (including both the physical and unphysical states). The physical state space is then picked out using these operators. The attraction of this approach is that a quantum description is maintained at all stages, thus allowing many familiar mathematical tools to be used in this generalized context. This philosophy can easily be extended to other approaches to quantization, such as the path integral, where one finds that the constraints alter the measure used on the extended states (see Ref. 3 and references therein).

There are, however, several types of problems associated with Dirac's procedure. The most serious is the difficulty in finding a factor ordering for the constraints that will allow for a consistent description of the physical states and observables. This has long been the bane of research into quantum gravity, where the (Hamiltonian) constraint is quadratic in momenta and there are structure functions instead of the familiar structure constants.[4] What is more surprising is that, even in finite-dimensional systems one cannot guarantee that a consistent factor ordering of the constraints will imply that one can recover the correct physical description of the theory.[5]

In a series of papers[6,7] Kuchar developed an alternative constraint quantization procedure for finite-dimensional systems that have first class constraints linear in momenta. He was able to solve the factor ordering problem explicitly and reproduce the quantized physical theory while maintaining covariance under a wide class of transformations. These successes were, however, at the expense of a Hilbert space structure on the extended state space. Operators are, therefore, self-adjoint only after reduction to the physical Hilbert space and hence it is no longer appropriate to talk about the constraints being self-adjoint.

This nominalistic solution to the ordering problems has its precursor in quantum gravity where there has been a long debate into the desired operator status assigned to the Hamiltonian constraint (see discussion in Ref. 6). There one can argue, on physical grounds, that since the Hamiltonian constraint is simply a manifestation of a "many-fingered time," and we know that the time variable cannot be represented by a self-adjoint operator,[8] we should not expect this constraint to be self-adjoint. However, that such a radical solution appears to be forced upon us for constraints related to gauge

[a] Supported by the Science and Engineering Research Council.
[b] Present address: Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom.

 0022-2488/89/020477-10$02.50 © 1989 American Institute of Physics 

invariance (and hence linear in momenta) is much harder to understand.

In Ref. 9 we presented a factor ordering prescription for the systems considered by Kuchar which, as well as satisfying all Kuchar's requirements, maintained an (indefinite) Hilbert space structure on the extended state space. This was achieved by using ghost variables. In this and the succeeding paper[10] we shall present an improved version of these orderings and give the details of how they work.

Ghost variables were originally introduced into constrained dynamics as a means for developing a path integral formulation for gauge theories when "covariant gauge fixing" was used.[11] We shall show that ghost variables can have a much wider use in canonical quantization. In the present paper, I, we shall concentrate on the classical aspects of ghosts and the structure of the quantum state space. In the following paper, II, the projection of physical states and the orderings for the physical observables will be derived. It will be shown that the ghost approach does not suffer from the failings pointed out in Ref. 5.

## II. CONSTRAINTS AND GHOSTS

### A. First class constraints

We shall be dealing with phase spaces of the form $T^*Q$, where $Q$ is an extended configuration space of dimension $N$. It is expected that the methods presented in this paper can be extended to more general symplectic manifolds. However, since our constraints will always be linear in momenta it is sensible to keep a cotangent bundle as basic. Locally $T^*Q$ is described by $2N$ canonically conjugate coordinates $Q^A, P_A$. Here we adopt the convention that capital latin indices will run over all the degrees of freedom on $T^*Q$, while greek indices will run over the number of constraints, which we denote by $k$ ($k<N$). Later we will have recourse to lowercase latin indices which will run over the true degrees of freedom.

The $k$ constraints will be denoted by $\varphi_\alpha$ and they are assumed to be independent, linear in momenta, and first class. Thus, locally in $T^*Q$, we can find smooth functions ${\varphi_\alpha}^A$ and ${C^\gamma}_{\alpha\beta}$ of the $Q^A$'s such that

$$\varphi_\alpha = {\varphi_\alpha}^A(Q^B)P_A, \tag{2.1}$$

and

$$\{\varphi_\alpha,\varphi_\beta\} = {C^\gamma}_{\alpha\beta}(Q^A)\varphi_\gamma. \tag{2.2}$$

Being independent implies that the constraint surface, given by $\varphi_\alpha = 0$, is a $(2N-k)$-dimensional submanifold of $T^*Q$. Note that, in (2.2), structure functions are allowed rather than just structure constants; we shall return to this later.

The dynamical objects of interest on $T^*Q$ are the physical observables. These are represented by equivalence classes of weakly invariant functions on $T^*Q$, where $F$ is said to be weakly invariant if

$$\{F,\varphi_\alpha\} = {F_\alpha}^\beta\varphi_\beta, \tag{2.3}$$

for some smooth function ${F_\alpha}^\beta$. The equivalence class structure is that of weak equivalence, i.e., equivalence when restricted to the constraint surface. So $F_1 \sim F_2$ iff $F_1 - F_2 = L^\alpha\varphi_\alpha$ for some $L^\alpha$.

Since we cannot expect to quantize all classical observables we shall follow Kuchar[6] and restrict our attention to special physical observables. So we denote by $Y,Z$ functions of the $Q^A$'s only; by $U,V$ those functions linear in the $P_A$'s; and by $H,K$ the functions quadratic in their dependence on the $P_A$'s. The equivalent lowercase letters will denote functions with the same momenta dependence, but now in terms of the physical degrees of freedom.

The system described by (2.1)–(2.3) has a number of symmetries that do not affect the true dynamical content of the theory. However, not all such transformations will preserve the linearity of the constraints and the classification of special observables. Since it is these structures that we wish to emphasize, we shall adopt Kuchar's philosophy of only concentrating on those symmetries that preserve these additional structures. Thus, for example, we replace invariance under the canonical transformations on $T^*Q$ with invariance under point transformations.

With these restrictions, there are four types of symmetries that we wish to preserve in the quantization. These are (a) invariance under general point transformations on $T^*Q$, i.e.,

$$Q^A \to Q^{A'}(Q^A), \tag{2.4a}$$

$$P_A \to P_{A'} = {Q^A}_{A'}P_A, \quad \text{where } {Q^A}_{A'} = \frac{\partial Q^A}{\partial Q^{A'}}; \tag{2.4b}$$

(b) invariance under general point transformations on the true degrees of freedom [which we can consider as a special case of (a)]; (c) invariance under weak changes to the special observables, i.e.,

$$H \to H + H^\alpha\varphi_\alpha,$$

where $H^\alpha$ is at most linear in momenta; and (d) invariance under rescaling of the constraints that preserve their independence and their linear structure, i.e.,

$$\varphi_\alpha \to \tilde{\varphi}_\alpha = {\Lambda_\alpha}^\beta\varphi_\beta, \tag{2.5}$$

for any invertible ${\Lambda_\alpha}^\beta(Q^A)$. We note that, by requiring invariance under this type of rescaling, we are forced to consider structure functions in (2.2).

It is worth noting a slight generalization in the classical setup presented here over that in Ref. 9. There we assumed a Riemannian structure on the extended configuration space $Q$. This was used to construct a kinetic energy term $G^{AB}P_AP_B$, where $G^{AB}$ is nondegenerate. All one actually needs though and, in particular, all that Kuchar assumed, is that on the true degrees of freedom the kinetic energy becomes $g^{ab}p_a\,p_b$, where $g^{ab}$ is degenerate. In practice, one is rarely given a degenerate $G^{AB}$ [see (1) in Sec. VII], but even if one is it is possible to exploit the weak invariance given in (c) above to remove the degeneracy (which, by assumption, is in the unphysical directions). Thus the Riemannian assumption in Ref. 9 corresponds to a privileged choice of kinetic energy term from the equivalence class of physical Hamiltonians. At the end of the day, though, such a choice does not affect the physical results. Here we shall show that it is possible to proceed without such a decomposition of the Hamiltonian. There is a price to pay for this extension of the results presented in Ref. 9 since we shall find that the mea-

sures used in the quantum state space will now depend on the choice of constraints.

The problematic invariance is that given by (2.5). Although it is true that such a rescaling does not affect the physical results of the theory, it is not a manifest symmetry since it is not compatible with the Poisson algebra on $T^*Q$. Hence there is no canonical transformation, giving (2.5), which we might try and extend to a unitary transformation in the quantum description. It is this fact that forced Kuchar to abandon the standard operator status of the constraints. One might argue then that we should simply drop this invariance from our list as it is not as natural on $T^*Q$ as the others. However, Kuchar's insistence on keeping it did lead to a correct physical quantization in contrast to Dirac's approach, where the rescaling invariance is never considered.

We shall show that, by extending the phase space in a fermionic direction, it is possible to implement (2.5) as an (even) canonical transformation and hence as a Hilbert space isomorphism in the quantum theory. If one accepts Kuchar's argument that such a rescaling invariance is essential to a consistent constraint quantization scheme, then, since making symmetries manifest has always been the working philosophy in physics, one must accept the conclusion that ghost variables are not just an artifact of covariant gauge fixing, but are fundamental to any constraint quantization formalism.

### B. Ghosts variables and constrained systems

In the context of canonical methods, ghost variables were first introduced by Batalin, Fradkin, and Vilkovisky (BFV) in an attempt to derive the covariant path integral for an arbitrary constrained system.[11] The general use of ghosts variables in field theory has, however, a much longer history.[12]

The essence of the BFV approach was to extend the phase space in a fermionic direction by adding, for each constraint $\varphi_\alpha$, a ghost (Grassman) variable $\eta^\alpha$ and its conjugate $\rho_\alpha$ [see (2) in Sec. VII]. The BRST charge $\Omega$ is then constructed where

$$\Omega = \varphi_\alpha \eta^\alpha + \tfrac{1}{2} C^\alpha_{\beta\gamma} \eta^\beta \eta^\gamma \rho_\alpha. \tag{2.6}$$

The function $\Omega$ has ghost number one and is Abelian with respect to the graded Poisson algebra, i.e.,

$$\{\Omega,\Omega\} = 0. \tag{2.7}$$

Using $\Omega$ we can construct a coboundary operator $\delta$ where $\delta \mathbf{F} = \{\Omega,\mathbf{F}\}$ for any function $\mathbf{F}$ on this graded phase space. Equation (2.7) then implies that $\delta^2 = 0$. When the constraints are related to a group action on $Q$ [i.e., structure constants in (2.2)] we can interpret $\delta$ as the Koszul resolution of the Lie algebra cohomology taking values in $C^\infty(T^*Q)$.[13] When structure functions are present one still has a foliation of the configuration space but now one can only interpret the cohomology of $\delta$ as the cohomology of these orbits, when restricted to the constraint surface. Otherwise a more general Koszul resolution is needed.[14]

Given a physical observable $F$, satisfying (2.3), one can always construct an even, ghost number zero function, denoted by $\mathbf{F}$, which satisfies

$$\delta \mathbf{F} = \{\Omega,\mathbf{F}\} = 0. \tag{2.8}$$

If $F_1$ is weakly equivalent to $F_2$ then $\mathbf{F}_1$ and $\mathbf{F}_2$ differ by a coboundary term, i.e.,

$$\mathbf{F}_1 = \mathbf{F}_2 + \{\Omega,\mathbf{X}\}, \tag{2.9}$$

from some $\mathbf{X} \in \Gamma^{-1}$.

Since $\Omega \in \Gamma^1$, we see that $\delta$: $\Gamma^r \to \Gamma^{r+1}$ and hence we have the ghost number complex,

$$\to \Gamma^{-2} \xrightarrow{\delta} \Gamma^{-1} \xrightarrow{\delta} \Gamma^0 \xrightarrow{\delta} \Gamma^1 \xrightarrow{\delta} \Gamma^2 \to . \tag{2.10}$$

One can show that the equivalence classes of functions given by (2.8) and (2.9), that is, the zeroth cohomology of this complex, is equivalent to the algebra of physical observables.[14]

It is clear that ghost variables have introduced a certain mathematical elegance into constrained dynamics. What we will show now is that this is not just a cosmetic facelift but represents a positive step forward in describing constrained systems within the symplectic category. In order to achieve this we need two results: the first of which is quite well known, the second is the *sine qua non* of our approach to constraint quantization.

**Theorem 2.1:** Given $T^*Q$, and a set of $k$ first class constraints $\varphi_\alpha$, there exists a rescaling matrix $\Lambda$ such that, locally, the new constraints $\tilde{\varphi}_\alpha$ are Abelian. Standard phase space methods then tell us that there exists a canonical transformation that will make the $\tilde{\varphi}_\alpha$ into the first $k$ momentum coordinates.

So, locally the constraints can always be arranged to be pure momenta. We shall refer to this as local trivialization of constraints. This is proved in all generality in Ref. 15. However, for linear constraints, there is a relatively simple geometric proof. The idea is to regard the constraints as generating vector fields on $Q$, i.e.,

$$\varphi_\alpha = \varphi_\alpha{}^A \partial_A, \tag{2.11}$$

where $\partial_A = \partial/\partial Q^A$. The first class nature of the $\varphi_\alpha$'s implies that

$$[\varphi_\alpha,\varphi_\beta] = C^\gamma{}_{\alpha\beta} \varphi_\gamma, \tag{2.12}$$

where $[\ ,\ ]$ is the Lie bracket on $Q$. This means that the $\varphi_\alpha$'s are surface forming, i.e., they generate a foliation of $Q$ by $k$-dimensional submanifolds. Any choice of basis vector fields on this foliation gives a valid choice of constraints on $T^*Q$. The trivialized constraints are then obtained by choosing a coordinate basis for the foliation. This has to be done in local patches and so the theorem may fail globally.

Theorem 2.1 is important because of the following result.

**Theorem 2.2:** Rescaling of constraints, as given in (2.5), can be interpreted as an even canonical transformation in the super phase space.

This is discussed infinitesimally in Ref. 15 and given its general setting in Ref. 16. The essence of the proof is to construct a generating function. The relevant function is

$$\mathbf{F}_3(\tilde{Q}^A,\tilde{\eta}^\alpha,P_B,\rho_\beta) = -\tilde{Q}^A P_A - (\Lambda^{-1})_\alpha{}^\beta \tilde{\eta}^\alpha \rho_\beta. \tag{2.13}$$

This gives the active canonical transformation ($\Lambda$ is a function of the $Q^A$'s only),

$$\tilde{Q}^A = Q^A, \tag{2.14a}$$

$$\tilde{P}_A = P_A + (\Lambda^{-1})_\alpha{}^\beta{}_{,A}\Lambda_\gamma{}^\alpha \eta^\gamma \rho_\beta, \tag{2.14b}$$

$$\tilde{\eta}^\alpha = \Lambda_\beta{}^\alpha \eta^\beta, \tag{2.14c}$$

$$\tilde{\rho}_\alpha = (\Lambda^{-1})_\alpha{}^\beta \rho_\beta, \tag{2.14d}$$

where $_{,A}$ has its usual interpretation as a derivative with respect to $Q^A$. This transformation can be generalized to $\Lambda(Q^A, P_B)$. In this case (2.13) is still valid but Eqs. (2.14) pick up extra terms. We shall concentrate purely on the restricted case (2.14) as it is all that is required to deal with linear constraints. Indeed, the existence of a quantum form of this transformation, in the general case of $\Lambda(Q^A, P_B)$, is far from clear and will be discussed in II.

It is straightforward to show that the transformation (2.14) sends the BRST charge $\Omega$, constructed from the old constraints $\varphi_\alpha$, to $\tilde{\Omega}$ constructed from the new constraints $\tilde{\varphi}_\alpha$. To see that it also transforms the physical observables correctly it is most convenient to use the ghost number complex (2.10). We now have two complexes, one constructed from the old constraints $\varphi_\alpha$, and one from the new constraints $\tilde{\varphi}_\alpha$. The above canonical transformation, which we shall denote by $R$, is clearly invertible and provides a homomorphism between these complexes via its pullback $R^*$, i.e.,

$$\begin{array}{ccccccccc} \rightarrow \tilde{\Gamma}^{-1} & \xrightarrow{\tilde{\delta}} & \tilde{\Gamma}^0 & \xrightarrow{\tilde{\delta}} & \tilde{\Gamma}^1 & \xrightarrow{\tilde{\delta}} & \tilde{\Gamma}^2 \rightarrow \\ \downarrow R^* & & \downarrow R^* & & \downarrow R^* & & \downarrow R^* \\ \rightarrow \Gamma^{-1} & \xrightarrow{\delta} & \Gamma^0 & \xrightarrow{\delta} & \Gamma^1 & \xrightarrow{\delta} & \Gamma^2 \rightarrow . \end{array}$$

Now, if $R^*$ is a chain mapping, that is,

$$\delta R^* = R^* \tilde{\delta}, \tag{2.15}$$

then the cohomology, and hence physical observables, described by the two complexes are equivalent. The transformation $R$ is easily seen to induce the above chain mapping by the following argument. Let $\mathbf{F} \in \Gamma^r$; then

$$\begin{aligned} \delta R^* \mathbf{F} &= \{\Omega, R^* \mathbf{F}\}, && \text{by definition of } \delta, \\ &= \{R^* \tilde{\Omega}, R^* \mathbf{F}\}, && \text{since } \Omega \text{ transforms correctly under } \mathbf{R}, \\ &= R^* \{\tilde{\Omega}, \mathbf{F}\}, && \text{since } R \text{ is a canonical transformation}, \\ &= R^* \tilde{\delta} \mathbf{F}, && \text{by definition of } \tilde{\delta}. \end{aligned}$$

The super phase space also allows weak changes of observables to be implemented as a canonical transformation.[15] It is, however, the transformation $R$ that is important to us. We shall show that $R$ becomes a Hilbert space isomorphism $\mathbf{R}$ in the quantum theory. The existence of this transformation makes it possible to elevate the symmetries (a)–(d) to the quantum theory without losing a Hilbert space structure in the process.

This transformation, in conjunction with Theorem 2.1, also provides a direct and powerful means of obtaining the correct operator orderings. The strategy is to formulate the orderings in the trivialized coordinate system and then use $R$, and general coordinate covariance on $Q$, to undo the trivialization. This is a different, more direct approach than that of Kuchar and is only possible because of Theorem 2.2. This, in turn, is only true once ghosts are added to the theory.

## III. THE PHYSICAL QUANTUM THEORY

After reduction to the true, independent degrees of freedom on $Q$, we will be left with the physical configuration space $q$, and its associated phase space $T^*q$. The term $q$ has dimension $n$ ( $= N - k$) and is endowed with a (nondegenerate) Riemannian metric, $g_{ab}$, induced by the kinetic energy part of the physical Hamiltonian $h$.

It is useful here to recall the salient features of the quantum theory on such a space. The quantization of the theory will be carried out in the Schrödinger picture, i.e., the physical quantum states will be the complex valued functions on $q$ that are square integrable with respect to the pairing

$$\langle \Psi | X \rangle = \int \Psi^*(q^a) X(q^a) |g|^{1/2}\, dq^1 \cdots dq^n, \tag{3.1}$$

where $|g| = \det[\, g_{ab}\,]$. It is a standard result of Riemannian geometry that the integral in (3.1) is coordinate covariant. This follows from the well known fact that

$$|g|^{1/2} \rightarrow |g'|^{1/2} = |q|\, |g|^{1/2}, \quad \text{where } |q| = \det[\, q^a{}_{a'}\,], \tag{3.2}$$

under a general coordinate transformation $q^a \rightarrow q^{a'}$ on $q$.

The quantum observables, corresponding to the classical ones on $T^*q$, must also satisfy coordinate covariance. In the quantum theory the transformation law (2.4) becomes

$$\hat{q}^a \rightarrow \hat{q}^{a'} = q^{a'}(q^a), \tag{3.3a}$$

$$\hat{p}_a \rightarrow \hat{p}_{a'} = \tfrac{1}{2}[\, q^a{}_{a'}, \hat{p}_a\,]_+, \tag{3.3b}$$

where $[\ ,\ ]_+$ denotes an anticommutator. This form of the classical transformation law is needed in order to keep the momentum operators Hermitian. In addition to coordinate covariance, the quantum observables must be Hermitian and they must give the correct classical limit. However, these three conditions do not uniquely determine the quantum theory. The standard quantization procedure, on a Riemannian manifold, amounts to choosing the simplest orderings of the classical observables that satisfy the above three requirements. We shall adopt this quantization which is

$$q^a \rightarrow \hat{q}^a = q^a, \tag{3.4a}$$

$$p_a \rightarrow \hat{p}_a = -i\hbar\left[\partial_a + \frac{1}{2}\frac{|g|^{1/2}{}_{,a}}{|g|^{1/2}}\right], \tag{3.4b}$$

$$u = u^a p_a \rightarrow \hat{u} = \tfrac{1}{2}[\, u^a, \hat{p}_a\,]_+, \tag{3.4c}$$

$$k = k^{ab} p_a p_b \rightarrow \hat{k} = |g|^{-1/4} \hat{p}_a k^{ab} |g|^{1/2} \hat{p}_b |g|^{-1/4}, \tag{3.4d}$$

where $\partial_a = \partial/\partial q^a$. The coordinate covariance of these expressions is easily checked using (3.2) and (3.3).

Equations (3.4) implement the classical Poisson alge-

bra as a commutator algebra, up to Van Hove obstructions.[1] These obstructions appear only in the commutators of quadratic variables with either a linear or another quadratic variable. The details of this quantization procedure and its associated Van Hove problems can be found in Ref. 7 and the references therein.

It is worth emphasizing that the Poisson brackets, not involving quadratic variables, do quantize according to the Dirac prescription,

$$\{\ ,\ \} \to -(i/\hbar)[\ ,\ ]. \tag{3.5}$$

It will be shown in Paper II that this is the reason why the unphysical directions in $Q$ do not give extra Van Hove obstructions without our BRST approach.

The task of constraint quantization is to find a practical quantization procedure on $Q$ (in our case the BRST extended $Q$) which reproduces the above physical quantum theory. By "a practical quantization procedure" we mean one that does not require the (very difficult) reduction, to the true degrees of freedom, to be carried out.

## IV. THE CONSTRAINED QUANTUM THEORY USING GHOSTS

In this section, we will outline the basic philosophy behind the use of ghosts in quantization. The idea is to quantize the classical theory in the "ghost equivalent" of the Schrödinger picture, i.e., the ghosts will become operators by the following prescription:

$$\eta^\alpha \to \hat{\eta}^\alpha = \eta^\alpha, \tag{4.1a}$$

$$\rho_\alpha \to \hat{\rho}_\alpha = -i\hbar \frac{\partial}{\partial \eta^\alpha}, \tag{4.1b}$$

where $\partial/\partial\eta^\alpha$ denotes right differentiation of Grassmann variables.[17] The quantum states upon which these operators act will be of the form

$$\Psi(Q^A,\eta^\alpha) = \Psi_0 + \sum_{m=1}^{k} \Psi_{\alpha_1\cdots\alpha_m}\eta^{\alpha_1}\cdots\eta^{\alpha_m}. \tag{4.2}$$

The $\Psi_\alpha$'s are totally antisymmetric in their indices and are normally taken to be square integrable functions over $Q$. The meaning of the statement "square integrable" depends, of course, on the choice of integration measure for $Q$. In Ref. 9 a Riemannian integration measure was used. This metric is given to us, in a contravariant form, from the kinetic energy piece of the Hamiltonian (i.e., the quadratic part of $H = G^{AB}P_AP_B$). To construct the Riemannian measure, it is necessary to invert this metric, but there is no guarantee that $G^{AB}$ is nondegenerate. However, by the arguments presented in Sec. II, the metric can be made nondegenerate by adding a suitably chosen combination of the constraints. There is nothing wrong with this approach. It is essentially what was done in Ref. 9. However, it is all a bit clumsy, the point being that $Q$ is not naturally a Riemannian manifold. It is instead a fiber bundle whose base space (the true physical configurations) is Riemannian. It would be preferable to have a covariant measure $d\mu$ for $Q$ which did not force unnatural geometric structures on $Q$. We have found such a measure and will use it in this paper. This means that the solution being presented here to the constraint quantization problem will be slightly different and, in fact, much neater than the one presented in Ref. 9. The details of this new measure will be dealt with in Sec. V A.

The pairing on the quantum states (4.2) is traditionally taken to be

$$\langle\Psi|X\rangle = (i)^{(1/2)k(k-1)}\int \Psi^* X\, d\eta^1\cdots d\eta^k\, d\mu, \tag{4.3}$$

where $d\eta^\alpha$ denotes Berezin integration.[17] It is not *a priori* obvious that the Berezin rules are the correct way to pair such ghost states. Indeed, since the aim of this paper is to characterize the ambiguity inherent in such a constraint quantization, we shall look critically at this choice of pairing in Sec. V B.

Expression (4.3) gives the states (4.2) the structure of an indefinite Hilbert space though, for brevity, we will refer to it (incorrectly) just as a Hilbert space. The physical states are then projected out of this space using the quantum BRST charge, i.e., the physical states are those satisfying

$$\hat{\Omega}\Psi = 0. \tag{4.4}$$

There is, of course, a redundancy in (4.4) due to the quantum analogy of (2.7), i.e.,

$$\hat{\Omega}^2 = 0. \tag{4.5}$$

This redundancy can be ignored if $\hat{\Omega}$ is Hermitian because then all such states have zero norm and so never contribute to physical results.

A problem with this cohomological description of physical states is that, although one can define $\hat{\Omega}$ as a self-adjoint operator on the state space given by expression (4.2) (where the coefficients are square integrable on $Q$), one cannot discuss cohomology on such a space since, in general, the spectrum of the constraints will almost always have a continuous part. Hence the only states satisfying (4.4) will be the zero state. We will show, in Sec. V C, how this can be overcome using a procedure similar to that used in rigged Hilbert space theory.

Having constructed a well defined extended state space, we will address, in Sec. VI, the problem of implementing the classical canonical transformation $R$, of Theorem 2.2, in the quantum theory. This will then complete our analysis of the quantum state space.

The remaining problem of projecting out the correct physical states, and ordering the physical observables, will be dealt with in Paper II.

## V. THE QUANTUM STATES AND THEIR PAIRING

### A. The measure on $Q$

To construct this measure, we need to introduce a very important object $\|\varphi\|$ taken in a modified form from Ref. 7. Here $\|\varphi\|$ is defined by the following two equations:

$$\varphi_{A_1\cdots A_n} = \delta_{A_1\cdots A_nB_1\cdots B_k}\varphi_1{}^{B_1}\cdots\varphi_k{}^{B_k}, \tag{5.1}$$

and [see (3) in Sec. VII]

$$\|\varphi\|^{-2} = (1/n!)\varphi_{A_1\cdots A_n}G^{A_1B_1}\cdots G^{A_nB_n}\varphi_{B_1\cdots B_n}, \tag{5.2}$$

where, in (5.1), we have contracted the constraints with the completely antisymmetric tensor density defined by $\delta_{1\cdots N}$

$= 1$. Here $\|\varphi\|$ is important because of the following three properties [these properties also guarantee that $\|\varphi\|$ is well defined, i.e., that the right-hand side of (5.2) is never zero].

(1) In the trivialized coordinate system, i.e., when the constraints are the first $k$ momenta,

$$\|\varphi\| = |g|^{1/2}, \tag{5.3}$$

where $|g|$ is the determinant of the physical metric, i.e., the metric in the directions $Q^{k+1}$ to $Q^N$.

(2) Under general coordinate transformations on $Q$, $\|\varphi\|$ transforms as a scalar density of weight 1, i.e.,

$$\|\varphi\| \to \|\varphi'\| = |Q|\,\|\varphi\|, \tag{5.4}$$

where $|Q| = \det[Q^A{}_{A'}]$.

(3) Under a rescaling of the constraints $\|\varphi\|$ has the transformation law

$$\|\varphi\| \to \|\tilde{\varphi}\| = |\Lambda|^{-1}\|\varphi\|, \tag{5.5}$$

where $|\Lambda| = \det[\Lambda_\alpha{}^\beta]$.

Here $\|\varphi\|$ can be interpreted geometrically in terms of the pullback, to $Q$, of the physical volume form.[7] It is therefore not surprising that $\|\varphi\|$ is important and, in fact, it will play a central role in our solution to the constraint quantization problem. The covariant measure for $Q$ is taken to be

$$d\mu = \|\varphi\| dQ^1 \cdots dQ^N. \tag{5.6}$$

Coordinate covariance of this measure is assured by (5.4).

Having chosen this measure, the momentum operators on $Q$ are forced to be

$$\hat{P}_A = -i\hbar\left[\partial_A + \frac{1}{2}\frac{\|\varphi\|_{,A}}{\|\varphi\|}\right]. \tag{5.7}$$

This is analogous to the statement that (3.4b) is the momentum operator associated with (3.1). To make this clear we state the following result.

**Theorem 5.1:** Assume that $\hat{P}_A$ must be of the form

$$\hat{P}_A = -i\hbar[\partial_A + f(Q^B)],$$

for some function $f(Q^B)$. Then (5.7) is the unique choice that makes $\hat{P}_A$ Hermitian with respect to (5.6). This operator also reproduces the canonical commutation relations and is coordinate covariant. The proof is straightforward.

We will now examine the measure on the ghosts and thereby construct the full pairing on the quantum states.

### B. The measure on the ghosts

In a path integral formulation with ghost variables the Berezin rules for integrating fermionic variables are always assumed. In our canonical approach we are faced with the task of introducing some pairing $\langle\Psi|X\rangle$ on the states of the form (4.2). Now consider the simple example where $Q = \mathbb{R}^N$ and the constraints are $P_1,\ldots,P_k$. Then,

$$\hat{\Omega} = \hat{P}_\alpha \eta^\alpha,$$

and this projects out states, by (4.4), that are of the form

$$\Psi^{\text{phys}} = \Psi_0(Q^{k+1},\ldots,Q^N) + \text{higher ghost number terms.} \tag{5.8}$$

The higher ghost number terms do not have any simple dependence on the physical directions $Q^{k+1},\ldots,Q^N$. Thus it appears to be $\Psi_0$ that contains the physical information in $\Psi^{\text{phys}}$ and so it would be natural for the pairing $\langle\Psi|X\rangle$ to contain a term of the form $\Psi_0{}^* X_0$. Thus one might imagine a pairing of the form

$$\langle\Psi|X\rangle = \int [\Psi_0{}^* X_0 + \Psi_1{}^* X_1 + \cdots + \Psi^*_{1\cdots k} X_{1\cdots k}]\, d\mu,$$

to be relevant in this operator language. However, the Berezin pairing has no such terms; instead,

$$\langle\Psi|X\rangle = (i)^{(1/2)k(k-1)} \int k![(-1)^{(1/2)k(k-1)}\Psi_0^* X_{1\cdots k} + \Psi^*_{1\cdots k} X_0]\, d\mu + \text{terms independent of } \Psi_0 \text{ or } X_0, \tag{5.9}$$

which makes it hard to see how the physical Hilbert space structure can emerge.

What we shall do now is to show that, even in this general operator formalism, one is forced to take the Berezin pairing. The problem of extracting a pairing on the physical states will then be related to the correct usage of a rigged Hilbert space.

To analyze the possible pairings we will start with the most general pairing and then impose Hermiticity relations for certain operators. This will restrict the possible choice of pairing and will, in fact, lead us uniquely to the Berezin measure for the ghosts. The most general pairing can be parametrized by, see Ref. 18,

$$\langle\Psi|X\rangle = \sum_{m,n=0}^{k} \int \Psi^*_{\alpha_1\cdots\alpha_m} X_{\beta_1\cdots\beta_n} I^{\alpha_1\cdots\alpha_m,\beta_1\cdots\beta_n}\, d\mu, \tag{5.10}$$

where the $I$'s are arbitrary functions on $Q$ that are antisymmetric in both their $\alpha$ and $\beta$ indices. To get information on the allowed $I$'s, we will require that $\hat{\Omega}$ and all the physical observables can be made Hermitian with respect to (5.10). This is only possible if

$$(\eta^\alpha)^\dagger = \eta^\alpha, \tag{5.11a}$$

and,

$$(\hat{\rho}_\alpha)^\dagger = -\hat{\rho}_\alpha. \tag{5.11b}$$

Condition (5.11a) follows from the first term in the classical expression (2.6). The ghost and constraints will commute with each other in the quantum theory and so $\eta^\alpha$ must be Hermitian for this term to be Hermitian (which is required for $\hat{\Omega}$ to be Hermitian). Equation (5.11b) follows because the BRST extension of most physical observables contain terms of the form $\eta^\alpha \rho_\beta$ $(\alpha \neq \beta)$. The corresponding quantum operators (4.1) will anticommute and so (5.11b) must be satisfied if physical observables are to be Hermitian.

Equations (5.11) impose very restrictive conditions on the allowed $I$'s in (5.10) as we now show in the following two technical results.

**Theorem 5.2:**

$$(\eta^\alpha)^\dagger = \eta^\alpha \Leftrightarrow I^{\beta_1\cdots\beta_m\alpha\gamma_1\cdots\gamma_n} = (-1)^m I^{\beta_1\cdots\beta_m\alpha,\gamma_1\cdots\gamma_n}, \quad \forall m,n. \tag{5.12}$$

*Proof:* It is easy to show that, using (5.10),

$$\langle X|\eta^\alpha\Psi\rangle = \sum_{m,n=0}^{k} \int X^*_{\beta_1\cdots\beta_m}\Psi_{\gamma_1\cdots\gamma_n} I^{\beta_1\cdots\beta_m\alpha\gamma_1\cdots\gamma_n}\, d\mu,$$

and

$$\langle \eta^\alpha X|\Psi\rangle = \sum_{m,n=0}^{k} \int X^*_{\beta_1\cdots\beta_m}\Psi_{\gamma_1\cdots\gamma_n} \times I^{\beta_1\cdots\beta_m\alpha,\gamma_1\cdots\gamma_n}(-1)^m\, d\mu.$$

Therefore,

$$(\eta^\alpha)^\dagger = \eta^\alpha \Leftrightarrow \langle X|\eta^\alpha\Psi\rangle = \langle \eta^\alpha X|\Psi\rangle, \quad \forall X,\Psi$$
$$\Leftrightarrow I^{\beta_1\cdots\beta_m,\alpha\gamma_1\cdots\gamma_n}$$
$$= (-1)^m I^{\beta_1\cdots\beta_m\alpha,\gamma_1\cdots\gamma_n}, \quad \forall m,n.$$

**Theorem 5.3:** $(\eta^\alpha)^\dagger = \eta^\alpha$ and $(\hat\rho_\alpha)^\dagger = -\hat\rho_\alpha \Leftrightarrow$ the pairing (5.10) is maximal in the sense that the only $I$'s allowed are those with exactly $k$ indices.

*Proof:* The previous theorem excludes the possibility of $I$ with greater than $k$ indices (simply take all the indices to one side of the comma and then use antisymmetry). To eliminate the case of less than $k$ indices observe that

$$\langle\Psi|\hat\rho_\alpha X\rangle = -i\hbar \sum_{m,n=0}^{k} (n+1) \int \Psi^*_{\beta_1\cdots\beta_m} \times X_{\alpha\gamma_1\cdots\gamma_n} I^{\beta_1\cdots\beta_m,\gamma_1\cdots\gamma_n}\, d\mu,$$

and

$$\langle\hat\rho_\alpha\Psi|X\rangle = +i\hbar \sum_{m,n=0}^{k} (m+1) \times \int \Psi^*_{\alpha\beta_1\cdots\beta_m} I^{\beta_1\cdots\beta_m,\gamma_1\cdots\gamma_n}\, d\mu.$$

Thus

$$(\hat\rho_\alpha)^\dagger = -\hat\rho_\alpha \Leftrightarrow \langle\Psi|\hat\rho_\alpha X\rangle = -\langle\hat\rho_\alpha\Psi|X\rangle, \quad \forall\Psi,X$$
$$\Leftrightarrow m\delta^{[\beta_1}_\alpha I^{\beta_2\cdots\beta_m],\gamma_1\cdots\gamma_n}$$
$$= nI^{\beta_1\cdots\beta_m,[\gamma_2\cdots\gamma_n}\delta^{\gamma_1]}_\alpha, \quad \forall m,n, \quad (5.13)$$

where [ ] denotes antisymmetrization of the enclosed indices. After some manipulation, the above condition reduces to,

$$(\hat\rho_\alpha)^\dagger = -\hat\rho_\alpha \Leftrightarrow \sum_{h=1}^{m} (-1)^{h+1}\delta^{\beta_h}_\alpha I^{,\beta_1\cdots\hat\beta_h\cdots\beta_m\gamma_1\cdots\gamma_n}$$
$$= \sum_{h=1}^{n} (-1)^{m+h}\delta^{\gamma_h}_\alpha I^{,\beta_1\cdots\beta_m\gamma_1\cdots\hat\gamma_h\cdots\gamma_n}, \quad \forall m,n. \quad (5.14)$$

The notation $\beta_1\cdots\hat\beta_h\cdots\beta_m$ means that the hatted index $\hat\beta_h$ is excluded from the list. Now, Eq. (5.14) must be satisfied for all possible choices of the $\beta$'s and $\gamma$'s; thus let us take $\beta_1 = \alpha$ and all the other $\beta$'s and $\gamma$'s different from $\alpha$. With this choice (5.14) becomes

$$I^{,\beta_2\cdots\beta_m\gamma_1\cdots\gamma_n} = 0. \quad (5.15)$$

Now, if $m+n-1 \leqslant k-1$ (i.e., the number of indices on $I$ is less than $k$) it is possible to choose all of $\beta_2,\ldots,\gamma_n$ different and so (5.15) is a nontrivial equation. Thus if $I$ has less than $k$ indices, it must be zero. It is straightforward to show that, when $I$ has exactly $k$ indices, Eqs. (5.11) hold.[19] This concludes the proof.

Theorems (5.2) and (5.3) tell us that Hermiticity of the physical observables (and the BRST charge) is enough to force the Berezin measure upon us. This means that the full pairing on the states (4.2) is,

$$\langle\Psi|X\rangle = (i)^{(1/2)k(k-1)} \times \int \Psi^* X\, d\eta^1\cdots d\eta^k \|\varphi\|\, dQ^1\cdots dQ^N, \quad (5.16)$$

and we shall use this for the rest of our work.

### C. The "rigged Hilbert space"

We have seen that when ghosts are introduced into constrained systems there are two immediate kinematical problems to be resolved. The first is that the BRST invariant states are not in the Hilbert space upon which the BRST charge is defined. The second is that the pairing introduced above (which was forced upon us by quite general arguments) looks like it will miss the physical states and hence it is not clear how the physical states can be equipped with their expected pairing. What we shall find is that, by a judicious use of rigged Hilbert space theory, both of these problems can be given a satisfactory solution.

The necessity of a rigged Hilbert space formulation for operators with a continuous spectrum is well known but often ignored in most discussions of constrained systems. Before we discuss why this is so, let us first recap the idea of a rigged Hilbert space.[20]

The key observation is that when one is presented with a Hilbert space $H$, one must realize that $H$ is the completion of some space $S$, and hence there is a natural triplet of densely nested spaces,

$$S \subset H \subset S^*, \quad (5.17)$$

where $S^*$ is the dual space to $S$ (i.e., the space of continuous, linear functions on $S$). The rigged Hilbert space (5.17) is often called a Gel'fand triplet.

The paradigm example of the application of these ideas to quantum mechanics already occurs in one-dimensional systems where $H = L^2(\mathbb{R},\mathbb{C})$. The self-adjoint operator $\hat p = -i\,d/dx$, which is defined on the subset of $H$ consisting of all absolutely continuous functions in $H$ whose derivatives are square integrable, has no eigenfunction at all in $H$. The function normally identified with the eigenfunction $(e^{ipx})$ actually belongs to $S^*(\mathbb{R})$, where $S^*(\mathbb{R})$ is the dual to $S(\mathbb{R})$, the space of smooth functions on $\mathbb{R}$ whose derivatives all vanish faster than any power of the coordinates at infinity.

This example can easily be extended to a (trivialized) set of constraints $\hat p_\alpha$, where $S = S(Q)$ are now the smooth functions on the extended configuration space with the appropriate fall off rate. Then, in the Dirac analysis, the solutions to $\hat p_\alpha|\text{phys}\rangle = 0$ do not give the physical Hilbert space but rather $(S_{\text{phys}})^*$, the distributional space in the physical Gel'fand triplet. However, one does not worry about this since one can identify $S_{\text{phys}}$ as the dual to $(S_{\text{phys}})^*$, since these spaces are reflexive. Finally $H_{\text{phys}}$, the physical Hilbert space, is constructed via the completion of $S_{\text{phys}}$.

An implementation of the rigged Hilbert space ideas to the ghost formulation of constrained systems has been presented in Ref. 21. The philosophy was to put the different

elements of the Gel'fand triplet into different parts of (4.2). Care is needed in this construction, though, since one needs $\hat{\eta}$, $\hat{\rho}$, and $\hat{\Omega}$ to be defined on the resulting state space. Now, these operators shift the ghost number of the states and hence, if the coefficient spaces are not all the same, these operators will take us out of the state space. In Ref. 9 we used a modified form of Ref. 21, where $\hat{\Omega}$ and the physical observables were well defined though $\hat{\rho}$ was not.

We now introduce the ghost Gel'fand triplet. Let $H$ be the Hilbert space of $L^2$ functions on $Q$ (with respect to the $\|\varphi\|$ measure) and let $S$ be the nuclear space introduced above, so we have the extended rigged Hilbert space given by (5.17). We now use this triplet to define the coefficient spaces for the nested ghost triplet,

$$S_{\mathrm{BFV}} \subset H_{\mathrm{BFV}} \subset (S_{\mathrm{BFV}})^*, \tag{5.18}$$

where $\Psi$ belongs to an element of this triplet if it has an expansion (4.2) where all the coefficient functions are in the appropriate part of the Gel'fand triplet (5.17). Expression (5.18) is not itself a rigged Hilbert space since the pairing on $H_{\mathrm{BFV}}$ given by (5.16) is not positive definite.

All the operators introduced in this paper are defined on $H_{\mathrm{BFV}}$. Thus in order to define the mapping $\hat{\mathbf{R}}$ we need only concern ourselves with its action on $H_{\mathrm{BFV}}$. In the following paper (II) we shall show how the kinematical problems discussed above are resolved in this setup via the use of self-dual embeddings of the physical Hilbert space into $(S_{\mathrm{BFV}})^*$.

## VI. QUANTUM RESCALING OF THE CONSTRAINTS

We showed, in Theorem 2.2, that the classical rescaling of constraints is a canonical transformation in the super phase space. This result leads to the expectation that the rescaling transformation (2.14) will become a Hilbert space isomorphism in the quantum theory. We will show, in this section, that this is indeed the case. It is this transformation that enables the factor orderings, for the physical observables, to be worked out.

Note that, because of (5.5), the pairing (5.16) is going to change when a rescaling transformation is applied. To deal with this properly, it is necessary to regard the extended state spaces, before and after the transformation, as different (they contain the same states but have a different pairing). We will denote the initial state space by $H_{\mathrm{BFV}}$ and final state space by $(H_{\mathrm{BFV}})_R$. In addition, we denote the pairing on $H_{\mathrm{BFV}}$ by $\langle\Psi|X\rangle$ and the pairing on $(H_{\mathrm{BFV}})_R$ by $\langle\Psi|X\rangle_R$. With this notation the quantum version (2.14) will take the form of a bijective mapping $\hat{\mathbf{R}}$ from $H_{\mathrm{BFV}}$ to $(H_{\mathrm{BFV}})_R$ which satisfies

$$\hat{\mathbf{R}}Q^A\hat{\mathbf{R}}^{-1} = Q^A, \tag{6.1a}$$

$$\hat{\mathbf{R}}\hat{P}_A\hat{\mathbf{R}}^{-1} = \hat{P}_A + \tfrac{1}{2}(\Lambda^{-1})_\alpha{}^\beta{}_{,A}\Lambda_\gamma{}^\alpha(\eta^\gamma\hat{\rho}_\beta - \hat{\rho}_\beta\eta^\gamma), \tag{6.1b}$$

$$\hat{\mathbf{R}}\eta^\alpha\hat{\mathbf{R}}^{-1} = \Lambda_\beta{}^\alpha\eta^\beta, \tag{6.1c}$$

$$\hat{\mathbf{R}}\hat{\rho}_\alpha\hat{\mathbf{R}}^{-1} = (\Lambda^{-1})_\alpha{}^\beta\hat{\rho}_\beta, \tag{6.1d}$$

where $\hat{\mathbf{R}}^{-1}$ is the inverse mapping to $\hat{\mathbf{R}}$ and $_{,A}$ has its usual interpretation as a derivative with respect to $Q^A$. On the right-hand side of (6.1b) we have taken the commutator ordering of $\eta^\gamma\rho_\beta$. This is necessary because this term must be Hermitian. It is also important to note that, because of (5.5), the momentum operators are going to change under a rescaling. This means that, in (6.1b), the $\hat{P}_A$ on the left-hand side is constructed using the old $\|\varphi\|$, whereas $\hat{P}_A$ on the right-hand side is constructed using the new $\|\tilde{\varphi}\|$.

We also require that $\hat{\mathbf{R}}$ be norm preserving, i.e,. for any two $\Psi, X \in H_{\mathrm{BFV}}$,

$$\langle\hat{\mathbf{R}}\Psi|\hat{\mathbf{R}}X\rangle_R = \langle\Psi|X\rangle. \tag{6.2}$$

This is the generalization of a unitary mapping to a mapping between different Hilbert spaces.

An operator satisfying all the above requirements does exist. It is most easily defined by its action on an arbitrary state of the form (4.2),

$$\hat{\mathbf{R}}\Psi = \sum_{m=0}^{k}\Psi_{\alpha_1\cdots\alpha_m}\Lambda_{\beta_1}{}^{\alpha_1}\cdots\Lambda_{\beta_m}{}^{\alpha_m}\eta^{\beta_1}\cdots\eta^{\beta_m}. \tag{6.3}$$

The mapping $\hat{\mathbf{R}}^{-1}$ is given by a similar expression but with $\Lambda^{-1}$ replacing $\Lambda$. This result is central to our method for constructing the quantum observables, so we will spend most of this section proving that (6.3) satisfies all the above statements.

*Proof of (6.1):* (6.1a) is trivial. To prove (6.1b) let $\Psi$ be an arbitrary state of the form (4.2) and observe that

$$\hat{\mathbf{R}}\hat{P}_A\hat{\mathbf{R}}^{-1}\Psi = \hat{\mathbf{R}}\hat{P}_A\sum_{m=0}^{k}\Psi_{\alpha_1\cdots\alpha_m}(\Lambda^{-1})_{\beta_1}{}^{\alpha_1}\cdots(\Lambda^{-1})_{\beta_m}{}^{\alpha_m}\eta^{\beta_1}\cdots\eta^{\beta_m}$$

$$= \hat{\mathbf{R}}\Bigg\{\sum_{m=0}^{k}(\hat{P}_A\Psi_{\alpha_1\cdots\alpha_m})(\Lambda^{-1})_{\beta_1}{}^{\alpha_1}\cdots(\Lambda^{-1})_{\beta_m}{}^{\alpha_m}\eta^{\beta_1}\cdots\eta^{\beta_m} - i\hbar\sum_{m=0}^{k}\Psi_{\alpha_1\cdots\alpha_m}\Big[(\Lambda^{-1})^{\alpha_1}_{\beta_1,A}\cdots(\Lambda^{-1})_{\beta_m}{}^{\alpha_m}$$

$$+\cdots+(\Lambda^{-1})_{\beta_1}{}^{\alpha_1}\cdots(\Lambda^{-1})^{\alpha_m}_{\beta_m,A}\Big]\eta^{\beta_1}\cdots\eta^{\beta_m}\Bigg\}$$

$$= \hat{P}_A\Psi - i\hbar\hat{\mathbf{R}}\Bigg\{\sum_{m=0}^{k}\Psi_{\alpha_1\cdots\alpha_m}\Big[(\Lambda^{-1})_{\beta_1}{}^{\alpha_1}{}_{,A}\cdots(\Lambda^{-1})_{\beta_m}{}^{\alpha_m}+\cdots+(\Lambda^{-1})_{\beta_1}{}^{\alpha_1}\cdots(\Lambda^{-1})_{\beta_m}{}^{\alpha_m}{}_{,A}\Big]\eta^{\beta_1}\cdots\eta^{\beta_m}\Bigg\}.$$

This expression can be reduced, via straightforward but technical manipulations, to give the following result:

$$\hat{\mathbf{R}}\hat{P}_A\hat{\mathbf{R}}^{-1} = \hat{P}_A + (\Lambda^{-1})_\alpha{}^\beta{}_{,A}\Lambda_\gamma{}^\alpha\eta^\gamma\rho_\beta. \tag{6.4}$$

In the above equations it is important to realize that $\hat{P}_A$ is constructed, via (5.7), using the old $\|\varphi\|$. Thus written more fully (6.4) reads,

$$\hat{\mathbf{R}}\hat{P}_A\hat{\mathbf{R}}^{-1} = -i\hbar[\partial_A + \tfrac{1}{2}(\|\varphi\|_{,A}/\|\varphi\|)] + (\Lambda^{-1})_\alpha{}^\beta{}_{,A}\Lambda_\gamma{}^\alpha\eta^\gamma\hat{\rho}_\beta.$$

This equation can be rewritten in the form

$$\hat{\mathbf{R}}\hat{\mathbf{P}}_A\hat{\mathbf{R}}^{-1} = -i\hbar[\partial_A + \tfrac{1}{2}(\|\tilde{\varphi}\|_{,A}/\|\tilde{\varphi}\|)] + \tfrac{1}{2}(\Lambda^{-1})_\alpha{}^\beta{}_{,A}\Lambda_\gamma{}^\alpha(\eta^\gamma\hat{\rho}_\beta - \hat{\rho}_\beta\eta^\gamma),$$

which is (6.1b).

The proofs of (6.1c) and (6.1d) are similar so we will only give one of them. As before, let $\Psi$ be an arbitrary state of the form (4.2) and observe that

$$\begin{aligned}\hat{\mathbf{R}}\hat{\rho}_\alpha\hat{\mathbf{R}}\Psi &= \hat{\mathbf{R}}\hat{\rho}_\alpha\left\{\sum_{m=0}^{k}\Psi_{\alpha_1\cdots\alpha_m}(\Lambda^{-1})_{\beta_1}{}^{\alpha_1}\cdots(\Lambda^{-1})_{\beta_m}{}^{\alpha_m}\eta^{\beta_1}\cdots\eta^{\beta_m}\right\}\\ &= -i\hbar\hat{\mathbf{R}}\left\{\sum_{m=1}^{k} m\Psi_{\alpha_1\cdots\alpha_m}(\Lambda^{-1})_\alpha{}^{\alpha_1}\cdots(\Lambda^{-1})_{\beta_{m-1}}{}^{\alpha_m}\eta^{\beta_1}\cdots\eta^{\beta_{m-1}}\right\}\\ &= -i\hbar\sum_{m=1}^{k} m\Psi_{\alpha_1\cdots\alpha_m}(\Lambda^{-1})_\alpha{}^{\alpha_1}\eta^{\alpha_2}\cdots\eta^{\alpha_m}.\end{aligned}$$

It is straightforward to show that

$$(\Lambda^{-1})_\alpha{}^\beta\hat{\rho}_\beta\Psi = -i\hbar\sum_{m=1}^{k} m\Psi_{\alpha_1\cdots\alpha_m}(\Lambda^{-1})_\alpha{}^{\alpha_1}\eta^{\alpha_2}\cdots\eta^{\alpha_m},$$

and hence (6.1d) follows.

*Proof of (6.2):* Let $\Psi, X$ be arbitrary states of the form (4.2). Observe that

$$\begin{aligned}\langle\hat{\mathbf{R}}\Psi|\hat{\mathbf{R}}X\rangle_R &= (-i)^{(1/2)k(k-1)}\int\left\{\left(\sum_{m=0}^{k}\Psi^*_{\alpha_1\cdots\alpha_m}\Lambda_{\beta_1}{}^{\alpha_1}\cdots\Lambda_{\beta_m}{}^{\alpha_m}\right)\right.\\ &\quad\left.\times\left(\sum_{n=0}^{k}X_{\alpha_1\cdots\alpha_n}\Lambda_{\beta_1}{}^{\alpha_1}\cdots\Lambda_{\beta_n}{}^{\alpha_n}\right)d\eta^1\cdots d\eta^k\frac{\|\varphi\|}{|\Lambda|}dQ^1\cdots dQ^N\right\}\\ &= (-i)^{(1/2)k(k-1)}\sum_{m=0}^{k}\int\Psi^*_{[\alpha_1\cdots\alpha_m}X_{\alpha_{m+1}\cdots\alpha_k]}\frac{\Lambda_{\beta_1}{}^{\alpha_1}\cdots\Lambda_{\beta_k}{}^{\alpha_k}}{|\Lambda|}\\ &\quad\times\eta^{\beta_m}\cdots\eta^{\beta_1}\eta^{\beta_{m+1}}\cdots\eta^{\beta_k}d\eta^1\cdots d\eta^k\|\varphi\|dQ^1\cdots dQ^N,\end{aligned}$$

where [ ] denotes antisymmetrization of the enclosed indices. This equation reduces easily to

$$\begin{aligned}&\langle\hat{\mathbf{R}}\Psi|\hat{\mathbf{R}}X\rangle_R\\ &\quad= (-i)^{(1/2)k(k-1)}\\ &\qquad\times\sum_{m=0}^{k}\int\Psi^*_{[1\cdots m}X_{m+1\cdots k]}k!\eta^m\cdots\eta^1\eta^{m+1}\cdots\eta^k\\ &\qquad\times d\eta^1\cdots d\eta^k\|\varphi\|dQ^1\cdots dQ^N.\end{aligned}$$

By a similar method it follows that

$$\begin{aligned}\langle\Psi|X\rangle &= (-i)^{(1/2)k(k-1)}\\ &\quad\times\sum_{m=0}^{k}\int\Psi^*_{[1\cdots m}X_{m+1\cdots k]}k!\eta^m\cdots\eta^1\eta^{m+1}\cdots\eta^k\\ &\quad\times d\eta^1\cdots d\eta^k\|\varphi\|dQ^1\cdots dQ^N,\end{aligned}$$

and so (6.2) follows.

Thus we have proved that (6.3) does, indeed, have all the required properties. We will now show that it is the only operator with these properties, apart form trivial modifications.

**Theorem 6.1:** $\hat{\mathbf{R}}$, defined by (6.3), is the unique solution of Eqs. (6.1) and (6.2), apart from a constant phase factor.

*Proof:* Assume that $\hat{\mathbf{R}}$ satisfies (6.1) and (6.2). Then, Eqs. (6.1a) and (6.1b) can be written more conveniently as

$$\hat{\mathbf{R}}Q^A = Q^A\hat{\mathbf{R}}, \tag{6.5a}$$

and

$$\hat{\mathbf{R}}\eta^\alpha = \Lambda_\beta{}^\alpha\eta^\beta\hat{\mathbf{R}}. \tag{6.5b}$$

These relations can now be used to commute $\hat{\mathbf{R}}$ through (4.2), i.e.,

$$\hat{\mathbf{R}}\Psi = \sum_{m=0}^{k}\Psi_{\alpha_1\cdots\alpha_m}\Lambda_{\beta_1}{}^{\alpha_1}\cdots\Lambda_{\beta_m}{}^{\alpha_m}\eta^{\beta_1}\cdots\eta^{\beta_m}\hat{\mathbf{R}}(1), \tag{6.6}$$

where 1 is being thought of as the state with $\Psi_0 = 1$ and the other $\Psi$'s zero. To determine the value of $\hat{\mathbf{R}}(1)$ observe that, because of (6.1c) and (6.1d), $\hat{\mathbf{R}}$ commutes with the ghost number operator [see (4) in Sec. VII]. This means that $\hat{\mathbf{R}}(1)$ is, at most, a function of the $Q^A$'s. Let

$$\hat{\mathbf{R}}(1) = f(Q^A).$$

The condition (6.2) puts major restrictions on $f$ as we now show. Let $\Psi_k$ be an arbitrary state with ghost number $k$, i.e.,

$$\Psi_k = \Psi_{\alpha_1\cdots\alpha_k}\eta^{\alpha_1}\cdots\eta^{\alpha_k}.$$

Observe that

$$\langle\Psi_k|\hat{\mathbf{R}}(1)\rangle_R = k!\int\Psi^*_{1\cdots k}f(Q^A)\frac{\|\varphi\|}{|\Lambda|}dQ^1\cdots dQ^N,$$

and

$$\langle\hat{\mathbf{R}}^{-1}(\Psi_k)|1\rangle = k!\int\Psi^*_{1\cdots k}[\hat{\mathbf{R}}^{-1}(1)]^*\frac{\|\varphi\|}{|\Lambda|}dQ^1\cdots dQ^k.$$

These two expressions must be equal, for all $\Psi_k$, because of (6.2). We can thus conclude that

$$f(Q^A)^* = \hat{\mathbf{R}}^{-1}(1).$$

From which it follows that

$$|f(Q^A)|^2 = 1,$$

i.e., $f(Q^A)$ is a phase factor. We have already proved that $f = 1$ satisfies all the required conditions. From this it can be seen that $f$ must be a constant otherwise (6.1b) will not be satisfied. This completes the proof.

There is one final point that is worth making about the quantum rescaling operator. Compare (6.3) with the corresponding expression in Ref. 9. There the operator was exactly the same, apart from a rather surprising factor of $|\Lambda|^{-1/2}$. This was due to the fact that a Riemannian measure was being used on $Q$ and so, the pairing was invariant under rescaling transformations. Thus this pairing transformation property appears to be essential if the more natural expression (6.3) is to give the correct rescaling operator.

## VII. NOTES

(1) It might be argued that a degenerate kinetic energy term is the paradigm rather than a rarity. After all, in gauge theories, it is the inability to construct a free propagator that is the initial indication of constraints. However, this degeneracy is caused by the primary constraints that can easily be remedied by imposing a temporal gauge condition. Alternatively, one could reinstate the primary constraints using a weakly equivalent Hamiltonian. For a suitable choice of $H^\alpha$ this would correspond to a covariant gauge fixing.

(2) We use the convention that $\{\eta^\alpha, \rho_\beta\} = -\delta^\alpha{}_\beta$. The space of homogeneous functions containing $p$ ghosts and $q$ conjugate ghosts will be denoted by $\Gamma^{pq}$. A function is said to have ghost number $r$ (written $\in\Gamma^r$) if $p - q = r$.

(3) Note that we have changed notation from Ref. 9. The $\|\varphi\|$ defined in (5.2) is the reciprocal of the corresponding quantity defined in Ref. 9.

(4) The ghost number operator is

$$\hat{g} = (i/\hbar)\eta^\alpha \hat{\rho}_\alpha.$$

The eigenspace of $\hat{g}$, corresponding to eigenvalue $r$, is $\Gamma^r$.


[1] L. Van Hove, Acad. R. Belg. Cl. Sci. Mem. Coll. **36**, 61 (1951).
[2] P. A. M. Dirac, *Lectures in Quantum Mechanics* (Yeshiva University, New York, 1964), and earlier papers by Dirac quoted therein.
[3] L. D. Faddeev, "Introduction to functional methods," in *Methods in Field Theory*, edited by R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1975).
[4] J. Anderson, in *Eastern Theoretical Physics Conference*, edited by M. E. Rose (Gordon and Breach, New York, 1963).
[5] K. Kuchar, Phys. Rev. D **35**, 596 (1987).
[6] K. Kuchar, Phys. Rev. D **34**, 3031 (1986).
[7] K. Kuchar, Phys. Rev. D **34**, 3044 (1986).
[8] W. Pauli, in *Handbuch der Physik Bd. XXIV, Teil 1*, edited by H. Geiger and K. Sheel (Springer, Berlin, 1933).
[9] D. McMullan and J. Paterson, Phys. Lett. B **202**, 358 (1988).
[10] D. McMullan and J. Paterson, J. Math. Phys. **30**, 487 (1989).
[11] E. S. Fradkin and G. A. Vilkovisky, CERN Report TH-2332, 1977, unpublished; I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).
[12] R. P. Feynman, Acta Phys. Pol. **24**, 697 (1963); B. S. DeWitt, Phys. Rev. **162**, 1195 (1967); L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967).
[13] D. McMullan, J. Math. Phys. **28**, 428 (1987).
[14] A. D. Browning and D. McMullan, J. Math. Phys. **28**, 438 (1987).
[15] M. Henneaux, Phys. Rep. **126**, 1 (1985).
[16] D. McMullan, in *The Physics of Phase Space*, edited by Y. S. Kim and W. W. Zachary (Springer, Berlin, 1987), p. 380.
[17] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[18] D. McMullan, Phys. Rev. D **33**, 2501 (1986).
[19] J. Paterson, Ph.D. thesis, University of Glasgow 1989.
[20] A. Bohm, *The Rigged Hilbert Space and Quantum Mechanics* (Springer, Berlin, 1978).
[21] P. Thomi, University of Bern preprint BUTP-87/16, 1987.

# Covariant factor ordering of gauge systems using ghost variables. II. States and observables

D. McMullan[a] and J. Paterson[b]
*Department of Physics and Astronomy, The University of Glasgow, Glasgow G12 8QQ, Scotland*

(Received 22 June 1988; accepted for publication 28 September 1988)

The aim of any constraint quantization procedure is to be able to recover the physical content of a system after quantizing on the extended state space (which will include both physical and unphysical configurations). Within the Dirac approach to such problems there is no guarantee that the resulting quantum theory will be equivalent to the quantized physical theory and one has to be content with a case by case analysis of its applicability. In this paper it is shown that for finite-dimensional systems with first class constraints linear in momenta, invariance under constraint rescaling and point transformations is sufficient to ensure a consistent quantization. As discussed in the preceding paper [J. Math. Phys. **30**, 477 (1989)], in order to become a manifest symmetry, constraint rescaling requires ghost variables. It is now shown how the associated Becchi–Rouet–Stora–Tyutin (BRST) charge is constructed and how it is used to describe the physical states and observables in the constrained system. The extension of this construction to more general constrained systems is also discussed.

## I. INTRODUCTION

In the preceding paper[1] [hereafter referred to as (I)] we presented a detailed discussion of the new kinematical structures present in constrained systems when ghost variables are used. We found that the introduction of these fermionic degrees of freedom enlarged the class of (even) canonical transformations, thus allowing for the construction of a generating function implementing constraint rescaling. For systems described by constraints linear in momenta we were able to elevate this invariance to the quantum theory. We now want to use these results to show how the constrained dynamics can be incorporated into such theories.

The attraction to a formalism that naturally allows for the rescaling of the constraints reflects the fact that a constrained system is kinematically characterized by a submanifold $C$ in an extended phase space $P$. For a first class system $C$ is coisotropic (see, for example, Ref. 2 for definitions from symplectic geometry). Associated with $C$ is the ring of functions $V(C)$ consisting of those smooth functions on $P$ that vanish when restricted to $C$. Now if $C$ is a smooth submanifold we can expect [see (1) in Sec. VI] to find $k$ functions $\varphi_\alpha$ (where $k$ is the codimension of $C$ in $P$) such that $C$ is given by the zero set of these constraints. These functions will then generate the ideal $V(C)$. Obviously, if $\Lambda_\alpha{}^\beta$ is invertible then $\tilde{\varphi}_\alpha = \Lambda_\alpha{}^\beta \varphi_\beta$ is an equally good choice of generators for this ideal. Thus we see that the ability to rescale constraints supports the view that it is the manifold $C$ that should directly enter into the quantization and not a particular parametrization of it.

The introduction of ghost variables allows for a classical constraint formalism that is independent of the basis of $V(C)$.[1,3] Indeed, one can formulate the homological construction implicit in these methods without ever introducing a basis of constraints.[4,5] However, in the standard approaches to quantization these geometric methods are, at best, difficult to implement. One might thus question the need for such aesthetic considerations.

In (I) we showed that a limited class of rescalings can be elevated to the quantum theory when the constraints are linear in momenta, where the rescalings being considered were those that preserved this particular momenta dependence. We now claim that for such systems this is a sufficient amount of covariance to ensure a consistent quantization. This is because one can now make use of the local trivialization results discussed in (I) to make manifest the local consistency of the procedure. In this way we shall show that the use of ghost variables avoids the problems found in the Dirac approach.[6]

The reliance on local trivialization does, however, open the door to the possibilities of global obstructions to these results (although not in our solution to the example discussed in Ref. 6). Within our approach, global complications can arise in both the characterization of physical states (which is cohomological) and the description of physical observables. Nontrivial global structures can be incorporated into the isolation of states via more restrictive prescriptions, i.e., restriction on the state's ghost number. For observables, though, one has to be more careful since, as well as giving a factor ordering that is Hermitian, one also needs to give the domain upon which the operator is self-adjoint and this *cannot* be done locally since often we will be dealing with unbounded operators.

Indeed, if one is going to take global structures at all seriously then the very nature of the quantization needs to be looked at more critically since the use of, say, the Schrödinger picture might be wholly inappropriate. We shall, therefore, just content ourselves with commenting on possible global effects. However, even for globally trivializable systems the results presented in this paper will be an improvement on the Dirac approach.

Since nature has not been benevolent enough to allow us

[a] Present address: Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom.
[b] Supported by the Science and Engineering Research Council.

 0022-2488/89/020487-11$02.50 © 1989 American Institute of Physics 

to restrict our attention exclusively to theories with linear (in momenta) constraints, any constraint formalism must allow for an extension to more general systems. Even in this finite-dimensional quantum mechanical context, theories with a more complicated constraint structure are very difficult to analyze. In particular, ordering prescriptions for quadratic functions are notoriously difficult to find. So, an interesting first step towards the study of such constrained systems would be to allow for more general rescalings of the linear constraints. However, in the conventional quantization procedures, even this extension is fraught with difficulties. Hence in order to develop some insight into such systems it is therefore useful to work within the prequantization stage of geometric quantization. The potential usefulness for such a scheme in the study of constrained systems has been advocated recently by several authors.[7,8] Hence we shall discuss the possible prequantization extensions of the constraint rescaling construction and, in particular, the relevance of this to the choice of constraint polarization.

The plan of this paper is as follows. In Sec. II we shall analyze in detail the characterization of physical states within the BRST formalism. Then, in Sec. III we shall address the dynamical problems and construct representations for the physical observables of the theory. Section IV will contain a discussion of the quantized Koszul complex that is used in the introduction of antighosts and the prequantized description of states. In Sec. V we shall discuss the possible extensions of these results to more general constrained theories. Finally, in the Appendix, we shall present our solution to the "quantum well of Orvieto."

## II. THE BRST CHARACTERIZATION OF PHYSICAL STATES

### A. The ghost philosophy

The need for a constraint formalism arose from the desire to avoid the classical reduction to the physical configurations. Thus the first task facing any constraint method is to give a prescription for recovering the physical results.

In the Dirac approach the physical states are characterized as those annihilated by the constraints. This is a very strong condition and reflects the belief that the physical states should be "gauge invariant." However, the actual implementation of this condition can be as involved as the original classical problem of reducing to the true degrees of freedom (this is especially true if structure functions are present) and, even then, one is not guaranteed to recover the expected physics.

Ghost variables allow for a weaker prescription for the recovery of physical results which will, in turn, guarantee consistency with the classically reduced quantization. In this approach one restricts attention to the BRST invariant states, i.e., those which satisfy

$$\hat{\Omega}\Psi = 0. \tag{2.1}$$

Such states are *not* directly equivalent to the physical states. In particular, since $\hat{\Omega}$ is taken to be nilpotent, any state of the form $\hat{\Omega}\chi$ satisfies (2.1). Since $\chi$ is arbitrary it is, at first, hard to see any remnant of the physical states in this construction.

The important additional condition that allows us to systematically ignore such problematic states is that $\hat{\Omega}$ is taken to be self-adjoint. From this we can deduce that states of the form $\hat{\Omega}\chi$ are "perpendicular" to the BRST invariant states and, in particular, they have zero norm. Hence in any particular calculation, we do not need to worry about them. So in order to see that the BRST methods can recover the physical states, we must show how the solutions to (2.1), not of the form $\hat{\Omega}\chi$, are directly related to the physical states.

It is clear that the discussion given above is cohomological in nature. What we want is a description of the cohomology groups of the complex described by (2.1). On top of this homological objective, we must also ensure that an expression for $\hat{\Omega}$ can be constructed that allows for such a description. Thus we must find an ordering for the terms in $\hat{\Omega}$ satisfying the following four conditions: (1) $\hat{\Omega}^\dagger = \hat{\Omega}$, (2) $\hat{\Omega}^2 = 0$, (3) $\hat{\Omega}$ determines the correct physical states, and (4) $\hat{\Omega}$ is covariant with respect to all the symmetries (a)–(d) of (I). Such an ordering for $\hat{\Omega}$ will be derived in Secs. II B and II C and the nature of solutions to (2.1) will be discussed.

### B. The solution in the trivialized coordinate system

In keeping with our general philosophy of first trivializing the constraints, solving the problem in that simple setup, and then rescaling back to the original system, we shall now give the solution to the BRST description of physical states in a trivialized system.

The pairing I(5.16) and the momentum operators I(5.7) take on a simple form when the constraints are the first $k$ momenta. These are

$$\langle \Psi | X \rangle = (i)^{(1/2)k(k-1)} \int \Psi^* X \, d\eta^1 \cdots d\eta^k |g|^{1/2} \, dQ^1 \cdots dQ^N \tag{2.2}$$

and

$$\hat{P}_\alpha = -i\hbar\partial_\alpha \quad (\alpha = 1,...,k), \tag{2.3a}$$

$$\hat{P}_a = -i\hbar\left[\partial_a + \tfrac{1}{2}(|g|^{1/2}{}_{,a}/|g|^{1/2})\right] \quad (a = k+1,...,N), \tag{2.3b}$$

where $|g|$ is the determinant of the physical metric. Another simplification of the trivialized system is the absence of any ordering ambiguities in the BRST charge. We can thus write

$$\hat{\Omega} = \hat{P}_\alpha \eta^\alpha = -i\hbar\partial_\alpha \eta^\alpha, \tag{2.4}$$

and study the solutions to Eq. (2.1) with this expression for $\hat{\Omega}$.

As pointed out in (I), the correct space upon which the solutions to (2.4) should be studied is $(S_{\mathrm{BFV}})^*$, the distributional part of the ghost triplet [see Sec. V C in (I)]. If we let $\Gamma^{*r}$ denote ghost number $r$ states in $(S_{\mathrm{BFV}})^*$, then $(\Gamma^*, \hat{\Omega})$ defines a complex given by

$$\Gamma^{*0} \xrightarrow{\hat{\Omega}} \Gamma^{*1} \xrightarrow{\hat{\Omega}} \cdots \xrightarrow{\hat{\Omega}} \Gamma^{*k}. \tag{2.5}$$

This complex is analogous to the de Rham (DR) complex familiar in differential geometry. If the coefficient space of distributions on $Q$ was replaced by smooth functions then it is straightforward to read off the cohomology. One finds

that the only nonvanishing cohomology group is $H^0{}_{DR}$, which describes those smooth functions on $Q$ that are constant in the unphysical directions. Clearly $C^\infty(Q) \subset S^*(Q)$ and hence we would expect at least an injection of the de Rham cohomology into the distributional complex given above. In fact, the cohomologies are isomorphic (see, for example, Chap. 3 in Ref. 9). Thus we can deduce that the only nonvanishing cohomology in the complex (2.5) is $H^0$, which, for these trivialized constraints, are those elements of $(S_{BFV})^*$ with ghost number zero and no dependence on $Q^1$, $Q^2$,...,$Q^k$. Thus $H^0$ is isomorphic to $(S_{phy})^*$, the distributional part of the physical Gel'fand triplet. As discussed in (I), given $(S_{phy})^*$ one can construct the full physical Gel'fand triplet and hence the physical state space.

In summary, for the trivialized set of constraints we can completely solve Eq. (2.1). What we find is those solutions not of the form $\hat{\Omega}\chi$ are directly related to the physical states. This is an ideal situation to be in but one which we do not expect to hold for more complicated constraints since, in general, there will be nonvanishing higher cohomology groups. In such a case we would need to restrict attention to the solutions of (2.1) with zero ghost number. We shall return to this point in Sec. IV. So, for the rest of this discussion we shall assume that we are given a globally trivializable set of constraints.

We have seen that invariance under the BRST transformation allows us to identify the physical states of the theory (up to coboundaries). So, we can consider $(S_{phy})^*$, and hence $H_{phy}$, as being embedded in $(S_{BFV})^*$. Then condition (2.1) just picks out this physical subset. However, there are *many* possible embeddings of $H_{phy}$ into $(S_{BFV})^*$ and we now need to discuss which embedding we should take.

There seems to be an obvious way to embed $H_{phy}$ into $(S_{BFV})^*$. Simply let $\Psi_{phy} \in H_{phy}$ be represented by $\Psi = \Psi_0 = \Psi_{phy}$ [a ghost number zero element of $(S_{BFV})^*$]. Clearly $\hat{\Omega}\Psi = 0$ and hence we have apparently solved the above four conditions on the isolation of physical states for this trivialized system. There is, though, a serious problem with this identification since, using (2.2), any state of this form will have zero norm. This is simply a consequence of the Berezin form for the ghost measure which will not pair a ghost number zero state with itself in a nontrivial way.

This is in marked contrast to the similar normalization problem that arises in the Dirac approach, where one finds that the physical states have an infinite norm due to the fact that they are not square integrable over the extended configuration space. What we shall do now is to show that between the Berezin result of zero and the Dirac result of infinity there is a halfway house where physical results can be calculated.

The problem in the above construction is that we only considered the set theoretic inclusion of $H_{phy}$ into $(S_{BFV})^*$, whereas, what is needed is a full Hilbert space embedding. This obviously raises two immediate questions. First, how do we find such an inclusion and, second, what will it mean since $(S_{BFV})^*$ is *not* itself a Hilbert space. Let us first address the problem of how subspaces of $(S_{BFV})^*$ can be given a traditional Hilbert space structure, then we shall introduce the embedding needed to solve the above problems.

As we have seen, the space $H_{BFV} \subset (S_{BFV})^*$, with the Berezin measure, is not really a Hilbert space. The reason being that it always pairs a ghost number $r$ state with one of ghost number $k - r$. A similar structure is seen in the study of differential forms on a $k$-dimensional orientable manifold, where one can use the volume element to pair $r$ forms with $k - r$ forms. There one is able to construct an inner product on the space of $r$ forms by pairing it with its dual.

Let us now proceed in a similar manner by defining the dual mapping on $H_{BFV}$. Let $\Psi \in H_{BFV}$ have ghost number $r$; we define its dual $\Psi'$ via

$$\Psi' = \Psi'_{\alpha_{r+1}\cdots\alpha_k}\eta^{\alpha_{r+1}}\cdots\eta^{\alpha_k},$$

where

$$\Psi'_{\alpha_{r+1}\cdots\alpha_k} = [(i)^{(3/2)k(k-1)}/r!]\Psi_{\alpha_1\cdots\alpha_r}\epsilon^{\alpha_k\cdots\alpha_{r+1}\alpha_1\cdots\alpha_r}. \tag{2.6}$$

We can now define an inner product $(\ ,\ )$ on the ghost number $r$ elements of $H_{BFV}$ by $(\Psi_1, \Psi_2) := \langle \Psi'_1, \Psi_2 \rangle$. Duality on a mixed ghost number state is then defined by taking the dual of the separate ghost number terms. Obviously, if one has a self-dual state then its Berezin pairing will give its norm. So, if $\Psi_0$ is a zero ghost number element of $H_{BFV}$ its dual has coefficients

$$\Psi_{k\cdots 1} = (i)^{(3/2)k(k-1)}\Psi_0. \tag{2.7}$$

We can thus construct a self-dual state $\Psi$ by adding this (ghost number $k$ state) to $\Psi_0$. The norm of the resulting state will be precisely the norm of $\Psi_0$ considered as a square integrable function on $Q$ (up to a normalization constant).

This is all well and good, but we really want to introduce a Hilbert space structure on the solutions to (2.1). Hence we need to extend the duality and pairing to such states. So, if $\Psi_0 \in H_{phy}$, we define its (distributional) dual in $(S_{BFV})^*$ by

$$\Psi_{k\cdots 1} = (i)^{(3/2)k(k-1)}\delta(Q^1)\cdots\delta(Q^k)\Psi_0. \tag{2.8}$$

Hence we can embed $H_{phys}$ into $(S_{BFV})^*$ to construct the self-dual state $\Psi$, where

$$\Psi = \Psi_0 + k!(i)^{(3/2)k(k-1)}\delta(Q^1)\cdots\delta(Q^k)\Psi_0\eta^k\cdots\eta^1. \tag{2.9}$$

We can now extend the Berezin pairing (2.2) to include such states, e.g., if $\Psi_0, X_0 \in H_{phys}$ the associated elements in $(S_{BFV})^*$ pair to give

$$\langle \Psi | X \rangle = 2(k!)\int \Psi_0^* X_0 |g|^{1/2} dQ^{k+1}\cdots dQ^N, \tag{2.10}$$

which is, up to a normalization, the desired physical result. Thus the self-dual solutions to (2.1) correctly characterize the physical states of the system.

It might be thought that this elaborate definition of self-dual embeddings of physical states into $(S_{BFV})^*$ has not solved anything since the ghost number $k$ part of the self-dual state is also a solution to (2.1). Hence by the vanishing cohomology argument, this term must be of the form $\hat{\Omega}\chi$ for some $\chi \in \Gamma^{*k-1}$ (indeed $\chi$ is easy to write down and will involve step functions). Therefore, such a term will give zero when paired with $\Psi_0$. This argument is false, though, because $\hat{\Omega}$ is *not* in general self-adjoint on $(S_{BFV})^*$ and, in

particular, for the term given above $\chi$ gives a surface contribution that destroys the vanishing norm argument; we shall return to this important point in Sec. IV.

In conclusion, for the trivialized set of constraints we have been forced to extend the Berezin pairing to allow a restricted class of distributional elements, on $Q$, to be paired. In doing this, care must be taken to ensure that the Hermiticity properties of all the basic operators still hold. In particular, the momentum operators will be Hermitian only if the distributions vanish at infinity. This requirement is satisfied by the self-dual states (2.9). In general, (2.9) will also be supplemented by states of the form $\hat{\Omega}\chi$. In order for these states to decouple we must impose the condition that $\chi$ vanishes at infinity (i.e., $\hat{\Omega}$ must be self-adjoint on these states).

This completes the solution in the trivialized coordinate system.

### C. The general solution

To turn the solution of the previous section into a fully covariant one there are two steps. First, keeping the constraints Abelian, $\hat{\Omega}$ and the self-dual condition must be written in a coordinate covariant manner. Second, the coordinate covariant solution must be boosted, using the operator $\hat{\mathbf{R}}$ of (I), to give a covariant solution for non-Abelian constraints. These two steps will now be carried out.

The first step is almost trivial. In a general coordinate system the Abelian constraints will be of the form

$$\tilde{\varphi}_\alpha = Q_\alpha{}^A \tilde{P}_A,$$

and $\hat{\tilde{\Omega}}$ becomes

$$\hat{\tilde{\Omega}} = \tfrac{1}{2}[Q_\alpha{}^A, \hat{\tilde{P}}_A]_+ \tilde{\eta}^\alpha.$$

In a similar manner the duality condition becomes

$$\Psi_{k\cdots 1} = (i)^{(3/2)k(k-1)}\delta(X^1)\cdots\delta(X^k)\Psi_0,$$

where $X^\alpha = Q^\alpha$ (of the trivialized coordinate system) is now being thought of as a general gauge fixing condition.

The final step to non-Abelian constraints proceeds straightforwardly to give

$$\hat{\Omega} = \hat{\mathbf{R}}^{\frac{1}{2}}[Q_\alpha{}^A, \hat{\tilde{P}}_A]_+ \tilde{\eta}^\alpha \hat{\mathbf{R}}^{-1},$$

which expands out, using I(6.1), to give

$$\hat{\Omega} = \tfrac{1}{2}[\varphi_\alpha{}^A, \hat{P}_A]_+ \eta^\alpha + \tfrac{1}{4}C^\alpha{}_{\beta\gamma}(\eta^\gamma \hat{\rho}_\alpha \eta^\beta - \eta^\beta \hat{\rho}_\alpha \eta^\gamma), \tag{2.11}$$

for a general set of non-Abelian constraints $\varphi_\alpha = \varphi_\alpha{}^A P_A$. This expression is fully covarient under coordinate and rescaling transformations and, by its method of construction, will project the correct physical states.

Equation (2.11) can also be obtained directly from the classical expression I(2.6) by parametrizing all possible orderings (e.g., the quantization of $\widehat{QP}$ would be $\epsilon_1 \hat{Q}\hat{P} + \epsilon_2 \hat{P}\hat{Q}$, where $\epsilon_1 + \epsilon_2 = 1$) and then demanding Hermiticity and nilpotency, etc., to fix the parameters.[10] This method has the advantage of showing that (2.11) is the unique ordering of I(2.6) which satisfies Hermiticity, nilpotency, and coordinate covariance.

The duality condition is boosted, using $\hat{\mathbf{R}}$, in the following way. Let $\Psi$ be a self-dual state for Abelian constraints, i.e.,

$$\Psi = \Psi_0 + k!(i)^{(3/2)k(k-1)}\delta(X^1)\cdots\delta(X^k)\Psi_0\tilde{\eta}^k\cdots\tilde{\eta}^1,$$

and observe that

$$\hat{\mathbf{R}}\Psi = \Psi_0 + k!(i)^{(3/2)k(k-1)}\delta(X^1)\cdots\delta(X^k)\Psi_0|\Lambda|\eta^k\cdots\eta^1.$$

To write this in a coordinate covariant manner we use the following result:

$$|\Lambda| = \det[\{X^\alpha, \varphi_\beta\}] = |\{X^\alpha, \varphi_\beta\}|.$$

This alternative form of $|\Lambda|$ is coordinate covariant. Thus the general form of the duality condition is

$$\Psi = \Psi_0 + k!(i)^{(3/2)k(k-1)}\delta(X^1)\cdots\delta(X^k)\Psi_0 \times|\{X^\alpha, \varphi_\beta\}|\eta^k\cdots\eta^1. \tag{2.12}$$

It is interesting to note that the term

$$\delta(X^1)\cdots\delta(X^k)|\{X^\alpha, \varphi_\beta\}|, \tag{2.13}$$

which enters into the duality condition, is familiar from the phase space path integral description of constrained systems,[11] where it enters as a modification to the measure. By standard arguments it is easy to show that (2.13) is invariant under infinitesimal changes to the gauge fixing conditions. In this operator approach we are forced to put this term into the duality condition, as opposed to the measure, since we require the BRST charge to be self-adjoint with respect to the Berezin measure introduced in I(5.16).

To summarize, Eqs. (2.11) and (2.12) are the solution to the kinematic aspects of constraint quantization. Together they project out the space of BRST invariant, self-dual states. When endowed with the pairing I(5.16) these states are isomorphic to the rigged Hilbert space of physical states. We will now discuss the dynamical parts of constraint quantization.

## III. THE QUANTUM OBSERVABLES

### A. The general approach

The task now is to find ordering prescriptions for all the special physical observables. These orderings must, of course, give the correct physical results [i.e., those given in Sec. III of (I)] when applied to the BRST invariant, self-dual states. In addition, if $\hat{\mathbf{F}}$ is a general physical observable, we require it to satisfy the following: (1) $[\hat{\mathbf{F}}, \hat{\Omega}] = 0$, (2) $\hat{\mathbf{F}}^\dagger = \hat{\mathbf{F}}$; and (3) $\hat{\mathbf{F}}$ is covariant under all the symmetries (a)–(d).

Condition (1) is essential for consistency of the theory. If it failed to be true $\hat{\mathbf{F}}$ would map physical states to unphysical states. For similar reasons, it is necessary for $\hat{\mathbf{F}}$ to preserve the self-dual condition, at least up to zero norm states.

Conditions (2) and (3) are not logically necessary as Kuchar pointed out.[12,13] It is only strictly necessary for $\hat{\mathbf{F}}$ to be Hermitian on physical states and covariant under symmetries that are lifts of symmetries from the true degrees of freedom. Having the full properties (2) and (3) is, nonetheless, very convenient and ghost methods allow them to be achieved.

As with the classical observables, the quantum observables have an equivalence class structure where $\hat{\mathbf{F}}$ and $\hat{\mathbf{G}}$ are equivalent if they differ by a coboundary term, i.e.,

$$\hat{\mathbf{F}} = \hat{\mathbf{G}} + [\hat{\mathbf{K}}, \hat{\mathbf{\Omega}}], \tag{3.1}$$

where $\hat{\mathbf{K}}$ is some ghost number minus one operator. Such coboundaries never contribute to any physical result, i.e., they vanish when paired with BRST invariant states. This equivalence class structure is consistent with the commutator algebra, i.e., if

$$\hat{\mathbf{F}} = \hat{\mathbf{F}}' + [\hat{\mathbf{A}}, \hat{\mathbf{\Omega}}],$$

and

$$\hat{\mathbf{G}} = \hat{\mathbf{G}}' + [\hat{\mathbf{B}}, \hat{\mathbf{\Omega}}],$$

then

$$[\hat{\mathbf{F}}, \hat{\mathbf{G}}] = [\hat{\mathbf{F}}', \hat{\mathbf{G}}'] + [\hat{\mathbf{C}}, \hat{\mathbf{\Omega}}],$$

for some $\hat{\mathbf{C}}$. This is an important point to which we shall return shortly.

The orderings of the physical observables will be obtained via the general strategy used in the discussion of the states, i.e., we start in the trivialized coordinate system where the constraints are pure momenta. In this coordinate system the special observables $\mathbf{F}$ can easily be split into their physical part plus a coboundary, i.e.,

$$\mathbf{F} = \mathbf{F}_{\text{phy}} + \{\mathbf{F}', \mathbf{\Omega}\}. \tag{3.2}$$

The quantization procedure is then

$$\mathbf{F}_{\text{phy}} \to \hat{\mathbf{F}}_{\text{phy}}, \tag{3.3a}$$

$$\{\mathbf{F}', \mathbf{\Omega}\} \to (-i/\hbar)[\hat{\mathbf{F}}', \hat{\mathbf{\Omega}}], \tag{3.3b}$$

where $\hat{\mathbf{F}}_{\text{phy}}$ is the required operator taken from I(3.4) and $\hat{\mathbf{F}}'$ is ordered to be anti-Hermitian. There are no Van Hove obstructions to (3.3b) because $\mathbf{F}'$ and $\mathbf{\Omega}$ are, at most, linear in the momenta. This is the reason why the unphysical directions do not produce extra Van Hove type obstructions (it is interesting to note that the above argument would fail for quadratic constraints).

The ordering (3.3) automatically guarantees $\hat{\mathbf{F}}$ to be Hermitian, commute with the BRST charge, preserve the self-dual condition, and give the required answers when paired with physical states. Also, because the equivalence class structure (3.1) is preserved by the commutator algebra, this prescription guarantees an implementation of the physical Poisson algebra as a commutator algebra up to unavoidable physical Van Hove obstructions.

The final step that gives $\hat{\mathbf{F}}$ in its general form, i.e., for non-Abelian constraints, is achieved by boosting (3.3) with the rescaling operator $\hat{\mathbf{R}}$.

There does not appear to be any way of guaranteeing that (3.3) will give an ordering covariant under all the symmetries (a)–(d). Fortunately, when the calculations are done for the special observables, the orderings are found to be expressible in a covariant form. This step depends crucially on the existence of $\|\varphi\|$ and its three properties I(5.3)–I(5.5).

The calculations will now be presented for each of the special physical observables.

### B. Configuration observables

Let $Y(Q^A)$ be a general physical, configuration observable. The function $Y$, being physical, must satisfy

$$\{Y, \varphi_\alpha\} = Y_\alpha{}^\beta \varphi_\beta, \tag{3.4}$$

for some $Y_\alpha{}^\beta$. However, the left-hand side of (3.4) depends only on the $Q^A$'s whereas the right-hand side has momentum dependence. Thus, to avoid contradiction, $Y_\alpha{}^\beta$ must be zero, i.e., physical configuration observables must satisfy

$$\{Y, \varphi_\alpha\} = 0. \tag{3.5}$$

This means that the BRST extension of $Y$ is trivial, i.e.,

$$\mathbf{Y} = Y. \tag{3.6}$$

In the trivialized coordinate system (3.5) implies that

$$\mathbf{Y} = Y(Q^{k+1}, \ldots, Q^N), \tag{3.7}$$

and so, following (3.3),

$$\hat{\mathbf{Y}} = \mathbf{Y} \tag{3.8}$$

will give the correct physical results. The term $\hat{\mathbf{R}}$ will not alter the ordering because of I(6.1a). Thus the quantization of configuration variables is trivial. Equation (3.8) is true in all coordinate systems and for all choices of basis constraints.

### C. Linear observables

Let $U = U^A(Q^B)P_A$ be a general linear, physical observable. Then, it will satisfy

$$\{U, \varphi_\alpha\} = U_\alpha{}^\beta \varphi_\beta. \tag{3.9}$$

The BRST extension of $U$ is

$$\mathbf{U} = U + U_\alpha{}^\beta \eta^\alpha \rho_\beta. \tag{3.10}$$

In the trivialized coordinate system this takes the simpler form

$$\begin{aligned}\hat{\mathbf{U}} &= U^A \tilde{P}_A + U^\beta{}_{,\alpha} \tilde{\eta}^\alpha \tilde{\rho}_\beta \\ &= U^a \tilde{P}_a - \{U^\beta \tilde{\rho}_\beta, \tilde{\mathbf{\Omega}}\}.\end{aligned} \tag{3.11}$$

In this equation capital latin indices range from $1,\ldots,N$; lowercase latin indices range from $k+1,\ldots,N$; and greek indices range from $1,\ldots,k$. Equation (3.11) is now in the form (3.2) and so (3.3) can be applied to give

$$\hat{\tilde{\mathbf{U}}} = \tfrac{1}{2}[U^a, \hat{\tilde{P}}_a]_+ + (i/\hbar)[U^\beta \hat{\tilde{\rho}}_\beta, \hat{\tilde{\mathbf{\Omega}}}]. \tag{3.12}$$

This expression expands out to give

$$\hat{\tilde{\mathbf{U}}} = \tfrac{1}{2}[U^A, \hat{\tilde{P}}_A]_+ + \tfrac{1}{2}U^\beta{}_{,\alpha}(\tilde{\eta}^\alpha \hat{\tilde{\rho}}_\beta - \hat{\tilde{\rho}}_\beta \tilde{\eta}^\alpha). \tag{3.13}$$

This form of $\hat{\tilde{\mathbf{U}}}$ is fully coordinate covariant but is restricted to Abelian constraints. The boosting to non-Abelian con straints, i.e., the computation of $\hat{\mathbf{R}}\hat{\tilde{\mathbf{U}}}\hat{\mathbf{R}}^{-1}$, is straightforward and gives

$$\hat{\mathbf{U}} = \tfrac{1}{2}[U^A, \hat{P}_A]_+ + \tfrac{1}{2}U_\alpha{}^\beta(\eta^\alpha \hat{\rho}_\beta - \hat{\rho}_\beta \eta^\alpha), \tag{3.14}$$

for the ordering of (3.10). This expression gives the desired physical results and satisfies (1)–(3).

### D. Quadratic observables

Let $K = K^{AB}(Q^C)P_A P_B$ be a general quadratic physical observable. Then, it satisfies

$$\{K, \varphi_\alpha\} = K_\alpha{}^\beta \varphi_\beta. \tag{3.15}$$

Unlike linear observables $K_\alpha{}^\beta$ has momentum dependence. In fact, it is linear in the momenta and so can be written as

$$K_\alpha{}^\beta = K_\alpha{}^{\beta A}(Q^B)P_A, \tag{3.16}$$

for some function $K_\alpha{}^{\beta A}(Q^B)$. This fact makes the BRST extension of quadratic observables more complex than that for linear observables. The BRST extension of $K$ is

$$\mathbf{K} = K + K_\alpha{}^\beta \eta^\alpha \rho_\beta + K_{\alpha\beta}{}^{\gamma\delta}\eta^\alpha\eta^\beta\rho_\gamma\rho_\delta. \tag{3.17}$$

The term $K_{\alpha\beta}{}^{\gamma\delta}$ is a function of the configuration variables only and is defined by

$$\{K, C^\gamma{}_{\alpha\beta}\} - \{K_\alpha{}^\gamma, \varphi_\beta\} + \{K_\beta{}^\gamma, \varphi_\alpha\} + C^\gamma{}_{\beta\epsilon}K_\alpha{}^\epsilon + C^\gamma{}_{\epsilon\alpha}K_\beta{}^\epsilon + C^\epsilon{}_{\alpha\beta}K_\epsilon{}^\gamma = K_{\alpha\beta}{}^{\gamma\delta}\varphi_\delta. \tag{3.18}$$

The details of this can be found in Ref. 14.

In the trivialized coordinate system $\mathbf{K}$ takes the simpler form

$$\tilde{\mathbf{K}} = K^{AB}\tilde{P}_A\tilde{P}_B + [2K^{\alpha\beta}{}_{,\alpha}\tilde{P}_a + K^{\gamma\beta}{}_{,\alpha}\tilde{P}_\gamma]\tilde{\eta}^\alpha\tilde{\rho}_\beta. \tag{3.19}$$

This can also be written as

$$\tilde{\mathbf{K}} = K^{ab}\tilde{P}_a\tilde{P}_b - \{(2K^{a\beta}\tilde{P}_a + K^{\gamma\beta}\tilde{P}_\gamma)\tilde{\rho}_\beta, \tilde{\Omega}\}, \tag{3.20}$$

thereby enabling the implementation of (3.3) to give

$$\hat{\tilde{\mathbf{K}}} = |g|^{-1/4}\hat{\tilde{P}}_a K^{ab}|g|^{1/2}\hat{\tilde{P}}_b|g|^{-1/4} + (i/2\hbar)[(2[K^{a\beta}, \hat{\tilde{P}}_a]_+ + [K^{\gamma\beta}, \hat{\tilde{P}}_\gamma]_+)\hat{\tilde{\rho}}_\beta, \hat{\tilde{\Omega}}]. \tag{3.21}$$

This then expands out into the form

$$\begin{aligned}\hat{\tilde{\mathbf{K}}} = {} & |g|^{-1/4}\hat{\tilde{P}}_A K^{AB}|g|^{1/2}\hat{\tilde{P}}_B|g|^{-1/4} \\ & - (\hbar^2/4)|g|^{-1/2}[2K^{a\alpha}{}_{,\alpha}|g|^{1/2}]_{,a} \\ & - (\hbar^2/4)|g|^{-1/2}[K^{\beta\alpha}{}_{,\alpha}|g|^{1/2}]_{,\beta} \\ & + \tfrac{1}{4}(2[K^{a\beta}{}_{,\alpha}, \hat{\tilde{P}}_a]_+ \\ & + [K^{\gamma\beta}{}_{,\alpha}, \hat{\tilde{P}}_\gamma]_+)(\tilde{\eta}^\alpha\hat{\tilde{\rho}}_\beta - \hat{\tilde{\rho}}_\beta\tilde{\eta}^\alpha).\end{aligned} \tag{3.22}$$

Using $\|\tilde{\varphi}\|$, and (3.16), this can be written coordinate covariantly as

$$\begin{aligned}\hat{\tilde{\mathbf{K}}} = {} & \|\tilde{\varphi}\|^{-1/2}\hat{\tilde{P}}_A K^{AB}\|\tilde{\varphi}\|\hat{\tilde{P}}_B\|\tilde{\varphi}\|^{-1/2} \\ & - (\hbar^2/4)\|\tilde{\varphi}\|^{-1}(K_\alpha{}^{\alpha A}\|\tilde{\varphi}\|)_{,A} \\ & + \tfrac{1}{4}[K_\alpha{}^{\beta A}, \hat{\tilde{P}}_A]_+(\tilde{\eta}^\alpha\hat{\tilde{\rho}}_\beta - \hat{\tilde{\rho}}_\beta\tilde{\eta}^\alpha).\end{aligned} \tag{3.23}$$

Finally, this expression is boosted, using $\hat{\mathbf{R}}$, to give the non-Abelian version. The result is

$$\begin{aligned}\hat{\mathbf{K}} = {} & \|\varphi\|^{-1/2}\hat{P}_A K^{AB}\|\varphi\|\hat{P}_B\|\varphi\|^{-1/2} \\ & - (\hbar^2/4)\|\varphi\|^{-1}(K_\alpha{}^{\alpha A}\|\varphi\|)_{,A} \\ & + \tfrac{1}{4}[K_\alpha{}^{\beta A}, \hat{P}_A]_+(\eta^\alpha\tilde{\rho}_\beta - \tilde{\rho}_\beta\eta^\alpha) \\ & + \tfrac{1}{2}K_{\alpha\beta}{}^{\gamma\delta}(\hat{\rho}_\gamma\eta^\alpha\eta^\beta\hat{\rho}_\delta - \hat{\rho}_\delta\eta^\alpha\eta^\beta\hat{\rho}_\gamma).\end{aligned} \tag{3.24}$$

This ordering, of the classical expression (3.17), gives the desired physical results and satisfies (1)–(3).

## IV. ANTIGHOSTS, PREQUANTIZATION, AND THE KOSZUL COMPLEX

### A. Self-dual states of definite ghost number

In Sec. II we showed how the BRST invariant states can be related to the physical states of a constrained system. For a trivializable set of constraints the identification was based on a vanishing cohomology result. In its turn, the cohomological input arose from the nilpotency and Hermiticity of the BRST charge $\hat{\Omega}$ on $H_{\mathrm{BFV}}$. However, we saw that the solutions to Eq. (2.1) required us to consider $\Psi$ not on $H_{\mathrm{BFV}}$ but, rather, on its distributional extension $(S_{\mathrm{BFV}})^*$. Now $\hat{\Omega}$ is still nilpotent on this space, but it is *not* in general Hermitian. Thus, in Sec. II we had to content ourselves with the identification of self-dual states of the form (2.9) with physical states, and then allow any coboundary $\hat{\Omega}\chi$ to be added as long as it is an element of $H_{\mathrm{BFV}}$. So, if $k = 2$, a state of the form $\Psi = \delta(Q_\alpha)\eta^\alpha$ is *not* allowed even though it is BRST invariant, normalizable, and a coboundary in $(S_{\mathrm{BFV}})^*$. Obviously if there are any interesting global structures then this prescription can become even more complicated.

In simple situations this analysis of physical states will suffice; it is clear what is meant by the solutions to (2.1) being physical. However, any field theoretic extension is at best problematic. What is more upsetting is that this description just is not very elegant. Somehow the rich cohomological structure implicit in the use of ghost variables is finding it difficult to survive quantization. What is needed is a more precise characterization of physical states that avoids the above implications.

What we do know (even in the globally nontrivial situation) is that it is the zeroth cohomology group of the ghost number complex (2.5) that describes the physical states. But restricting attention to the ghost number zero solutions to (2.1) will not do since the Berezin pairing forces us to look at self-dual states of the form (2.9) that have *no* definite ghost number. Also, if the Berezin pairing is used, it will evaluate states like the $\delta(Q_\alpha)\eta^\alpha$ ones, so they need to be excluded by hand. To overcome this we cannot just change the Berezin pairing since, in (I), we showed that it is the unique pairing giving the correct Hermiticity assignment to the ghosts and conjugate ghosts.

In order to solve the problem that self-dual states do not necessarily have a definite ghost number we shall extend the definition of ghost number by introducing new constraints and their ghosts. The need for such an extension is already seen in field theory, although the motivation then is to maintain covariance. There the ghost variables are associated with the secondary first class constraints (which encodes the non-Abelian Gauss law) while the antighosts are the conjugate fields to the "primary" ghosts associated with the primary constraints (which are the conjugate momenta to the time component of the gauge field).

So let us now introduce the extra bosonic configurations $(\lambda^\alpha, \pi_\alpha)$ and the ghosts $(\bar{\eta}^\alpha, \bar{\rho}_\alpha)$, where the $\pi_\alpha(\bar{\rho}_\alpha)$ are conjugate to the $\lambda^\alpha(\bar{\eta}^\alpha)$. To avoid confusion, at this stage of our presentation, we shall refer to the ghosts $\bar{\eta}^\alpha$ as the primary ghosts and $\eta^\alpha$ as the secondary ghosts. Thus we have enlarged the classical phase space to one of dimension $(2n + 2k; 4k)$. To recover the physical system we consider the momenta $\pi_\alpha$ as additional constraints. Hence the new extended BRST charge $\Omega_{\mathrm{ext}}$ is

$$\Omega_{\mathrm{ext}} = P_\alpha\eta^\alpha + \pi_\alpha\bar{\eta}^\alpha, \tag{4.1}$$

where we are using a trivialized set of "secondary" constraints. The overall ghost number of any function will now be the difference between the number of ghosts (primary and

secondary) and conjugate ghosts (again, primary and secondary).

If we now quantize this extended system using the Schrödinger representation, or configuration polarization, as used in (I) for the system with just the secondary constraints, then all the states will have positive ghost number and we will be no better off than before. To overcome this we shall use a different polarization for the primary constraints and their ghosts. So, in addition to I(4.1), we require

$$\pi_\alpha \to \hat{\pi}_\alpha = \pi_\alpha, \quad \lambda^\alpha \to \hat{\lambda}^\alpha = i\hbar \frac{\partial}{\partial \pi_\alpha} \tag{4.2a}$$

and

$$\bar{\rho}_\alpha \to \hat{\bar{\rho}}_\alpha = i\bar{\rho}_\alpha, \quad \bar{\eta}^\alpha \to \hat{\bar{\eta}}^\alpha = -\hbar \frac{\partial}{\partial \bar{\rho}_\alpha}. \tag{4.2b}$$

Quantum states will now be of the form

$$\Psi(Q^A, \pi_\alpha, \eta^\alpha, \bar{\rho}_\alpha) = \Psi_0 + \sum_{r,s=1}^{k} \Psi_{\alpha_1 \cdots \alpha_r}{}^{\beta_1 \cdots \beta_s} \eta^{\alpha_1} \cdots \eta^{\alpha_r} \bar{\rho}_{\beta_1} \cdots \bar{\rho}_{\beta_s}. \tag{4.3}$$

Which can clearly be grouped into states of definite ghost number and, in particular, $\Psi$ will have ghost number zero if $r = s$.

The choice of the momentum polarization in (4.2a) is not central to this discussion; we could easily proceed with any other. The operator assignment for the primary ghosts in (4.2b) is unavoidable if the Berezin pairing is used (now extended to the $2k$ fermionic variables $\eta^\alpha, \bar{\rho}_\alpha$) and $\hat{\Omega}_{\text{ext}}$ is to be Hermitian. Note that this assignment is consistent with the discussion of antighosts presented above since the primary conjugate ghost $\bar{\rho}_\alpha$ is anti-Hermitian, which is the Hermiticity assignment of the antighost field.[15] Thus from now on we shall simply refer to the $\bar{\rho}_\alpha$'s as the antighosts and the $\eta^\alpha$'s as the ghosts.

The quantized charge $\hat{\Omega}_{\text{ext}}$ is easy to write down and we see that it is indeed nilpotent, Hermitian, and increases the ghost number of a state by 1. What is more interesting is that the ghost number complex for $\hat{\Omega}_{\text{ext}}$, which is analogous to (2.5) for $\hat{\Omega}$, is now a double complex. To see this let $\Gamma^{r,s}$ denote those states with $r$ ghosts and $s$ antighosts, so the net ghost number for such states is $r - s$. We can decompose the action of $\hat{\Omega}_{\text{ext}}$ on $\Gamma^{r,s}$ by writing $\hat{\Omega}_{\text{ext}} = \delta_0 + \delta_1$, where

$$\delta_0 = -\hbar \pi_\alpha \frac{\partial}{\partial \bar{\rho}_\alpha} \tag{4.4}$$

and

$$\delta_1 = -i\hbar \partial_\alpha \eta^\alpha. \tag{4.5}$$

So $\delta_0$: $\Gamma^{r,s} \to \Gamma^{r,s-1}$ and $\delta_1$: $\Gamma^{r,s} \to \Gamma^{r+1,s}$. Hence $(\hat{\Omega}_{\text{ext}})^2 = 0$ implies

$$\delta_0{}^2 = 0, \quad \delta_0\delta_1 + \delta_1\delta_0 = 0, \quad \delta_1{}^2 = 0. \tag{4.6}$$

The part of the complex given by $\delta_1$ is clearly reflecting the contribution of the original ghost complex (2.5). The new part is that coming from $\delta_0$. From (4.4) and (4.6) we can identify $\delta_0$ as the quantum version of the Koszul complex that was central to the classical analysis of ghost systems.[3–5,16] In that case we had the strong result that $\delta_0$ gave a resolution of the ghost complex, i.e., $\delta_0 F = 0 \Rightarrow F = \delta_0 G$, for some $G$. Later on we shall see what happens to this result after quantization. Let us first, though, show how a state of the form (2.9) can be modified to be invariant under $\hat{\Omega}_{\text{ext}}$, self-dual, and have zero ghost number.

If we simply try to multiply the top term in (2.9) by $k$ antighosts we will have a zero ghost number state, but it will not be BRST invariant or normalizable. The resolution to this is to consider the state

$$\Psi = \Psi_0 + (2k)!(i)^{3k(2k-1)}\delta(Q^1)\cdots\delta(Q^k)\delta(\pi_1)\cdots \times\delta(\pi_k)\eta^k\cdots\eta^1\bar{\rho}_k\cdots\bar{\rho}_1. \tag{4.7}$$

This is normalizable, BRST invariant, and has ghost number zero. Later we shall discuss in what sense $\Psi$ uniquely satisfies these conditions. Before doing that let us first investigate another situation where the quantized Koszul complex arises.

### B. Prequantization

Geometric quantization offers an effective first step in the general analysis of the transition from a classical to a quantum description of a system. This is because the process of quantization is now split into two steps. First, a system is prequantized; then, when possible, this is elevated to a full quantum description.[2] The intermediate step of prequantization gives a faithful representation of the classical Poisson algebra as an operator algebra. The expected Van Hove type of obstruction is neatly sidestepped since the operator representation is now reducible. Indeed, the states are now taken as square integrable functions on the whole phase space. The transition to the full quantization then involves a selection of a suitable polarization and measure on these extended states. Although this last step is fraught with all the expected difficulties inherent in any quantization, various authors[7,8] have suggested that a prequantum analysis of constrained systems is useful, especially related to any discussion of polarizations. Hence we shall discuss a possible extension of geometric quantization to the ghost analysis of constrained systems; an independent motivation for this is that we can then allow more general rescalings to the trivial constraints, and hence some insight into the structure of systems with quadratic constraints might emerge.

In geometric quantization the symplectic structure of the phase space plays an important role. On the bosonic part of the phase space we have the symplectic form [see (2) in Sec. VI] $\omega = dQ^A\, dP_A$. This allows one to associate a Hamiltonian vector field $X_f$ with each smooth function $f$ on the phase space via its inner product with $\omega$, i.e.,

$$l_{X_f}\omega = df. \tag{4.8}$$

The Poisson bracket $\{f,g\}$ is then defined by $\{f,g\} = -X_f g$ and it is straightforward to show that

$$[X_f, X_g] = -X_{\{f,g\}}. \tag{4.9}$$

Prequantization involves the construction of a prequantum, Hermitian operator $O_f$, acting on the Hilbert space of square integrable functions on the phase space, such that

$$[O_f, O_g] = i\hbar O_{\{f,g\}}. \tag{4.10}$$

An expression for $O_f$ is

$$O_f = i\hbar X_f + (f - l_{X_f}\theta), \tag{4.11}$$

where $\theta = P_A\, dQ^A$ is such that $d\theta = -\omega$. Other choices for the symplectic potential $\theta$ are possible and reflect the adaptability of $O_f$ to a particular polarization.

The extension of this construction to a graded phase space is quite straightforward,[17] the only subtle point for us being that the use of the Berezin measure precludes the automatic assignment of a defined Hermitian structure to the prequantum operators. So, let us now take as our graded symplectic form [see (2) in Sec. VI] $\omega = dQ^A\, dP_A + d\eta^\alpha\, d\rho_\alpha$. Then, given a graded function $\mathbf{F}$, we define the graded Hamiltonian vector field $X_\mathbf{F}$ just as in the bosonic case (4.8). Thus we find

$$X_\mathbf{F} = \frac{\partial \mathbf{F}}{\partial P_A}\frac{\partial}{\partial Q^A} - \frac{\partial \mathbf{F}}{\partial Q^A}\frac{\partial}{\partial P_A} - (-1)^\mathbf{F}\frac{\partial \mathbf{F}}{\partial \rho_\alpha}\frac{\partial}{\partial \eta^\alpha} - (-1)^\mathbf{F}\frac{\partial \mathbf{F}}{\partial \eta^\alpha}\frac{\partial}{\partial \rho_\alpha}, \tag{4.12}$$

which can be used to define a graded Poisson bracket and will satisfy the graded version of (4.9). Then, just as in the bosonic case above, we can construct $O_\mathbf{F}$, the graded prequantum operator corresponding to the graded function $\mathbf{F}$. Using the graded extension to (4.11), with $\theta = P_A\, dQ^A + d\eta^\alpha \rho_\alpha$, we find

$$O_\mathbf{F} = -i\hbar X_\mathbf{F} + \mathbf{F} - P_A\frac{\partial \mathbf{F}}{\partial P_A} - \rho_\alpha\frac{\partial \mathbf{F}}{\partial \rho_\alpha} \tag{4.13}$$

and this operator satisfies the graded version of (4.10).

Applying this prequantum prescription to the BRST charge (2.4), we get

$$O_\Omega = -i\hbar\eta^\alpha\partial_\alpha - i\hbar P_\alpha\frac{\partial}{\partial \rho_\alpha}. \tag{4.14}$$

Clearly, just as in (4.4) and (4.5), $O_\Omega$ can be decomposed into a $\delta_0$ and a $\delta_1$. Hence we will again get a double complex when acting on the prequantum states. Even though $O_\Omega$ is not Hermitian [cf. (4.4)] its zeroth cohomology group should still be used to describe the physical states since that was the classical use of the BRST charge. So again we see the need to address the effect quantization has had on the classical Koszul resolution.

### C. The quantized Koszul complex

Sections IV A and IV B have shown us that we need to understand the structure of the solutions to the equation $\delta_0\mathbf{F} = 0$ with with $\mathbf{F}\in\Gamma^{r,s}$. So, in terms of the distributional coefficients, we are interested in the general solution to

$$\varphi_{\alpha_1}F^{\alpha_1\alpha_2\cdots\alpha_s} = 0, \tag{4.15}$$

where $\varphi_\alpha$ is a set of $k$ independent constraints that act on the distributional coefficients of $\mathbf{F}$ by multiplication (we are, without loss of generality, taking $r = 0$).

In the classical version of this problem, when $\mathbf{F}$ is a smooth function on the phase space, all solutions to (4.15) are of the form $\mathbf{F} = \delta_0\mathbf{G}$. This result was important in the proof that the BRST charge could be used to recover the constrained dynamics. We shall now show that this result does *not* hold in the quantum theory where the smooth functions are replaced by distributions.

Just as in Ref. 16, we can study the solutions to (4.15) by an inductive argument on the number of indices $s$. We shall only give the first step in the inductive proof since it contains all the important steps; the extension to general $s$ is then a matter of combinatorics. Consider the distributions $F^\alpha$ that satisfy

$$\varphi_\alpha F^\alpha = 0. \tag{4.16}$$

We will prove that $F^\alpha$ must be of the form

$$F^\alpha = F^{\alpha\beta}\varphi_\beta + R^\alpha\delta(\varphi_\alpha), \tag{4.17}$$

where $F^{\alpha\beta}$ is a distribution antisymmetric in its indices and $R^\alpha$ is a distribution with *no dependence* on $\varphi_\alpha$ [there is *no* summation in the $R^\alpha\delta(\varphi_\alpha)$ term]. As in Ref. 16, this result is proven by induction on the zero sets of $\varphi_\alpha$. Let $Z_r$ denote the zero set of $(\varphi_r,\ldots,\varphi_k)$; then, on $Z_2$, (4.16) is

$$\varphi_1 F^1 = 0. \tag{4.18}$$

We now make use of the following well known result for distributions (see, for example, Ref. 18 for a proof).

*Lemma:* If $g$ is a distribution and $yg = 0$, then $g = h\delta(y)$ where $h$ is a distribution independent of $y$.

Applying this result to (4.18) gives $F^1 = R^1\delta(\varphi_1)$, thus establishing (4.17) on $Z_2$. Now assume (4.17) is true on $Z_r$ ($r\geqslant 2$). Then, on $Z_{r+1}$ we must have

$$F^\alpha = \sum_{\beta=1}^{r} F^{\alpha\beta}\varphi_\beta + R^\alpha\delta(\varphi_\alpha) + A^\alpha\varphi_{r+1} + B^\alpha, \tag{4.19}$$

for $\alpha = 1,\ldots,k$. In this equation $F^{\alpha\beta}$ is a distribution antisymmetric in its indices. $R^\alpha$ can depend on any of the $\varphi_\beta$'s apart from $\varphi_\alpha$ itself, $A^\alpha$ is a distribution, as is $B^\alpha$, but $B^\alpha$ cannot be written in the form $C^\alpha\varphi_{r+1}$ for some $C^\alpha$. On $Z_{r+1}$ (4.16) and (4.19) give

$$\sum_{\alpha=1}^{r} A^\alpha\varphi_\alpha\varphi_{r+1} + \sum_{\alpha=1}^{r} B^\alpha\varphi_\alpha + \varphi_{r+1}F^{r+1} = 0, \tag{4.20}$$

which implies that $B^\alpha$ can be written in the form $C^\alpha\varphi_{r+1}$ and hence it must be zero. Therefore we can write (4.20) as

$$\varphi_{r+1}\left[\sum_{\alpha=1}^{r} A^\alpha\varphi_\alpha + F^{r+1}\right] = 0.$$

Using the lemma we can conclude that

$$F^{r+1} = -\sum_{\alpha=1}^{r} A^\alpha\varphi_\alpha + R^{r+1}\delta(\varphi_{r+1}), \tag{4.21}$$

where $R^{r+1}$ is any distribution that does not depend on $\varphi_{r+1}$. Equations (4.19) and (4.21) can now be easily manipulated to establish (4.17) on $Z_{r+1}$ and so, by induction, (4.17) is the general solution of (4.16).

Thus we see that $\delta_0\mathbf{F} = 0 \Rightarrow \mathbf{F} = \delta_0\mathbf{F}' + \delta(\varphi_\alpha)\mathbf{F}^\alpha$ and hence the Koszul complex has nontrivial homology. This means that the classical argument that the BRST charge picks out the physical states as the *only* nontrivial cohomology is in need of modification in the quantum theory. Note that although the lack of a Koszul resolution holds irrespective of any global properties of the constraints (such as trivializability), the obstructions *are* still trivial in the $\delta_1$ cohomology. Hence, in the trivialized case, the zeroth cohomo-

logy group of the BRST charge is directly related to the physical states.

As an example let us consider the situation when antighosts are being used, if $k=2$, the state $\delta(\pi_1)\delta(Q_2)\eta^1\bar{\rho}_2+\delta(\pi_2)\delta(Q_1)\eta^2\bar{\rho}_1$ is normalizable, invariant, and has ghost number zero. It is a coboundry via the triviality of the ghost complex but it has nonzero norm due to surface terms. It is not clear to us how to interpret such states. One might argue that we should just avoid them and treat this as an extra condition on the states. However, as the example above shows, they are not too wild as states go. In fact, duality seems to force them to be concentrated on the physical configurations. Also, they have no pairing with the more natural physical states (4.7) and this is preserved by the physical observables. So possibly they could be interpreted as giving another copy of the physical states. The example above could then be thought of as representing the constant (distributional) physical state.

## V. DISCUSSION AND CONCLUSIONS

In these two papers we set out to solve explicitly, in a covariant manner, the constraint quantization problem for finite-dimensional gauge theories. Our solution to this problem differs from that suggested by Kuchar in that we have maintained a full extended state space structure. Our approach has relied on the ability to rescale constraints, via unitary operators, within the ghost description of such systems. These rescaling transformations are known to exist classically for all possible rescalings. What we have shown, in the first of these papers, is that they also exist quantum mechanically, at least for rescalings that depend on the configuration space variables. In this second paper we have shown that the existence of this quantum rescaling, when coupled with the local trivialization theorem, led to a natural way of deriving the orderings for all special observables and the BRST charge. These orderings are coordinate and rescaling covariant and avoid all Van Hove problems, except for the inevitable ones in the physical directions. The reliance on local trivialization can lead to global problems, in particular, the domains for these operators needs to be investigated separately. Surprisingly, most of the difficulties we found in solving this problem were not in the ordering of observables, but in the construction of the state space.

The standard approach to BRST methods has always emphasized the elegant and powerful cohomological aspects of its formulation. In practice we have found that these cohomological ideas require extreme care in being implemented. This occurs first because one is forced to work with distributional wave functions, which can cause the BRST charge to be no longer Hermitian, and, second, because one must work with the Berezin pairing that requires the introduction of duality in order to pair states. The general definition of a distributional dual is particularly tricky since, as is well known, products of distributions do not exist in general. However, for the BRST invariant states we have been able to define a meaningful dual.

The above gives a summary of these two papers and, we feel, shows that ghost variables have an important role to play in *all* aspects of constrained dynamics. Let us now discuss possible extensions of the methods presented in this paper.

There has been considerable debate as to whether I(3.4d) should contain scalar curvature terms in addition to the Laplace–Beltrami operator. There is no reason why our solution could not be extended to include such terms. The only difference would be in the ordering prescription, (3.24), for quadratic observables.

There is also considerable physical interest in constraints that are quadratic in momenta and we would like to see a similar ghost analysis of such systems. It is not clear, though, how the constraint rescaling argument should be used in this situation. If the philosophy is that the only rescalings allowed should be those that *preserve* the momenta dependence of the constraints then we are again in the situation covered in these papers with $\Lambda=\Lambda(Q^A)$. However, we cannot then expect to work initially in a trivialized system where the constraints are just a set of momenta, and hence have no factor ordering problems. This step in the argument needs to be replaced with some simple, generic form of quadratic constraints. The problem would then reduce to quantizing this simpler "canonical" set of constraints. Even if such a set could be found and quantized, the boosting is now going to hit possible Van Hove types of ordering problems. In particular, the step from (3.2) to (3.3) would almost certainly hit such an obstruction.

An alternative approach would be to continue to use the trivialization to pure momenta, but now attempt to boost back to the original system via $\Lambda(Q^A, P_A)$. It is known that such boosts exist classically, so it would be nice to know what happens quantum mechanically. We must expect, though, that any attempt to write such transformations down will encounter a whole new set of difficult ordering problems, especially since we need both $\Lambda$ and $\Lambda^{-1}$. For instance, if $\Lambda$ depends linearly on momenta then $\Lambda^{-1}$ will be a rational function of the momenta. This is just the Van Hove problem surfacing in a different guise. To get some insight into this approach the methods of geometric quantization seem attractive since the ordering problems are avoided initially.

Within a prequantum description we can expect to completely solve the problem of boosting with $\Lambda(Q^A,P_A)$.[19] On top of this, we expect that the use of ghost methods will resolve a problem pointed out in Ref. 8 related to the choice of constraint polarizations. In Ref. 7 conditions were introduced which ensured that, at the prequantum level, a polarization would be compatible with a given set of constraints. Within these geometric methods, any construction that only works with a particular parametrization of the system is an anathema. Yet, since constraint rescaling is not normally a canonical transformation, the conditions on polarizations discussed in Ref. 7 are dependent on the particular bases of constraints chosen. This unsatisfactory situation is removed when ghost variables are used since, as we have stressed in these papers, now constraint rescaling is a canonical transformation. As an application of this one can give a constructive method for finding a suitable polarization for quadratic (or worse) constraints by simply trivializing, introducing the vertical polarization in the trivialized system, then boost-

ing *both* the observables and the polarization. The details of this approach are currently under investigation.[19]

Another interesting extension of the methods presented in these papers is to formulate the solution using path integrals.[20] In particular, it would be instructive to see how the term (2.13), in the duality condition, relates to the Faddeev determinant in the path integral measure. A path integral formulation of these methods would also assist in any attempts to extend this work to constrained field theories.

Any extension of the ideas presented in these papers to field theory must address the problems of renormalizability. In particular, the ability to rescale constraints will most certainly be severely restricted since manifest Lorentz invariance will be lost in general. A particularly interesting area in field theory where these ideas might have some importance is in the analysis of anomalous theories. As is well known, anomalies imply Schwinger terms in the constraint algebra. These extra terms are nontrivial in the sense that field redefinitions cannot remove them. Thus it would be interesting to find out the effect rescalings can have on them and hence on the Ward identities of the theory.

## VI. NOTES

(1) If $C$ can be identified as the zero map of an equivariant momentum map then we are guaranteed to find a set of globally defined set of constraints. Such a situation arises in Yang–Mills theory.

(2) We always assume that when two differential forms are multiplied together then the product is the wedge product. The mixed grading between fermionic and differential form structures is such that if $\omega$ (respectively, $\mu$) is a graded $n$- ($m$-) form then $\omega\mu = (-1)^{nm}(-1)^{\omega\mu}\mu\omega$. The exterior derivative is the right exterior derivative.


## ACKNOWLEDGMENTS

We have benefited from conversations with Tim Blackwell, Andrew Liddle, and John Rawnsley.


## APPENDIX: THE QUANTUM WELL OF ORVIETO

We will now illustrate our solution by means of an example first suggested by DeWitt,[21] and developed in detail by Kuchar.[6,12] The example consists of a nonrelativistic particle moving in flat, three-dimensional space subject to the gauge group of helical motions. That is, if $(X,Y,Z)$ is the natural, global coordinate system for $Q$, then all points lying on the helices

$$X(\tau) = X(0)\cos(\tau) + Y(0)\sin(\tau),$$

$$Y(\tau) = -X(0)\sin(\tau) + Y(0)\cos(\tau),$$

$$Z(\tau) = Z(0) + \tau$$

($\tau \in \mathbb{R}$) are gauge equivalent. In phase space language this problem is described by the Hamiltonian

$$H = (P_R)^2 + (1/R^2)(P_\Theta)^2 + (P_Z)^2, \tag{A1}$$

and by one constraint

$$\varphi = P_Z - P_\Theta, \tag{A2}$$

where we have expressed everything in cylindrical polar coordinates $(R,\Theta,Z)$. The BRST formulation of this problem is

$$\mathbf{\Omega} = (P_Z - P_\Theta)\eta \tag{A3}$$

and

$$\mathbf{H} = H. \tag{A4}$$

To do the quantization it is necessary to compute $\|\varphi\|$. This is straightforward and gives

$$\|\varphi\| = R/[1 + R^2]^{1/2}. \tag{A5}$$

With this result the quantization procedure can be applied and gives

$$\hat{\mathbf{\Omega}} = -i\hbar[\partial_Z - \partial_\Theta]\eta \tag{A6}$$

and

$$\hat{\mathbf{H}} = -\hbar^2[(1/R^2)\partial_\Theta^2 + \partial_Z^2 + \partial_R^2 + (1/R[1 + R^2])\partial_R]. \tag{A7}$$

The self-dual, BRST invariant states are of the form

$$\Psi = \Psi_0(R,\Theta, + Z) + \delta(\Theta)\Psi_0\eta, \tag{A8}$$

where $\Theta = 0$ has been chosen as the gauge fixing condition. To confirm that this is the correct quantum theory let $\Psi^1$ and $\Psi^2$ be two such states and observe that

$$\langle\Psi^1|\hat{\mathbf{H}}|\Psi^2\rangle = -\hbar^2\int \Psi_0^1\left[\frac{\partial^2}{\partial R^2} + \left[1 + \frac{1}{R^2}\right]\frac{\partial^2}{\partial(\Theta + Z)^2} + \frac{1}{R[1 + R^2]}\frac{\partial}{\partial Z}\right]\Psi_0^2 \times \frac{R}{[1 + R^2]^{1/2}}\,dR\,d(\Theta + Z). \tag{A9}$$

The true degrees of freedom for the quantum well is a curved manifold described by coordinates $q^1 = R$ and $q^2 = \Theta + Z$, and with metric

$$[g^{ab}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 + 1/(q^1)^2 \end{bmatrix} \tag{A10}$$

(see Ref. 6 for the details). From this it is easy to check that (A9) is the correct physical result.


[1] D. McMullan and J. Paterson, J. Math. Phys. **30**, 477 (1989).
[2] N. M. J. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).
[3] A. D. Browning and D. McMullan, J. Math. Phys. **28**, 438 (1987).
[4] J. Stasheff, University of North Carolina preprint, 1987.
[5] M. Dubois-Violette, preprint L.P.T.H.E.-ORSAY 87/23, 1987.
[6] K. Kuchar, Phys. Rev. D **35**, 596 (1987).
[7] A. Ashtekar and M. Stillerman, J. Math. Phys. **19**, 1319 (1986).
[8] M. Blau, preprint SISSA-87/87/EP, 1987.
[9] P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1979).
[10] J. Paterson, Ph.D. thesis, University of Glasgow, 1989.
[11] L. P. Faddeev, in *Methods in Field Theory*, edited by R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1975).
[12] K. Kuchar, Phys. Rev. D **34**, 3031 (1986).
[13] K. Kuchar, Phys. Rev. D **34**, 3044 (1986).

[14]M. Henneaux, Phys. Rep. **126**, 1 (1985).
[15]T. Kugo and I. Ojima, Suppl. Prog. Theor. Phys. **66**, 1 (1979).
[16]D. McMullan, Imperial College preprint TP/83-84/21, 1984 (unpublished).
[17]B. Kostant, in *Differential Geometric Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977).
[18]M. J. Lighthill, *Introduction to Fourier Analysis and Generalised Functions* (Cambridge U.P., Cambridge, 1958).
[19]D. McMullan and J. Paterson (unpublished).
[20]J. Paterson (unpublished).
[21]B. S. DeWitt, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979).

# The geometry of gauge symmetry breaking in the superstring context

Brett McInnes
*Department of Mathematics, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511, Republic of Singapore*

(Received 19 April 1988; accepted for publication 12 October 1988)

The Higgs mechanism has a very natural global formulation: given a principal fiber bundle $(P,M,G)$, with structural group $G$, a subbundle with the unbroken subgroup as structural group (with the aid of the Higgs fields themselves) can be constructed. In superstring theory, the fundamental $E_8$ symmetry, as well as the grand unification symmetry, are broken in a very different way. The principal purpose of this work is to construct a global formulation of the superstring gauge vacuum, without Higgs fields.

## I. HIGGS FIELDS AND GLOBAL GAUGE STRUCTURE

The Higgs mechanism is widely regarded as one of the least satisfactory elements of conventional gauge theory. Indeed, to answer the question, "Why are gauge symmetries broken?" with "Because the Higgs fields have nonzero vacuum expectation values," is not very enlightening, since we have no *other* reason (theoretical or experimental) to introduce such fields. The principal objection to the Higgs mechanism, then, is that it requires the introduction of extraneous fields having no clear relationship to the basic ideas of gauge theory itself.

Before we dismiss the Higgs mechanism, however, let us remind ourselves of its virtues, which—somewhat unexpectedly—are particularly in evidence at the global level. Gauge fields are (pullbacks of) connections on principal bundles. To say that a gauge group $G$ is "broken" to a closed subgroup $H$ means that we are given a principal bundle $(P,M,G)$, and we can find (or construct) a principal bundle $(Q,M,H)$ in such a way that connections on these bundles are related to each other. The natural way to relate one principal bundle to another is via a bundle homomorphism,[1] that is, a pair of maps $\phi: Q\to P$, $\psi: H\to G$, where $\psi$ is a group homomorphism and $\phi$ satisfies $\phi(qh)=\phi(q)\psi(h)$, for all $q\in Q$, $h\in H$. Evidently it is also natural to require that the induced mapping on $M$ be the identity map, and that $\psi$ should be a monomorphism. Thus we obtain the usual formulation of symmetry breaking: $Q$ is taken to be a *subbundle* of $P$. Later we shall argue that (in a certain context) this is not the only possible formulation; but let us retain it for the present.

The interesting point here is this: given a principal bundle $(P,M,G)$ and a subgroup $H$ of $G$, it is certainly not the case that, in general, $P$ will admit *any* subbundle with structural group $H$. A decision on this point will often involve a careful investigation, making use of obstruction theory.[2] But the usual procedure is not to do this—rather, one simply assumes that the desired subbundle exists. In other areas of gauge theory (such as magnetic monopole theory) such a procedure would not work; why, then, is it permitted here? There are two main reasons.

The first is that by postulating the existence of Higgs fields with nonvanishing vacuum expectation values, we automatically ensure the existence of the relevant subbundle. That is, the local symmetry breaking mechanism (scalar fields, etc.) itself takes care of the global structure—a remarkable result. Given $(P,M,G)$ and $H$ as above, the manifold $P/H$ is an associated bundle of $P$ with standard fiber $G/H$. A Higgs field $\sigma$ is essentially (see Ref. 3 for details) a cross section of $P/H$. The subbundle $Q$ can then be defined as the set of all $p\in P$ that satisfy $pH=\sigma(\pi(p))$, where $\pi: P\to M$ is the projection. The situation here is closely analogous to the problem of defining spinors in general relativity. To do this, one needs a Lorentz subbundle of the frame bundle; again, this subbundle does not always exist; but one of the fundamental postulates of general relativity (the global existence of the metric tensor) automatically solves the problem, since it permits the construction of the bundle of orthonormal frames. (Note that the metric tensor, which plays the role of "Higgs field" here, is the canonical example of a field with nonzero "vacuum expectation value.")

The second point to note is that even if the appropriate subbundle structure did not arise naturally, it could nevertheless always be *constructed*. Let $(Q,M,H)$ be any principal $H$ bundle, and let $H$ be a subgroup of $G$. Then there always exists[4] a principal $G$ bundle having $Q$ as subbundle: we merely define $P=(Q\times G)/H$. Thus if—as is often the case—we are not particular as to the precise nature of the manifold $P$, then we can always choose our initial bundle in a way that is compatible with the requirements of symmetry breaking.

In superstring theory,[5] the fundamental $E_8$ gauge symmetry (actually $E_8\times E_8$, but we shall ignore the second $E_8$ henceforth) is broken in a very different way, as is the $E_6$ grand unification group. Here the base manifold is (Minkowski space) $\times K$, where $K$ is a six-dimensional Ricci-flat Kähler manifold. The $E_8$ gauge curvature form is equated to the curvature form of $K$, which takes its values in $\mathcal{SU}(3)$, the Lie algebra of $SU(3)$. This breaks $E_8$ down to $E_6$. The $E_6$ gauge curvature form vanishes, but its holonomy group is nontrivial, and this further[6] breaks $E_6$ down to some strong + electroweak group.

Now what we have described here is the "local" symmetry-breaking structure. It has clear advantages over the Higgs mechanism: in particular, the objects that are nontrivial in the vacuum are gauge field strengths and holonomy groups, and these are items that are implicitly present in any gauge theory. But now we must ask: how should we interpret all this at the global level? We raise this point for two main reasons. First, the theory of magnetic monopoles and instan-

 0022-2488/89/020498-08$02.50 © 1989 American Institute of Physics 

tons, and many subsequent developments, have shown the importance of global questions in gauge theory. Ultimately, for example, the study of monopoles (which involves global aspects of symmetry breaking) in the superstring context will require a global formulation of symmetry breaking by holonomy groups. Second, the method itself cannot be properly implemented unless due attention is paid to global questions, since it is indeed the holonomy *groups*, not the corresponding curvatures, that break symmetries.

The Higgs mechanism gives rise to a definite fiber bundle structure for the gauge vacuum. What is the analogous structure here? The central technical question is: what does it mean to "equate" the linear curvature of $K$ to the gauge curvature, when these forms are defined, *a priori*, on two different bundles? Initially we shall attempt to formulate this along the lines of the Higgs mechanism. The result is not very satisfactory, and so we shall subsequently propose a different approach.

## II. SYMMETRY BREAKING BY HOLONOMY GROUPS: SUBBUNDLE FORMULATION

Let $S$ be the $E_8$ principal bundle over $K$, and let $H_K$ be the holonomy bundle of the linear connection on $K$. If we are to equate the curvatures on these bundles, we must relate them in some way. There are many ways of doing so, as we shall see later. Of these, the closest possible relationship is obtained by taking $H_K$ to be a subbundle of $S$. The process of "equating the curvatures" undoubtedly reflects some very deep property of gravitation in the superstring context, and so we may feel justified in making this assumption. Proceeding on this basis, we assume that the embedding of the holonomy group of $K$ (denoted $\Phi_K$) in $E_8$ has been specified. Now the linear connection of $K$ is a connection on $H_K$: so we must ask how this connection is related to the one on $S$. Again we must stress that there is no single way of establishing such a relationship. The simplest procedure (which we shall be compelled to modify later, however) is as follows.

As stated earlier, subbundles are embedded in bundles by means of a particular type of bundle homomorphism. It is one of the fundamental properties of connections that they can be "pushed forward" by bundle homomorphisms. That is, if $(P_1,M_1,G_1)$ and $(P_2,M_2,G_2)$ are principal bundles, and if $\phi: P_1 \to P_2$, $\psi: G_1 \to G_2$ define a bundle homomorphism that induces a diffeomorphism $M_1 \to M_2$, then any connection form $\omega_1$ on $P_1$ determines a connection form $\omega_2$ on $P_2$, the two being related[1] by $\phi^*\omega_2 = \tilde{\psi}\omega_1$, where $\tilde{\psi}$ is the algebra homomorphism induced by $\psi$. Thus any connection on a subbundle automatically induces a connection on the larger bundle. Hence if $H_K$ is a subbundle of $S$, then the linear connection on $K$ induces a connection on $S$, and it now makes sense to equate this induced connection to the $E_8$ gauge connection on $S$. This provides a very simple rigorous formulation of the breakdown of the $E_8$ symmetry.

Problems begin to arise, however, when we come to consider the residual gauge symmetry. As we shall explain later, this "unbroken" group is a group of vertical automorphisms (i.e., homomorphisms of $S \to S$ that induce the identity map on $K$), isomorphic to the centralizer of $\Phi_K$ in $E_8$ [denoted $C(\Phi_K)$]. This group is our grand unification group,[5] and so we must have yet another gauge bundle, say $(R,K,C(\Phi_K))$ over $K$, and this bundle too should be related to $S$ in some way. Again, we should prefer it to be a subbundle of $S$. But now we are in danger of burdening $S$ with too many conditions. We have already assumed that $S$ admits $H_K$ as a subbundle, for reasons explained earlier; this is relatively harmless, since in the last resort we know that such an $S$ can *always* be constructed once the embedding of $\Phi_K$ in $E_8$ is specified. This method does not, however, allow us to construct a bundle with *two* different given bundles as subbundles. Nor is it the case that, in general, a principal bundle $(P,M,G)$ with a connection having holonomy group $H$ must admit a $C(H)$ subbundle. (As a counterexample, suppose that $P$ is connected, that $M$ is paracompact, and that $\dim M > 1$. Then there exists[1] a connection on $P$ with holonomy group $G$. If the above statement were true, then every such bundle would admit a subbundle with structural group $Z(G)$, the center of $G$; in particular, if $G$ is centerless [i.e., $Z(G) = 1$, the group consisting of a single element], every $G$ bundle would be trivial. Thus, for example, since $E_8$ is centerless, all $E_8$ bundles would be trivial over paracompact manifolds. This is nonsense.) In short, there is a price to be paid for dispensing with the Higgs fields: no longer can we guarantee the existence of the desired global gauge structure, at least not if we use the present formulation.

For the moment, however, let us assume that the $E_8$ bundle $S$ does admit both $H_K$ and $R$ as subbundles, and that the gauge connection on $S$, $\omega_S$, is induced by the linear connection $\omega_K$ on $H_K$. Then how is the gauge connection on $R$ determined? Although connections can be "pushed," they cannot in general be pulled back; nevertheless, we can proceed as follows. Let $(P,M,G)$ be a principal bundle with a connection $\omega_P$, and let $(Q,M,H)$ be a subbundle. Let $\mathcal{G}$, $\mathcal{H}$ be the corresponding Lie algebras, and express $\mathcal{G}$ as $\mathcal{G} = \mathcal{H} \oplus \mathcal{M}$. Let $\omega_Q$ be the $\mathcal{H}$ component of the restriction of $\omega_P$ to $Q$. (Here, "restriction" means that $\omega_P$ is to be evaluated only on vectors that are tangential to $Q$.) Then we have the following result.

*Lemma 1:* Let either $G$ or $H$ be compact. Then $\omega_Q$ defines a connection on $Q$.

*Proof:* It is easy to show[1] that $\omega_Q$ defines a connection on $Q$ provided that $\mathrm{Ad}(H)\mathcal{M} = \mathcal{M}$. Suppose that $G$ is compact. Then[7] $G$ admits a bi-invariant metric, hence $\mathcal{G}$ admits an $\mathrm{Ad}(G)$ invariant inner product. Define $\mathcal{M}$ as the orthogonal complement of $\mathcal{H}$, and let $f \in \mathcal{H}$, $m \in \mathcal{M}$. Then if $h \in H$, $\langle f, \mathrm{Ad}(h)m \rangle = \langle \mathrm{Ad}(h^{-1})f, m \rangle = 0$, hence $\mathrm{Ad}(h)m \in \mathcal{M}$. Suppose instead that $H$ is compact. Then from the representation theory of compact groups[8] it follows that $\mathrm{Ad}(H)$ is completely reducible to a direct sum of irreducible representations; this yields the stated result.

Here, and in nearly all other applications, we use only compact groups, and so the above construction does yield a connection on the subbundle. (Probably the only physically interesting example of the failure of this construction is in the gauge theory of the Poincaré group,[9] which is neither compact nor semisimple.)

Let us assume, then, that the (vacuum) gauge connection on $R$ is obtained as the $\mathcal{C}$ component of the restriction of $\omega_S$ to $R$. [Here, $\mathcal{C}$ is the algebra of $C(\Phi_K)$.] Now, ac-

cording to our general description of gauge symmetry breaking in superstring theory, this connection (denoted $\omega_R$) should be flat, i.e., it should have vanishing curvature. Let us attempt to prove this.

Let $p \in H_K$ and $\bar{p} \in R$ be two points in the same fiber of $S$. Then there exists $g \in E_8$ such that $\bar{p} = pg$. (Here, of course, we are identifying $H_K$ and $R$ with their images in $S$, as usual.) Let $\bar{X}$ be any tangent vector (to $R$) at $\bar{p}$, and let $X$ be the unique tangent vector at $p$ such that $R_{g*}X = \bar{X}$, where $R_g$ denotes the action of $E_8$ on $S$. This $X$ certainly exists, but it may *not* be tangential to $H_K$—an important point. Now we have

$$\omega_R(\bar{p})(\bar{X}) = \mathscr{C}\,\omega_S(pg)(R_{g*}X)$$
$$= \mathscr{C}\ \mathrm{Ad}(g^{-1})\omega_S(p)(X),$$

where $\mathscr{C}$ denotes "$\mathscr{C}$ component of." Care is required at this point: $\omega_S(p)$ equals $\omega_K(p)$ only when acting on vectors tangential to $H_K$, and we do not know whether this is true of $X$. The correct procedure is as follows. If $\mathscr{E}_8$ denotes the algebra of $E_8$, then we have a decomposition $\mathscr{E}_8 = \mathscr{SU}(3) \oplus \mathscr{C} \oplus \mathscr{N}$, where $\mathscr{SU}(3)$ is the Lie algebra of the holonomy group $\Phi_K$ [which, as we shall see later, need not be globally isomorphic to SU(3)], $\mathscr{C}$ is as above, and $\mathscr{N}$ is the orthogonal complement (as in the proof of Lemma 1). We now have the following lemma.

*Lemma 2:* Let $X$ be a tangent vector to $S$ at $p$. Then if $X^0$ is the component of $X$ that is tangential to $H_K$, we have $X = X^0 + A^*(p) + B^*(p)$, where $A \in \mathscr{C}$, $B \in \mathscr{N}$, and where the asterisk denotes the usual algebra homomorphism induced by the action of $E_8$ on $S$.

*Proof:* Let $\{e_\mu,\ \mu = 1,...,6\}$ be a basis of the tangent space at $x = \pi(p) \in K$, let $\{e_i,\ i = 1,...,248\}$ be a basis of $\mathscr{E}_8$, and let $\sigma$ be a local cross section of $H_K$ with $\sigma(x) = p$. Then $\{\sigma_*(e_\mu), e_i^*(p)\}$ gives a basis of the $S$ tangent space at $p$. The result now follows by choosing $\{e_i\}$ in the obvious way.

Using this lemma, we find

$$\omega_S(p)(X) = \omega_K(p)(X^0) + A + B,$$

and so

$$\omega_R(\bar{p})(\bar{X}) = \mathscr{C}\ \mathrm{Ad}(g^{-1})[\omega_K(p)(X^0) + A + B].$$

Now, in fact, $\mathrm{Ad}(g^{-1})\,\omega_K(p)(X^0)$ has no $\mathscr{C}$ component. To see this, note that $E_8$ is connected, hence it suffices to show that $[u, \omega_K(p)(X^0)]$ is orthogonal to $\mathscr{C}$ (for all $u \in \mathscr{E}_8$) with respect to the Cartan–Killing form $k$ on $\mathscr{E}_8$. (Here $k$ is nondegenerate and indeed negative definite, since $E_8$ is compact and simple.) This is straightforward:

$$k([u, \omega_K(p)(X^0)], c) = k(u, [\omega_K(p)(X^0), c]) = 0,$$

for any $c \in \mathscr{C}$, since $[\mathscr{SU}(3), \mathscr{C}] = 0$. (Note that $\mathscr{SU}(3)$ and $\mathscr{C}$ are in fact orthogonal with respect to $k$; this can be shown easily by noting that any semisimple Lie algebra equals its own derived algebra, so $[\mathscr{SU}(3), \mathscr{SU}(3)] = \mathscr{SU}(3)$.)

Thus $\omega_K$ does not contribute directly to $\omega_R$; we have simply $\omega_R(\bar{p})(\bar{X}) = \mathscr{C}\ \mathrm{Ad}(g^{-1})(A + B)$. This might lead one to expect that $\omega_R$ must be flat. In order to compute the curvature of $\omega_R$, we need the following result.

*Lemma 3:* Let the notation be as in Lemma 1. Let $q \in Q$ and let $X, Y$ be tangential to $Q$ at $q$. If $G$ is compact (so that $\omega_Q$ is a connection) let $\Omega_Q$ be the curvature of $\omega_Q$. Then, with an obvious notation,

$$\Omega_Q(q)(X,Y) = \mathscr{H}\Omega_P(q)(X,Y)$$
$$-\tfrac{1}{2}\mathscr{H}[\mathscr{M}\omega_P(q)(X), \mathscr{M}\omega_P(q)(Y)].$$

*Proof:* A straightforward computation using the structural equation for $\omega_P$, together with $[\mathscr{H}, \mathscr{M}] \subseteq \mathscr{M}$ [which follows from $\mathrm{Ad}(H)\mathscr{M} = \mathscr{M}$—see Lemma 1].

Now, in our case, this lemma implies that $\omega_R$ *need not* be flat. For if we let $\bar{X}, \bar{Y}$ be tangential to $R$ at $\bar{p} = pg$, then

$$\mathscr{C}\,\Omega_S(\bar{p})(\bar{X},\bar{Y}) = \mathscr{C}\ \mathrm{Ad}(g^{-1})\Omega_S(p)(X,Y)$$
$$= \mathscr{C}\ \mathrm{Ad}(g^{-1})\Omega_S(p)(X^0,Y^0),$$

where we have used Lemma 2 and noted that curvature forms annihilate vertical vectors. But $\Omega_S(p)(X^0,Y^0) = \Omega_K(p)(X^0,Y^0)$, which takes its values in $\mathscr{SU}(3)$; hence by the reasoning used above, $\mathrm{Ad}(g^{-1})\Omega_S(p)(X^0,Y^0)$ has no $\mathscr{C}$ component. Hence the term corresponding to $\mathscr{H}\Omega_P(q)(X,Y)$ in Lemma 3 is indeed absent here. But the second term on the right-hand side may not be zero. It will simplify matters, without altering our essential point, if we take $p = \bar{p}$ (i.e., we assume that $H_K$ and $R$ intersect at this point). Then setting

$$X = X^0 + A_x^*(p) + B_x^*(p),$$
$$Y = Y^0 + A_y^*(p) + B_y^*(p),$$

where $A_x, A_y \in \mathscr{C}$, and $B_x, B_y \in \mathscr{N}$, we find that the object corresponding to $\mathscr{M}\omega_p(q)(X)$ is just $\omega_K(p)(X^0) + B_x$, and so the term in question is

$$-\tfrac{1}{2}\mathscr{C}[\omega_K(p)(X^0) + B_x, \omega_K(p)(Y^0) + B_y].$$

We know that, for all $u \in \mathscr{E}_8$, $[u, \omega_K(p)(X^0)]$ has no $\mathscr{C}$ component; so this reduces to $-\tfrac{1}{2}\mathscr{C}[B_x, B_y]$, and Lemma 3 gives us $\Omega_R(p)(X,Y) = -\tfrac{1}{2}\mathscr{C}[B_x, B_y]$, which is not zero in general. (We can certainly find $n_1, n_2 \in \mathscr{N}$ such that $[n_1, n_2]$ has a nonzero $\mathscr{C}$ component. But, suppose the contrary. Then for all $c \in \mathscr{C}$ and $n_1, n_2 \in \mathscr{N}$, we would have

$$k(c, [n_1, n_2]) = 0 = k([c, n_1], n_2)$$

(where $k$ is the $\mathscr{E}_8$ Cartan–Killing form) so that $[c, n_1]$ has no $\mathscr{N}$ component. Similarly it has no $\mathscr{SU}(3)$ component, so it is an element of $\mathscr{C}$. But then $[c, u] \in \mathscr{C}$ for all $u \in \mathscr{E}_8$, which contradicts the fact that $\mathscr{E}_8$ is simple.)

We conclude, therefore, that flat connections on bundles do not necessarily induce flat connections on subbundles. Of course, the induced connection *may* be flat, but it does not seem to be possible to ensure this in any particularly natural way. Again we find that this feature of the vacuum has to be introduced as an additional assumption.

It would seem, then, that the usual formulation of symmetry breaking in terms of subbundles is not satisfactory in the present context. The assumption that the grand unification bundle is a subbundle of the $E_8$ bundle is particularly troublesome. A more flexible formulation will now be proposed.

## III. SYMMETRY BREAKING BY HOLONOMY GROUPS: BUNDLE SPLICING

Before proceeding to our proposal, let us consider the following points.

(a) It is desirable that the formulation be *constructive:* we do not wish to make ad hoc assumptions as to the existence of the structures to be employed.

(b) The symmetry breaking proceeds in two stages: first, $E_8$ is broken to the grand unification group $C(\Phi_K)$, and then $C(\Phi_K)$ is broken by some (discrete) holonomy group. The transition from the first stage to the second will change the holonomy group of the connection on the $E_8$ bundle, but should not disturb the "equality" of linear curvature with gauge curvature. During the first stage, the holonomy group of the grand unification bundle must be trivial, and the holonomy group of the $E_8$ bundle should be just $\Phi_K$. This is necessary for internal consistency, since we always suppose [see (d) below] that the grand unification group is $C(\Phi_K)$.

(c) We shall need some analog of Lemma 1 in order to relate connections on various bundles.

(d) We state here some technical results that are used repeatedly below.

Underlying all approaches to symmetry breaking of this type is the idea that if a gauge field is nontrivial in the vacuum, then the symmetry is broken to the subgroup that commutes with the corresponding holonomy group. The formal statement is as follows.

*Proposition 4:* Let $(P,M,G)$ be a principal bundle with connection form $\omega$, having holonomy group $\Phi$. Then the group of vertical automorphisms $\mu$ that preserve $\omega$ (in the sense that $\mu^*\omega = \omega$) is isomorphic to the centralizer, $C(\Phi)$, of $\Phi$ in $G$.

*Proof:* See Refs. 10 and 11.

Thus the residual gauge symmetry is indeed $C(\Phi_K)$ in our case, where as before $C(\Phi_K)$ denotes the centralizer of $\Phi_K$ in $E_8$. Note in particular that the formal result pertains to the holonomy *group*, not to the curvature.

A second technical point, which cannot be fully analyzed here, concerns the structure of $\Phi_K$, the holonomy group of $K$. As is well known, the *connected component* of $\Phi_K$ (the *restricted* linear holonomy group of $K$) is isomorphic to $SU(3)$. But this information is not sufficient to compute $C(\Phi_K)$: we need the full global structure. The relevant result is as follows.

*Proposition 5:* Let $K$ be a compact $n$-dimensional Kähler manifold with nonzero Euler characteristic and vanishing Ricci tensor. Then the holonomy group of the linear connection generated by the Kähler metric is contained in

$$S_m U(n) = \{u \in U(n) \text{ such that } \det u \in \mathbb{Z}_m\},$$

where $m$ is a fixed integer such that there exists a homomorphism from the fundamental group of $K$ onto $\mathbb{Z}_m$.

*Proof:* See Ref. 12.

In our case, of course, the group will be $S_m U(3)$, with $SU(3)$ as the special case $m = 1$. The distinction between $SU(3)$ and $S_m U(3)$ as holonomy groups (over which we have no control—it depends on the structure of $K$) is of great importance because these groups have *different* centralizers in $E_8$. The computation of centralizers is an intrinsically global problem; elementary techniques for dealing with it will be discussed elsewhere. For the present the following will suffice.

*Proposition 6:* If $C(\ )$ denotes centralizers in $E_8$, then

$$C(SU(3)) = E_6,$$

$$C(S_m U(3)) = U(1) \times SO(10), \quad m \neq 1,$$

where equality means global isomorphism.

*Proof:* Reference 13.

These, then, are our candidates for grand unification groups. Note that while $SU(3)$ and $S_m U(3)$ are locally isomorphic—indeed, the former is the connected component of the latter, for all $m$—the same is not true of their centralizers. We shall continue to denote the holonomy group of $K$ by $\Phi_K$, and its centralizer by $C(\Phi_K)$; bear in mind that $C(\Phi_K)$ may not be $E_6$, and that it may not be semisimple.

One final technical point remains to be discussed; it is very elementary, but is so often neglected that a discussion may be justified. Let $G_1$ and $G_2$ be any two groups, and let $\psi$: $G_1 \to G_2$ be a homomorphism with kernel Ker $\psi$ and image $\psi(G_1)$. Then the homomorphism theorem of elementary group theory states that Ker $\psi$ is a normal subgroup of $G_1$, and that $G_1/\text{Ker } \psi$ is isomorphic to $\psi(G_1)$, which is, of course, a subgroup of $G_2$. To see the importance of this for our purposes, let $G$ be any group with a subgroup $H$, and let $C(H)$ be the centralizer of $H$ in $G$ [*or* any subgroup of $C(H)$ containing the center of $H$]. Now the map $\psi$: $H \times C(H) \to G$ given by $\psi$: $(h,c) \to hc$ is a homomorphism with kernel consisting of pairs $(z,z^{-1})$ for all $z \in Z(H)$, the center of $H$. This group is isomorphic to $Z(H)$, and so we find that $[H \times C(H)]/Z(H)$ is a subgroup of $G$. For example, the centralizer of $SU(n)$ in $U(n)$ is $U(1)$. This $U(1)$ has subgroups, $\mathbb{Z}_{mn}$, containing the center, $\mathbb{Z}_n$, of $SU(n)$; hence, for all $m$, the groups $[\mathbb{Z}_{mn} \times SU(n)]/\mathbb{Z}_n$ are subgroups of $U(n)$. These are, in fact, precisely the subgroups denoted by $S_m U(n)$ earlier.

Now suppose that $H \times C(H)$ is connected and that $Z(H)$ is finite (as is often the case when $H$ and $G$ are Lie groups). Then $H \times C(H)$ is *not a subgroup* of $G$. For in that case $H \times C(H)$and $[H \times C(H)]/Z(H)$ are locally isomorphic and are both connected, and one knows that a given subalgebra of the Lie algebra of a Lie group can be the Lie algebra of only one connected subgroup.[14] Thus, in general, if we wish to use $H$ and $C(H)$ to construct a subgroup of $G$, it is not sufficient merely to take $H \times C(H)$: we must also factor out the center.

We may now begin to construct the global background for this approach to symmetry breaking.

### A. The construction

One of the drawbacks of the subbundle formulation was the difficulty of constructing a principal bundle having two specified bundles as subbundles. As before, let $H_K$ be the linear holonomy bundle of $K$, and let $(R,K,C(\Phi_K))$ be the grand unification bundle. Then the product manifold $H_K \times R$ is a principal $\Phi_K \times C(\Phi_K)$ bundle over $K \times K$. If we restrict ourselves to the submanifold consisting of pairs $(u,r) \in H_K \times R$ such that $\pi_H(u) = \pi_R(r)$ (where $\pi_H$, and $\pi_R$ are the respective projections of $H_K$ and $R$), then we obtain[15] what is often called the "spliced" bundle, $H_K + R$. This is a $\Phi_K \times C(\Phi_K)$ bundle over the diagonal subspace of $K \times K$, the set of pairs $(x,x) \in K \times K$, which is obviously diffeomorphic to $K$.

Having manufactured a *single* bundle, $H_K + R$, from $H_K$ and $R$, our objective now is to extend to an $E_8$ bundle in the ordinary way. But this cannot be done directly, since, as we already know, $\Phi_K \times C(\Phi_K)$ is not usually a subgroup of $E_8$. (For example, $[SU(3) \times E_6]/\mathbb{Z}_3$ is a subgroup of $E_8$, but $SU(3) \times E_6$ is not.) We need the following lemma.

*Lemma 7:* Let $(P,M,G)$ be a principal bundle, and let $Z$ be any finite normal subgroup of $G$. Then $(P/Z, M, G/Z)$ is a principal bundle.

*Proof:* This $P/Z$ can be constructed from its transition functions, defined as the composites of the transition functions of $P$ with the projection $G \to G/Z$. Verification of the cyclic condition on the transition functions is straightforward. The action of $G/Z$ on $P/Z$, given by $gZ: pZ \to pgZ$, is free because the action of $G$ on $P$ is free.

In our case, $Z(\Phi_K)$ is always finite, and so if we denote $[\Phi_K \times C(\Phi_K)]/Z(\Phi_K)$ by $\Phi_K \cdot C(\Phi_K)$, and $[H_K + R]/Z(\Phi_K)$ by $H_K \cdot R$, then by Lemma 7 we find that $H_K \cdot R$ is a $\Phi_K \cdot C(\Phi_K)$ bundle over $K$. This latter group is indeed a subgroup of $E_8$, and so we can now extend $H_K \cdot R$ to an $E_8$ bundle $(S,K,E_8)$ in the usual way (so that $H_K \cdot R$ is a subbundle of $S$.) This is how we propose to construct the gauge vacuum for superstring theory.

Before moving on, we should point out that the admittedly convenient notation $H_K \cdot R$ may be misleading if it suggests that $H_K$ and $R$ are necessarily subbundles of $H_K \cdot R$. That is not the case. A full analysis of this question is not necessary here, but some discussion is required because the relationship between these bundles has a bearing on the corresponding connection theory.

Let $F$ and $H$ be commuting subgroups of a group $G$, and let $\theta$: $H \to F$ be a homomorphism; then the map $h \to h\theta(h)$ is a homomorphism from $H$ into $H \cdot F$. The kernel is a subgroup of the center of $H$ consisting of elements $z$ with $\theta(z) = z^{-1}$. Thus the homomorphism will embed $H$ in $H \cdot F$ as a subgroup if and only if $\theta(z) = z^{-1} \Rightarrow z =$ identity, for all $z \in H \cap F$. There is always at least one such homomorphism (the one that maps all elements of $H$ to the identity), but usually there will be many others, since the restriction on $\theta$ is rather weak.

Now we can try to adapt this idea to show that, if $(P,M,H)$ and $(Q,M,F)$ are principal bundles over a manifold $M$, then $P$ can be a subbundle of $P \cdot Q$. Let $\theta$: $P \to Q$ be a bundle homomorphism with corresponding group homomorphism also denoted $\theta$. Suppose that $\theta$ is vertical (i.e., that it induces the identity map on $M$) and that $\theta$: $H \to F$ satisfies $\theta(z) = z^{-1} \Rightarrow z =$ identity for all $z \in H \cap F$. Then the map $h \to h\theta(h)$ is a monomorphism of $H$ into $H \cdot F$, as above. Define $\bar{f}$: $P \to P + Q$ by $\bar{f}(p) = (p, \theta(p))$, and let

$$\pi: P + Q \to P \cdot Q = [P + Q]/H \cap F$$

be the projection. Then the map $f$: $P \to P \cdot Q$ defined by $f = \pi \circ \bar{f}$ is a vertical homomorphism. Thus $P$ is a subbundle of $P \cdot Q$, just as $H$ is a subgroup of $H \cdot F$.

There is one major difference between the two cases, however: whereas a homomorphism from $H$ to $F$ always exists, the same is not true of homomorphisms from $P$ to $Q$.

*Lemma 8:* Let $\theta$: $(P,M,H) \to (Q,M,F)$ be a vertical homomorphism. Then $Q$ is reducible to a subbundle with structural group $\theta(H)$.

*Proof:* Given $x \in M$, let $p$ be any element of $P$ with $\pi_P(p) = x$. Then $x \to \theta(p)\theta(H)$ is a well-defined global cross section of $Q/\theta(H)$, the associated bundle of $Q$ with standard fiber $F/\theta(H)$. Hence $Q$ is reducible to a $\theta(H)$ subbundle.

Reducibility to a $\theta(H)$ subbundle is a very severe restriction on $Q$. In general, therefore, one cannot expect $P$ to be a subbundle of $P \cdot Q$. (The same applies, of course, to $Q$.) Returning to the case of $H_K \cdot R$, take the simplest case, $\Phi_K = SU(3)$, $C(\Phi_K) = E_6$. Then it is not difficult to show that there exist *no* bundle homomorphisms from $R$ to $H_K$, since $E_6$ is a simple group, and thus its only normal subgroups are $E_6$, $\mathbb{Z}_3$ (its center), and 1 (the group consisting of one element). By the homomorphism theorem, then, the only homomorphism from $E_6$ to $SU(3)$ is the one that maps all elements to the identity [since obviously $E_6$ and $E_6/\mathbb{Z}_3$ cannot be subgroups of $SU(3)$]. But then Lemma 8 implies that $H_K$ is trivial. This is impossible, since we always assume that the Euler characteristic of $K$ is nonzero (it is proportional to the number of particle generations), whereas all characteristic classes of the frame bundle of $K$ would vanish if $H_K$ were trivial. Hence we certainly cannot expect $R$ to be a subbundle of $H_K \cdot R$. On the other hand, the possibility that $H_K$ could be a subbundle of $H_K \cdot R$ is not so completely obstructed, since there is nothing to prevent the gauge bundle from being (for example) trivial; indeed, an important example does occur later, in Proposition 12. However, this is a special case. In general, we must *not* assume that either $H_K$ or $R$ is a subbundle of $H_K \cdot R$. How, then, are connections on these various bundles related?

### B. Construction of the $E_8$ connection

Let $\omega_K$ be the linear connection on $K$, and let $\omega_R$ be any connection on $R$. As we do not have homomorphisms from $H_K$ or $R$ into $H_K + R$, neither of these connections alone induces a connection on $H_K + R$. However, we can use a combination of the two, as follows. Define $f_H$: $H_K + R \to H_K$ by $f_H$: $(h,r) \to h$, and similarly $f_R$: $H_K + R \to R$; these are bundle homomorphisms in an obvious way. Then it is easy to show[1] that $f_H^* \omega_K + f_R^* \omega_R$ [where the $+$ denotes addition in the Lie algebra of $\Phi_K \times C(\Phi_K)$] is a connection form on $H_K + R$.

Now, as before, let $\pi$: $H_K + R \to H_K \cdot R$ be the projection; it, too, is a bundle homomorphism with corresponding group homomorphism $\pi$: $\Phi_K \times C(\Phi_K) \to \Phi_K \cdot C(\Phi_K)$. (This last is in fact a covering homomorphism, and so the algebra homomorphism $\tilde{\pi}$ is an isomorphism in our case.) Thus $\pi$ "pushes" $f_H^* \omega_K + f_R^* \omega_R$ to a connection $\omega_{HR}$ on $H_K \cdot R$. Finally, $H_K \cdot R$ is a subbundle of the $E_8$ bundle $S$, and so $\omega_{HR}$ induces a connection on $S$.

To summarize, then, we have a well-defined procedure whereby the $E_8$ gauge connection is generated by $\omega_K$ and $\omega_R$; this procedure parallels the construction of $S$. In fact, the process works both ways, because, since $E_8$ is compact, any connection on $S$ generates one on $H_K \cdot R$ (Lemma 1). We now need the following result.

*Proposition 9:* Let $(P,M,G)$ be a principal bundle, let $Z$

be a finite subgroup of the center of $G$, and let $\omega$ be a connection form on $P/Z$. If $\pi: P \to P/Z$, $\pi: G \to G/Z$ is the projection homomorphism, and $\tilde{\pi}$ is the corresponding algebra isomorphism, then $\bar{\omega} = \tilde{\pi}^{-1}\pi^*\omega$ is a connection form on $P$.

*Proof:* Taking the derivative of $\mathrm{Ad}(gZ)\circ\pi = \pi\circ\mathrm{Ad}(g)$ (maps from $G$ to $G/Z$), we obtain $\tilde{\pi}^{-1}\circ\mathrm{Ad}(gZ) = \mathrm{Ad}(g)\circ\tilde{\pi}^{-1}$. Now let $V$ be a tangent vector at $p\in P$, and let $g\in G$. Then

$$\bar{\omega}(R_g V) = \tilde{\pi}^{-1}\omega(\pi_* R_g V) = \tilde{\pi}^{-1}\omega(R_{gZ}\pi_* V),$$

since $\pi\circ R_g = R_{gZ}\circ\pi$. Thus

$$\bar{\omega}(R_g V) = \tilde{\pi}^{-1}\,\mathrm{Ad}(g^{-1}Z)\omega(\pi_* V) = \mathrm{Ad}(g^{-1})\tilde{\pi}^{-1}\omega(\pi_* V) = \mathrm{Ad}(g^{-1})\bar{\omega}(V).$$

Next, let $A^*(p)$ be a vertical vector at $p$, where $A$ is a tangent vector at the identity of $G$. If $\sigma_p$: $G\to P$ is defined by $\sigma_p$: $g\to pg$, then (Ref. 1) $A^*(p) = \sigma_{p*}A$. Thus

$$\bar{\omega}(A^*(p)) = \tilde{\pi}^{-1}\omega(\pi_*\sigma_{p*}A).$$

But, for any $g\in G$,

$$\pi(\sigma_p g) = pgZ = pZgZ = \sigma_{\pi(p)}(\pi g),$$

so that

$$\bar{\omega}(A^*(p)) = \tilde{\pi}^{-1}\omega((\tilde{\pi}A)^*(\pi(p))) = A.$$

Any algebra-valued one-form with these properties defines a connection on $P$, so this completes the proof.

Returning to our case, we see that the connection on $H_K\cdot R$ induces a connection on $H_K + R$. Finally, the homomorphisms $f_H$ and $f_R$ can be used to push this connection to $H_K$ and $R$ separately.

We conclude this section with a result on holonomy groups.

*Proposition 10:* Let the connection $\omega_S$ on $S$ be constructed, as above, from $\omega_K$ and $\omega_R$. Then the holonomy group of $\omega_S$ is a subgroup of $[\Phi_K\times\Phi_R]/\Phi_K\cap\Phi_R$, where $\Phi_R$ is the holonomy group of $\omega_R$.

*Proof:* It can be shown[1] that the holonomy group of the connection on $H_K + R$ is a subgroup of $\Phi_K\times\Phi_R$, while the holonomy group of $\omega_{HR}$ is the image of this subgroup upon projection by $\pi$: $\Phi_K\times C(\Phi_K)\to\Phi_K\cdot C(\Phi_K)$, that is, it is a subgroup of $\pi(\Phi_K\times\Phi_R) = [\Phi_K\times\Phi_R]/\Phi_K\cap\Phi_R$. The result now follows from the fact that, by its construction, $\omega_S$ is reducible to $\omega_{HR}$.

*Remark:* Note the word *subgroup*: one cannot, in general, prove that the groups are equal, though of course that can happen. (The problem, essentially, is that a disconnected Lie group has proper subgroups locally isomorphic to itself.)

## IV. THE SYMMETRY BREAKING "MECHANISM"

As remarked at the beginning of Sec. III, the symmetry breaking proceeds in two stages: first $E_8\to E_6$ [or $U(1)\times SO(10)$], and then down to the strong + electroweak group.

During *both* stages of symmetry breaking, the linear curvature of $K$ is to be "equated" to the $E_8$ gauge curvature. What can this mean? Since the connection on $S$ is reducible to $\omega_{HR}$, we can concentrate on this last. Its curvature form, $\Omega_{HR}$, is a form on $H_K\cdot R$. The linear curvature of $K$, $\Omega_K$ is a form on $H_K$. Obviously $\Omega_{HR}$ cannot be equated to $\Omega_K$, but we can proceed instead as follows. There are two homomorphisms defined on $H_K + R$, namely, $\pi$: $H_K + R\to H_K\cdot R$ and $f_H$: $H_K + R\to H_K$. The two-forms $\pi^*\Omega_{HR}$ and $f_H^*\Omega_K$ now have the same domain, and they will also take their values in the same algebra [that of $\Phi_K\cdot C(\Phi_K)$] if we consider $\tilde{\pi}f_H^*\Omega_K$ instead of $f_H^*\Omega_K$. We therefore propose to interpret "equating the linear curvature of $K$ to the gauge curvature of $E_8$" to mean

$$\pi^*\Omega_{HR} = \tilde{\pi}f_H^*\Omega_K.$$

This equation has the following welcome consequence.

*Lemma 11:* The above equation is valid if and only if the connection on $R$ is flat.

*Proof:* The curvature of the connection $f_H^*\omega_K + f_R^*\omega_R$ on $H_K + R$ is just $f_H^*\Omega_K + f_R^*\Omega_R$. Now, by definition, $\omega_{HR}$ is induced by this connection via $\pi$; hence, in general, we have $\pi^*\Omega_{HR} = \tilde{\pi}[f_H^*\Omega_K + f_R^*\Omega_R]$. [It is worth noting at this point that, because $\mathcal{SU}(3)$ is simple, the algebra of $\Phi_K\times C(\Phi_K)$ is the *direct* sum of the respective algebras, so that $\tilde{\pi}f_H^*\Omega_K$ is the $\mathcal{SU}(3)$ component of the right side of this equation.] Suppose now that the stated equation holds. Then since $\tilde{\pi}$ is an isomorphism, $f_R^*\Omega_R = 0$. Now $f_R$ is surjective: given any $r\in R$, let $h$ be any element of $\pi_H^{-1}(x)$, where $x = \pi_R(r)$. Then $(h,r)\in H_K + R$ is projected to $r$ by $f_R$. Thus $f_{R*}$ is surjective, and so $\Omega_R = 0$. The converse is obvious.

Thus, within the present formulation, the process of "equating the curvatures" does indeed force the $R$ connection to be flat. On the other hand, *any* flat connection on $R$ will be satisfactory, even if the holonomy group is not trivial. Proposition 10 shows that if the holonomy group of $\omega_R$ is not trivial, then the holonomy group of $\omega_S$ may not coincide with $\Phi_K$, even though the curvatures have been "equated."

We may now describe the first stage of symmetry breaking, in which $E_8$ breaks to the grand unification group. Here, the curvatures are "equated," and, *in addition*, $\Phi_K$ is trivial. In this case, the holonomy group of $\omega_S$ must coincide precisely with $\Phi_K$—otherwise the whole method would not be consistent, since we have always assumed that the grand unification group is $C(\Phi_K)$. Proposition 10 is insufficiently precise to allow us to verify this; we need the following result.

*Proposition 12:* If $\Phi_R$ is trivial and $K$ is connected and paracompact, then the holonomy bundles of $\omega_S$ are isomorphic to $H_K$, and its holonomy group is precisely $\Phi_K$.

*Proof:* We begin by proving that, in this particular case, $H_K$ is a subbundle of $H_K\cdot R$ (and therefore of $S$). Given any $r\in R$, let $P_r$ be the submanifold of $R$ consisting of all points which can be connected to $r$ by a horizontal curve. Then[1] $P_r$ is a principal bundle over $K$, with structural group $1$ = the holonomy group of $\omega_R$. Then given any $x\in K$, there exists a unique $\sigma(x)\in P_r$ with $\pi_R\sigma(x) = x$. Clearly, $x\to\sigma(x)$ defines a global cross section of $R$ (which must therefore be trivial). If $V$ is any tangent vector at $x$, then $\sigma_* V$ is horizontal; hence $\sigma^*\omega_R = 0$.

Now, as in the discussion before Lemma 8 above, define a bundle homomorphism $\bar{f}$: $H_K\to H_K + R$ by $\bar{f}(u) = (u,\sigma(\pi_H u))$, with corresponding group homomorphism $\Phi_K\to\Phi_K\times C(\Phi_K)$ given by $g\to(g,e)$, where $e$ is the identity element of $C(\Phi_K)$. Then the homomorphism $\pi\circ\bar{f} = f$ em-

beds $H_K$ in $H_K \cdot R$ as a subbundle. Now the connection $\omega_{HR}$ on $H_K \cdot R$ is induced by the connection $f_H^* \omega_K + f_R^* \omega_R$ on $H_K + R$, and so we have

$$\pi^* \omega_{HR} = \tilde{\pi}[f_H^* \omega_K + f_R^* \omega_R].$$

Applying $\bar{f}^*$ to both sides, one obtains

$$f^* \omega_{HR} = \tilde{\pi}[(f_H \circ \bar{f})^* \omega_K + (f_R \circ \bar{f})^* \omega_R].$$

The definition of $\bar{f}$ yields $f_H \circ \bar{f} =$ identity map on $H_K$, and $(f_R \circ \bar{f})^* = (\sigma \circ \pi_H)^* = \pi_H^* \circ \sigma^*$, which annihilates $\omega_R$. Hence, noting that the algebra homomorphism corresponding to $g \to (g,e)$ is just $A \to (A,O) = A + 0$, where $A$ is any element of the Lie algebra of $\Phi_K$, we obtain finally $f^* \omega_{HR} = \tilde{f} \omega_K$. That is, if $\Phi_R$ is trivial, then $\omega_{HR}$ is just the connection induced on $H_K \cdot R$ by $\omega_K$ via the embedding of $H_K$ in $H_K \cdot R$ as a subbundle. Hence the holonomy bundles of $\omega_{HR}$ are isomorphic to those of $\omega_K$, that is, to $H_K$. The result now follows from the fact that (by definition) $\omega_S$ is reducible to $\omega_{HR}$.

One more point remains to be considered before we complete our description of the first stage of symmetry breaking. The bundle $H_K$ is "given"—it is a specific subbundle of the unitary frame bundle of $K$. The same is not true of $R$, however; how can we ensure that it exists? This is easily answered at this point. From the proof of Proposition 12, $R$ must be a trivial bundle, so we can simply take $R = K \times C(\Phi_K)$. The connection with $\Phi_R = 1$ can be identified with the canonical flat connection[1] on such a bundle, defined by taking the horizontal subspaces to be tangential to the submanifolds of the form $K \times \{g\}$, $g \in C(\Phi_K)$.

We now turn to the second stage of symmetry breaking, in which $C(\Phi_K)$ is broken to the strong + electroweak group. We need to construct an $R$ with a connection $\omega_R$ which is still flat, but which no longer has a trivial holonomy group $\Phi_R$. The possibilities for $\Phi_R$ are limited by the fact that, for any flat connection, there exists[1] a homomorphism from the fundamental group $\pi_1(K)$ onto the holonomy group; hence $\Phi_R$ must have the form $\pi_1(K)/N$ for some normal subgroup $N$ of $\pi_1(K)$. Henceforth we suppose that some definite choice of $\Phi_R$ [and of its embedding as a subgroup of $C(\Phi_K)$] has been made—one does this, of course, bearing in mind that $\Gamma(\Phi_R)$, the centralizer of $\Phi_R$ in $C(\Phi_K)$, is to be the strong + electroweak group.

Clearly we should construct $R$ in exactly the same way that we constructed the $E_8$ bundle $S$. For this, we need a principal bundle $H_2$ with a connection having holonomy group $\Phi_R$, and another principal bundle $R_2$ (the strong + electroweak gauge bundle) which should be trivial and have a connection with trivial holonomy group. We can take $R_2 = K \times \Gamma(\Phi_R)$, but $H_2$ requires some discussion (since, unlike $H_K$, it is not "given"). It can be shown, from its assumed properties (nonzero Euler characteristic, etc.), that $K$ is paracompact and has a finite fundamental group, and we may assume that it is connected. Let $\bar{K}$ be the universal covering space of $K$: it is (by definition) a connected[16] manifold which can be regarded as a principal bundle over $K$ with structural group $\pi_1(K)$. From Lemma 7 we now find that $\bar{K}/N$ is a principal bundle over $K$ with structural group $\pi_1(K)/N$, for any normal subgroup $N$ of $\pi_1(K)$. Now choose $N$ so that $\pi_1(K)/N$ is the selected group $\Phi_R$—as we know, this can always be done. Set $H_2 = \bar{K}/N$. Then $(H_2, K, \Phi_R)$ is a principal bundle such that (i) $K$ is paracompact and dim $K > 1$, and (ii) $H_2$ is connected. For any such principal bundle, it can be shown[1] that there exists a connection having holonomy group equal to the structural group of the bundle. Thus, we have constructed a bundle with all the desired properties. (Note that we could not have taken $H_2$ simply as $K \times \Phi_R$, for then $H_2$ would not be connected.)

The construction of $R$ is now clear: we define it as the $C(\Phi_K)$ extension of $H_2 \cdot R_2$. This completes our interpretation of gauge symmetry breaking in superstring theory. (Of course, the electroweak group must again be broken, but that occurs in an entirely different way.)

## V. CONCLUSION

The problem of understanding the relationship between gauge theory and gravitation—interpreted as (extended) space-time structure, *not* as "just another field"—is the central question for any unified field theory. Attempts have been made to regard gravitation as gauge theory (of the Poincaré group) and to reduce gauge theory to gravitation (Kaluza–Klein), but neither approach has led to any conspicuous success. One hopes that superstring theory represents a new departure in this respect, but the position remains unclear as yet. There can be no doubt, however, that gravitation has new roles to play in this theory: for example, it is responsible for the breakdown of the $E_8$ symmetry. In this work, we have found that this development requires a radical reformulation of the global aspects of gauge theory, particularly in the description of the breakdown of the grand unification symmetry. We propose to construct the $E_8$ gauge bundle by (i) splicing the linear holonomy bundle of $K$ with the grand unification bundle to obtain a principal bundle $H_K + R$ with structural group $\Phi_K \times C(\Phi_K)$, then (ii) factoring out $Z(\Phi_K)$ to obtain a bundle $H_K \cdot R$ with structural group $\Phi_K \cdot C(\Phi_K)$ which is a subgroup of $E_8$, and finally (iii) extending to an $E_8$ bundle. The linear connection $\omega_K$ on $K$, combined with any gauge connection $R$, gives a well-defined connection on the $E_8$ bundle; the curvature of this connection can be "equated" to that of $\omega_K$ by pulling both back to $H_K + R$, provided that the curvature of the $R$ connection is zero. The bundle $R$ can itself be constructed in a very similar way, though care must be taken regarding questions of existence.

For the sake of concreteness, and because of its topical interest, gauge symmetry breaking in superstring theory has been the main subject of our discussion. Clearly, however, the above construction can be extended to a complete global formulation of symmetry breaking by holonomy groups. The principal peculiarity of the superstring case is the fact that the linear holonomy bundle is "given"—we do not need to construct it. In order to apply the method to (for example) a purely gauge-theoretic context, one would need to investigate a number of questions. If we wish to break a gauge group $G$ down to a subgroup $H$, then we must first ask: does $G$ admit another subgroup, say $J$, such that the centralizer of $J$ in $G$ is equal to $H$? Not every subgroup of a Lie group can be thus represented—for example, it can be shown[13] that if $G = SU(n)$ and $H = SU(m)$, $m < n$, then no

such $J$ exists. In such a case, "symmetry breaking by holonomy groups" will obviously not work. But even if $J$ does exist, we must still ask whether $J$ can be regarded as the holonomy group of some connection on some principal bundle. (This is not trivial, particularly if, as is often the case, $J$ is not connected.) We intend to return to these questions elsewhere.

In conclusion, it should be said that our understanding of the holonomy group approach to symmetry breaking is by no means complete. Here we have been concerned almost exclusively with questions of *construction*—we have not explained, for example, how the embedding of $\Phi_R$ in $C(\Phi_K)$ is selected. This must be decided by physical considerations: see Ref. 17 for some work in this direction.


[1]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[2]C. J. Isham, Phys. Lett. B **102**, 251 (1981).
[3]A. Trautman, Czech. J. Phys. B **29**, 107 (1979).
[4]D. Husemoller, *Fibre Bundles* (Springer, New York, 1975).
[5]P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B **258**, 46 (1985).
[6]E. Witten, Nucl. Phys. B **258**, 75 (1985).
[7]W. Poor, *Differential Geometric Structures* (McGraw-Hill, New York, 1981).
[8]V. S. Varadarajan, *Lie Groups, Lie Algebras, and their Representations* (Prentice–Hall, London, 1974).
[9]B. T. McInnes, Class. Quantum Grav. **1**, 1 (1984).
[10]B. Booss and D. D. Bleecker, *The Atiyah—Singer Index Formula and Gauge—Theoretic Physics* (Springer, New York, 1985).
[11]A. E. Fischer, Commun. Math. Phys. **113**, 231 (1987).
[12]B. T. McInnes, Class. Quantum Grav. **5** (1988).
[13]B. T. McInnes, "Group theoretic aspects of the Hosotani mechanism," N.U.S. Math Preprint 346.
[14]S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978).
[15]D. Bleecker, *Gauge Theory and Variational Principles* (Addison–Wesley, London, 1981).
[16]F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Scott, Foresman, London, 1971).
[17]A. T. Davies and A. McLachlan, Phys. Lett. B **200**, 305 (1988).

# Membranes in string theory, trees, the Weil conjectures, and the Ramanujan numbers

Bernard Grossman
*Rockefeller University, New York, New York 10021*

(Received 16 September 1988; accepted for publication 23 November 1988)

It is suggested that the factorization of the inverse partition function for 26-dimensional bosonic string theory can be interpreted as the existence of an 11-dimensional membrane. The geometry of diffusion on trees is found to be especially useful, with the $p \to 1$ limit corresponding to ordinary diffusion in the plane. Finally, the interpretation of the cohomological information in the Weil conjectures in terms of supersymmetric quantum mechanics and the arithmetic of the semiclassical limit is outlined.

## I. INTRODUCTION

In Ref. 1, the author suggested that $p$-adic analysis may lead to theories of membranes. These membranes would be described by geometrical $p$-adically complete membranes that correspond to the geometric rational or real membranes that are usually thought of as the observable world. In number theory, one counts points in a space over a finite field. Assembling the number of points over all finite field extensions into the zeta function of Artin and Mazur leads to cohomological information about the space over the complex numbers.[2] (This is the essence of the Weil conjectures as proved by Deligne.) Since the Artin–Mazur zeta function can be expressed as a fixed-point theorem for the Frobenius map (determined by an element of the Galois group of the finite-field extension), we conjectured that one can use a type of supersymmetric sigma model to calculate the cohomology of spaces. These more exotic sigma models might be able to be interpreted as membrane theories. Moreover, the Artin–Mazur zeta function could be multiplied together using the notion of adèles to obtain a modular form which satisfies a functional relation. (In the case of elliptic curves, the product of zeta functions is called the Hasse–Weil $L$ function. The functional relation for the Hasse–Weil zeta function is still unproved for a general elliptic function over the relationals, although it has been proved for many such curves.)

Gervais[3] then noticed that the fact that the algebraic completion of the $p$-adic numbers is infinite dimensional (as opposed to the complex one-dimensional extension of the real numbers) might be interpreted as the possibility of higher-dimensional membrane theories with scattering amplitudes that are generalizations of the Virasoro–Shapiro amplitude (which is analogous to the complex extension of the reals).

In this paper, we will link these two approaches by showing that the factors in the zeta function signifying nontrivial cohomology can be interpreted in terms of the representation of a $p$-adic group.[4] Given a $p$-adic group [$\mathrm{SL}_2(Q_p)$ will be our standard example], one can obtain arbitrarily high-dimensional cohomology by algebraic extensions, just as Gervais claimed to describe membranes by $p$-adic algebraic extensions.

Furthermore, we will use $p$-adic group theory to understand the factorization of the inverse bosonic string partition function. We interpret this factorization in terms of the existence of an 11-dimensional membrane. This factorization may be relevant for one-loop string amplitude. Our interpretation of this factorization depends on the use of Hecke operators which generate $p$-adic trees. The generating functional for a tree becomes the Poisson kernel for two-dimensional electrodynamics in the $p \to 1$ limit.[5] Therefore, propagation along a tree is analogous to the Green's function method for two-dimensional electrodynamics. Tree geometry is a new form of geometry analogous to the hyperbolic geometry of the upper-half plane. One can consider families of trees to define a supersymmetric theory on trees such that the fixed point of the family is given by the eigenvalue of an operator related to the Hecke operator, i.e., Frobenius operator.

## II. ALGEBRAIC EXTENSIONS[3]

We first review the relation of $p$-adic numbers of membranes according to Ref. 2. Recall that the algebraic completion of the $p$-adic numbers is an infinite-dimensional vector space over the $p$-adic numbers. This is in contrast to the complex plane, which is the algebraic completion of the real numbers and only a two-dimensional vector space over them.

As a concrete example of a four-dimensional extension of the $p$-adic numbers, consider the equation $x^4 + 1 = 0$. Denote $j$ as a primitive root, i.e., a root of the equation that cannot be written as another root raised to some integer power. Then an arbitrary element of the extension $Q_p(j)$ has the form

$$Z = \alpha + j\beta,$$
$$\alpha = x + j^2 y, \quad \beta = r + j^2 s.$$

The Galois group is generated by the permutations $j \to j^3$, $j \to j^5$, and $j \to j^7$, which changes $Z$ to $Z_3$, $Z_5$, and $Z_7$, respectively:

$$Z_3 = \alpha^* + j^3\beta,$$
$$Z_5 = \alpha + j^5\beta,$$
$$Z_7 = \alpha^* + j^7\beta^*,$$
$$\alpha^* = x - j^2 y, \quad \beta^* = r - j^2 s.$$

Finally, we have a norm invariant under the Galois group:

 0022-2488/89/020506-06$02.50 © 1989 American Institute of Physics 

$$|z|^4 = ZZ_3Z_5Z_7 = (x^2 - y^2 + 2rs)^2 + (r^2 - s^2 - 2xy)^2 .$$

If $|z| = 1$, we obtain that $Z$ can be written in terms of three angles, $\alpha$, $\beta$, and $\theta$:

$$\alpha = \sqrt{[\cos(\theta_1 - \theta_2)/\cos(\theta_3 - \theta_1)]}\, e^{i\theta_{3/2}} ,$$

$$\beta = \sqrt{[\sin(\theta_3 - \theta_2)/\sin(\theta_3 - \theta_1)]}\, e^{i\theta_{1/2}} .$$

This is the crux of the matter. The membrane obtained from the algebraic extension forms an irreducible representation of a continuous group as well as of the Galois group. A similar phenomenon occurs in varieties over finite fields.

Give a curve or variety over a finite field, like $Z_p$, one is interested in those vectors $x = (x_1,...,x_N)$ that lie in a finite extension of the field, like $F_{q=p^L}$. The points on the variety are obtained as fixed points of the $L$ th iterate of the Frobenius map $F^L$, where

$$F = (x_1,...,x_N) \to (x_1^p,...,x_N^p) ,$$

$$F^L = (x_1,...,x_N) \to (x_1^q,...,x_N^q) .$$

This follows from the definite of $F_q$ as the set of points with $Z^q = Z$. The $i$th cohomology elements are obtained from the $i$-dimensional irreducible representation of the Frobenius operator. They can be understood as something like the membranes we considered earlier. To see how these arise, we consider more aspects of number theory.

## III. THE WEIL CONJECTURES: CALCULATING COHOMOLOGY BY COUNTING POINTS

In number theory, if one is given a curve (or a variety of curves) defined by a polynomial (or set of polynomials) in several variables, with coefficients in a finite field $F_{q=p^N}$, one is interested in counting points that lie on this curve. The points do not all lie in the finite field, but in an algebraic extension. It is interesting to consider all algebraic extensions $F_{q^N}$ of $F_q$ and count the number of points $N_L$ on the curve that lie in $F_{q^L}$. One then forms the Artin–Mazur zeta function

$$Z(T) = \exp\left(\sum_{i=0}^{\infty} \frac{N_L T^L}{L}\right).$$

Then the Weil conjectures, which hold for an arbitrary smooth variety, as proved by Deligne, state that

(1) $Z(T)$ is a rational function;

(2) $Z((q^n T)^{-1}) = q^{n\epsilon/2} T^{\epsilon} Z(T)$, for some integer $\epsilon$;

(3) $Z(T) = \left(\prod_{i=1}^{n} P_{2i-1}(T)\right)\left(\prod_{i=0}^{n} P_{2i}(T)\right)^{-1}$;

where each $P_i(T)$ is a polynomial over the integers

$$P_0(T) = 1 - T, \quad P_{2n}(T) = 1 - q^n T,$$

and

$$P_i(T) = \prod_j (1 - \alpha_{ij} T) ,$$

$$|\alpha_{ij}) = q^{1/2}, \quad 1 \leqslant i \leqslant 2n - 1 .$$

(4) Each $P_i(T)$ determines the $i$th cohomology group of the variety over the complex number. Moreover, $P_i(T)$ is the characteristic polynomial of a Frobenius operator acting in the $i$th cohomology basis. The Frobenius operator is defined on the vector $x = (x_1,...,x_N)$ mapped into $(x_1^q,...,x_N^q)$. The vectors $x$ that lie in $F_q$ are fixed points of this map.

As an example, we look at the projective $(N-1)$-space, i.e., the space of all $N$ vectors as above with $x_i \in F_q$, modulo multiplication by the nonzero elements $\lambda \in F_q^*$.

We would like to count the points in $F_{N-1}$ over $q = p^l$ for each integer $L$:

$$N_L = (q^N - 1)/(q - 1) = 1 + q + \cdots + q^{N-1} .$$

Notice that the cohomology of $Cp^{N-1}$ is determined by a two-form $x$. The fact that the total Chern class is $1 + x + \cdots + x^{N-1}$ is not a coincidental similarity with $N_L$ as we see below:

$$Z(T) = \exp\left(\sum \frac{N_L T^L}{L}\right)$$

$$= \exp\left(\sum \frac{\Sigma p^{Li} T^L}{L}\right)$$

$$= \exp\left(-\sum \log(1 - p^i T)\right)$$

$$= \prod_{i=0}^{N-1} \frac{1}{(1 - p^i T)} ;$$

thus $P_{2i}(T) = 1 - p^i T$, $P_{2i-1}(T) = 1$. Therefore there is cohomology in only the even dimensions, as is well known.

We note that exactly the same zeta function is obtained by looking at the curve $\Sigma_{i=1}^{N} x_i^2 - x_0^2 = 0$ over $Z_p$. One can consider this calculation as similar to a covariant method of calculating the zeta function. It would seem that this would lead to a supersymmetry or BRST symmetry, just as the introduction of ghosts leads to this symmetry in gauge theories. Namely, one can introduce a vector $k$ so that $k^2 = 0$ in the $(+ + \cdots + -)$ norm. Then we define cohomology by those $x$ so that $k \cdot x = 0$ but $x \neq ky$ for some $y$. For $N = 2$, we choose $k = (0,1,1)$. Then for $x$ we have $(a,b,b)$ and $(-a,b,b)$, which are not equal to 0. This defines the two-dimensional space on which Tr $F = p$. That is to say, it defines an irreducible representation.

Finally, we note that the zeta function for an elliptic curve over the rationals can be shown to have the following form for $T = p^{-s}$:

$$Z(s) = [1 - a(p)p^{-s} + p^{1-s}]/(1 - p^{-s})(1 - p^{1-s}) ,$$

$$a(p) = \alpha_p + \bar{\alpha}_p = 2 \text{ real}(\alpha_p) ,$$

$$|\alpha_p| = p^{1/2} .$$

[Note that the expression $a(p)$ as the real part of $\alpha_p$ is like the separation of a string into right and left movers.] The denominator is rather trivial; as expected it is that obtained from $P^1(Z_p)$. However, the numerator is interesting. The inverse of the numerator can be expanded in a power series:

$$(1 - 2a(p)p^{-s} + p^{1-s})^{-1} = \sum a(p^l) p^{-ls} ,$$

where

$$a(p)a(p^l) = a(p^{l+1}) + pa(p^{l-1}) .$$

If one takes the adelic product over all primes $p$ (ignoring the subtlety of bad reduction), one obtains $\Sigma a(n)n^{-s}$. This function is conjectured to be the Mellin transform of a modular form.[2] We will need some definitions in order to proceed.

## IV. MODULAR FORMS, HECKE OPERATORS,[6] AND TREES[7]

Let $H$ denote the upper-half plane of the complex plane $C$. Then $\mathrm{PSL}_2(R)$, which is $\mathrm{SL}_2(R)$ mod$(\pm 1)$ is the group of all analytic automorphisms of $H$ with discrete subgroup $\mathrm{PSL}_2(Z)$, called the modular group. Let $k$ be an integer. Then $f$ is a modular form of weight $2k$ if $f$ is a holomorphic function of $H$ and

$$f(z) = (cz+d)^{-2k} f[(az+b)/(cz+d)]$$

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(Z)$. Since the elements $S = z \to -1/z$ and $T = z \to z+1$ generate the modular group, one must only check that

$$f(z+1) = f(z) ,$$

$$f(-1/z) = z^{2k} f(z) .$$

If $f$ vanishes at infinity, it is called a cusp form.

We can identify modular functions of weight $2k$ with lattice functions of weight $2k$ as follows: given a lattice, for example, the two-dimensional lattice

$$\Gamma(\omega_1,\omega_2) = Z\omega_1 \oplus Z\omega_2 ,$$

with basis $\{\omega_1,\omega_2\}$, we define a lattice function

$$F(\omega_1,\omega_2) = (\omega_2)^{-2k} f(\omega_1/\omega_2) ,$$

where $f$ is a modular function of weight $2k$. Then

$$F(\lambda\omega_1,\lambda\omega_2) = \lambda^{-2k} F(\omega_1,\omega_2) .$$

Given two cusp forms of weight $2k$, $k>0$, then there is a measure, called the Weil–Peterson measure,

$$\mu(f,g) = f(z)\bar{g}(z) y^{2k} (dx\,dy)/y^2 , \quad z = x+iy ,$$

that is invariant under $\mathrm{SL}_2(Z)$ and bounded on $H/\mathrm{SL}_2(Z)$. This measure defines a Hermitian, positive definite inner product on $M_k^0$, the vector space of cusp forms of weight $2k$. Note that one can define a graded algebra

$$M = \sum_{k=0}^{\infty} M_k^0 .$$

Within the Hilbert space $L_2(H/\mathrm{SL}_2(Z))$ defined by the Weil–Peterson metric, there exists a class of Hermitian operators, called Hecke operators. These operators play an essential role in breaking up the Hilbert space into $p$-adic spaces which are simpler. The Hecke operators are best described in terms of lattice correspondences. We then use the relation between modular forms and lattice functions to define the Hecke operators on modular forms.

Consider the set of lattices in the complex plane. Given a lattice $\Gamma$, the Hecke operator $T(n)$ denotes a correspondence that transforms $\Gamma$ to a sum of its sublattices $\Gamma'$ of index $(\Gamma:\Gamma') = n$:

$$T(n)\Gamma = \sum_{(\Gamma:\Gamma')} \Gamma' .$$

The number of such lattices $\Gamma'$ is equal to the number of subgroups of order $n$ in $(Z_n/Z)^2$. If $n$ is a prime $p$, this number is $p+1$, the number of points on the projective line over a field of $p$ elements. This last result can be seen by considering $\Gamma$ as generated by $(1,0)$ and $z = (x,y)$. Then $\Gamma'$ is of index $p$ if it is generated by $(1/p)$ $(a_1x, a_1y + a_2)$ and $(1,0)$ with $(a_1,a_2)$ an element of the projective line over $Z/pZ$. The projective line contains $p+1$ points corresponding to $(1,j)$, $j = 0,1,\ldots,p-1$, and $(0,1)$. We also need to define the homothety operators $R_\lambda\Gamma = \lambda\Gamma$. We then have the following algebra:

$$R_\lambda R_\mu = R_{\lambda\mu} ,$$

$$R_\lambda T(n) = T(n) R_\lambda ,$$

$$T(m)T(n) = T(mn) , \quad (m,n) = 1 ,$$

$$T(p)T(p^n) = T(p^{n+1}) + pT(p^{n-1})R_p ,$$

for $p$ prime.

Consider the set of integral matrices $S_n$, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) ad = n$, $a \geqslant 1$, $0 \leqslant b < d$. Then there is a one-to-one onto map of the set $\Gamma(n)$ of sublattices of index $n$ in $\Gamma$ with the matrices $S_n$ given by a mapping of the basis $(\omega_1,\omega_2)$ for $\Gamma$ to the basis of $\Gamma(n)$, $(a\omega_1 + b\omega_2, d\omega_2)$.

Once we have the isomorphism between sublattices and matrices, the relation between the Hecke lattice correspondences and Hecke matrix operators is immediate. If $f(z)$ is a modular form of weight $2k$, then

$$T(n) f(z) = n^{2k-1} \sum_{\substack{a>1,\ ad=n \\ 0<b<d}} d^{-2n} f\left(\frac{az+b}{d}\right),$$

so that, for $p$ prime,

$$T(p)T(p^n) f(z) = [T(p^{n+1}) + p^{2k-1} T(p^{n-1})] f(z) ,$$

where $T(n)$ is a Hermitian operator with respect to the Weil–Peterson metric.

We define new correspondence operators $\Theta(p^n)$, which makes immediate the relation to the underlying tree geometry:

$$\Theta_0 = 1 ,$$

$$\Theta(p) = T(p) ,$$

$$T(p^n) = \sum_{0<i<n/2} \Theta(p^{n-2i}) = O(p^n) + T(p^{n-2}) .$$

Thus, when $k = 1$, we have

$$[\Theta(p)]^2 = \Theta(p^2) + (p+1)\Theta_0 ,$$

$$\Theta(p)\Theta(p^n) = \Theta(p^{n+1}) + p\Theta(p^{n-1}) , \quad n \geqslant 2 .$$

One can then define a $p$-adic tree, where each vertex corresponds to a lattice. If it is the lattice corresponding to a choice origin, $\Theta(p)$ gives a correspondence to the $(p+1)$ lattices contained in $\Gamma$ with index $p$, and $\Theta(p^n)$ gives a correspondence to those vertices with ultrametric valuation $n$.

Just as there is a hyperbolic geometry defined in the unit disk $x^2 + y^2 \leqslant 1$ with respect to the metric $(dx^2 + dy^2)/(1 - x^2 - y^2)^2$, so that geodesics through a point do not intersect, so, too, the geodesics through the vertex representing $\Gamma$, for example, either coincide or never intersect.[8]

The tree also makes clear the relation of Hecke operators to representations of $p$-adic groups. Just as the integers label the cosets of the circle within the real numbers by the winding number, so, too, do the $p$-adic valuations, i.e., the integers $n$ given by the $p$-adic norm $p^{-n}$, label cosets of the $p$-adic integers $Z_p$ with $Q_p$. However, for $\mathrm{PGL}_2(Q_p)$, one has the double cosets

$$\mathrm{GL}_2(Z_p)\backslash\begin{pmatrix} p^n & 0 \\ 0 & p^{-n}\end{pmatrix}/\mathrm{GL}_2(Z_p)$$

with the $\mathrm{GL}_2(Z_p)$ operating on two different vertices (lattices) that are separated by $2n$ ultrametric lengths. We can then describe the tree as $\mathrm{PGL}_2(Q_p)/\mathrm{PGL}_2(Z_p)$, the $p$-adic analog of the upper-half plane. The Hecke operators $\Theta(p^n)$ denote correspondences between vertices separated by $n$ ultrametric lengths. Hecke operators are in $\mathrm{GL}_2$, not $\mathrm{SL}_2$, since they have nonunit determinant.

## V. SPHERICAL FUNCTIONS[4]

In a disconnected space, like a tree, there are not differential operators. However, the Hecke operator plays the role of a Laplacian. It generates a diffusion process along the tree.[8] It is interesting to construct the analog of spherical harmonic and radial functions. We will construct in this section the analog of radial functions. These are called spherical functions, invariant under the operators representing $\mathrm{SL}_2(Z_p)$ in the double-coset representation.

In the two-dimensional space $(x_1,x_2)\equiv x$ we define the norm

$$|x| = \max(|x_1|,|x_2|).$$

Every function invariant under the compact subgroup $\mathrm{SL}_2(Z_p)$ is of the form $F(|x|)$. We can represent the total group in the space of function $|x|^{s-1}$ corresponding to the character $\pi(x) = |x|^s$. If we define

$$f_0 = \sqrt{p/(p+1)}\,|x|^{s-1},$$

then an irreducible representation $T_\pi(g)$ invariant under $\mathrm{SL}_2(Z_p)$ is determined by the spherical function

$$\begin{aligned}\phi(g) &= (T(g)f_0, f_0)\\ &= \phi(\delta),\end{aligned}$$

if $g = u_1\delta u_2$ in the double coset representation where

$$\delta = \begin{pmatrix} p^{-n} & 0 \\ 0 & p^n\end{pmatrix} \quad n\geqslant 0.$$

If $s = ip$, an imaginary number, we can determine the representation of the principal series

$$(f_1,f_2)\equiv\int f_1(t,1)\,\overline{f_2(t,1)}\,dt,$$

$$\begin{aligned}\phi(\delta) &= \frac{p}{p+1}\int[\max(p^n|t|,p^{-n})]^{s-1}[\max(|t|,1)]^{-s-1}\,dt\\ &= \frac{p}{p+1}\left\{p^{-n(s-1)}\int_{|t|<p^{-2n}}dt + p^{n(s-1)}\int_{p^{-2n}\leqslant|t|<1}dt + p^{n(s-1)}\int_{|t|>1}|t|^{-2}\,dt\right\}\\ &= \frac{p}{p+1}\left\{p^{-ns-n} + (1-p^{-1})p^{-n}\frac{p^{ns}-p^{-ns}}{1-p^{-s}} + p^{ns-n-1}\right\}\\ &= p^{-n}\left\{p^{1/2}\frac{\sin(n+\frac{1}{2})\theta - p^{-1/2}\sin(n-\frac{1}{2})\theta}{\sin(\theta/2)(p^{1/2}+p^{-1/2})}\right\},\end{aligned}$$

where $p^s = e^{i\theta}$.

Moreover, the spherical function is related to the eigenfunction of $\Theta(p)$ with eigenvalue $2p^{1/2}\cos\theta$, namely, the function that, at the $n$th level in the tree, is given by

$$F_n(2p^{1/2}\cos\theta) = p^{+n/2}\left\{\frac{p\sin(n+1)\theta - \sin(n-1)\theta}{p\sin\theta}\right\}.$$

Note that the limit $p\to 1$ leads to the Chebyshev polynomial

$$C_n(\cos\theta) = (1/2^{n-1})\cos(n\theta)$$

with generating functional that is essentially the Poisson kernel[9] for two-dimensional electrostatics:

$$\frac{1-t^2}{1-2t\cos\theta+t^2} = \sum C_n(\cos\theta)(2t)^n.$$

We, therefore, see that the $p\to 1$ limit defines a continuum. The generating functional for $F^n(2p^{1/2}\cos\theta)$ is

$$(1-t^2)/(1-2tp^{1/2}\cos\theta + pt^2).$$

If we replace $t$ by $p^{-s}$, we recover the Euler factor in the Hasse–Weil zeta function discussed earlier. The numerator is just a trivial Euler factor related to the Riemann zeta function. We, therefore, see that the characteristic polynomial for the Frobenius operator in the one-dimensional cohomology of an elliptic curve, i.e., loop space, is a generating functional for particular angles $\theta_p$ for a $p$-adic tree. This yields the simplest example of a modular form whose Mellin transform factorizes over the primes and satisfies a functional relation

$$L(s) = (2\pi)^{-s/2}\Gamma\left(\frac{s}{2}\right)\times\prod_p(1-a(p)p^{-s}+p^{1-s})^{-1} = \pm L(2-s).$$

## VI. FACTORIZATION OF MODULAR FORMS

A necessary and sufficient condition for a cusp form $f$ to have a Mellin transform that factorizes over the primes is that the Fourier series

$$f(\tau) = \sum_{n=1}^{\infty} a_n q^n, \quad q = e^{i\pi\tau},$$

be an eigenfunction of all the Hecke operators $T(n)$ with eigenvalue $a_n$. If $n$ is a cusp form of weight $2k$, then, as a consequence of the multiplicative properties of the Hecke operators, one has that the Mellin transform of $f$ is

$$\sum a_n n^{-s} = \prod_p (1 - 2a(p)p^{-s} + p^{2k-1-s})^{-1}.$$

As an example, we consider the inverse of the partition function in bosonic string theory. This is a cusp form of weight 12 in a space $M_{12}^0$ of dimension 1. We, therefore, have that the inverse partition function $(2\pi)^{-12}\Delta$ is an eigenfunction of $T(n)$ for each $n$ with eigenvalue $T(n)$

$$(2\pi)^{-12}\Delta = q\prod_n (1-q^n)^{24} = \sum_{n=0}^{\infty} \tau(n) q^n,$$

$$\tau(n)\tau(m) = \tau(nm), \quad \text{for } (n,m) = 1,$$

$$\tau(p)\tau(p^n) = \tau(p^{n+1}) + p^{11}\tau(p^{n-1}), \quad \text{for prime } p.$$

Moreover, the Mellin transform

$$\sum_{n=0}^{\infty} \tau(n) n^{-s} = \prod_p (1 - \tau(p)p^{-s} + p^{11-2s})^{-1}.$$

It has been proved that there exists a variety whose zeta function has a $P_{11}(T)$ divisible by the right-hand side. By the Weil conjectures, $|\tau(p)| \leqslant 2p^{11/2}$. That is to say, the eigenvalues of the Hecke operators are obtained by taking the trace of the Frobenius operator on the cohomology of the variety,

$$\tau(p) = \operatorname{Tr} F = \operatorname{Tr}\begin{pmatrix} \exp(i\theta_p) & 0 \\ 0 & \exp(-i\theta_p) \end{pmatrix}, \quad \text{for } \theta_p,$$

that give the Ramanujan numbers $\tau(p)$. If we are to express the polynomial $p_{11}(s) = (1 - \tau(p)p^{-s} + p^{11-2s})^{-1}$ in terms of a tree-generating function $(1 - a(p)p^{-s} + p^{1-s})^{-1}$, we need to solve an 11-order integral polynomial equation requiring an 11-dimensional vector space over the $p$-adic tree. We need to express $2p^{11/2}\cos\theta_p$ as an 11th-degree polynomial in terms of $2p^{1/2}\cos(\theta_p)/11$, which is possible by definition of the Chebyshev polynomials. Moreover, the inverse partition function behaves as if there is an 11-dimensional membrane. We can relate the inverse partition function to string theory by considering it as a partition function for ghosts. It is, therefore, suggested that an interesting phenomenon, perhaps an 11-dimensional membrane condensate, can occur in such a theory. This would be described $p$-adically in terms of an 11-dimensional tree structure. [To give another indication of the existence of this membrane, we note that $\tau(p) = \sigma_{11}(p) \bmod 691$, where $\sigma_{11}(p) = 1 + p^{11}$. Moveover, 691 is the numeration of $\frac{1}{6}B_6$, where $B_6$ is a Bernoulli number. This number divides the order of the number of diffeomorphism classes of exotic 23-spheres which bound parallelizable manifolds and is therefore associated with a global anomaly.]

We mention incidentally that the Einstein series

$$E_k(z) = \left(1 - \frac{2k}{B_n}\sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n\right) 2\zeta(k),$$

where $B_k$ is the $k$ th Bernoulli number and

$$\sigma_{k-1}(n) = \sum_{d/n} d^{k-1}$$

has a Mellin transform that factorizes with factor $(1 - \sigma_{k-1}(p)p^{-s} + p^{k-1-2s})^{-1}$. This is interesting because the anomaly-generating functional for superstrings is given in terms of the $E_k(z)$. Nontrivial cohomology occurs in dimensions 0 and $2(k-1)$ corresponding to the factors $(1-p^{-s})^{-1}$ and $(1-p^{k-1-s})^{-1}$.

## VII. SUPERSYMMETRY, INDEX THEOREMS, AND THE WEIL CONJECTURES

We now return to the idea mentioned at the beginning. The Artin–Mazur zeta function for an arbitrary smooth variety over a finite field can be expressed as a fixed point theorem for the Frobenius operator, where

$$P_i(T) = \det(1 - FT|H^i|).$$

Where the cohomology group $H^i$ defines an $i$-dimensional irreducible representation. By our recently developed knowledge of the index theorem in terms of quantum mechanics, it would seem that there should exist a $p$-adic form for quantum mechanics that yields the polynomial $P_i(T)$.

While we have not yet succeeded in defining the above conjectural quantum mechanics, we will use this section to suggest some ideas about the matter. First, the definition of a quantum mechanics requires a time. This time is a parameter for a loop because the boundary conditions that request supersymmetry are periodic for bosons and fermions. For the Hasse–Weil zeta function, there is a natural definition of time, that is, the angle in $\alpha_p = p^{1/2}e^{i\theta_p}$. This angle defines a dual to a lattice, the lattice associated with the elliptic curve. As one rotates $\theta_p$, one considers arbitrary characters for the tree functions $f_0 = \sqrt{[p/(p+1)]}\,|x|^{s-1}$. The quantum mechanical time is then a circle in the Hilbert space of tree functions. The bosons are scalar functions and the fermions are one-forms on this loop space. The supersymmetry between the bosons and fermions is a Parisi–Sourlas[10] type supersymmetry, and therefore associated with $SL_2$.

The theory should be independent of the parametrization of the loop space. We, therefore, should include in the quantum mechanics a delta function

$$\delta(\dot{f}(t) + \eta)\left|\frac{\delta\eta}{\delta t}\right| = \delta(\dot{f}(t) + \eta)\det\frac{\partial}{\partial t},$$

where $\eta$ is some source, with a distribution, that needs to be integrated over. In such a way, we can obtain a supersymmetry theory $\frac{1}{2}\dot{f}(t)^2 + \psi^*(\partial\psi/\partial t)$ (for Gaussian noise). This method has lead to a consideration of the arithmetic content of semiclassical quantum mechanics.[11]

In brief, the semiclassical limit of quantum mechanics is defined in terms of integral cohomology, i.e., the Bohr–Sommerfeld quantization rule counts the number of orbits.[12] One can, therefore, consider this limit arithmetically.

One possible application for the future is to calculate the cohomology of vector bundles over a Riemann surface. Using equivariant Morse theory, Atiyah and Bott[13] have calculated the Poincaré series for stable vector bundles

$$P_t\{\mathrm{Map}(M\to BU(n))\} = \left(\prod_{k=1}^{n}(1+t^{2k-1})^{2g}\right) \times\left(\left\{\prod_{k=1}^{n-1}(1-t^{2k})^2\right\}(1-t^{2n})\right)^{-1}.$$

Amazingly enough, they have also calculated the number of vector bundles over Riemann surfaces over finite fields using number theory and the Weil conjectures. The formula for this number is

$$q^{(n^2-1)(g-1)}Z_M(s=2)\cdots Z_M(s=n)$$

for a genus $g$ Riemann surface $M$ over finite field $F_q$, whose zeta function is $Z_M(s)$. Atiyah and Bott noticed that if they replaced $t^{-2}$ by $q$ and $-t^{-1}$ by $\omega_i$ (a zero of the zeta function) in the Poincaré series, they recovered the number theoretical quantity. This suggests a future direction. Namely, one tries to find a topological Lagrangian like that used to calculate the Floer groups and Donaldson polynomials.[14] Then define this Lagrangian for Riemann surfaces over finite fields to calculate number theoretical quantities. We have suggested that the topological Lagrangian for elliptic curves should be defined over loop space on trees. In this case, the eigenvalues of the Frobenius operator $F$ are the periods of the elliptic curve. This becomes topological if we consider it as the holonomy for the loop space of a tree. Notice that the really interesting quantity is the numerator, because this contains the holonomy. The denominator corresponds to periodic even-dimensional cohomology.

A natural choice for the topological Lagrangian is a generalization of the Lagrangian of Frampton and Okada[15] that leads to the operator

$$p^{\square}F.$$

The characteristic polynomial

$$\det(1-p^{\square}F)$$

is what occurs in the Weil zeta function.

## ACKNOWLEDGMENT

This work was supported in part by the Department of Energy under Contract No. DE-AC02-87ER40325.Task B.

[1]B. Grossman, Phys. Lett. B **197**, 101 (1987). Other considerations of $p$-adic physics have been given in I. V. Volovich, Classical Quantum Gravity **4**, 183 (1987); P. G. D. Freund and B. Witten, Phys. Lett. B **199**, 191 (1987); P. Frampton and Y. Okada, Phys. Rev. Lett. **60**, 484 (1988); G. Parisi, to be published in Int. Mod. Phys. Lett.

[2]A. Weil, Bull. Am. Math. Soc. **55**, 497 (1978); N. M. Katz, "An overview of Deligne's proof of the Reimann hypothesis for varieties over finite fields," in *Proceedings of Symposia in Pure Mathematics* (Am. Math. Soc., Providence, RI, 1976), Vol. 28, p. 275.

[3]J. L. Gervais, Phys. Lett. B **201**, 306 (1988).

[4]I. M. Gelfand, M. I. Graev, and I. I. Piatetskii-Shapiro, *Representation Theory and Automorphic Function* (Saunders, Philadelphia, 1969).

[5]B. Spokoiny, "Quantum geometry of non-arithmetic particles and strings," Landau preprint, considers $p\to 1$ in another context.

[6]J. P. Serre, *A Course in Arithmetic* (Springer, New York, 1978).

[7]J. P. Serre, *Trees* (Springer, New York, 1980).

[8]P. Cartier, *Geometrie et analyse sur les arbre, Lecture Notes in Mathematics* (Springer, Berlin, 1973), Vol. 317, p. 123.

[9]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).

[10]G. Parisi nd N. Sourlas, Nucl. Phys. B **206**, 321 (1982).

[11]B. Grossman, in preparation.

[12]M. F. Atiyah and R. Bott, Topology **23**, 1 (1984).

[13]M. F. Atiyah and R. Bott, Philos. Trans. R. Soc. London, Ser. A **308**, 523 (1982).

[14]E. Witten, *Topological Quantum Field Theory*, IAS; L. Baulieu and I. Singer, preprint LPTHE and Ecole Normale Superier preprints; B. Grossman and L. Baulieu, Rockefeller preprint.

[15]P. H. Frampton and Y. Okada, Phys. Rev. D **37**, 3077 (1988).

# Supersymmetric embedding of a model of a quantum harmonic oscillator interacting with infinitely many bosons

Asao Arai
*Department of Mathematics, Hokkaido University, Sapporo, 060, Japan*

(Received 8 April 1988; accepted for publication 21 September 1988)

A mathematically rigorous analysis is given on supersymmetric embedding of a model of a one-dimensional quantum harmonic oscillator interacting with infinitely many bosons moving in the $s$-dimensional space $\mathbb{R}^s$. The model is exactly soluble. By a rigorous explicit construction of supersymmetric quantum field theories, it was proven that the Hamiltonian of the model is supersymmetrically embeddable if it is non-negative. The index of the Dirac–Kähler-type operators associated with the supercharges of the supersymmetric quantum field theories is computed.

## I. INTRODUCTION

In the course of developing supersymmetric quantum mechanics, as initiated by Witten,[1-3] it has been found that a number of quantum mechanical Hamiltonians have supersymmetric extensions.[1-14] In particular, in Ref. 5 a general formalism for constructing supersymmetric quantum mechanical Hamiltonians was given, from which it is seen that some classes of quantum mechanical Hamiltonians of the Schrödinger type have supersymmetric extensions.

In view of a generalization of these results, it is natural to ask the following question: Given a quantum mechanical Hamiltonian, under what conditions does it have a supersymmetric extension? We shall say that a Hamiltonian is supersymmetrically embeddable if it has a supersymmetric extension (for a mathematically precise definition, see Definition 2.2).

The problem of supersymmetric embedding is interesting both physically and mathematically: It has been hoped that supersymmetry may give a fundamental framework for a unification of elementary-particle interactions. From this point of view, it may be natural to ask if the usual quantum theoretical models have something to do with supersymmetry. Namely, this leads to the question of supersymmetric embedding. On the other hand, the Hamiltonian of a supersymmetric quantum theoretical model is given as the square of a Dirac-type operator (a "supercharge"), while the Hamiltonian of a usual (bosonic) quantum theoretical model is an operator of the Schrödinger type. Therefore, from a mathematical point of view, the problem of supersymmetric embedding may be regarded as one that investigates a "hidden" structure (i.e., the supersymmetric structure) associated with a Schrödinger-type operator. This is also interesting in relation to index theorems (e.g., Refs. 15 and 16).

In Ref. 17, Gozzi showed that the one-dimensional Schrödinger Hamiltonian $-d^2/dx^2 + U(x), x\in\mathbb{R}$, is supersymmetrically embeddable if it has a nodeless ground-state wavefunction. This result was extended to the three-dimensional case in Ref. 18 and to an arbitrary $n$-dimensional case in Ref. 19, where it was shown that every $n$-dimensional scalar Hamiltonian of the form

$$h = -\sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2} + U(x), \quad x = (x_1,...,x_n)\in\mathbb{R}^n, \tag{1.1}$$

is supersymmetrically embeddable if it has a strictly positive ground-state wave function. (In Ref. 19, the boundedness from below of $h$ is assumed. However, this is not necessary; in fact, one can deduce it from the assumption that $h$ has a strictly positive eigenfunction. This is easily seen by tracing the "proof" of the supersymmetric embeddability of $h$.) Thus, as far as quantum mechanical Hamiltonians of the form (1.1) are concerned, the problem of supersymmetric embedding has been entirely solved, at least on a formal level. [From a mathematically rigorous point of view, in addition to the existence of a strictly positive eigenfunction $\Omega$, some regularity conditions have to be imposed on $\Omega$ for the formal scheme of supersymmetric embedding to be justified. A mathematically rigorous analysis of the problem has been given in Ref. 20 as an application of an abstract mathematical theory, with a class of symmetric operators acting in $L^2(M,\mu)$, where $(M,\mu)$ is an abstract measure space.]

As a next stage, it is interesting to study supersymmetric embedding of Hamiltonians in quantum field theory. In a previous paper,[21] the present author considered $n$-component quantum real scalar field Hamiltonians of the form

$$H = -\sum_{j=1}^{n} \int_{\mathbb{R}^d} \frac{\delta^2}{\delta\phi_j(x)^2}\,dx + U(\phi) \tag{1.2}$$

in the Schrödinger representation of the canonical commutation relations,[22] where $\phi(x) = (\phi_1(x),...,\phi_n(x))$, $x\in\mathbb{R}^d$, is the time-zero field on the $d$-dimensional space $\mathbb{R}^d$ and $U(\phi)$ is a real-valued functional of $\phi$ denoting a potential, and formally showed by an infinite-dimensional extension of the method used in Ref. 19 that $H$ is supersymmetrically embeddable if it has a strictly positive eigenfunctional.

However, the question of how to give a mathematically rigorous basis to the formal scheme has been left open. [A class of models in which $U$ is a polynomial type with a special form[23] was discussed rigorously in Ref. 24 in the case $n = 1$ and $d = 1,2$. However, this is a rather tractable case once one employs results of constructive quantum field theory on $P(\phi)_d$ models (e.g., Ref. 25 and references therein; cf., also, Ref. 26).]

In this paper, we start a mathematically rigorous analysis on supersymmetric embedding of quantum field models. As a first step, we consider an exactly soluble model, which describes a quadratic interaction of a one-dimensional quan-

 0022-2488/89/020512-09$02.50 © 1989 American Institute of Physics 

tum harmonic oscillator with infinitely many scalar bosons and has been discussed in quantum statistical physics (e.g., Ref. 27 and references therein; in this context, the model is called the *RWA oscillator*). The Hamiltonian of the model is not exactly of the form (1.2) [see (3.9) in Sec. III]. Therefore, the method of Ref. 21 does not apply in its original form. The method taken in the present paper is as follows. We first "solve" the model exactly. The key to this is to find a canonical transformation[28] by which the Hamiltonian is diagonalized and its spectrum completely determined. This has been done in a previous paper.[29] Then, by employing a method for supersymmetric extensions of a free scalar field, which consists of a direct infinite-dimensional extension of the method for supersymmetric extensions of the harmonic oscillator,[30,31] we construct a supersymmetric quantum field theory (SSQFT) in which a reduction of the supersymmetric Hamiltonian to a subspace of state vectors yields the Hamiltonian of the model. In this way, it is proved that the Hamiltonian of the model is supersymmetrically embeddable if it is bounded from below (non-negative as a matter of fact). In fact, under the condition, we shall construct two kinds of SSQFT depending on the range of the parameters contained in the Hamiltonian. (It is known that a supersymmetric extension of a given Hamiltonian is not necessarily unique. See, e.g., Refs. 6–13, 30, and 31.) A big difference between the two theories is in the spectrum of the Hamiltonian and, related to it, in the fermionic degrees of freedom (see Secs. V and VI). However, in both of the SSQFT's the bosons and the oscillator do not interact with the fermions.

We remark that the method in the present paper applies also to other exactly soluble models with quadratic interactions (e.g., Ref. 32–39).

This paper is organized as follows. In Sec. II, we first recapitulate an abstract axiomatic mathematical formulation of supersymmetric quantum theory (SSQT) for a rigorous approach, which was proposed first in Ref. 40 and then in Ref. 41 with a reformulation in terms of sequilinear forms. Then we give a mathematically precise definition of supersymmetric embeddability of quantum Hamiltonians. In Sec. III, we define the model by giving its Hamiltonian. In Sec. IV, as preliminaries to Secs. V and VI, some technical facts taken from Ref. 29 are presented and some additional formulas are proved. In Secs. V and VI, we construct explicitly SSQFT's and prove the supersymmetric embeddability of the model. In Sec. VII, some remarks are given. In particular, we discuss index problems related to the supersymmetric quantum field models constructed in Secs. V and VI.

## II. DEFINITION OF SUPERSYMMETRIC QUANTUM THEORY AND SUPERSYMMETRIC EMBEDDING

Following Refs. 40 and 41, we first give an axiomatic mathematical formulation of SSQT, which applies to both supersymmetric quantum mechanics and SSQFT.

*Definition 2.1:* Let $N \geqslant 1$ be a fixed integer. A SSQT with $N$-supersymmetry is a quadruple $\{\mathscr{H}, \{Q_j\}_{j=1}^N, H_{\rm SS}, N_{\rm F}\}$ consisting of a Hilbert space $\mathscr{H}$, a set of self-adjoint operators $\{Q_j\}_{j=1}^N$ (supercharges), self-adjoint operators $H_{\rm SS}$ (supersymmetric Hamiltonian), and $N_{\rm F}$ (fermion number operator) satisfying the following properties.

(a) The Hilbert space $\mathscr{H}$ is decomposed into two mutually orthogonal closed subspaces $\mathscr{H}_\pm$:

$$\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$$

and, for all $\Psi_\pm \in \mathscr{H}_\pm$,

$$N_F \Psi_\pm = \pm \Psi_\pm .$$

(b) The operator $H_{\rm SS}$ is related to $Q_j$, $j = 1,...,N$, by

$$H_{\rm SS} = Q_1^2 = Q_2^2 = \cdots = Q_N^2 .$$

(c) The operators $Q_j$, $j = 1,...,N$, satisfy the anticommutation relations in the sense of sesquilinear form:

$$(Q_i\Psi, Q_j\Phi) + (Q_j\Psi, Q_i\Phi) = 0 , \quad i \neq j ,$$
$$\Psi, \Phi \in D(Q_i) \cap D(Q_j) ,$$

where $(\ ,\ )$ is the inner product of $\mathscr{H}$ and $D(A)$ denotes the domain of operator $A$.

(d) For all $\Psi$ and $\Phi$ in $D(Q_j)$,

$$(N_{\rm F}\Psi, Q_j\Phi) + (Q_j\Psi, N_F\Phi) = 0 , \quad j = 1,2,...,N . \tag{2.1}$$

Some remarks are in order. Condition (a) is equivalent to the fact that $N_{\rm F}$ has a purely discrete spectrum with eigenvalues $\pm 1$. The subspaces $\mathscr{H}_+$ and $\mathscr{H}_-$ are called the subspaces of bosonic and fermionic states, respectively. Condition (b) implies that the supersymmetric Hamiltonian $H_{\rm SS}$ is non-negative and

$$D(Q_1) = \cdots = D(Q_N) = D(H_{\rm SS}^{1/2}) ,$$

which follows from the spectral theorem for a self-adjoint operator. Since $N_{\rm F}$ is bounded and self-adjoint, Eq. (2.1) is equivalent to

$$(Q_j\Psi, N_{\rm F}\Phi) = -(\Psi, N_{\rm F}Q_j\Phi) , \quad \Psi, \Phi \in D(Q_j) .$$

Hence, for every $\Phi \in D(Q_j)$, $N_{\rm F}\Phi$ is in $D(Q_j)$, that is, $N_{\rm F}$: $D(Q_j) \to D(Q_j)$ and

$$Q_j N_{\rm F} + N_{\rm F} Q_j = 0 , \quad j = 1,...,N ,$$

on $D(Q_j)$. For some abstract results derived from the above definition of a SSQT, see, e.g., Refs. 40–42.

*Definition 2.2:* Let $A$ be a self-adjoint operator acting in a Hilbert space. We say that $A$ is supersymmetrically embeddable if it is unitarily equivalent to a reduced part of the supersymmetric Hamiltonian of a SSQT.

*Remark:* Since a supersymmetric Hamiltonian is non-negative [condition (b)], every supersymmetrically embeddable operator is non-negative. This implies that an operator that is not non-negative is never supersymmetrically embeddable. However, for a self-adjoint operator $A$ that is bounded from below, the "renormalized" operator $A - E_0$ may be supersymmetrically embeddable, where $E_0$ is the infimum of the spectrum of $A$.

## III. DEFINITION OF THE MODEL

In this section we define the model mentioned in Sec. I. For a mathematical generality, we assume that the bosons move in the $s$-dimensional Euclidean space $\mathbb{R}^s$. Further, in this paper we shall confine ourselves to the case with a cutoff interaction. Thus the Hilbert space $\mathscr{F}$ of the state vectors is given by the tensor product of $L^2(\mathbb{R})$ and the Boson Fock space $\mathscr{F}_{\rm B}(L^2(\mathbb{R}^s))$ over $L^2(\mathbb{R}^s)$:

$$\mathscr{F} = L^2(\mathbb{R}) \otimes \mathscr{F}_{\rm B}(L^2(\mathbb{R}^s)) . \tag{3.1}$$

Let $\omega_1$ be a non-negative, strictly monotone increasing, and continuously differentiable function on $(0,\infty)$ satisfying the condition that $\omega_1(t)\to\infty$ as $t\to\infty$. Then we define the rotation invariant function $\omega$ on $\mathbb{R}^s$ by

$$\omega(k)=\omega_1(|k|)\,,\quad k\in\mathbb{R}^s\,, \tag{3.2}$$

which physically denotes the energy of one free boson with momentum $k$. Regarding $\omega$ as a non-negative self-adjoint multiplication operator in $L^2(\mathbb{R}^s)$, we define the boson-free Hamiltonian $H_{0B}$ acting in $\mathscr{F}_B(L^2(\mathbb{R}^s))$ by

$$H_{0B}=d\Gamma(\omega)\,, \tag{3.3}$$

where $d\Gamma(\omega)$ is the second quantization of $\omega$ (see Ref. 43, §X.7, p. 208); the formal expression is given as

$$H_{0B}=\int\omega(k)b(k)^*b(k)\,dk\,, \tag{3.4}$$

where $b(k)$, $k\in\mathbb{R}^s$, is the operator-valued distribution kernel of the boson annihilation operator acting in $\mathscr{F}_B(L^2(\mathbb{R}^s))$.

The free Hamiltonian $h_0$ of the harmonic oscillator, which acts in $L^2(\mathbb{R})$, is given by

$$h_0=\frac{1}{2}\left(-\frac{d^2}{dq^2}+\omega_0^2q^2-\omega_0\right)=\omega_0a^*a\,,\quad q\in\mathbb{R}\,, \tag{3.5}$$

where $\omega_0>0$ is a constant parameter denoting the frequency and $a$ is the annihilation operator for the harmonic oscillator defined by

$$a=\frac{1}{\sqrt{2}}\left(\sqrt{\omega_0}q+\frac{1}{\sqrt{\omega_0}}\frac{d}{dq}\right). \tag{3.6}$$

We denote by $b(f)$, $f\in L^2(\mathbb{R}^s)$, the (smeared) boson annihilation operator:

$$b(f)=\int b(k)f(k)\,dk\,. \tag{3.7}$$

The Hamiltonian $H$ that defines the model is given by

$$H=I\otimes H_{0B}+h_0\otimes I+a\otimes b(\rho)^*+a^*\otimes b(\rho)\,, \tag{3.8}$$

where $\rho\in L^2(\mathbb{R}^s)$ denotes a cutoff function and $I$ denotes identity. We shall assume that

$$\text{(A1)}\quad \int\frac{|\rho(k)|^2}{\omega(k)}dk<\infty\,,\quad \int\omega(k)^2|\rho(k)|^2\,dk<\infty\,.$$

Then we have the following proposition.

*Proposition:* The Hamiltonian $H$ is essentially self-adjoint on every core for $I\otimes H_{0B}+h_0\otimes I$.

*Proof:* See Ref. 29, Proposition 2.1. □

*Remark:* By passing to the Schrödinger representation of the canonical commutation relations[22,25] (or the $Q$-space representation[43]), where the time-zero field

$$\phi(x)=\int(2(2\pi)^s\omega(k))^{-1/2}(b(k)+b(-k)^*)\,e^{ikx}\,dk$$

is realized as a distributional multiplication operator, one can see that $H$ can be regarded as an infinite-dimensional Schrödinger operator. In fact, $H$ has the following formal expression:

$$\begin{aligned}H=\frac{1}{2}\Big\{&-\int\frac{\delta^2}{\delta\phi(x)^2}dx+\int[\phi(x)L(x-y)\phi(y)\\&-\delta(x-y)M(x-y)]\,dx\,dy\Big\}\\&+\frac{1}{2}\left(-\frac{\partial^2}{\partial q^2}+\omega_0^2q^2-\omega_0\right)\\&+\sqrt{\omega_0}q\int\phi(x)(\hat{\omega}^{1/2}\hat{\rho})(x)\,dx\\&-\frac{1}{\sqrt{\omega_0}}\frac{\partial}{\partial q}\int(\hat{\omega}^{-1/2}\hat{\rho})(x)\frac{\delta}{\delta\phi(x)}dx\,.\end{aligned} \tag{3.9}$$

Here

$$L(x)=\int\omega(k)^2e^{ikx}\,dk\,,\quad M(x)=\int\omega(k)e^{ikx}\,dk\,,$$

$\hat{f}$ denotes the Fourier transform of $f$, and the operator $\hat{\omega}$ acting in $L^2(\mathbb{R}^s)$ is defined by

$$(\hat{\omega}f)(x)=(2\pi)^{-s/2}\int\omega(k)\hat{f}(k)\,e^{-ikx}\,dk\,.$$

The spectral property of $H$ is analyzed in Ref. 29.

## IV. SOME TECHNICAL FACTS

In this section we summarize some of the technical facts discussed in Ref. 29, which are needed in Secs. V and VI.

Let $\omega_1$ be as in Sec. III,

$$m=\inf_{t>0}\omega_1(t)\geqslant 0\,, \tag{4.1}$$

and put

$$\mathbb{C}_m=\mathbb{C}\setminus[m,\infty)\,. \tag{4.2}$$

In the cut plane $\mathbb{C}_m$, we define the function

$$D(z)=-z+\omega_0+\int_{\mathbb{R}^s}\frac{|\rho(k)|^2}{z-\omega(k)}dk\,,\quad z\in\mathbb{C}_m\,. \tag{4.3}$$

Henceforth, throughout the paper, we assume the following (A2)–(A4) in addition to (A1):

$$\text{(A2)}\quad \sup_{\substack{s>0\\ t\in[m,\infty)}}|D(t\pm i\epsilon)+t\pm i\epsilon-\omega_0|<\infty\,,$$

$$\text{(A3)}\quad \inf_{\substack{s>0\\ t\in[m,\infty)}}|D(t\pm i\epsilon)|>0\,,$$

$$\text{(A4)}\quad \omega_1'(|k|)^{-1/2}|k|^{(s-1)/2}\rho\in L^\infty(\mathbb{R}^s)\,.$$

It follows from (4.3) that $D(z)$ is analytic in $\mathbb{C}_m$ and

$$D'(z)=-1-\int_{\mathbb{R}^s}\frac{|\rho(k)|^2}{(z-\omega(k))^2}dk\,,\quad z\in\mathbb{C}_m\,. \tag{4.4}$$

In particular, we have $D(t)<0$ for all $t\in(-\infty,m)$. Therefore, $D(t)$ is monotone decreasing in $t\in(-\infty,m)$ and hence the limit

$$\alpha_m\equiv\lim_{t\uparrow m}D(t) \tag{4.5}$$

exists. It follows from (A2) and (A3) that $\alpha_m\neq 0,-\infty$. (In Ref. 29, $\alpha_m$ is denoted as $d_m$.) We shall denote the inner product and the norm of $L^2(\mathbb{R}^s)$ by $(\cdot\,,\cdot)_2$ (linear in the left vector) and $\|\cdot\|_2$, respectively.

*Lemma 4.1:* (a) If $\alpha_m > 0$, then $D(z)$ has no zeros in $\mathbb{C}_m$.

(b) If $\alpha_m < 0$, then $D(z)$ has a unique simple zero $\nu \in (-\infty, m)$.

In particular, if $m > 0$ and $\omega_0 \geqslant \|\omega^{-1/2}\rho\|_2^2$ (resp. $m \geqslant 0$ and $\omega_0 < \|\omega^{-1/2}\rho\|_2^2$), then $0 \leqslant \nu < m$ (resp. $\nu < 0$).

*Proof:* An elementary exercise. (In Ref. 29, $\nu$ is denoted as $\omega_A$.) □

*Remark:* The spectral property of $H$ drastically changes according to the sign of $\alpha_m$ (see Ref. 29, Theorem 3.2). We shall see that this fact is related to the nonuniqueness of a SSQT into which $H$ is embedded (see Secs. V and VI).

Let

$$u_1(t) = \omega_1'(t)^{1/2} t^{-(s-1)/2}, \quad t > 0, \tag{4.6}$$

and define

$$u(k) = u_1(|k|), \quad k \in \mathbb{R}^s. \tag{4.7}$$

For $\epsilon > 0$, we define a linear operator $G_\epsilon$ by

$$(G_\epsilon f)(k) = \int \frac{u(k)u(k')f(k')}{\omega(k) - \omega(k') + i\epsilon}\, dk'. \tag{4.8}$$

It is proved that for every $\epsilon > 0$, $G_\epsilon$ is a bounded linear operator on $L^2(\mathbb{R}^s)$ and that the strong limit

$$s\text{-}\lim_{\epsilon \downarrow 0} G_\epsilon \equiv G \tag{4.9}$$

exists. One can easily see that the limits

$$\lim_{\epsilon \downarrow 0} D(t \pm i\epsilon) \equiv D_\pm(t) \tag{4.10}$$

exist for a.e. $t \in (m, \infty)$ and are not zero by (A3). Hence we can define the function

$$F(k) = \rho(k)/D_+(\omega(k)). \tag{4.11}$$

[In Ref. 29, $F(k)$ is denoted as $Q(k)$.]

As already seen, we have $D'(t) < 0$ for all $t < m$. In the case $\alpha_m < 0$, for convenience we define the constant $c_0 > 0$ by

$$c_0^2 = -1/D'(\nu). \tag{4.12}$$

Further, we introduce the function

$$f_0(k) = c_0\rho(k)/[\omega(k) - \nu]. \tag{4.13}$$

[In Ref. 29, $f_0(k)$ is denoted as $\psi_A(k)$.]

In Ref. 29, it is proved that the operator $T$ given by

$$Tf = f - Fu^{-1}Gu^{-1}\rho f, \quad f \in L^2(\mathbb{R}^s), \tag{4.14}$$

is bounded on $L^2(\mathbb{R}^s)$. By $\theta(t)$ we denote the Heaviside function: $\theta(t) = 1$ for $t > 0$ and $\theta(t) = 0$ for $t < 0$.

*Lemma 4.2:* The following formulas hold:

$$T^*T = I - \theta(-\alpha_m)(\cdot, f_0)_2 f_0, \tag{4.15}$$

$$TT^* + (\cdot, F)_2 F = I, \tag{4.16}$$

$$\|F\|_2^2 = 1 - \theta(-\alpha_m)c_0^2, \tag{4.17}$$

$$T^*F + \theta(-\alpha_m)c_0 f_0 = 0, \tag{4.18}$$

$$T\rho + (\omega - \omega_0)F = 0, \tag{4.19}$$

$$[T^*, \omega]f = -(f, F)_2\rho, \quad f \in D(\omega), \tag{4.20}$$

where $D(\omega)$ denotes the domain of $\omega$ as a multiplication operator in $L^2(\mathbb{R}^s)$. Also,

$$Tf_0 + c_0 F = 0, \tag{4.21}$$

$$\|f_0\|_2^2 + c_0^2 = 1. \tag{4.22}$$

[Equations (4.21) and (4.22) are meaningful only for the case $\alpha_m < 0$.]

*Proof:* See Ref. 29, Corollary 4.9 and Lemmas 4.11 and 4.12. □

We shall need more formulas.

*Lemma 4.3:* The following formulas hold:

$$T^*\omega Tf = \omega f - \theta(-\alpha_m)\nu(f, f_0)_2 f_0, \quad f \in D(\omega), \tag{4.23}$$

$$T^*\omega F + \theta(-\alpha_m)c_0\nu f_0 = -\rho, \tag{4.24}$$

$$\|\omega^{1/2}F\|_2^2 + \theta(-\alpha_m)\nu c_0^2 = \omega_0. \tag{4.25}$$

*Proof:* We first note that (4.20) implies that $T$: $D(\omega) \to D(\omega)$. Then, by replacing $f$ in (4.20) by $Tf$ and using (4.15) and (4.18), we have

$$T^*\omega Tf = \omega f - \theta(-\alpha_m)(f, f_0)_2(\omega f_0 - c_0\rho).$$

On the other hand, it is easy to see that

$$\omega f_0 - c_0\rho = \nu f_0. \tag{4.26}$$

Thus (4.23) follows.

Equation (4.19) implies that $F$ is in $D(\omega)$. Then, by putting $f = F$ in (4.20) and using (4.17), (4.18), and (4.26), we obtain (4.24).

To prove (4.25), we take the inner product of (4.19) with $F$. Then, using (4.17) and (4.18), we have

$$\|\omega^{1/2}F\|_2^2 = \omega_0 - \theta(-\alpha_m)c_0^2\left(\omega_0 + \int \frac{|\rho(k)|^2}{\nu - \omega(k)}\, dk\right).$$

Since $D(\nu) = 0$, we see that

$$\omega_0 + \int \frac{|\rho(k)|^2}{\nu - \omega(k)}\, dk = \nu.$$

Thus (4.25) follows. □

## V. SUPERSYMMETRIC EMBEDDING OF $H$— THE CASE $\alpha_m > 0$

Before discussing the supersymmetric embedding of $H$, we remark that if $\alpha_m > 0$, then we have

$$\sigma_p(H) = \{0\}, \quad \sigma_{\text{sing}}(H) = \phi,$$

$$\sigma_{\text{ac}}(H) = \{\omega(k) \mid k \in \mathbb{R}^s\} = [m, \infty),$$

where $\sigma_p$ (resp. $\sigma_{\text{sing}}, \sigma_{\text{ac}}$) denotes the point (resp. singular continuous, absolutely continuous) spectrum and the multiplicity of the eigenvalue zero is 1. In particular, $H$ is nonnegative. These results are proved in Ref. 29.

We first construct a SSQFT. In order to do so, we need to introduce fermionic degrees of freedom.

Let $\mathcal{F}_F(L^2(\mathbb{R}^s))$ be the Fermion Fock space over $L^2(\mathbb{R}^s)$:

$$\mathcal{F}_F(L^2(\mathbb{R}^s)) = \bigoplus_{n=0}^{\infty} \Lambda^n(L^2(\mathbb{R}^s)), \tag{5.1}$$

where $\Lambda^n(L^2(\mathbb{R}^s))$ is the $n$-fold antisymmetric tensor product of $L^2(\mathbb{R}^s)$ (e.g., Ref. 44, Sec. II.4) [$\Lambda^0(L^2(\mathbb{R}^s)) \equiv \mathbb{C}$]. Let $\psi(f)$, $f \in L^2(\mathbb{R}^s)$, be the (smeared) Fermion annihilation operator on $\mathcal{F}_F(L^2(\mathbb{R}^s))$, so that the anticommutation relations

$$\{\psi(f), \psi(g)^*\} = (f, g)_2, \quad \{\psi(f), \psi(g)\} = 0,$$

$$f, g \in L^2(\mathbb{R}^s), \tag{5.2}$$

hold, where $\{A,B\}\equiv AB+BA$.

Let

$$\mathscr{H}=\mathscr{F}\otimes\mathscr{F}_{\mathrm{F}}(L^2(\mathbb{R}^s))\,. \tag{5.3}$$

This will be the Hilbert space of state vectors for the SSQFT. Let

$$\mathscr{H}^{(n)}=\mathscr{F}\otimes\Lambda^n(L^2(\mathbb{R}^s)) \tag{5.4}$$

and

$$\mathscr{H}_+=\bigoplus_{n=0}^{\infty}\mathscr{H}^{(2n)}\,,\quad \mathscr{H}_-=\bigoplus_{n=0}^{\infty}\mathscr{H}^{(2n+1)}\,. \tag{5.5}$$

Then, $\mathscr{H}$ is identified as

$$\mathscr{H}=\mathscr{H}_+\oplus\mathscr{H}_-\,. \tag{5.6}$$

Let

$$\phi_0(q)=(\omega_0/\pi)^{1/4}e^{-\omega_0 q^2/2}\,,\quad q\in\mathbb{R}\,, \tag{5.7}$$

which is the normalized ground-state function of $h_0$ in $L^2(\mathbb{R})$ (i.e., $h_0\phi_0=0$). We denote by $\Omega_{\mathrm{B}}$ and $\Omega_{\mathrm{F}}$ the Fock vacuum in $\mathscr{F}_{\mathrm{B}}(L^2(\mathbb{R}^s))$ and $\mathscr{F}_{\mathrm{F}}(L^2(\mathbb{R}^s))$, respectively. We define the vector $\Omega\in\mathscr{H}$ by

$$\Omega=\phi_0\otimes\Omega_{\mathrm{B}}\otimes\Omega_{\mathrm{F}}\,. \tag{5.8}$$

A closed linear operator $A$ in $L^2(\mathbb{R})$ [resp. $\mathscr{F}_{\mathrm{B}}(L^2(\mathbb{R}^s)),\mathscr{F}_{\mathrm{F}}(L^2(\mathbb{R}^s))$] can be extended to a closed linear operator in $\mathscr{H}$ as $A\otimes I\otimes I$ (resp. $I\otimes A\otimes I, I\otimes I\otimes A$). We shall denote the extension by the same symbol. Henceforth, we shall omit the symbol $\otimes$ in operator tensor products if there would be no confusions.

Let $\mathscr{H}_0$ be the subspace in $\mathscr{H}$ spanned algebraically by vectors of the form

$$\begin{aligned}&\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad\equiv a^{*n}b(f_1)^*\cdots b(f_p)^*\psi(g_1)^*\cdots\psi(g_r)^*\Omega\,,\\&\qquad n,p,r\geqslant 0\,,\quad f_1,\ldots f_p,g_1,\ldots,g_r\in D(\omega)\,.\end{aligned} \tag{5.9}$$

It is easy to see that $\mathscr{H}_0$ is dense in $\mathscr{H}$. We define the operator $d$ on $\mathscr{H}_0$ by

$$\begin{aligned}&d\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad=\sum_{j=1}^{p}\psi(\omega^{1/2}Tf_j)^*\Psi_n(f_1,\ldots,\hat{f}_j,\ldots,f_p;g_1,\ldots,g_r)\end{aligned} \tag{5.10}$$

and extending by linearity to all vectors in $\mathscr{H}_0$, where the operator $T$ is given by (4.14) and $\hat{f}$ indicates omission of $f$.

We denote by $H_{0\mathrm{F}}$ the second quantization of $\omega$ (the fermion free Hamiltonian) in $\mathscr{F}_{\mathrm{F}}(L^2(\mathbb{R}^s))$. We put

$$H_0=H_{0\mathrm{B}}+H_{0\mathrm{F}}\,, \tag{5.11}$$

which is a non-negative self-adjoint operator in $\mathscr{H}$.

*Lemma 5.1:* (a) For every $r\geqslant 0$, $d$ maps $\mathscr{H}^{(r)}\cap\mathscr{H}_0$ into $\mathscr{H}^{(r+1)}$ with

$$d^2\restriction\mathscr{H}^{(r)}\cap\mathscr{H}_0=0\,. \tag{5.12}$$

(b) The operator $d$ is closable and the adjoint $d^*$ is given by

$$\begin{aligned}&d^*\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad=\sum_{i=1}^{r}(-1)^{i-1}\Psi_n(T^*\omega^{1/2}g_i,f_1,\ldots,f_p;g_1,\ldots,\hat{g}_i,\ldots,g_r),\\&\qquad n,p,r\geqslant 0\,,\quad f_1,\ldots,f_p,g_1,\ldots,g_r\in D(\omega)\,.\end{aligned} \tag{5.13}$$

(c) The following anticommutation relations hold on $\mathscr{H}_0$:

$$\{d,\psi(f)^*\}=0\,,\quad f\in L^2(\mathbb{R}^s)\,, \tag{5.14}$$

$$\{d,\psi(f)\}=b(T^*\omega^{1/2}f)\,,\quad f\in D(\omega^{1/2})\,, \tag{5.15}$$

$$\{d,d^*\}=H_0-\psi(\omega^{1/2}F)^*\psi(\omega^{1/2}F)\,. \tag{5.16}$$

*Proof:* (a) The first assertion is obvious from definition (5.10). To prove (5.12), we see that

$$\begin{aligned}&d^2\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad=\sum_{i\neq j}^{p}\Psi_n(f_1,\ldots,\hat{f}_i,\ldots,\hat{f}_j,\ldots f_p;\omega^{1/2}Tf_i,\omega^{1/2}Tf_j,g_1,\ldots,g_r).\end{aligned}$$

On the other hand, $\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)$ is symmetric (resp. antisymmetric) in $(f_1,..,f_p)$ [resp. $(g_1,\ldots,g_r)$]. Hence the rhs of the above equation must vanish. Therefore we obtain (5.12).

(b) Equation (5.13) follows from definition (5.10) and the anticommutation relations (5.2). This shows at the same time that $D(d^*)$, the domain of $d^*$, is dense with $\mathscr{H}_0\subset D(d^*)$. Hence, by a general criterion (e.g., Ref. 44, Theorem VIII.1), $d$ is closable.

(c) We have, from definition (5.10),

$$\begin{aligned}&\{d,\psi(f)^*\}\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad=\sum_{j=1}^{p}\{\psi(\omega^{1/2}Tf_j)^*,\psi(f)^*\}\\&\qquad\times\Psi_n(f_1,\ldots,\hat{f}_j,\ldots,f_p;g_1,\ldots,g_r).\end{aligned}$$

By (5.2), the rhs vanishes. Hence (5.14) follows.

Using (5.2) and the fact

$$\psi(f)\Omega=0\,,\quad f\in L^2(\mathbb{R}^s)\,, \tag{5.17}$$

we have

$$\begin{aligned}&\{d,\psi(f)\}\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad=\sum_{j=1}^{p}(T^*\omega^{1/2}f,f_j)_2\Psi_n(f_1,\ldots\hat{f}_j,\ldots f_p;g_1,\ldots,g_r)\,.\end{aligned} \tag{5.18}$$

On the other hand, by the canonical commutation relation

$$[b(f),b(g)^*]=(f,g)_2\,,\quad f,g\in L^2(\mathbb{R}^s)\,, \tag{5.19}$$

and the fact

$$b(f)\Omega=0\,,\quad f\in L^2(\mathbb{R}^s)\,, \tag{5.20}$$

we see that the rhs on (5.18) is equal to $b(T^*\omega^{1/2}f)\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)$. Hence (5.15) follows.

By direct computations, we have

$$\begin{aligned}&\{d,d^*\}\Psi_n(f_1,\ldots,f_p;g_1,\ldots,g_r)\\&\quad=\sum_{i=1}^{r}\Psi_n(f_1,\ldots,f_p;g_1,\ldots,\omega^{1/2}TT^*\omega^{1/2}g_i,\ldots,g_r)\\&\qquad+\sum_{j=1}^{p}\Psi_n(f_1,\ldots,T^*\omega Tf_j,\ldots,f_p;g_1,\ldots,g_r)\,.\end{aligned} \tag{5.21}$$

Then (4.16) and (4.23) give (5.16). □

Based on (b) of Lemma 5.1, we shall denote the closure of $d\restriction\mathscr{H}_0$ by the same symbol.

Let

$$Q=d-a\psi(\omega^{1/2}F)^*\,. \tag{5.22}$$

*Lemma 5.2:* For every $r\geqslant 0$, $Q$ maps $\mathscr{H}^{(r)}\cap\mathscr{H}_0$ into $\mathscr{H}^{(r+1)}$ and

$$Q^2 \upharpoonright \mathscr{H}^{(r)} \cap \mathscr{H}_0 = 0 . \tag{5.23}$$

Further, $Q$ is closable and

$$Q^* = d^* - a^* \psi(\omega^{1/2} F) \tag{5.24}$$

on $\mathscr{H}_0$.

*Proof:* The first half follows from Lemma 5.1 (a), the fact that

$$\psi(f)^*: \Lambda^{(r)}(L^2(\mathbb{R}^s)) \to \Lambda^{r+1}(L^2(\mathbb{R}^s)),$$

(5.14) and (5.2). It is obvious that $D(Q^*) \supset \mathscr{H}_0$ with (5.24). In particular, $D(Q^*)$ is dense and hence, by a general criterion (e.g., Ref. 44, Theorem VIII.1), $Q$ is closable. □

We shall denote the closure of $Q \upharpoonright \mathscr{H}_0$ by the same symbol.

*Lemma 5.3:* The operator $Q$ (resp. $Q^*$) maps $D(Q)$ [resp. $D(Q^*)$] into itself with

$$Q^2 = 0 , \quad Q^{*2} = 0 . \tag{5.25}$$

*Proof:* We need only to prove that, for all $\Psi$ in $D(Q)$ and $\Phi$ in $D(Q^*)$,

$$(Q\Psi, Q^*\Phi) = 0 . \tag{5.26}$$

For this purpose, we first consider the case $\Psi \in \mathscr{H}_0$. Then (5.26) follows from (5.23). Since $\mathscr{H}_0$ is a core for $Q$ by definition, we can extend the result to all $\Psi$ in $D(Q)$ by a limiting argument. □

Let

$$Q_1 = Q^* + Q , \tag{5.27}$$

$$Q_2 = i(Q^* - Q) . \tag{5.28}$$

*Lemma 5.4:* For each $j = 1,2$, $Q_j$ is closed symmetric on $D(Q) \cap D(Q^*)$. Further we have

$$\{Q_1, Q_2\} = 0 \tag{5.29}$$

in the sense of sesquilinear form on $D(Q) \cap D(Q^*)$.

*Proof:* The symmetricity of $Q_j$ is obvious. For the closedness, we need only note that

$$\|Q_j \Psi\|^2 = \|Q^* \Psi\|^2 + \|Q\Psi\|^2, \quad \Psi \in D(Q) \cap D(Q^*) ,$$

since we have (5.26). Equation (5.29) follows from (5.26). □

The operator

$$\tilde{H}_0 = h_0 + H_0 \tag{5.30}$$

is a non-negative self-adjoint operator and $\mathscr{H}_0$ is a core for $\tilde{H}_0$ (cf. Ref. 44, §VIII.10).

*Proposition 5.5:* For each $j = 1,2$, $Q_j$ is essentially self-adjoint on every core for $\tilde{H}_0$, in particular, on $\mathscr{H}_0$.

*Proof:* The operators $Q_j$, $j = 1,2$, are given explicitly as follows:

$$Q_1 = d + d^* - a^* \psi(\omega^{1/2} F) - a \psi(\omega^{1/2} F)^* , \tag{5.31}$$

$$Q_2 = i(d^* - d) - i a^* \psi(\omega^{1/2} F) + i a \psi(\omega^{1/2} F)^* \tag{5.32}$$

on $\mathscr{H}_0$. By (5.12) and (5.16), we have

$$(d + d^*)^2 = [i(d^* - d)]^2 = H_0 - \psi(\omega^{1/2} F)^* \psi(\omega^{1/2} F)$$

on $\mathscr{H}_0$. It follows from this relation and the boundedness of $\psi(f)$, $f \in L^2(\mathbb{R}^s)$, that $Q_j$ is $\tilde{H}_0$ bounded (in fact, it is $\tilde{H}_0^{1/2}$ bounded). Further, one can show by direct computations that

$$|(Q_j \Psi, \tilde{H}_0 \Psi) - (\tilde{H}_0 \Psi, Q_j \Psi)|$$

$$\leqslant c \|(\tilde{H}_0 + I)^{1/2} \Psi\|^2 , \quad \Psi \in \mathscr{H}_0 ,$$

with a constant $c > 0$. Therefore, by the Glimm–Jaffe–Nelson commutator theorem (Ref. 43, §X.5, Ref. 25, §19.4), we obtain the desired result. □

Let $P_+$ (resp. $P_-$) be the orthogonal projection from $\mathscr{H}$ onto $\mathscr{H}_+$ and put

$$N_F = P_+ - P_- . \tag{5.33}$$

*Lemma 5.6:* For each $j = 1,2$, $N_F$ maps $D(Q_j)$ into itself and

$$\{N_F, Q_j\} = 0 \tag{5.34}$$

on $D(Q_j)$.

*Proof:* We first prove (5.34) on $\mathscr{H}_0$. Then, by a limiting argument using the fact that $\mathscr{H}_0$ is a core for $Q_j$ (Proposition 5.5), we obtain the desired result. □

*Lemma 5.7:* We have

$$Q_1^2 = Q_2^2 = H + H_{0F} \tag{5.35}$$

as the operator equality.

*Proof:* By (5.27), (5.28), and (5.25), we have

$$Q_1^2 = Q_2^2 = \{Q, Q^*\} \tag{5.36}$$

on $\mathscr{H}_0$. On the other hand, by (5.15), (5.16), and (5.2), we have

$$\{Q, Q^*\} = \tilde{H}_0 - a^* b(T^* \omega F) - ab(T^* \omega F)^*$$

$$+ a^* a \|\omega^{1/2} F\|_2^2$$

on $\mathscr{H}_0$. Hence (5.23) and (5.24) give (5.35) on $\mathscr{H}_0$. Since $\mathscr{H}_0$ is a core for the self-adjoint operator $H + H_{0F}$ and $Q_j^2$ is self-adjoint, we obtain the desired result. □

Let

$$H_{SS} = H + H_{0F} . \tag{5.37}$$

Then (5.6), (5.29), and Lemmas 5.6 and 5.7 immediately give the following proposition.

*Proposition 5.8:* The quadruple $\{\mathscr{H}, \{Q_1, Q_2\}, H_{SS}, N_F\}$ is a SSQT with $N = 2$ supersymmetry.

Further, we have the following lemma.

*Lemma 5.9:* The supersymmetric Hamiltonian $H_{SS}$ is reduced by every $\mathscr{H}^{(r)}$, $r \geqslant 0$ and we have

$$H = H_{SS}^{(0)} , \tag{5.38}$$

where $H_{SS}^{(r)}$ is the reduced part of $H_{SS}$ to $\mathscr{H}^{(r)}$.

*Proof:* Let $P^{(r)}$ be the orthogonal projection from $\mathscr{H}$ onto $\mathscr{H}^{(r)}$ and $\Psi$ be in $D(H_{SS})$. Then, by the definition of $H + H_{0F}$ [which is the closure of $H \otimes I + I \otimes H_{0F}$ on $D(H) \otimes D(H_{0F})$], there exists a sequence $\{\Psi_n\} \subset D(H) \otimes D(H_{0F})$ such that $\Psi_n \to \Psi$ and $H_{SS} \Psi_n \to H_{SS} \Psi$. Hence $P^{(r)} \Psi_n \to P^{(r)} \Psi$ and $P^{(r)} H_{SS} \Psi_n \to P^{(r)} H_{SS} \Psi$. On the other hand, it follows from the reducibility of $H_{0F}$ by $\Lambda^r(L^2(\mathbb{R}^s))$ that $P^{(r)} \Psi_n \in D(H_{SS})$ and

$$P^{(r)} H_{SS} \Psi_n = H_{SS} P^{(r)} \Psi_n .$$

Hence, by the closedness of $H_{SS}$, we have $P^{(r)} \Psi \in D(H_{SS})$ and $P^{(r)} H_{SS} \Psi = H_{SS} P^{(r)} \Psi$. Thus, $H_{SS}$ is reduced by $\mathscr{H}^{(r)}$. Then using the fact

$$H_{0F}\Omega_F = 0$$

and (5.37), we obtain (5.38). □

Proposition 5.8 and Lemma 5.9 yield the following theorem.

**Theorem 5.10:** Let $\alpha_m > 0$. Then the Hamiltonian $H$ given by (3.8) is supersymmetrically embeddable into the SSQT $\{\mathscr{H},\{Q_1,Q_2\},H_{SS},N_F\}$.

As seen from (5.37), the bosonic and fermionic degrees of freedom are uncoupled in $H_{SS}$. Therefore, in the SSQT constructed above, the bosons do not interact with the fermions.

## VI. SUPERSYMMETRIC EMBEDDING OF $H$— THE CASE $\alpha_m < 0$

We first note the following proposition.

*Proposition 6.1:* Let $\alpha_m < 0$ and $\omega_0 < \|\omega^{-1/2}\rho\|_2^2$. Then $H$ is not supersymmetrically embeddable.

*Proof:* It is proved in Ref. 29 that under the assumption of Proposition 6.1, the spectrum of $H$ is the whole real line $\mathbb{R}$. Thus the assertion follows (see the Remark after Definition 2.2). □

Henceforth we consider the case $\alpha_m < 0$, $m > 0$, and $\omega_0 > \|\omega^{-1/2}\rho\|_2^2$. In this case, it is proved in Ref. 29 that $0 < \nu < m$ and

$$\sigma_p(H) = \{n\nu\}_{n=0}^{\infty}, \quad \sigma_{sing}(H) = \phi,$$
$$\sigma_{ac}(H) = [m,\infty),$$

where the multiplicity of each eigenvalue $n\nu$ is 1. In particular, $H$ is non-negative. We shall show that $H$ is supersymmetrically embeddable into a SSQFT. Since the method is quite parallel to the preceding case $\alpha_m > 0$, we shall give only the outline.

In the present case, the Hilbert space of state vectors for the SSQFT is taken as

$$\mathscr{H} = \mathscr{F} \otimes \mathscr{F}_F(\mathbb{C} \oplus L^2(\mathbb{R}^s)), \tag{6.1}$$

where $\mathscr{F}_F(\mathbb{C} \oplus L^2(\mathbb{R}^s))$ is the Fermion Fock space over $\mathbb{C} \oplus L^2(\mathbb{R}^s)$ [cf. (5.3)]. We shall denote by $\Psi(\lambda,f)$, $(\lambda,f) \in \mathbb{C} \oplus L^2(\mathbb{R}^s)$, the fermion annihilation operator on $\mathscr{F}_F(\mathbb{C} \oplus L^2(\mathbb{R}^s))$ and put

$$\psi(f) = \Psi(0,f), \tag{6.2}$$

$$a_F = \Psi(1,0). \tag{6.3}$$

Then $\psi(\cdot)$ satisfies (5.2) and the following anticommutation relations hold:

$$\{a_F, a_F^*\} = I, \quad a_F^2 = 0, \tag{6.4}$$

$$\{\psi(f), a_F\} = 0 = \{\psi(f), a_F^*\}, \quad f \in L^2(\mathbb{R}^s). \tag{6.5}$$

Let $\Omega_F$ be the Fock vacuum in $\mathscr{F}_F(\mathbb{C} \oplus L^2(\mathbb{R}^s))$ and define the vectoor $\Omega$ in $\mathscr{H}$ by the rhs of (5.8). Let $\mathscr{H}_0$ be the subspace spanned algebraically by vectors of the form

$$a^{*n} b(f_1)^* \cdots b(f_p)^* \psi(g_1)^* \cdots \psi(g_r)^* a_F^{*q} \Omega,$$
$$n,p,r \geqslant 0, \quad q = 0,1, \quad f_1,\ldots,f_p, g_1,\ldots,g_r \in D(\omega).$$

Then, as in the preceding case $\alpha_m > 0$, we define the operators $Q_1$ and $Q_2$ on $\mathscr{H}_0$ by (5.27) and (5.28) with

$$Q = d - a\psi(\omega^{1/2}F)^* + \nu^{1/2} a_F^* (b(f_0) - c_0 a), \tag{6.6}$$

where $d$ is defined by (5.10). One can show that $Q \upharpoonright \mathscr{H}_0$ is closable (cf. Lemma 5.2); we shall denote the closure by the same symbol. Further, in the present case also Lemmas 5.3 and 5.4 hold.

Let

$$K_0 = h_0 + H_0 + \nu a_F^* a_F. \tag{6.7}$$

Then we have the following proposition.

*Proposition 6.2:* Each $Q_j$, $j = 1,2$, is essentially self-adjoint on every core for $K_0$, in particular, on $\mathscr{H}_0$.

*Proof:* Similar to the proof of Proposition 5.5. □

Let $\Lambda_j^r(L^2(\mathbb{R}^s))$, $j = 0,1$, $r \geqslant 0$, be the closure of the subspace generated by vectors of the form

$$\psi(f_1)^* \cdots \psi(f_r)^* a_F^{*j} \Omega_F, \quad f_1,\ldots,f_r \in L^2(\mathbb{R}^s).$$

Then we have

$$\mathscr{F}_F(\mathbb{C} \oplus L^2(\mathbb{R}^s)) = \bigoplus_{r=0}^{\infty} [\Lambda_0^r(L^2(\mathbb{R}^s)) \oplus \Lambda_1^r(L^2(\mathbb{R}^s))]$$

and hence

$$\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_- \tag{6.8}$$

with

$$\mathscr{H}_+ = \bigoplus_{r=0}^{\infty} \mathscr{F} \otimes [\Lambda_0^{2r}(L^2(\mathbb{R}^s)) \oplus \Lambda_1^{2r+1}(L^2(\mathbb{R}^s))], \tag{6.9}$$

$$\mathscr{H}_- = \bigoplus_{r=0}^{\infty} \mathscr{F} \otimes [\Lambda_0^{2r+1}(L^2(\mathbb{R}^s)) \oplus \Lambda_1^{2r}(L^2(\mathbb{R}^s))]. \tag{6.10}$$

We denote by $P_\pm$ the orthogonal projection from $\mathscr{H}$ onto $\mathscr{H}_\pm$ and put

$$N_F = P_+ - P_-. \tag{6.11}$$

Then one can show in the same way as in Lemma 5.6 that $N_F: D(Q_j) \to D(Q_j)$, $j = 1,2$, and that (5.34) holds.

*Lemma 6.3:* We have

$$Q_1^2 = Q_2^2 = H + H_{0F} + \nu a_F^* a_F \tag{6.12}$$

as the operator equality.

*Proof:* As in the proof of Lemma 5.7, we have (5.36). By direct computation, we have

$$\begin{aligned}\{Q,Q^*\} = H_0 &- a^* b(T^*\omega F + \nu c_0 f_0) \\ &- ab(T^*\omega F + \nu c_0 f_0)^* \\ &- \nu^{1/2} a_F \psi(\omega^{1/2}(Tf_0 + c_0 F))^* \\ &- \nu^{1/2} a_F^* \psi(\omega^{1/2}(Tf_0 + c_0 F)) \\ &+ (\|\omega^{1/2}F\|_2^2 + \nu c_0^2) a^* a \\ &+ \nu(\|f_0\|_2^2 + c_0^2) a_F^* a_F\end{aligned}$$

on $\mathscr{H}_0$. Then by (4.21), (4.22), (4.24), and (4.25), we obtain (6.12) on $\mathscr{H}_0$. Since $\mathscr{H}_0$ is a core for the self-adjoint operator $H + H_{0F} + \nu a_F^* a_F$, the result can be extended to operator equality. □

Let

$$H_{SS} = H + H_{0F} + \nu a_F^* a_F. \tag{6.13}$$

Then the above results show that $\{\mathscr{H},\{Q_1,Q_2\},H_{SS},N_F\}$ is a SSQT with $N = 2$ supersymmetry.

Let

$$\mathscr{H}^{(0)} = \mathscr{F} \otimes \{\Omega_F\} \subset \mathscr{H}_+. \tag{6.14}$$

Then, in the same way as in Lemma 5.9, one can prove that $H_{SS}$ given by (6.13) is reduced by $\mathscr{H}^{(0)}$ and

$$H = H_{SS} \upharpoonright \mathscr{H}^{(0)}, \tag{6.15}$$

the reduced part of $H_{SS}$ to $\mathscr{H}^{(0)}$. Thus we conclude that $H$ is supersymmetrically embeddable.

As seen from (6.13), in the SSQT constructed above, the bosons do not interact with the fermions.

## VII. CONCLUDING REMARKS

### A. A renormalized case

Let us consider the case where a parameter is renormalized so that $H$ is always non-negative. Namely, let

$$\tilde{\omega}_0 = \omega_0 + m + \int \frac{|\rho(k)|^2}{\omega(k) - m}\, dk$$

and $H_{ren}$ be $H$ with $\tilde{\omega}_0$ in place of $\omega_0$. Then it is shown[29] (or easy to see) that $H_{ren}$ is always non-negative. Further, we see, by replacing $\omega_0$ in everything by $\tilde{\omega}_0$, that the discussion in Sec. V works and that $H_{ren}$ is supersymmetrically embeddable independent of the range of the parameters.

### B. Exceptional cases

For a complete analysis of supersymmetric embedding of the present model, it still remains for us to consider the two exceptional cases: the case with $\alpha_m = 0$ and the case where $\alpha_m < 0$, $m > 0$, and $\omega_0 = \|\omega^{-1/2}\rho\|_2^2$. In both cases, $H$ is non-negative[29] and, in the same way as in Secs. V and VI, we can show that $H$ is supersymmetrically embeddable. Thus we have proved that $H$ is supersymmetrically embeddable if it is non-negative.

### C. Index problem

Let $\{\mathscr{H}, \{Q_j\}_{j=1}^N, H_{SS}, N_F\}$ be a SSQT (Definition 2.1). Then one can show[40,41] that $H_{SS}$ is reduced by $\mathscr{H}_\pm$. Let $H_{SS,\pm}$ be the reduced part of $H_{SS}$ to $\mathscr{H}_\pm$. Then the *Witten index* $I_W(H_{SS})$ is defined by

$$I_W(H_{SS}) = \dim \ker H_{SS,+} - \dim \ker H_{SS,-}\,, \tag{7.1}$$

which physically means the number of bosonic zero-energy states minus the number of fermionic zero-energy states (e.g., Refs. 1–3 and 15 and 16). It is well known[2] that $I_W(H_{SS})$ is the index of an operator: Since $Q_j$ maps $D(Q_j) \cap \mathscr{H}_\pm$ into $\mathscr{H}_\mp$ [property (d) in Definition 2.1], one can define the operator $Q_{j+} : D(Q_j) \cap \mathscr{H}_+ \to \mathscr{H}_-$ by

$$Q_{j+} = Q_j \restriction D(Q_j) \cap \mathscr{H}_+\,. \tag{7.2}$$

The index of $Q_{j+}$ is defined by

$$\text{index } Q_{j+} = \dim \ker Q_{j+} - \dim \ker Q_{j+}^*\,, \tag{7.3}$$

provided that at least one of $\dim \ker Q_{j+}$ and $\dim \ker Q_{j+}^*$ is finite (if both are finite and Ran $Q_{j+}$ is closed, then $Q_{j+}$ is called Fredholm; in this case, index $Q_{j+}$ is the Fredholm index of $Q_{j+}$ (e.g., Ref. 45)]. Then it is easy to see[2] that

$$I_W(H_{SS}) = \text{index } Q_{j+}\,,\quad j = 1,\ldots,N\,. \tag{7.4}$$

The supercharges defined by Secs. V and VI may be regarded as infinite-dimensional Kähler–Dirac-type operators. It is interesting to compute their indices as a special case of the index problem in infinite-dimensional manifolds[26,46] as in the following proposition.

*Proposition:* (a) Let $\alpha_m > 0$ and let $\mathscr{H}_\pm$ and $Q_j$, $j = 1,2$, be given by (5.5), (5.31), and (5.32), respectively. Then

$$\text{index } Q_{j+} = 1\,,\quad j = 1,2\,. \tag{7.5}$$

(b) Let $\alpha_m < 0$, $m > 0$, $\omega > \|\omega^{-1/2}\rho\|_2^2$, and let $\mathscr{H}_\pm$ be given by (6.9) and (6.10). Let $Q_j$, $j = 1,2$, be defined by (5.27) and (5.28) with $Q$ given by (6.6). Then

$$\text{index } Q_{j+} = 1\,,\quad j = 1,2\,. \tag{7.6}$$

*Proof:* (a) By (5.37) and the non-negativity of $H$ and $H_{0F}$ (see the beginning of Sec. V), $H_{SS}\Psi = 0$ [$\Psi \in D(H_{SS})$] implies

$$H\Psi = 0\,, \tag{7.7}$$

$$H_{0F}\Psi = 0\,. \tag{7.8}$$

It is well known that, in $\mathscr{F}_F(L^2(\mathbb{R}^s))$, every vector annihilated by $H_{0F}$ is a constant multiple of the Fock vacuum $\Omega_F$. Therefore Eq. (7.8) implies that

$$\Psi = \text{const}\, \Phi \otimes \Omega_F \in \mathscr{H}^{(0)} \subset \mathscr{H}_+\,. \tag{7.9}$$

On the other hand, it is proved in Ref. 29 that, in $\mathscr{F}$, $H\Phi = 0$ if and only if $\Phi$ is a constant multiple of $\phi_0 \otimes \Omega_B$. Combining this fact with (7.9), we obtain $\Psi = \text{const}\,\Omega$, with $\Omega$ defined by (5.8). Thus we have proved that

$$\dim \ker H_{SS,+} = 1\,,\quad \dim \ker H_{SS,-} = 0\,, \tag{7.10}$$

which, together with (7.4), give (7.5).

(b) In this case also, we can prove (7.10) in the same way as in part (a). Thus (7.6) follows. □

## ACKNOWLEDGMENTS

This work was supported in part by Grants-In-Aid No. 63740063 and No. 62460001 for science research from the Ministry of Education, Japan.

[1] E. Witten, "Dynamical breaking of supersymmetry," Nucl. Phys. B **188**, 513 (1981).
[2] E. Witten, "Constraints on supersymmetry breaking," Nucl. Phys. B **202**, 253 (1982).
[3] E. Witten, "Supersymmetry and Morse theory," J. Differ. Geom. **17**, 661 (1982).
[4] P. Salomonson and J. W. Holten, "Fermionic coordinates and supersymmetry in quantum mechanics," Nucl. Phys. B **196**, 509 (1982).
[5] M. de Crombrugghe and V. Rittenberg, "Supersymmetric quantum mechanics," Ann. Phys. (NY) **151**, 99 (1983).
[6] E. D'Hoker and L. Vinet, "Supersymmetry of the Pauli equation in the presence of a magnetic monopole," Phys. Lett. B **137**, 72 (1984).
[7] E. D'Hoker and L. Vinet, "Spectrum (super-) symmetries of particles in a Coulomb potential," Nucl. Phys. B **260**, 79 (1985).
[8] S. Fubini and E. Rabinovici, "Superconformal quantum mechanics," Nucl. Phys. B **245**, 17 (1984).
[9] A. B. Balantekin, "Accidental degeneracies and supersymmetric quantum mechanics," Ann. Phys. (NY) **164**, 277 (1985).
[10] M. Claudson and M. Halpern, "Supersymmetric ground state wave functions," Nucl. Phys. B **250**, 689 (1985).
[11] L. F. Urrutia and E. Hernández, "Long-range behavior of nuclear forces as a manifestation of supersymmetry in nature," Phys. Rev. Lett. **51**, 755 (1983).
[12] J. C. D'Olivo, L. F. Urrutia, and F. Zertuche, "Study of a three-dimensional quantum-mechanical supersymmetric model with nucleon–nucleon-type interaction," Phys. Rev. D **32**, 2174 (1985).
[13] J. C. D'Olivo, L. F. Urrutia, and F. Zertuche, "Scattering regime in supersymmetric quantum mechanics," Phys. Lett. B **159**, 307 (1985).
[14] As an interesting mathematical aspect of supersymmetric quantum theo-

ry, relation to index theorems has also been discussed extensively. See, e.g., Refs. 2, 3, 15, and 16.
[15]L. Alvrez-Gaumé, "Supersymmetry and the Atiyah–Singer index theorem," Commun. Math. Phys. **90**, 161 (1983).
[16]D. Friedan and P. Windey, "Supersymmetric derivation of the Atiyah–Singer index and the chiral anomaly," Nucl. Phys. B **235**, 395 (1984).
[17]E. Gozzi, "Ground-state wave-function 'representation,' " Phys. Lett. B **129**, 432 (1983).
[18]J. Camboa and J. Zanelli, "Supersymmetric non-relativistic quantum mechanics," Phys. Lett. B **165**, 91 (1985).
[19]O. Castaños, J. C. D'Olivo, S. Hojman, and L. F. Urrutia, "Supersymmetric embedding of arbitrary $n$-dimensional scalar Hamiltonians," Phys. Lett. B **174**, 307 (1986).
[20]A. Arai, "A criterion for the boundedness from below with a class of symmetric operators and its applications," J. Math. Anal. Appl. (in press).
[21]A. Arai, "Formal aspects of supersymmetric embedding of Hamiltonians in quantum scalar field theories," Lett. Math. Phys. **15**, 275 (1988).
[22]G. Rosen, *Formulation of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).
[23]E. Gozzi, "Functional-integral approach to Parisi–Wu stochastic quantization: Scalar theory," Phys. Rev. D **28**, 1922 (1983).
[24]A. Arai, "On mathematical construction of supersymmetric quantum field theory associated with Parisi–Wu stochastic quantization," MPI (München) preprint MPI-PAE/PTh, 60/84 (1984), unpublished.
[25]J. Glimm and A. Jaffe, *Quantum Physics* (Springer, Berlin, 1981).
[26]A. Arai, "Path integral representation of the index of Kähler–Dirac operators on an infinite dimensional manifold," J. Funct. Anal. (in press).
[27]E. Braun, "Irreversible behavior of a quantum harmonic oscillator coupled to a heat bath," Physica A **129**, 262 (1985).
[28]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[29]A. Arai, "Spectral analysis of a quantum harmonic oscillator coupled to infinitely many scalar bosons," J. Math. Anal. Appl. (in press).
[30]J. Beckers and V. Hussin, "Dynamical supersymmetries of the harmonic oscillators," Phys. Lett. A **118**, 319 (1986).
[31]J. Beckers, D. Dehin, and V. Hussin, "Symmetries and supersymmetries of the quantum harmonic oscillator," J. Phys. A Math. Gen. **20**, 1137 (1987).
[32]E. M. Henley and W. Thirring, *Elementary Quantum Field Theory* (McGraw-Hill, New York, 1962).
[33]F. Schwabl and W. Thirring, "Quantum theory of laser radiation," Ergeb. Exakt. Naturwiss. **36**, 219 (1964).
[34]L. Rosen, "Renormalization of the Hilbert space in the mass shift model," J. Math. Phys. **13**, 918 (1972).
[35]W. J. Eachus and L. Streit, "Exact solution of the quadratic interaction Hamiltonian," Rep. Math. Phys. **4**, 161 (1973).
[36]A. Arai, "On a model of a harmonic oscillator coupled to a quantized, massless, scalar field. I, II," J. Math. Phys. **22**, 2539, 2549 (1981).
[37]A. Arai, "A note on scattering theory in non-relativistic quantum electrodynamics," J. Phys. A Math. Gen. **16**, 49 (1983).
[38]A. Arai, "Rigorous theory of spectra and radiation for a model in quantum electrodynamics," J. Math. Phys. **24**, 1896 (1983).
[39]A. Arai, "Perturbation of embedded eigenvalues in Fock spaces: A general class of exactly soluble models," *Hokkaido University Preprint Series in Mathematics*, No. 21 (February, 1988).
[40]A. Arai, "Supersymmetry and singular perturbations," J. Funct. Anal. **60**, 378 (1985).
[41]H. Grosse and L. Pittner, "Supersymmetric quantum mechanics defined as sesquilinear forms," J. Phys. A Math. Gen. **20**, 4265 (1987).
[42]A. Arai, "Some remarks on scattering theory in supersymmetric quantum systems," J. Math. Phys. **28**, 472 (1987).
[43]M. Reed and B. Simon, *Methods of Modern Mathematical Physics Vol. II: Fourier Analysis, Self-adjointness* (Academic, New York, 1975).
[44]M. Reed and B. Simon, *Methods of Modern Mathematical Physics Vol. I: Functional Analysis* (Academic, New York, 1972).
[45]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976).
[46]A. Jaffe, A. Lesnieski, and J. Weitsman, "Index of a family of Dirac operators on loop space," Commun. Math. Phys. **112**, 75 (1987).

# The algebraical structure of the electromagnetic tensor and description of charged particles moving in the strong electromagnetic field

Yu. A. Rylov
*Institute for Problems in Mechanics, Academy of Sciences of the USSR, Prospect Vernadskogo 101, Moscow 117526, USSR*

(Received 2 October 1987; accepted for publication 24 August 1988)

In the event space the electromagnetic tensor eigenvectors generate two sets of invariant baselines, which are tangent to these eigenvectors. In the case of a slowly changing electromagnetic field the charged particles are shown to move along a helix around these baselines (the guiding center approximation). In the case of strong radiation damping (a strong field and a small mass) the helical motion of the particle transforms into the motion along the baseline (the massless approximation). In the self-consistent electromagnetic field the gas of charged massless particles can contain two-phase regions (the capture regions). Inside the capture regions the gas can consist of two phases: (1) a dynamical phase (DP) moving with the speed of light, and (2) a statical phase (SP) moving with speed less than that of light. An understanding of the two-phase character of the ultrarelativistic electron–positron gas is very important for investigation of magnetosphere processes in pulsars.

## I. INTRODUCTION

In this paper the case of a charged particle in the strong electromagnetic field is considered when the particle Lorentz factor is very large and the effect of the radiational deceleration is essential. Such a situation arises in some astrophysical problems. For instance, it takes place in the pulsar magnetosphere,[1] where the magnetic field **H**, the electric field **E**, and the electron Lorentz factor $\gamma$ achieve values of the order $H\sim 10^{12}$ G, $E\sim 10^{10}$ V/cm, and $\gamma\sim 10^{10}$.

In the strong slowly changing electromagnetic field the direction of the four-velocity vector $u^i$ $(i=0,1,2,3)$ of the charged particle tends asymptotically (time $t\to\infty$) to one of eigenvectors (eigendirections) of the electromagnetic tensor $F_{ik}$. At large $t$ the particle velocity is directed along one of the eigenvectors, and the particle moves along some baseline, determined by the electromagnetic tensor only. The baseline is tangent to one of eigenvectors at each point.

The electromagnetic tensor eigenvectors are used as a natural basis for decomposing velocity and other vectors. Different versions of this idea were used in many papers dealing with the relativistic particle motion in the electromagnetic field.[2–4] Bertotti[2] used the decomposition of the electromagnetic tensor in two simple bivectors for describing the relativistic plasma in the strong magnetic field. In Ref. 3 the eigenvectors of the electromagnetic tensor were used for numerical calculation of the particle motion in the pulsar electromagnetic field. Fradkin[4] showed that the four-space is separated into mutually orthogonal two-flats spanned by the field eigenvectors associated, respectively, with real or imaginary eigenvalues. He used this circumstance for describing the charged particle guiding center motion. Vandervoort[5] obtained equations for relativistic guiding center motion. Practically, he also used the properties of eigenvectors of the electromagnetic tensor, although in not as explicit a form as Bertotti or Fradkin. Littlejohn[6] used the eigenvector properties for calculating the magnetic moment of the charged particle.

In the present paper the motion of relativistic charged particles (electrons) in the guiding center approximation and in the massless approximation is considered. The relativistical equation for the guiding center motion are considered in a series of papers.[4–7] Our contribution to this field is the derivation of equations that take into account the radiation damping. It is essential for ultrarelativistic particles moving in the strong electromagnetic field. The case where the strong radiation damping suppresses the particle gyration is especially interesting. In this case the equations simplify essentially, and the massless approximation arises. The simplicity of equations in the massless approximation enables us to describe effectively the ultrarelativistic charged particle moving in the self-consistent electromagnetic field. It is done in application to the pulsar magnetosphere.

In Sec. II the algebraic structure of the electromagnetic tensor is considered. Section III is devoted to derivation of the guiding center motion equation. The particle motion in the strong electromagnetic field is considered in Sec. IV. In Sec. V the massless particle motion in the self-consistent electromagnetic field is considered. In Sec. VI the established motion of the massless charged particles in the pulsar magnetosphere is considered.

## II. THE ALGEBRAICAL STRUCTURE OF THE ELECTROMAGNETIC TENSOR

The classical motion of the particle of the mass $m$ and charge $e$ in the given electromagnetic field is described by

$$\frac{dx^j}{d\tau}=u^j,\quad j=0,1,2,3,$$
$$\frac{du^j}{d\tau}=-\frac{e}{mc}F^j{}_k u^k+g^j,\quad j=0,1,2,3, \tag{2.1}$$

 0022-2488/89/020521-16$02.50 © 1989 American Institute of Physics 

where $x^j = (x^0,x^1,x^2,x^3)$ are the particle coordinates in the event space, $c$ is the speed of light, $\tau$ is the proper time,

$$d\tau = (1/c)\sqrt{g_{jk}\,dx^j\,dx^k}\,, \tag{2.2}$$

$g_{jk}$ is the metric tensor, $u^j$ is the four-velocity of the particle,

$$u_j u^j = u^j g_{jk} u^k = c^2, \tag{2.3}$$

$$F^j{}_k = g^{jl}F_{lk}, \quad g^{jl}g_{lk} = \delta^j{}_k \tag{2.4}$$

is the electromagnetic tensor, and $g^j$ is the radiational deceleration. It is defined by the expression

$$g^j = -\frac{2e^3}{3m^2c^4}\left(\delta^j_k - \frac{1}{c^2}u^j u_k\right) \times\left(\frac{dF^k{}_l}{d\tau} - \frac{e}{mc}F^k{}_r F^r{}_l\right)u^l, \quad j = 0,1,2,3. \tag{2.5}$$

Here and later on the summation is made on like Latin superscripts and subscripts from 0 to 3.

In the Galilean coordinate system, where $x^0 = t$ is the time, and $\mathbf{x} = (x^1,x^2,x^3)$ is the position vector of the particle, one has

$$g_{jk} = \left\|\begin{vmatrix} c^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix}\right\|,$$

$$g^{jk} = \left\|\begin{vmatrix} c^{-2} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix}\right\|,$$

$$g = \det\|g_{jk}\| = -c^2, \tag{2.6}$$

$$u^i = \{\gamma,\gamma\mathbf{v}\}, \quad \gamma = \left(1 - \frac{\mathbf{v}^2}{c^2}\right)^{-1/2}, \quad \mathbf{v} = \frac{d\mathbf{x}}{dt}, \tag{2.7}$$

$$F^j{}_k = g^{jl}F_{lk} = \left\|\begin{vmatrix} 0 & -E_1/c & -E_2/c & -E_3/c \\ -cE_1 & 0 & -H_3 & H_2 \\ -cE_2 & H_3 & 0 & -H_1 \\ -cE_3 & -H_2 & H_1 & 0 \end{vmatrix}\right\|, \tag{2.8}$$

where $\mathbf{E} = (E_1,E_2,E_3)$ and $\mathbf{H} = (H_1,H_2,H_3)$ are the electrical field and the magnetic one, respectively, and $\mathbf{v}$ is the particle velocity.

Let us decompose the four-vector $u^i$ over eigenvectors $u^i_{(l)}$ $(l = \pm 1, \pm 2)$ of the electromagnetic tensor $F^j{}_k$:

$$F^j{}_k u^k_{(l)} = \lambda_{(l)} u^j_{(l)}, \quad l = \pm 1, \pm 2, \quad j = 0,1,2,3. \tag{2.9}$$

Here indices in parentheses number eigenvectors $u^i_{(l)}$ and their eigenvalues $\lambda_{(l)}$. There is no summation on $(l)$ in (2.9). Later on the summation over indices in parentheses will be denoted by the symbol of summation, whereas for summation on like Latin superscripts and subscripts from 0 to 3 the symbol of summation will be omitted.

Using decomposition over basic vectors $u^j_{(l)}$,

$$u^j = \sum_{l=\pm 1,\pm 2} c a_l u^j_{(l)}, \quad j = 0,1,2,3, \tag{2.10}$$

one can express the Lorentz force acceleration and the radiational deceleration in the simple form

$$-\frac{e}{mc}F^j{}_k u^k = -\sum_{l=\pm 1,\pm 2}\frac{e\lambda_{(l)}}{mc}a_l u^j_{(l)}, \quad j = 0,1,2,3, \tag{2.11}$$

$$g^j = \frac{2e^3}{3m^2c^4}\sum_{l=\pm 1,\pm 2}\left(1 - \sum_{s,n=\pm 1,\pm 2} a_s a_n u^j_{(s)} u_{(n)k}\right) \times u^k_{(l)}\left(\frac{d}{d\tau} - \frac{e}{mc}\lambda_{(l)}\right)\lambda_{(l)} a_l c, \tag{2.12}$$

and transform the system (2.1) to variables $(x,a)$, where the system becomes more simple.

The eigenvalues $\lambda_{(l)}$ have the form

$$\begin{aligned} -\lambda_{(1)} &= \lambda_{(-1)} \\ &= \lambda \\ &= \sqrt{\tfrac{1}{2}\left[\sqrt{(\mathbf{H}^2 - \mathbf{E}^2)^2 + 4(\mathbf{E},\mathbf{H})^2} - (\mathbf{H}^2 - \mathbf{E}^2)\right]}\,, \\ -\lambda_{(2)} &= \lambda_{(-2)} \\ &= i\nu \\ &= i\sqrt{\tfrac{1}{2}\left[\sqrt{(\mathbf{H}^2 - \mathbf{E}^2)^2 + 4(\mathbf{E},\mathbf{H})^2} + \mathbf{H}^2 - \mathbf{E}^2\right]}\,, \end{aligned} \tag{2.13}$$

and depend only on the electromagnetic field invariants $\mathbf{H}^2 - \mathbf{E}^2$, $(\mathbf{E},\mathbf{H})^2$. The four, generally speaking, different eigenvalues arise as a result of different combinations of signs of the two radicals, which enter into the first of expressions (2.13).

The eigenvectors $u^j_{(l)}$ $(l = \pm 1, \pm 2)$ are defined by the equation (2.9) to within the constant factor. As far as all eigenvectors are null,

$$u^j_{(l)} g_{jk} u^k_{(l)} = 0, \quad l = \pm 1, \pm 2, \tag{2.14}$$

they cannot be normalized routinely. Let us normalize them as follows:

$$\begin{aligned} &u^j_{(l)} g_{jk} u^k_{(-l)} = B_l, \\ &l = 1,2, \quad u^0_{(l)} = u^0_{(-l)} > 0, \quad l = 1,2, \\ &B_1 = \tfrac{1}{2}, \quad B_2 = -\tfrac{1}{2}. \end{aligned} \tag{2.15}$$

Then

$$\begin{aligned} &u^j_{(l)} = \{(1/2c)\sqrt{\sigma_{|l|}/\mu}, \mathbf{u}_l\}, \quad l = \pm 1, \pm 2, \\ &\mathbf{u}_l = (1/2\sqrt{\sigma_1\mu})(l(\lambda\,\mathbf{E} + s_1\nu\mathbf{H}) + \mathbf{E}\times\mathbf{H}), \quad l = \pm 1, \\ &\mathbf{u}_l = (1/2\sqrt{\sigma_2\mu})(i(l/|l|)(\nu\mathbf{E} - s_1\lambda\,\mathbf{H}) + \mathbf{E}\times\mathbf{H}), \\ &\quad l = \pm 2, \end{aligned} \tag{2.16}$$

where

$$\begin{aligned} &\mu = \sqrt{(\mathbf{H}^2 - \mathbf{E}^2)^2 + 4(\mathbf{E},\mathbf{H})^2}\,, \\ &s_1 = (\mathbf{E},\mathbf{H})/|(\mathbf{E},\mathbf{H})|, \\ &\sigma_1 = \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{H}^2 + \mu), \quad \sigma_2 = \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{H}^2 - \mu). \end{aligned} \tag{2.17}$$

One can see from (2.16) that $\mathbf{u}_1$ and $\mathbf{u}_{-1}$ are real, but $\mathbf{u}_2$ and $\mathbf{u}_{-2}$ are complex conjugate each to other:

$$\mathbf{u}_2^* = \mathbf{u}_{-2}, \quad (u^j_{(2)})^* = u^j_{(-2)}, \quad j = 0,1,2,3, \tag{2.18}$$

where an asterisk means complex conjugation. According to (2.16) the scalar product $b_{kl}$ has the form

$$b_{kl} = u^j_{(k)} g_{js} u^s_{(l)} = B_{|l|}\delta_{k,-l}, \quad k,l = \pm 1, \pm 2. \tag{2.19}$$

In the case $\mu = \lambda^2 + \nu^2 \neq 0$ the vectors (2.16) are linear independent and form a basis in the event space. If $\lambda^2 + \nu^2 = 0$ [$\mathbf{E}^2 = \mathbf{H}^2$, $(\mathbf{E},\mathbf{H}) = 0$] then the four eigenvectors are degenerate into one,

$$u^j_{(0)} = \{1/c, (\mathbf{E}\times\mathbf{H})/|\mathbf{E}\times\mathbf{H}|\}, \tag{2.20}$$

and do not form a basis in the event space.

In the case $\lambda^2 + \nu^2 = 0$ the electromagnetic tensor represents a simple bivector (a skew product of two vectors)

$$F^{jk} = u^j_{(0)} u^k_{(5)} - u^j_{(5)} u^k_{(0)},$$
$$\lambda^2 + \nu^2 = 0, \quad j,k = 0,1,2,3, \tag{2.21}$$

where

$$u^j_{(5)} = (0, -\mathbf{E}). \tag{2.22}$$

In the general case $\lambda^2 + \nu^2 \neq 0$, $F^{jk}$ can be represented as a sum of two orthogonal real simple bivectors $F^{jk}_{\mathscr{M}}$ and $F^{jk}_{\mathscr{E}}$, which will be referred to as the magnetic bivector and electric one, respectively:

$$F^{jk} = F^{jk}_{\mathscr{M}} + F^{jk}_{\mathscr{E}}, \quad j,k = 0,1,2,3, \tag{2.23}$$

where

$$F^{jk}_{\mathscr{E}} = \lambda e^{jk}_{\mathscr{E}} = -\frac{\lambda}{B_1}\sum_{l=\pm 1}\frac{l}{|l|}u^j_{(l)}u^k_{(-l)},$$
$$F^{jk}_{\mathscr{M}} = \nu e^{jk}_{\mathscr{M}} = -\frac{i\nu}{B_2}\sum_{l=\pm 2}\frac{l}{|l|}u^j_{(l)}u^k_{(-l)}. \tag{2.24}$$

Here $e^{jk}_{\mathscr{E}}$ and $e^{jk}_{\mathscr{M}}$ are unit simple bivectors, which determine two mutually orthogonal two-dimensional directions; $e^{jk}_{\mathscr{E}}$ and $e^{jk}_{\mathscr{M}}$ will be referred to as the $\mathscr{E}$ direction and $\mathscr{M}$ direction, respectively. Orthogonality of $e^{jk}_{\mathscr{M}}$ and $e^{sl}_{\mathscr{E}}$ follows from (2.19) and (2.24):

$$e^{jk}_{\mathscr{M}} g_{ks} e^{sl}_{\mathscr{E}} = 0, \quad j,l = 0,1,2,3. \tag{2.25}$$

The decomposition (2.23) and (2.24) is invariant with respect to coordinate transformations. It follows from the representation (2.24) of bivectors through the eigenvectors and eigenvalues of the tensor $F^{jk}$.

It should be noted that the transformation properties under the Lorentz transformation of the $u^j_{(l)}$ ($l = \pm 1, \pm 2$) defined by (2.9) and (2.15) are not those of the vector. One can define the $u^j_{(l)}$ ($l = \pm 1, \pm 2$)as vectors under proper Lorentz transformations, substituting the gauge (2.15) by the gauge

$$u^j_{(s)} g_{jk} u^k_{(-s)} = B_{|s|}, \quad s = 1,2, \quad B_1 = \tfrac{1}{2}, \quad B_2 = -\tfrac{1}{2},$$
$$u^0_{(l)} > 0, \quad l = \pm 1, \tag{2.26}$$

where the $u^j_{(s)}$ are real for $s = \pm 1$ and complex for $s = \pm 2$. In this case the expressions (2.23) and (2.24) remain valid. In the case $\lambda \neq 0$, $\nu \neq 0$, (2.9) and (2.26) determine the $u^j_{(s)}$ ($s = \pm 1, \pm 2$) only to within the gauge transformations

$$u^j_{(s)} \to \tilde{u}^j_{(s)} = u^j_{(s)}/\alpha_s,$$
$$u^j_{(-s)} \to \tilde{u}^j_{(-s)} = \alpha_s u^j_{(-s)}, \quad s = 1,2, \tag{2.27}$$

where $\alpha_1$ is positive, $\alpha_2$ is complex, $\alpha_1 > 0$, and $\alpha_2 \neq 0$. In the case, when $\lambda = 0$ or $\nu = 0$, the class of gauge transformations (2.27) becomes wider.

The noncovariant gauge (2.15) will be used further, and the $u^j_{(s)}$ ($s = \pm 1, \pm 2$) will be referred to as vectors, although each of them determines only direction in the event space.

The substitution of the two-dimensional direction by the orthogonal one is realized by means of dual conjugation:

$$\overset{d}{F}{}^{ij} = \tfrac{1}{2}\varepsilon^{ijkl}F_{kl}, \quad \varepsilon_{0123} = -\sqrt{-g} = -c,$$
$$\varepsilon^{0123} = 1/\sqrt{-g} = 1/c, \quad g = \det\|g_{ik}\|, \tag{2.28}$$

where $\varepsilon^{ijkl}$ is the Levi-Civita pseudotensor, which is antisymmetric over all indices.

If one of tensors $F^{ij}$, $F^{ij}_{\mathscr{M}}$, or $F^{ij}_{\mathscr{E}}$ is not equal to zero, then others can be expressed through it and its dual tensor. For instance,

$$F^{jk}_{\mathscr{M}} = \cos\psi(F^{jk}\cos\psi + \overset{d}{F}{}^{jk}\sin\psi),$$
$$F^{jk}_{\mathscr{E}} = -\sin\psi(-F^{jk}\sin\psi + \overset{d}{F}{}^{jk}\cos\psi), \tag{2.29}$$
$$F^{jk}_{\mathscr{E}} = -\tan\psi\,\overset{d}{F}{}^{jk}_{\mathscr{M}}, \quad F^{jk} = F^{jk}_{\mathscr{M}} - \tan\psi\,\overset{d}{F}{}^{jk}_{\mathscr{M}}, \tag{2.30}$$

where $\psi$ is an angle that determines the contribution of the two-dimensional directions $F^{jk}_{\mathscr{M}}$ and $F^{jk}_{\mathscr{E}}$ to $F^{jk}$. The $\psi$ is defined by the expressions

$$\cos(2\psi) = \frac{(FF)}{\mu} = \frac{\mathbf{H}^2 - \mathbf{E}^2}{\mu},$$
$$\sin(2\psi) = \frac{(\overset{d}{F}\overset{d}{F})}{\mu} = \frac{2(\mathbf{E},\mathbf{H})}{\mu}, \tag{2.31}$$

$$\mu = \sqrt{(FF)^2 + (\overset{d}{F}\overset{d}{F})^2} = \sqrt{(\mathbf{H}^2 - \mathbf{E}^2)^2 + 4(\mathbf{E},\mathbf{H})^2},$$
$$(FF) \equiv 2F_{jk}F^{jk} = \mathbf{H}^2 - \mathbf{E}^2,$$
$$(\overset{d}{F}\overset{d}{F}) \equiv 2F_{jk}\overset{d}{F}{}^{jk} = 2(\mathbf{E},\mathbf{H}).$$

The angle $\psi$ is an invariant, and it is a function of the invariants $\mathbf{E}^2 - \mathbf{H}^2$, $(\mathbf{E},\mathbf{H})^2$. The magnitudes of the bivectors $F^{jk}_{\mathscr{M}}$ and $F^{jk}_{\mathscr{E}}$ are described by the relations

$$(F_{\mathscr{M}}F_{\mathscr{M}}) = \mu\cos^2\psi, \quad (F_{\mathscr{E}}F_{\mathscr{E}}) = -\mu\sin^2\psi. \tag{2.32}$$

The eigenvalues $\lambda_{(l)}$ $l = (\pm 1, \pm 2)$ of the tensor $F^{jk}$ can be represented as analytical functions of the angle $\psi$:

$$\lambda_{(l)} = -\frac{l}{|l|}\sqrt{\mu}\sin\psi, \quad l = \pm 1;$$
$$\lambda_{(l)} = -i\sqrt{\mu}\frac{l}{|l|}\cos\psi, \quad l = \pm 2, \tag{2.33}$$

i.e., $\lambda_{(l)}$ are analytical functions of $\psi,\mu$ at finite values of $\psi,\mu$ everywhere except the point $\mu=0$.

Let us refer to the real eigenvectors $u^j_{(s)}$ $(s=\pm 1)$ as base vectors. The base vectors determine two invariant directions at each point of the event space. Let us refer to lines tangent to field of the base vector $u^j_{(s)}$ $(s=\pm 1)$ as the baselines $L_{(s)}$ $(s=\pm 1)$.

The coordinate system whose coordinate lines $x^1$ and $x^2$ coincide with baselines $L_{(1)}$ and $L_{(-1)}$ is the coordinate system distinguished by the electromagnetic field.

If the electromagnetic field is stationary, then the spatial components $\mathbf{u}_1$ and $\mathbf{u}_{-1}$ of the base vectors do not depend on time. In this case the lines $l_{(s)}$ $(s=\pm 1)$ tangent to the field of the vector $\mathbf{u}_{(s)}$ $(s=\pm 1)$ can be built into the configuration space. Such a line will be referred to as the base trajectory $l_{(s)}$ $(s=\pm 1)$.

The world line of the particle moving along the base trajectory $l_{(s)}$ with the velocity $2c(\sigma_1/\mu)^{1/2}\mathbf{u}_s$ realizes the base line $L_{(s)}$. It is worthwhile to remember some ambiguity of the baseline definition, which is connected with ambiguity of numerating the baseline by means of the index $s=\pm 1$. In reality, enumerating the real eigenvectors $u^j_{(s)}$ $(s=\pm 1)$ by means of the sign of the eigenvalue $s=\operatorname{sgn}(-\lambda_{(s)})$ and enumerating by means of the analytical continuation (2.33) lead, generally speaking, to different results. For instance, according to (2.33),

$$\lambda_{(1)}=-\sqrt{\mu}\sin\psi,\quad \lambda_{(-1)}=\sqrt{\mu}\sin\psi. \tag{2.34}$$

The $\lambda_{(1)}$ is negative for $-\pi/2<\psi<\pi/2$, but it is positive for $\pi/2<\psi<3\pi/2$.

## III. THE GUIDING CENTER APPROXIMATION

The complex basis $\mathscr{B}_c$ of vectors $u^j_{(l)}$ $(l=\pm 1,\pm 2)$ is convenient for description of ultrarelativistic particle motion. For describing nonrelativistic or slightly relativistic motion another basis is preferable. It will be used along with the basis $\mathscr{B}_c$. Let us define the real orthonormal basis $B_r$, constituted by vectors

$$\begin{aligned}
&e^j_{(0)}\equiv u^j_{(3)}=u^j_{(1)}+u^j_{(-1)},\\
&e^j_{(1)}\equiv u^j_{(-3)}=u^j_{(1)}-u^j_{(-1)},\\
&e^j_{(2)}\equiv u^j_{(4)}=u^j_{(2)}+u^j_{(-2)},\\
&e^j_{(3)}\equiv u^j_{(-4)}=-i(u^j_{(2)}-u^j_{(-2)}),
\end{aligned} \tag{3.1}$$

$$\begin{aligned}
u^j_{(3)}&=\left\{\frac{1}{c}\sqrt{\frac{\sigma_1}{\mu}},\frac{\mathbf{E}\times\mathbf{H}}{\sqrt{\sigma_1\mu}}\right\},\\
u^j_{(4)}&=\left\{\frac{1}{c}\sqrt{\frac{\sigma_2}{\mu}},\frac{\mathbf{E}\times\mathbf{H}}{\sqrt{\sigma_2\mu}}\right\},\\
u^j_{(-3)}&=\left\{0,\frac{\lambda\mathbf{E}+s_1\nu\mathbf{H}}{\sqrt{\sigma_1\mu}}\right\},\\
u^j_{(-4)}&=\left\{0,\frac{\nu\mathbf{E}-s_1\lambda\mathbf{H}}{\sqrt{\sigma_2\mu}}\right\}.
\end{aligned} \tag{3.2}$$

Further, the electromagnetic field is supposed to change slightly, i.e., for any field quantity $F$ the following condition is fulfilled:

$$\left|\frac{1}{\omega_0}\frac{dF}{dt}\right|\lesssim\epsilon|F|,\quad \epsilon\ll 1, \tag{3.3}$$

where $\omega_0$ is the gyration frequency. Formally, the electromagnetic field $F_{ik}$ is supposed to be a function of a slow variable $X^i=\epsilon x^i$, $\epsilon\ll 1$:

$$F_{ik}=F_{ik}(X)=O(1),\quad \frac{\partial F_{ik}}{\partial x^j}=\epsilon\frac{\partial F_{ik}}{\partial X^j}=O(\epsilon). \tag{3.4}$$

The $\epsilon$ is a small formal parameter of the asymptotic expansion.

Let us introduce designations

$$\omega=\frac{e\nu}{mc},\quad D=\frac{2e^3}{3m^2c^4},\quad t_{\rm rad}=\frac{3m^3c^5}{4e^4\mu}=\frac{3m^3c^5}{4e^4(\lambda^2+\nu^2)}, \tag{3.5}$$

where $\omega$ is the invariant gyration frequency in the coordinate system, where $\mathbf{E}\|\mathbf{H}$, and $t_{\rm rad}$ is a characteristic radiation damping time. The longitudinal electric field $E_\parallel=(\mathbf{E},\mathbf{H})/H$ is supposed to be small as compared with the magnetic field $H$ $(\lambda\ll\nu)$. The radiation damping is supposed to be small in the sense that the change of the gyration magnitude in the gyration period is small:

$$\begin{aligned}
&\lambda/\nu=\epsilon_1\alpha(X),\quad \alpha=O(1),\quad \epsilon_1\ll 1,\\
&1/\omega t_{\rm rad}=\epsilon_1\alpha_{\rm rad}=2\epsilon_1(1+\epsilon_1^2\alpha^2)\alpha_\nu,\quad D\nu=\epsilon_1\alpha_\nu,\\
&\alpha_{\rm rad}=O(1),\quad \alpha_\nu=O(1).
\end{aligned} \tag{3.6}$$

It follows from Eqs. (2.3), (2.10), and (2.15) that

$$a_2^*=a_{-2},\quad a_1a_{-1}-|a_2|^2=1. \tag{3.7}$$

Let us substitute the (2.10)–(2.12) into Eq. (2.1). One obtains

$$\frac{da_l}{d\tau}=\epsilon_1\omega\left(l\alpha-\frac{1}{2}\alpha_{\rm rad}|a_2|^2\right)a_l+\epsilon f_l,\quad l=\pm 1, \tag{3.8}$$

$$\frac{da_l}{d\tau}=\omega\left(i\frac{l}{|l|}-\epsilon_1\frac{\alpha_{\rm rad}}{2}(1+|a_2|^2)\right)a_l+\epsilon f_l,\quad l=\pm 2, \tag{3.9}$$

where

$$\begin{aligned}
f_l=&-c\sum_{r,s=\pm 1,\pm 2}\frac{a_ra_s}{B_l}R_{(-l,r,s)}(1+\epsilon_1\alpha_\nu(\alpha_s-\alpha_l))\\
&-\epsilon_1\alpha_\nu a_l\sum_{r=\pm 1,\pm 2}ca_ru^j_{(r)}\\
&\times\frac{1}{\nu}\frac{\partial}{\partial X^j}(\alpha_l\nu),\quad l=\pm 1,\pm 2,
\end{aligned} \tag{3.10}$$

$$\begin{aligned}
&R_{(l,r,s)}=u_{(l)k}u^j_{(r)}\frac{\partial}{\partial X^j}u^k_{(s)},\\
&l,r,s=\pm 1,\pm 2,\pm 3,\pm 4,
\end{aligned} \tag{3.11}$$

$$\begin{aligned}
&\alpha_l=\frac{\lambda_{(l)}}{\nu},\quad l=\pm 1,\pm 2;\\
&\alpha_l=-\epsilon_1\alpha l/|l|,\quad l=\pm 1;\\
&\alpha_l=-il/|l|,\quad l=\pm 2.
\end{aligned} \tag{3.12}$$

It follows from Eqs. (3.11), (2.18), and (3.1) that

$$R_{(l,r,s)} = -R_{(s,r,l)}, \quad R_{(l,r,l)} = 0,$$
$$s,r,l = \pm 1, \pm 2, \pm 3, \pm 4. \tag{3.13}$$

Using the designations

$$a_2 = a_g e^{i\vartheta}, \quad a_{-2} = a_g e^{-i\vartheta}, \quad a_g = |a_2|, \tag{3.14}$$

Eqs. (3.9) can be rewritten in the form

$$\frac{d \ln a_g^2}{d\tau} = -\epsilon_1 \omega \alpha_{\text{rad}} (1 + a_g^2) + \epsilon f_0, \tag{3.15}$$

$$\frac{d\vartheta}{d\tau} = \omega + \epsilon f_\vartheta, \tag{3.16}$$

where

$$f_0 = \sum_{l=\pm 2} \frac{a_l f_{-l}}{|a_2|^2}, \quad f_\vartheta = \frac{i}{2} \sum_{l=\pm 2} \frac{l a_l f_{-l}}{|l|\,|a_2|^2}. \tag{3.17}$$

Let us consider a transformation from variables $a_l$ to variables $\chi_1, \chi_2, \vartheta$:

$$a_l = a_l(X,\chi,\vartheta_1) = \frac{e^{l(\alpha(X)\vartheta_1 + \chi_l)}}{\sqrt{1 - e^{-\alpha_{\text{rad}}(X)\vartheta_1 - \chi_2}}}, \quad l = \pm 1,$$
$$a_2 = a_g(X,\chi,\vartheta_1)e^{i\vartheta}, \quad a_{-2} = a_g(X,\chi,\vartheta_1)e^{-i\vartheta}, \tag{3.18}$$

$$a_g = a_g(X,\chi,\vartheta_1) = \frac{e^{-(1/2)(\alpha_{\text{rad}}\vartheta_1 + \chi_2)}}{\sqrt{1 - e^{-\alpha_{\text{rad}}\vartheta_1 - \chi_2}}},$$
$$\vartheta = \omega(X)\tau, \quad \vartheta_1 = \epsilon_1 \vartheta. \tag{3.19}$$

It is easy to verify that the relations (3.18) satisfy Eqs. (3.7) identically. If the electromagnetic field does not depend on $X$ and $\chi = \{\chi_1, \chi_2\} = \text{const}$, the relations (3.18) and (3.19) realize the general solution of Eqs. (3.8) and (3.9). For this reason the substitution of relations (3.18) into Eqs. (3.8) and (3.9) leads to equations of the form

$$\frac{d\chi}{d\tau} = \epsilon \mathbf{X}(X,\chi,\vartheta_1,\vartheta) + O(\epsilon^2), \quad \mathbf{X} = O(1),$$
$$\frac{d\vartheta}{d\tau} = \omega(X) + \epsilon \Theta'(X,\chi,\vartheta_1,\vartheta) + O(\epsilon^2), \quad \Theta' = O(1). \tag{3.20}$$

It follows from (3.19) and (3.20) that $\chi$, $\vartheta_1$ are slow variables. Substituting Eqs. (2.10) into the first of Eqs. (2.1), one obtains, for the variable $X = \epsilon x$,

$$\frac{dX}{d\tau} = \epsilon c \Bigg[ \sum_{l=\pm 1} a_l(X,\chi,\vartheta_1) u_{(l)}(X) + \sum_{l=\pm 2} a_l(X,\chi,\vartheta_1,\vartheta) u_{(l)}(X) \Bigg], \tag{3.21}$$

where $\vartheta$ is the only fast variable among the variables $X$, $\chi$, and $\vartheta$. Our purpose is to obtain such a transformation of the system of Eqs. (3.20) and (3.21) that time derivatives of slow variables $X$, $\chi$, and $\vartheta_1$ do not depend on the fast variable $\vartheta$.

Let us represent $X$ in the form

$$X = Y + X_{\text{os}}, \tag{3.22}$$

where $Y$ is a regular component of $X$ and $X_{\text{os}}$ is an oscillatory one. The decomposition is determined by equations

$$\frac{dY}{d\tau} = \epsilon c \Bigg\{ \sum_{l=\pm 1} a_l(Y,\chi,\vartheta_1) u_{(l)}(Y) + \epsilon \sum_{l=\pm 2} X_{\text{os}}^j \frac{\partial}{\partial Y^j} (a_l(Y,\chi,\vartheta_1,\vartheta) u_{(l)}(Y)) \Bigg\} + \epsilon^2 V_g(Y,\chi,\vartheta_1) + O(\epsilon^3),$$
$$\frac{dX_{\text{os}}}{d\tau} = \epsilon c \Bigg\{ \sum_{l=\pm 2} a_l(Y,\chi,\vartheta_1,\vartheta) u_{(l)}(Y) + \epsilon \sum_{l=\pm 1} X_{\text{os}}^j \frac{\partial}{\partial Y^j} (a_l(Y,\chi,\vartheta_1) u_{(l)}(Y)) \Bigg\} - \epsilon^2 V_g(Y,\chi,\vartheta_1) + O(\epsilon^3), \tag{3.23}$$

which are compatible with Eqs. (3.21) and (3.22). The quantity $V_g = O(1)$ will be determined later from the condition that Eqs. (3.23) realize decomposition of $X$ into regular component $Y$ and oscillatory $X_{\text{os}}$ to within $\epsilon$. At $V_g = 0$ such a decomposition is realized only in the zeroth approximation.

Let the linear operator $\mathscr{L} = \mathscr{L}_{X,\epsilon_1}$ depending on parameters $X$, $\epsilon_1$, and the averaging operation $\langle \cdot \rangle_\vartheta$ be defined in the space of quasiperiodical functions $Z$ of $\vartheta$, which can be represented in the form

$$Z(\vartheta_1,\vartheta) = \sum_{n=-\infty}^{\infty} Z_n(\vartheta_1) e^{in\vartheta}, \quad \vartheta = \epsilon_1 \vartheta, \quad \epsilon_1 \ll 1. \tag{3.24}$$

The action of the linear operator $\mathscr{L} = \mathscr{L}_{X,\epsilon_1}$ on $Z$ is defined by its action on the Fourier components $Z_n$:

$$\tilde{Z} = \mathscr{L} Z = \mathscr{L}_{X,\epsilon_1} Z,$$
$$\tilde{Z}_n(X,\vartheta_1) = \hat{L}_{X,\epsilon_1} Z_n(\vartheta_1) = \frac{1 - \delta_{n0}}{\omega(X)(in + \epsilon_1 \partial/\partial\vartheta_1)} Z_n(\vartheta_1), \tag{3.25}$$

where the $\tilde{Z}_n$ are Fourier components of $\tilde{Z}$.

The averaging operation over $\vartheta$,

$$\langle Z(X,\vartheta_1,\vartheta) \rangle_\vartheta \equiv \frac{1}{2\pi} \int_{\vartheta - \pi}^{\vartheta + \pi} Z(X,\vartheta_1,\vartheta')\, d\vartheta' = Z_0(X,\vartheta_1), \tag{3.26}$$

is the extraction operation of zeroth Fourier component.

The solution of the second equation (3.23) has the form

$$X_{\text{os}} = \epsilon c \sum_{l=\pm 2} u_{(l)}(Y) \mathscr{L}_{Y,\epsilon_1} a_l(Y,\chi,\vartheta_1,\vartheta) + O(\epsilon^2). \tag{3.27}$$

Substitution of (3.27) into the first of Eqs. (3.23) leads to an equation of the type

$$\frac{dY}{d\tau} = \epsilon U(Y,\chi,\vartheta_1) + \epsilon^2 V(Y,\chi,\vartheta_1,\vartheta) + \epsilon^2 V_g(Y,\chi,\vartheta_1) + O(\epsilon^3), \tag{3.28}$$

$$U(Y,\chi,\vartheta_1) = c \sum_{l=\pm 1} a_l(Y,\chi,\vartheta_1) u_{(l)}(Y),$$
$$U, V = O(1). \tag{3.29}$$

Substitution of $X = Y + X_{os}$, $X_{os} = O(\epsilon)$ transforms the system (3.20) into a system of the same type with the function $\Theta'$ of another form. So, one can merely replace $X$ in Eqs. (3.20) by $Y$:

$$\frac{d\chi}{d\tau} = \epsilon \mathbf{X}(Y,\chi,\vartheta_1,\vartheta) + O(\epsilon^2),$$
$$\frac{d\vartheta}{d\tau} = \omega(Y) + \epsilon\Theta(Y,\chi,\vartheta_1,\vartheta) + O(\epsilon^2). \tag{3.30}$$

The functions $Z = \{V,\mathbf{X},\Theta\}$ are quasiperiodical functions of the argument $\vartheta$ in the sense that they can be represented in the form (3.24). The Fourier components $Z_n$ change slowly with changing $\vartheta$, because, according to Eq. (3.19), $\vartheta_1 = \epsilon\vartheta$ is a slow variable.

For removing the fast variable $\vartheta$ let us make the change of variables[8]

$$\chi = \bar{\chi} + \epsilon\mathscr{L}_{Y,\epsilon_1}\mathbf{X}(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}),$$
$$\vartheta = \bar{\vartheta} + \epsilon\mathscr{L}_{Y,\epsilon_1}\Theta(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}), \tag{3.31}$$
$$Y = \bar{Y} + \epsilon^2\tilde{V}(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}_1).$$

The function $\tilde{V}$ is defined by the relations

$$\tilde{V}(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}) = \mathscr{L}_{\bar{Y},\epsilon_1}\Big\{V(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}) + \frac{\partial U}{\partial\bar{\chi}}(\bar{Y},\bar{\chi},\bar{\vartheta}_1)\mathscr{L}_{\bar{Y},\epsilon_1}\mathbf{X}(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}) + \epsilon_1\frac{\partial U}{\partial\bar{\vartheta}_1}(\bar{Y},\bar{\chi},\bar{\vartheta}_1)\mathscr{L}_{\bar{Y},\epsilon_1}\Theta(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta}) \times + V_g(\bar{Y},\bar{\chi},\bar{\vartheta}_1)\Big\} + O(\epsilon), \tag{3.32}$$

$$\epsilon V_g(\bar{Y},\bar{\chi},\bar{\vartheta}_1) = \Big\langle \sum_{l=\pm 2} a_l(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta})u_{(l)}(\bar{Y}) + \epsilon\sum_{l=\pm 1} X^j_{os}\frac{\partial}{\partial\bar{Y}^j} \times (a_l(\bar{Y},\bar{\chi},\bar{\vartheta}_1)u_{(l)}(\bar{Y}))\Big\rangle_{\bar{\vartheta}}. \tag{3.33}$$

Such a determination of $V_g(Y,\chi,\vartheta_1)$ leads to

$$\left\langle\frac{dX_{os}}{d\tau}\right\rangle_{\bar{\vartheta}} = O(\epsilon^3),$$
$$\left\langle\frac{dX}{d\tau}\right\rangle_{\bar{\vartheta}} = \left\langle\frac{dX_{os}}{d\tau} + \frac{dY}{d\tau}\right\rangle_{\bar{\vartheta}} = \left\langle\frac{dY}{d\tau}\right\rangle_{\bar{\vartheta}} + O(\epsilon^3). \tag{3.34}$$

The substitution (3.31) transforms Eqs. (3.28) and (3.30) into the form

$$\frac{d\bar{Y}}{d\tau} = \epsilon U(\bar{Y},\bar{\chi},\bar{\vartheta}_1) + \epsilon^2\bar{V}(\bar{Y},\bar{\chi},\bar{\vartheta}_1) + \epsilon^2 V_g(\bar{Y},\bar{\chi},\bar{\vartheta}_1) + O(\epsilon^3),$$
$$\frac{d\bar{\chi}}{d\tau} = \epsilon\bar{\mathbf{X}}(\bar{Y},\bar{\chi},\bar{\vartheta}_1) + O(\epsilon^2), \tag{3.35}$$
$$\frac{d\bar{\vartheta}_1}{d\tau} = \epsilon_1(\omega(\bar{Y}) + \epsilon\bar{\Theta}(\bar{Y},\bar{\chi},\bar{\vartheta}_1)) + O(\epsilon_1\epsilon^2),$$

where the functions $\bar{Z} = \{\bar{V},\bar{\mathbf{X}},\bar{\Theta}\}$ are obtained from the functions $Z = \{V,\mathbf{X},\Theta\}$ by means of averaging over $\vartheta$ with $\vartheta_1$ being fixed:

$$\bar{Z}(\bar{Y},\bar{\chi},\bar{\vartheta}_1) = \langle Z(\bar{Y},\bar{\chi},\bar{\vartheta}_1,\bar{\vartheta})\rangle_{\bar{\vartheta}}. \tag{3.36}$$

Thus the system (3.35) is a system of equations for averaged variables $\bar{Y}$, $\bar{\chi}$, and $\bar{\vartheta}$, which are slow variables. The system (3.35) is obtained from the system (3.28) and (3.30) by averaging over the variable $\vartheta$ at fixed slow variable $\vartheta_1$. The system (3.35) is more adequate for numerical integration than (3.28) and (3.30), because it contains only slow variables.

Further, only averaged variables will be used. For brevity the overbar over symbols of these variables will be omitted.

Now let us turn from variables $Y,\chi_1,\chi_2,\vartheta_1$ to variables $Y,a_1,a_{-1},\vartheta$. Let us change the variables $\chi$ by variables $a_1,a_g$, using relations (3.18). Then the second of Eqs. (3.35) transforms into some equations for $a_1,a_g$. As far as the transformation from $\chi_1,\chi_2$ to $a_1,a_{-1}$, $a_g$ contains only slow variables; these equations can be obtained by averaging Eqs. (3.8) and (3.15) over $\vartheta$ at fixed $\vartheta_1$. Averaging Eqs. (3.8), (3.15), and (3.16), one obtains

$$\frac{da_l}{d\tau} = \epsilon_1\omega\Big(l\alpha - \frac{1}{2}\alpha_{rad}a_g^2\Big)a_l + \epsilon\langle f_l\rangle + O(\epsilon^2), \quad l = \pm 1, \tag{3.37}$$

$$\frac{d\ln a_g^2}{d\tau} = -\epsilon_1\omega\alpha_{rad}(1 + a_g^2) + \epsilon\langle f_0\rangle + O(\epsilon^2), \tag{3.38}$$

$$\frac{d\vartheta_1}{d\tau} = \epsilon_1\omega(Y) + \epsilon_1\epsilon\langle f'_\vartheta\rangle + O(\epsilon_1\epsilon^2), \tag{3.39}$$

$$f'_\vartheta = f_\vartheta + \sum_{l=\pm 2} a_l u^j_{(l)}\frac{\partial\omega(Y)}{\partial Y^j}, \tag{3.40}$$

where angular brackets denote the averaging operation (3.26), and $f_\vartheta$ is determined by Eq. (3.17).

The variable $y^j = Y^j/\epsilon$ can be treated as a coordinate of the guiding center. After calculation the first of Eqs. (3.35) can be represented in the form

$$\frac{1}{c}\frac{dy^k}{d\tau} = \sum_{l=\pm 1} a_l u^k_{(l)}(y) + \epsilon\sum_{s,l=\pm 2}\Big\langle(\mathscr{L}_{Y,\epsilon_1}a_s)u^j_{(s)}\frac{\partial}{\partial Y^j}(a_l u^k_{(l)}(Y))\Big\rangle_\vartheta + \frac{\epsilon}{c}V^k_g(Y,\chi,\vartheta) + O(\epsilon^2), \quad k = 0,1,2,3, \tag{3.41}$$

$$\frac{1}{c}V^k_g(Y,\chi,\vartheta_1) = \frac{1}{2}\sum_{l=\pm 2}u^k_{(l)}(Y)\Big\{(1 + a_g^2)\Big\langle a_l\mathscr{L}_{Y,\epsilon_1}\frac{f_0}{1 + a_g^2}\Big\rangle_\vartheta + il\langle a_l\mathscr{L}_{Y,\epsilon_1}f'_\vartheta\rangle_\vartheta\Big\} + O(\epsilon), \quad k = 0,1,2,3. \tag{3.42}$$

In Eqs. (3.41) and (3.42) the quantities $a_1$, $a_g$, $f_0$, and $f'_\vartheta$ are known functions of the arguments $Y$, $\chi$, $\vartheta_1$, and $\vartheta$. They

are defined by the relations (3.18), (3.17), and (3.40). The variables $\chi$ should be eliminated by means of relations (3.18) after calculating the operator $\mathscr{L}$ action and averaging in relation (3.42).

The relation (3.42) has been obtained as follows. Let us substitute relations (3.31) into the second of Eqs. (3.23). Let us expand the obtained expression over powers of $\epsilon$ and use the fact that in (3.31) $\mathbf{X}$ and $\Theta$ are $\epsilon^{-1}\, d\chi/d\tau$ and $\epsilon^{-1}(d\vartheta/d\tau - \omega)$, respectively. These derivatives can be expressed through derivatives $da_l/d\tau$ by means of (3.18). Using Eqs. (3.8) and (3.9) for $da_l/d\tau$ and (3.34), one obtains, after calculations, the expression (3.42).

In the case when $\epsilon_1 \leqslant \epsilon$, the derivative $\partial/\partial\vartheta_1$ in the operator (3.25) produces the terms of $O(\epsilon^2)$ in Eq. (3.41). In this case the operator (3.25) reduces to the multiplication operator, and calculation of (3.41) and (3.42) is simplified.

In the case when $\epsilon_1 = \epsilon^{1/n}$, and $n \geqslant 2$, the variables $\chi$ are slower variables than $a_l$. It can be obtained by comparing (3.37) and (3.38) with (3.35). In this case the numerical integration of the system (3.37)–(3.40) is less effective than that of the system (3.35).

The mean momentum $\langle p^j \rangle$ of the particle can be defined as

$$\langle p^i \rangle = \left\langle \sum_{l=\pm 1,\pm 2} mcu^i_{(l)} a_l(Y,\chi,\vartheta_1,\vartheta) \right\rangle_{\bar{\vartheta}}, \tag{3.43}$$

where $Y, \chi$, and $\vartheta$ are functions (3.31) of the averaged variables $\bar{Y}, \bar{\chi}$, and $\bar{\vartheta}$, the functions $a_l$ being determined by Eq. (3.18). The angular brackets denote the averaging (3.26) over the averaged variable $\bar{\vartheta}$. Using (3.21) and (3.34), on obtains

$$p^i_{\mathrm{GC}} = \langle p^i \rangle_{\bar{\vartheta}} = mc\frac{dy^i}{d\tau} + O(\epsilon^2), \quad i = 0,1,2,3 \tag{3.44}$$

where $y^i$ satisfies Eq. (3.41). It means that the guiding center can be considered as a particle of the mass

$$m_{\mathrm{GC}} = \frac{1}{c}\sqrt{\langle p^i \rangle \langle p_i \rangle}$$

$$= \frac{1}{c}\sqrt{p^i_{\mathrm{GC}} g_{ik} p^k_{\mathrm{GC}}} = m\sqrt{1 + a_g^2} + O(\epsilon), \tag{3.45}$$

whose world line is described by Eq. (3.41) and whose momentum is directed along the four-velocity $dy^i/d\tau$.

It is used to describe the guiding center motion[7] in terms of variables $M, \varepsilon, p_\parallel$, and $\mathbf{v}_\perp$, where $M$ is the magnetic moment of the particle gyration, $\varepsilon$ is the mean energy of the particle, $p_\parallel$ is the momentum component along the magnetic field $\mathbf{H}$, and $\mathbf{v}_\perp$ is the guiding center velocity component orthogonal to $\mathbf{H}$. Let us transform the system (3.37)–(3.39) and (3.41) to such a form, where comparison with the conventional system of equations for the relativistic guiding center can be produced. For simplicity, let us take

$$\alpha_{\mathrm{rad}} = (\epsilon/\epsilon_1)\alpha'_{\mathrm{rad}} = O(\epsilon/\epsilon_1). \tag{3.46}$$

Such a supposition allows one to substitute the small parameter $\epsilon_1$ before $\alpha_{\mathrm{rad}}$ by the small parameter $\epsilon$ ($\epsilon \leqslant \epsilon_1$).

Let us pass in Eqs. (3.37) from the complex basis $\mathscr{B}_c$ to the real orthonormal basis $\mathscr{B}_r$ defined by (3.1) and (3.2). One has

$$a_3 = \tfrac{1}{2}(a_1 + a_{-1}), \quad a_{-3} = \tfrac{1}{2}(a_1 - a_{-1}),$$

$$\sum_{l=\pm 1} a_l u^k_{(l)} = \sum_{l=\pm 3} a_l u^k_{(l)}. \tag{3.47}$$

Using the corollary

$$\sum_{s=\pm 2} R_{(l,s,-s)} = -\frac{1}{2} u^k_{(l)} \frac{\partial}{\partial Y^k} \ln\mu + \epsilon_1 \frac{2\pi}{c} \frac{l\alpha}{|l|\mu} u^k_{(l)} j_k, \quad l = \pm 1, \tag{3.48}$$

of the Maxwell equations

$$\partial_k F^{ik} = (4\pi/c) j^i, \quad \overset{d}{\partial}_k F^{ik} = 0, \quad i = 0,1,2,3, \tag{3.49}$$

one obtains from Eq. (3.37)

$$\frac{da_l}{d\tau} = \epsilon_1 \alpha\omega a_{-l} - \epsilon\left[\frac{1}{2}\omega\alpha'_{\mathrm{rad}} a_g^2 a_l + \frac{cl}{|l|} a_{-l} \sum_{s=\pm 3} a_s R_{(l,s,-l)} - \frac{l}{2|l|} c a_g^2 u^j_{(l)} \frac{\partial \ln\mu}{\partial Y^j}\right] - \epsilon_1\epsilon \frac{2\pi\alpha l\nu}{\mu|l|} \times a_g^2 u^k_{(-l)} j_k + O(\epsilon^2), \quad l = \pm 3. \tag{3.50}$$

Here $j^i$ is the four-current density generating the electromagnetic field $F^{ik}$.

Let us introduce the generalized magnetic moment

$$M = \frac{ma_g^2 c^2}{2\sqrt{\mu}} = \frac{mc^2}{2\sqrt{\mu}}(a_3^2 - a_{-3}^2 - 1) = \frac{M_0}{\sqrt{1 + \epsilon_1^2\alpha^2}} = M_0 + O(\epsilon_1^2), \tag{3.51}$$

where $M_0$ is the magnetic moment. It is defined by means of relation $M_0 H$ = transverse energy in the coordinate system, where $\mathbf{E} \| \mathbf{H}$ and $\nu = \mathbf{H}$. Combining both of Eqs. (3.50), one obtains

$$\frac{dM}{d\tau} = -\frac{1}{2}\epsilon\alpha'_{\mathrm{rad}}\omega\left(1 + \frac{2\sqrt{\mu}M}{mc^2}\right)M - \epsilon_1\epsilon\frac{4\pi\alpha\nu M}{c\mu} \times \sum_{l=\pm 3} a_l u^k_{(-l)} j_k + O(\epsilon^2). \tag{3.52}$$

It follows from Eq. (3.52) that $M = \mathrm{const} + O(\epsilon^2)$, if one can neglect radiation damping ($\alpha_{\mathrm{rad}} = 0$) and the particle moves in vacuum, where $j^i = 0$. The usual magnetic moment $M_0$ conserves only in the case of small $E_\parallel$ [$E_\parallel = O(\epsilon)$, cf. Ref. 6].

If radiation damping is essential, then at $t \to \infty$ $M$ tends to some $M_{\min} = O(\epsilon^2)$. The value of $M_{\min}$ can be obtained from (3.9) in the form

$$M_{\min} = \frac{mc^2}{2\sqrt{\mu}}(a_g^2)_{\min},$$

$$(a_g^2)_{\min} = \left|\frac{2\epsilon c}{\omega} \sum_{s,l=\pm 3} a_l a_s R_{(2,l,s)}\right|^2. \tag{3.53}$$

In order to obtain the result (3.53), it is sufficient to substitute $a_2 = a_g \exp(i\omega\tau)$ in Eq. (3.9). Supposing $a_g = O(\epsilon)$ and neglecting small terms, the integration of the obtained equation for $a_g$ yields the result (3.53).

After transformation Eq. (3.41) is reduced to the form

$$\frac{1}{c}\frac{dy^k}{d\tau} = u^k_{\mathscr{G}} + \epsilon(w^k_1 + w^k_2) + O(\epsilon^2), \quad k = 0,1,2,3, \tag{3.54}$$

$$u^k_{\mathscr{G}} = \sum_{l=\pm 3} a_l u^k_{(l)}, \quad v^k_{\mathscr{G}} = \sum_{l=\pm 3} a_l \frac{l}{|l|} u^k_{(l)}, \tag{3.55}$$

$$w^k_1 = -\frac{mcu_{\mathscr{G}i}}{e\nu\epsilon}\frac{d}{d\tau}e^{ki}_{\mathscr{M}} + \frac{M\sqrt{\mu}}{e\nu}\left(\Pi^k_{\mathscr{G}i}\frac{\partial}{\partial Y^j}e^{ij}_{\mathscr{M}} - \frac{1}{\nu}e^{kj}_{\mathscr{M}}\frac{\partial\nu}{\partial Y^j}\right), \tag{3.56}$$

$$w^k_2 = \frac{\epsilon_1\alpha mc^2}{e\nu(1+4\epsilon_1^2\alpha^2)}\left\{(u^j_{\mathscr{G}}v^l_{\mathscr{G}} + v^j_{\mathscr{G}}u^l_{\mathscr{G}})\frac{\partial}{\partial Y^j}\Pi^k_{\mathscr{M}\cdot l} + 2\epsilon_1\alpha(u^j_{\mathscr{G}}u^l_{\mathscr{G}} + v^j_{\mathscr{G}}v^l_{\mathscr{G}})\frac{\partial}{\partial Y^j}e^{k\cdot}_{\mathscr{M}l}\right\}, \tag{3.57}$$

where $e^{kj}_{\mathscr{M}}$ and $e^{kj}_{\mathscr{G}}$ are the unit bivectors, describing the $\mathscr{M}$ direction and $\mathscr{G}$ direction, respectively,

$$e^{kj}_{\mathscr{M}} = \frac{1}{\nu}F^{kj}_{\mathscr{M}} = \sum_{s=\pm 4}\frac{s}{|s|}u^k_{(s)}u^j_{(-s)} = \sum_{s=\pm 2} 2i\frac{s}{|s|}u^k_{(s)}u^j_{(-s)},$$

$$e^{kj}_{\mathscr{G}} = \frac{1}{\lambda}F^{kj}_{\mathscr{G}} = \sum_{s=\pm 3}\frac{s}{|s|}u^k_{(s)}u^j_{(-s)} = -\sum_{s=\pm 1} 2\frac{s}{|s|}u^k_{(s)}u^j_{(-s)}, \tag{3.58}$$

$\Pi^{kj}_{\mathscr{M}}$ and $\Pi^{kj}_{\mathscr{G}}$ are the projection operators[4] onto the $\mathscr{M}$ direction and the $\mathscr{G}$ direction, respectively,

$$\Pi^{kj}_{\mathscr{M}} = -\sum_{s=\pm 4}u^k_{(s)}u^j_{(s)} = \sum_{s=\pm 2}\frac{1}{B_s}u^k_{(s)}u^j_{(-s)},$$

$$\Pi^{kj}_{\mathscr{G}} = \sum_{s=\pm 3}u^k_{(s)}u^j_{(s)}\frac{s}{|s|} = \sum_{s=\pm 1}\frac{1}{B_s}u^k_{(s)}u^j_{(-s)}. \tag{3.59}$$

The following relations hold:

$$e^{k\cdot}_{\mathscr{M}j}e^{j\cdot}_{\mathscr{M}l} = -\Pi^{k\cdot}_{\mathscr{M}l}, \quad e^{k\cdot}_{\mathscr{G}j}e^{j\cdot}_{\mathscr{G}l} = \Pi^{k\cdot}_{\mathscr{G}l}, \quad e^{k\cdot}_{\mathscr{M}j}e^{j\cdot}_{\mathscr{G}l} = 0, \tag{3.60}$$

$$\Pi^{k\cdot}_{\mathscr{M}j}\Pi^{j\cdot}_{\mathscr{M}l} = \Pi^{k\cdot}_{\mathscr{M}l}, \quad \Pi^{k\cdot}_{\mathscr{G}j}\Pi^{j\cdot}_{\mathscr{G}l} = \Pi^{k\cdot}_{\mathscr{G}l}, \quad \Pi^{k\cdot}_{\mathscr{G}j}\Pi^{j\cdot}_{\mathscr{M}l} = 0,$$

$$\Pi^{k\cdot}_{\mathscr{M}j} + \Pi^{k\cdot}_{\mathscr{G}j} = \delta^k_j. \tag{3.61}$$

All terms in relations (3.54)–(3.57) have a covariant form. They have the estimation $w_1 = O(1)$, $w_2 = O(\epsilon_1)$. The first term in Eq. (3.56) describes the inertial effects connected with the curvature of the baselines. Because of the Maxwell equation (3.49), and (3.58), the second term of Eq. (3.56) can be represented in the form

$$\frac{M\sqrt{\mu}}{e\nu}\Pi^{k\cdot}_{\mathscr{G}i}\frac{\partial}{\partial Y^j}e^{ij}_{\mathscr{M}} = \frac{M\sqrt{\mu}}{e\nu^2}\Pi^{k\cdot}_{\mathscr{G}i}\frac{\partial}{\partial Y^j}F^{ij}_{\mathscr{M}} = \frac{M\sqrt{\mu}}{e\nu^2}\Pi^{k\cdot}_{\mathscr{G}i}\left(\frac{4\pi}{c}j^i - \frac{\partial}{\partial Y^j}F^{ij}_{\mathscr{G}}\right) = \frac{M\sqrt{\mu}}{e\nu^2}\Pi^{k\cdot}_{\mathscr{G}i}\left(\frac{4\pi}{c}j^i - \epsilon_1\alpha\nu\frac{\partial}{\partial Y^j}e^{ij}_{\mathscr{G}}\right), \tag{3.62}$$

where $j^i$ is the four-current generating the electromagnetic field. Usually $j^i = 0$ in the volume, where the particle moves. Then the expression (3.62) is of $O(\epsilon_1)$. The last term in (3.56) depends on the magnetic field magnitude gradient, because $\nu$ is the magnitude of the magnetic field in the coordinate system, where $\mathbf{E}\|\mathbf{H}$. This term is responsible for the gradient drift of the particle. It is the only term containing derivatives of the electromagnetic field magnitude. Other terms depend only on derivatives of the unit bivector $e^{ij}_{\mathscr{M}}$, which describes the $\mathscr{M}$ direction. For the small longitudinal electric field $E_{\|} = O(\epsilon_1)$, $w_2$ is of the order $\epsilon_1$. In this case $w_2/w_1 = O(\epsilon_1)$. But formally the expressions (3.56) and (3.57) are valid for any $\epsilon_1$. In particular, it is valid in the case when $E^2 > H^2$, $\lambda/\nu = \epsilon_1\alpha > 1$.

Thus expressions (3.54)–(3.57) can be used in the case when $E_{\|} = O(1)$, $H = O(1)$, $\epsilon_1 \simeq 1$. But at $\epsilon_1 = 1$ the $\vartheta_1 = \epsilon_1\vartheta$ will be a fast variable, and numerical integration of the system (3.35) will not be as effective as in the case $\epsilon_1 \ll 1$.

One concludes from Eqs. (3.45), (3.44), (3.51), and (3.54) that the "mass of the guiding center"

$$m_{\mathrm{GC}} = m\sqrt{1+\frac{2M\sqrt{\mu}}{mc^2}}\times\left(1+\epsilon\frac{(M\sqrt{\mu}/e\nu)u_{\mathscr{G}i}(\partial/\partial Y^j)e^{ij}_{\mathscr{M}}}{1+2M\sqrt{\mu}/mc^2}\right)+O(\epsilon^2). \tag{3.63}$$

Let us consider the case of a weak electromagnetic field when the radiation damping can be neglected ($\alpha_{\mathrm{rad}} = 0$) and the four-current, generating the electromagnetic field, vanishes in the volume, where the particle moves:

$$j^i = 0, \quad H = O(1), \quad E_{\|} = (\mathbf{E},\mathbf{H})/|H| = O(\epsilon),$$
$$\mathbf{E}_{\perp} = \mathbf{E} - \mathbf{H}E_{\|}/H = O(1), \quad \epsilon_1 = \epsilon. \tag{3.64}$$

Using the designations

$$\mathbf{e}_E = \frac{\mathbf{E}_{\perp}}{|\mathbf{E}_{\perp}|}, \quad \mathbf{e}_H = \frac{\mathbf{H}}{|\mathbf{H}|},$$
$$\mathbf{e}_D = \mathbf{e}_E \times \mathbf{e}_H, \quad g = \sqrt{1-\frac{E_{\perp}^2+E_{\|}^2}{H^2}},$$
$$E = \sqrt{E_{\perp}^2+E_{\|}^2} = E_{\perp} + O(\epsilon^2), \quad s_1 = \frac{(\mathbf{E},\mathbf{H})}{|(\mathbf{E},\mathbf{H})|}, \tag{3.65}$$

and definitions (2.13) and (2.15), one obtains, for the vectors (3.1) of the basis $\mathscr{B}_r$

$$u^j_{(3)} = \left\{\frac{1}{cg}, \frac{E}{Hg}\mathbf{e}_D\right\} + O(\epsilon^2),$$
$$u^j_{(4)} = \left\{\frac{E}{cHg}, \frac{1}{g}\mathbf{e}_D\right\} + O(\epsilon^2),$$
$$u^j_{(-3)} = \left\{0, s_1\mathbf{e}_H + \epsilon\alpha\frac{E}{H}\mathbf{e}_E\right\} + O(\epsilon^2),$$
$$u^j_{(-4)} = \left\{0, \mathbf{e}_E - \epsilon\frac{s_1}{H}\alpha\mathbf{e}_H\right\} + O(\epsilon^2). \tag{3.66}$$

Let us transform the system of equations (3.50), (3.52), and (3.54) to the variables $M$, $p_\parallel$, and $\mathbf{v}_\perp$, which are defined by the relation (3.51) and by the relations

$$p_\parallel = \left(\mathbf{e}_H, \frac{d\mathbf{y}}{d\tau}\right)mc = mca_{-3} + O(\epsilon),$$
$$\frac{d\mathbf{y}_\perp}{d\tau} = \mathbf{v}_\perp = \frac{1}{\gamma_{GC}}\left(\frac{d\mathbf{y}}{d\tau} - \mathbf{e}_H\left(\mathbf{e}_H, \frac{d\mathbf{y}}{d\tau}\right)\right). \tag{3.67}$$

Here $\gamma_{GC}$ is the Lorentz factor of the guiding center,

$$\gamma_{GC} = \gamma_0 + \epsilon\gamma_1 + O(\epsilon^2) = \frac{dy^0}{d\tau}$$
$$= \frac{dt}{d\tau} = \frac{a_3}{g} + \epsilon\gamma_1 + O(\epsilon^2), \tag{3.68}$$

and $t = y^0$ is the time.

Integration of Eq. (3.52) (at $\epsilon = \epsilon_1$) leads to the result

$$M = \text{const} + O(\epsilon^2). \tag{3.69}$$

In terms of the variables in (3.67), Eqs. (3.50) and (3.54) take the form

$$\frac{dp_\parallel}{dt} = eE_\parallel + mc\gamma_0\mathbf{u}_D\frac{d\mathbf{e}_H}{dt} - \frac{M}{\gamma_0}(\mathbf{e}_H, \nabla)gH + O(\epsilon^2), \tag{3.70}$$

$$\frac{1}{c}\frac{d\mathbf{y}_\perp}{dt} = \mathbf{e}_H \times \left\{-\frac{E}{H}\mathbf{e}_E + \frac{E_\parallel v_\parallel}{Hg^2c}\mathbf{u}_D + \frac{m\gamma_0 c}{eHg^2}\left(\frac{v_\parallel}{c}\frac{d\mathbf{e}_H}{dt} + \frac{d\mathbf{u}_D}{dt}\right) + \frac{M}{eHg^2\gamma_0}\left(\nabla + \mathbf{u}_D\frac{1}{c}\frac{\partial}{\partial t}\right)(Hg)\right\} - \mathbf{u}_D\frac{\epsilon\gamma_1}{\gamma_0^2} + O(\epsilon^2). \tag{3.71}$$

Here the following designations are used:

$$\mathbf{u}_D = \frac{E}{H}\mathbf{e}_D, \quad v_\parallel = \frac{p_\parallel}{m\gamma_0} = \frac{ca_{-3}}{a_3}g,$$
$$\gamma_0 = \frac{1}{g}\sqrt{1 + \frac{p_\parallel^2}{m^2c^2} + 2g\frac{HM}{mc^2}} = \frac{a_3}{g},$$
$$\epsilon\gamma_1 = -\frac{mcE}{eg^2H^2}\left[\mathbf{e}_E\frac{d\mathbf{u}_D}{dt} + \frac{v_\parallel}{c}\mathbf{e}_E\frac{d\mathbf{e}_H}{dt} + \frac{M}{mc}(\mathbf{e}_E, \nabla)(Hg)\right],$$
$$\nabla = \left\{\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}, \frac{\partial}{\partial y^3}\right\}. \tag{3.72}$$

Let us compare the obtained equations (3.69)–(3.71) with the corresponding equations presented in the book by Northrop.[7] Equations (3.69) and (3.70) coincide with the corresponding equations (1.78) and (1.77) of Ref. 7. But Eq. (3.71), as compared with the corresponding Eq. (1.76) of Ref. 7, has an extra term $-\epsilon\mathbf{u}_D\gamma_1/\gamma_0^2$. This term vanishes in the coordinate system where $E_\perp = 0$. Then Eq. (3.71) coincides with the corresponding equation (1.68) of Ref. 7. Apparently, the loss of the extra term in Northrop's paper is connected with the fact that the Lorentz transformation from the system, where $E_\perp = 0$, is produced only to within $\epsilon^0$, whereas in our case it is produced to within $\epsilon$.

## IV. STRONG SELF-CONSISTENT ELECTROMAGNETIC FIELD. MASSLESS APPROXIMATION

Let the electromagnetic field be strong and the radiation damping be essential in the sense that

$$\gamma t_{rad} \ll \frac{L}{c} \quad \text{or} \quad \left(\frac{mc^2}{e\sqrt{\mu}L}\right)^2 \ll \frac{e^2}{\varepsilon L}, \tag{4.1}$$

where $L$ is the space scale, $T = L/c$ is the time scale, and $\varepsilon$ is the particle energy. If the condition (4.1) is fulfilled, then the particle magnetic moment $M$ quickly approaches its asymptotic value $M_{min} = O(\epsilon^2)$, determined by Eq. (3.53). In this case the magnetic moment $M$ can be considered to have its asymptotic value $M_{min} = O(\epsilon^2)$. Formally, the terms containing $M$ can be omitted in Eqs. (3.50) and (3.54) as the terms of the order $\epsilon^2$.

Let us introduce variables $\varepsilon$ and $p_\parallel$ by means of

$$\varepsilon = mc^3a_3u^0_{(3)} = mc^2a_3\sqrt{\sigma_1/\mu},$$
$$p_\parallel = mca_{(-3)}|\mathbf{u}_{(-3)}| = mca_{-3},$$
$$\varepsilon = c^2p^0 + O(\epsilon). \tag{4.2}$$

The variable $\varepsilon$ is the particle energy in the zeroth approximation, and $p_\parallel$ is the momentum component along the direction defined by the unit spacelike vector $u^k_{(-3)}$. If $E_\parallel = O(\epsilon)$ [$\lambda/\nu = O(\epsilon)$], then according to (3.66) $p_\parallel$ distinguishes slightly from the momentum projection $p_H = (\mathbf{e}_H, \mathbf{p})$ on the direction $\mathbf{e}_H = \mathbf{H}/H$,

$$p_\parallel - (\mathbf{p}, \mathbf{e}_H) = O(\epsilon^2), \quad E_\parallel = O(\epsilon). \tag{4.3}$$

Let us transform Eqs. (3.50) and (3.54) to the dependent variables $\varepsilon$, $p_\parallel$, and $\mathbf{y}$ and to the independent variable $y^0 = t$ instead of $\tau$. Setting $M = 0$ in Eqs. (3.50) and (3.54), one obtains

$$\frac{d\varepsilon}{dt} = \frac{p_\parallel c}{\varepsilon}\sqrt{\frac{\sigma_1}{\mu}}\left[e\lambda c\left(1 - \epsilon\frac{\gamma_1}{\gamma_0}\right) - \varepsilon u^k_{(3)}v^l\frac{\partial}{\partial y^l}u_{(-3)k}\right] - \varepsilon v^l\frac{\partial}{\partial y^l}\ln\sqrt{\frac{\sigma_1}{\mu}} - \dot{\varepsilon}_{rad} + O(\epsilon^2), \tag{4.4}$$

$$\frac{dp_\parallel}{dt} = \sqrt{\frac{\mu}{\sigma_1}}\left[e\lambda\left(1 - \epsilon\frac{\gamma_1}{\gamma_0}\right) - \frac{\varepsilon}{c}u^k_{(3)}v^l\frac{\partial}{\partial y^l}u_{(-3)k}\right] - \dot{\varepsilon}_{rad}\frac{p_\parallel}{\varepsilon} + O(\epsilon^2), \tag{4.5}$$

$$\frac{dy^k}{dt}=\left(1-\epsilon\frac{\gamma_1}{\gamma_0}\right)v^k-\frac{\varepsilon}{e\nu c}v_i\frac{d}{dt}e^{ki}_{\mathcal{M}}$$
$$+\frac{\varepsilon\nu^{-1}\lambda}{ec(\nu^2+4\lambda^2)}$$
$$\times\left\{-\nu v_l(v^j e^{il}_{\mathcal{g}}+v^i e^{jl}_{\mathcal{g}})\frac{\partial}{\partial y^j}\Pi^k_{\mathcal{M}i}\right.$$
$$\left.+2\lambda(v^iv^j+e^{il}_{\mathcal{g}}e^{iq}_{\mathcal{g}}v_lv_q)\frac{\partial}{\partial y^j}e^k_{\mathcal{M}i}\right\}$$
$$+O(\epsilon^2),\quad k=0,1,2,3, \tag{4.6}$$

where

$$v^i=\left\{1,c\sqrt{\frac{\mu}{\sigma_1}}\mathbf{u}_3+\frac{p_\parallel c^2}{\varepsilon}\mathbf{u}_{-3}\right\} \tag{4.7}$$

$$\dot{\varepsilon}_{\rm rad}=\frac{1}{2}\epsilon\nu c\alpha_{\rm rad}(a_g)^2_{\min}$$
$$=\frac{2\epsilon^2c^2e\nu}{\omega^2t_{\rm rad}}\left|\sum_{l,s=\pm3}R_{(2,s,l)}a_sa_l\right|^2, \tag{4.8}$$

and the Lorentz factor $\gamma_0+\epsilon\gamma_1$ is determined from Eq. (4.6) for $k=0$, $dy^0/dt=1$, $\gamma_0=\varepsilon/mc^2$. Equations (4.4)–(4.6) depend on the particle mass only through $\dot{\varepsilon}_{\rm rad}$, which is formally of $O(\epsilon^2)$.

If $\lambda\neq0$ and the electromagnetic field is strong, then as a result of $e\lambda L\gg mc^2$, the particle is ultrarelativistic, generally speaking, because it is accelerated by the longitudinal electric field $E_\parallel$. In terms of $a_3$ and $a_{-3}$, relation (3.7) can be represented in the form,

$$a_{-3}=sa_3\sqrt{1-[1+(a_g^2)_{\min}]/a_3^2},\quad s=\operatorname{sgn}(a_{-3}). \tag{4.9}$$

In the ultrarelativistic case, $a_3\gg1$, but $(a_g^2)_{\min}$ cannot be too large, because if $(a_g^2)_{\min}\gg1$, then according to (4.4) and (4.8) the characteristic damping time of energy is

$$t_c=\frac{\varepsilon}{\dot{\varepsilon}_{\rm rad}}=\frac{\gamma t_{\rm rad}}{(a_g^2)_{\min}}\ll\gamma t_{\rm rad}\ll\frac{L}{c} \tag{4.10}$$

and the particle energy decreases very rapidly until the $(a_g^2)_{\min}$ becomes small. If $(a_g^2)_{\min}\ll a_3^2$ and $a_3\gg1$, then it follows from (4.9) that

$$a_{-3}=sa_3. \tag{4.11}$$

Relation (4.11) is an approximate one. It is valid asymptotically at $a_3\gg1$. Let us consider relation (4.11) as an exact one. It follows from (4.11), (4.2), and (3.45) that

$$p_\parallel=s(\varepsilon/c)\sqrt{\mu/\sigma_1},\quad s=\operatorname{sgn}(p_\parallel), \tag{4.12}$$

$$m_{\rm GC}=(1/c^2)\sqrt{\varepsilon^2\mu/\sigma_1-p_\parallel^2c^2}=O(\epsilon). \tag{4.13}$$

The guiding center mass $m_{\rm GC}$ is zero to within $\epsilon^0$. We shall reference the approximation (4.11) or (4.12) as the massless approximation. It is adequate for a strong electromagnetic field, where the particles are relativistic, and the term $mc^2$ can be neglected as compared with the particle energy $\varepsilon$.

At the condition (4.12) Eqs. (4.4) and (4.5) stop to be independent ones. In the massless approximation Eqs. (4.4)–(4.6) take the form,

$$\frac{dp_\parallel}{dt}=e\lambda\sqrt{\frac{\mu}{\sigma_1}}\left(1-\epsilon\frac{\gamma_1}{\gamma_0}\right)-sp_\parallel u^k_{(3)}v^j_{(s)}\frac{\partial}{\partial y^j}u_{(-3)k}$$
$$-\frac{8}{3}\epsilon^2\frac{e^2\mu}{\nu}\sqrt{\frac{\mu}{\sigma_1}}s|R_{(2,s,s)}|^2\left(\frac{2p_\parallel}{mc}\right)^4+O(\epsilon^2), \tag{4.14}$$

$$\frac{dy^k}{dt}=v^k_{(s)}+\frac{sp_\parallel}{e(\nu^2+4\lambda^2)}\sqrt{\frac{\sigma_1}{\mu}}v^i_{(s)}v^j_{(s)}\frac{\partial}{\partial y^j}$$
$$\times\{2\lambda s(\Pi^k_{\mathcal{M}i}-\Pi^0_{\mathcal{M}i}v^k_{(s)})-\nu(e^k_{\mathcal{M}i}-e^0_{\mathcal{M}i}v^k_{(s)})\} \tag{4.15}$$

where $v^k_{(s)}$ $(s=\pm1)$ are the tensor $F^{ik}$ eigenvectors, defined by

$$v^k_{(s)}=2c\sqrt{\frac{\mu}{\sigma_1}}u^k_{(s)},\quad \mathbf{v}_s=2c\sqrt{\frac{\mu}{\sigma_1}}\mathbf{u}_s,$$
$$s=\pm1,\quad(\mathbf{v}_s,\mathbf{v}_{s'})=c^2\delta_{ss'},\quad s,s'=\pm1, \tag{4.16}$$

and

$$\epsilon\frac{\gamma_1}{\gamma_0}=\frac{sp_\parallel}{e}\sqrt{\frac{\sigma_1}{\mu}}\frac{1}{\nu^2+4\lambda^2}v^i_{(s)}v^j_{(s)}$$
$$\times\frac{\partial}{\partial y^j}[2\lambda s\Pi^0_{\mathcal{M}i}-\nu e^0_{\mathcal{M}i}].$$

Equation (4.15) does not contain the real mass $m$ of the particle. Equation (4.14) contains $m$ only through the last term, which describes the radiation deceleration. Formally it is of the order $\epsilon^3$, when $(a_g^2)_{\min}=O(\epsilon^2)$. But it is the only dissipative term that prevents $p_\parallel$ from the strong increase. According to Eq. (4.15) the particle moves along the baseline $L_{(s)}$ with the speed of light (in the zeroth-order approximation).

If $\lambda=0$ [$(\mathbf{E},\mathbf{H})=0$ and $\mathbf{E}^2<\mathbf{H}^2$] in some region, then two eigenvalues of $F^{ik}$ coincide, $\lambda_{(1)}=\lambda_{(-1)}=\lambda=0$, and a degeneration arises. At $\lambda=0$ Eqs. (4.4)–(4.5) have an additional solution $\varepsilon=p_\parallel c=0$. This solution corresponds to an indefinite velocity component $(p_\parallel c^2/\varepsilon)\mathbf{u}_{-3}$, which corresponds to arbitrary motion along the direction $\mathbf{u}_{-3}=\nu\mathbf{H}/\sqrt{\sigma_1\mu}$. For this reason Eq. (4.6) is not transformed into Eq. (4.15). Instead one obtains

$$\frac{dy}{dt}=\frac{1}{2\cosh\chi_0}(e^{\chi_0}\mathbf{v}_1+e^{-\chi_0}\mathbf{v}_{-1})+O(\epsilon^2),$$
$$\varepsilon=0,\quad p_\parallel=0,\quad\tanh\chi_0=\frac{p_\parallel c}{\varepsilon}\sqrt{\frac{\sigma_1}{\mu}}, \tag{4.17}$$

where $\chi_0$ is an arbitrary parameter.

The case where $(\mathbf{E},\mathbf{H})=0$, $\mathbf{E}^2<\mathbf{H}^2$ inside some region of the event space is a common occurrence if the electromagnetic field generated by moving charged particles is taken into account. For instance, let the external elecromagnetic field $F^{ik}$ be stationary and $(\mathbf{E},\mathbf{H})=0$ at some two-dimensional surface $\Gamma_1$ of the configurational space. Then the potential $\Phi$ along the magnetic line $L$ has an extremum at the point $P$ of intersection of $L$ and $\Gamma_1$. If for certain, there is a maximum of $\Phi$ at $P$, then the projection of the base trajectories $l_{(1)}$ onto the plane $\Sigma_2$ orthogonal to the vector $\mathbf{v}_1+\mathbf{v}_{-1}$ has the form represented in Fig. 1.

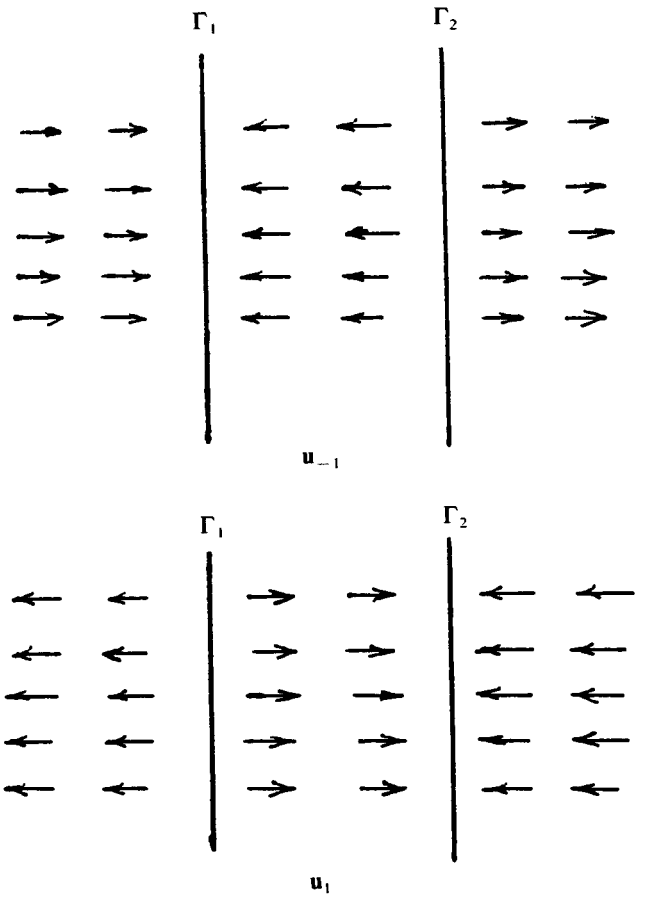



FIG. 1. The structure of the base vector field before the formation of capture regions. The negative charges are captured near $\Gamma_1$ and the positive charges are captured near $\Gamma_2$.

The massless particles move for most part in such a way that $\mathrm{sgn}(p) = \mathrm{sgn}(dp/dt) = \mathrm{sgn}(e)$. Then, according to Eq. (4.15), the positively charged particles move essentially in the direction of the vector $\mathbf{v}_1$ and the negative ones move in the direction $\mathbf{v}_{-1}$. In the case shown in Fig. 1, electrons accumulate in vicinity of $\Gamma_1$, and positrons accumulate in the vicinity of $\Gamma_2$. Accumulation (capture) of charged particles lasts until the field generated by captured particles changes the direction of the base vectors $\mathbf{v}_1$ in such a way that they would be directed similarly in the vicinity of $\Gamma_1$. In other words, (E,H) does not change the sign on the base trajectory.

Inside the capture region the captured particles are placed in such a way that $(\mathbf{E},\mathbf{H}) = 0$, because the Lorentz force can vanish only in this case. The parameter $\chi_0$ describing the particle motion inside the capture region is determined by the motion of the capture region boundaries and by the manner of the particle accumulation.

In addition to the captured massless particles of $p_\parallel = 0$ inside the capture region there may be particles with $p_\parallel \neq 0$, whose motion is described by Eqs. (4.14) and (4.15) with $\lambda = 0$.

Further, the motion of particles in the strong self-consistent electromagnetic field $F^{ik}$ will be considered. It means that the $F^{ik}$ consists of the external electromagnetic field and of the field generated by moving particles. The particle motion is supposed to be collisionless. The massless approximation seems to be most convenient for the description of such a motion, because in the zeroth approximation, the particle velocity (4.15) and the electrical four-current generated by particles depend on the momentum $p_\parallel$ through the parameter $s = \mathrm{sgn}(p_\parallel) = \pm 1$ only. In some cases, when the particles are only accelerated, the sign of $p_\parallel$ does not change, and the velocity $dy/dt$ does not depend on $p_\parallel$. In this case, Eq. (4.14) can be ignored in the calculation of the self-consistent electromagnetic field, and the equations of the particle motion are reduced to the equation

$$\frac{dy}{dt} = \mathbf{v}_s + O(\epsilon), \quad s = \mathrm{sgn}(e), \quad \lambda \neq 0,$$
$$\frac{dy}{dt} = \frac{1}{2\cosh\chi_0}(e^{\chi_0}\mathbf{v}_1 + e^{-\chi_0}\mathbf{v}_{-1}) + O(\epsilon), \quad \lambda = 0, \tag{4.18}$$

where $\chi_0$ is determined from the Maxwell equations and the condition $\lambda = 0$.

Further, for describing the particle motion one uses the zeroth approximation (4.18). The next approximation (4.15) and (4.17) should be used only thereafter, when the self-consistent electromagnetic field will be calculated in the zeroth approximation. In the massless approximation the particles of $\varepsilon > 0$ and those of $\varepsilon = 0$ should be considered as different phases.

Thus in the massless approximation there are two phases of each species: (1) the dynamical phase (DP) of electrons (or positrons) moving with the speed of light, and (2) the statical phase (SP) of electrons (or positrons) moving with speed less than that of light. The dynamical phase and the static one can turn into one another. The conversion DP→SP takes place when an external electric field stopping DP appears on the path of the DP. Part of the DP stops, and its electric field neutralizes the external electric field, which prevents the DP motion. The capture region with $(\mathbf{E},\mathbf{H}) = 0$ arises. The remaining part of the DP continues to move. The conversion SP→DP takes place when the external electric field, which forms the capture region, becomes weak. A smaller amount of the SP is necessary for its neutralization. Hence part of the SP converts into the DP and escapes from the capture region. A description of the conversions SP↔DP needs a consideration of moving charged particles in the self-consistent electromagnetic field.

## V. THE PROPERTIES OF THE PHASES INSIDE AND AT THE BOUNDARY OF THE CAPTURE REGION

The motion of the charged particles in the given electromagnetic field is considered established inside some spatial region $V$ if (E,H) does not change the sign along each particle trajectory. For the established motion the event space is separated into regions of different kinds: (1) DP regions (the regions filled with DP only), (2) capture regions (regions filled with DP and SP), and (3) vacuum regions, where there are neither DP nor SP.

The capture regions can be separated into SP regions, where there is only SP, and leaky capture regions, where the DP moves through SP.

Let $\Omega$ be a four-dimensional capture region in the event space, $\Gamma$ be its three-dimensional spacelike boundary, and $n^j$ be the external normal vector to $\Gamma$, i.e., the normal directed outside from $\Omega$. Then

$$n^j = \{n^0, \mathbf{n}\}, \quad n_j n^j = -1. \tag{5.1}$$

At the boundary $\Gamma$ the Maxwell equations (3.49) take the form

$$[n_k F^{ik}]_\Gamma = (4\pi/c)\, j^i_{\text{sur}}, \quad \left[n_k \overset{d}{F}{}^{ik}\right]_\Gamma = 0, \tag{5.2}$$

where the square brackets denote the jump at the boundary $\Gamma$,

$$[F^{ik} n_k]_\Gamma \equiv (F^{ik}|_{\Gamma_+} - F^{ik}|_{\Gamma_-}) n_k, \quad i = 0,1,2,3, \tag{5.3}$$

where $\Gamma_+$ denotes the external side of $\Gamma$ and $\Gamma_-$ denotes the internal one, and $j^i_{\text{sur}}$ is the surface four-current on the boundary $\Gamma$. The $\Gamma$ is defined by the equation

$$\Gamma\colon\ \Gamma(x) = 0, \quad n_i = \pm \frac{\partial_i \Gamma}{\sqrt{|\partial^k \Gamma\, \partial_k \Gamma|}}. \tag{5.4}$$

The sign is chosen so that the normal will be external. The volume current $j^i$ on $\Gamma$ is expressed through $j^i_{\text{sur}}$ by means of

$$j^i(x) = j^i_{\text{sur}} \delta(\Gamma(x)) |\sqrt{\partial^k \Gamma\, \partial_k \Gamma}|, \tag{5.5}$$

where $\delta$ denotes Dirac's $\delta$ function.

Convoluting the first of Eqs. (5.2) with $n_i$, one obtains

$$n_i j^i_{\text{sur}}|_\Gamma = 0. \tag{5.6}$$

Let there be two charged species numbered by the index $s = \text{sgn}(e) = \pm 1$, where $e$ is the charge of the particle of the given species. Let, for instance, $s = -1$ represent electrons and $s = 1$ represent positrons.

If $\Gamma$ satisfies the condition

$$n_i v^i_{(s)}|_\Gamma \neq 0, \quad s = \pm 1, \tag{5.7}$$

then the surface current vanishes:

$$j^i_{\text{sur}}|_\Gamma = 0, \quad i = 0,1,2,3. \tag{5.8}$$

This is connected with the fact that, in the massless approximation,

$$j^k = \rho_1 v^k_{(1)} + \rho_{-1} v^k_{(-1)}, \quad k = 0,1,2,3. \tag{5.9}$$

When condition (5.7) holds, the relation (5.6) can be fulfilled, if only $\rho_1 = \rho_{-1} = 0$, i.e., under condition (5.8). Thus if the baselines $L_{(s)}$ $(s = \pm 1)$ cross $\Gamma$, then the surface charge and the surface current vanish on $\Gamma$, because they are destroyed by the repulsion of particles that can move across $\Gamma$.

If the $\Gamma$ is formed from the baselines and

$$n_i v^i_{(s)}|_\Gamma = 0, \quad \text{for } s = 1 \text{ or } s = -1, \tag{5.10}$$

then the surface four-current $j^i_{\text{sur}}$ does not vanish, generally speaking.

Let us assume condition (5.7) to be fulfilled and the $F^{ik}$ to be continuous on $\Gamma$. In this case the tangential derivatives of $F^{ik}$ are continuous on $\Gamma$. Only normal derivatives

$$\frac{\partial F^{ik}}{\partial n} \equiv n^l \partial_l F^{ik} = \left(n^0 \frac{\partial}{\partial t} + (\mathbf{n}, \nabla)\right) F^{ik}, \quad i,k = 0,1,2,3, \tag{5.11}$$

can be discontinuous. Substituting (5.9) into (3.49) and taking into account the continuity of $F^{ik}$, $v^j_{(s)}$ $(s = \pm 1)$ on $\Gamma$, one obtains

$$n_k \left[\frac{\partial F^{jk}}{\partial n}\right]_\Gamma = \frac{4\pi}{c} ([\rho_1]_\Gamma v^j_{(1)} + [\rho_{-1}]_\Gamma v^j_{(-1)}), \tag{5.12}$$

$$n_k \left[\frac{\partial \overset{d}{F}{}^{jk}}{\partial n}\right]_\Gamma = 0, \quad j = 0,1,2,3. \tag{5.13}$$

In the coordinate system $K_0$, where the boundary $\Gamma$ is at rest ($n_0 = 0$, $\mathbf{n}^2 = 1$), and the gauge (2.15) is used, (5.12) and (5.13) take the form

$$\left[\frac{\partial E_n}{\partial n}\right]_\Gamma = 4\pi[\rho_1 + \rho_{-1}]_\Gamma, \quad \left[\frac{\partial \mathbf{E}_t}{\partial n}\right]_\Gamma = 0,$$

$$\left[\frac{\partial H_n}{\partial n}\right]_\Gamma = 0,$$

$$\left[\frac{\partial \mathbf{H}_t}{\partial n}\right]_\Gamma = -\frac{4\pi}{H^2}\{[\rho_1 - \rho_{-1}]_\Gamma s_1 \nu(\mathbf{n}\times\mathbf{H}) + [\rho_1 + \rho_{-1}]_\Gamma (\mathbf{E}H_n - \mathbf{H}E_n),$$

$$[\rho_1 + \rho_{-1}]_\Gamma (\mathbf{E}\times\mathbf{H}, \mathbf{n}) + [\rho_1 - \rho_{-1}]_\Gamma s_1 \nu H_n = 0, \tag{5.14}$$

where indices $n$ and $t$ denote, respectively, the normal component and the tangential one:

$$E_n = (\mathbf{E}, \mathbf{n}), \quad \mathbf{E}_t = \mathbf{E} - \mathbf{n}(\mathbf{n}, \mathbf{E}). \tag{5.15}$$

In particular, it follows from (5.14) that

$$\left[\frac{\partial}{\partial n}(\mathbf{H}^2 - \mathbf{E}^2)\right]_\Gamma = 0. \tag{5.16}$$

It results from (2.31), (5.16), and the continuity of $(\mathbf{E},\mathbf{H})$ on $\Gamma$ that

$$\left[\frac{\partial \mu}{\partial n}\right]_\Gamma = 0. \tag{5.17}$$

Thus the invariants $\mathbf{H}^2 - \mathbf{E}^2$ and $\mu$ and their first derivatives are continuous on $\Gamma$.

Let there be two species: positrons ($s = 1$) and electrons ($s = -1$). Let us consider an established flow of charged particles. The term "established" means, first, that $(\mathbf{E},\mathbf{H})$ does not change its sign along the basic line and, second, that each particle moves in such a way that the electric field does not decelerate the particles. In this case the particle motion equation has the form (4.18) with $s = \text{sgn}(e)$ being a number of the species. Thus, for established flow inside the DP region ($\lambda \neq 0$), the positrons move only along $L_{(1)}$ and the electrons move only along $L_{(-1)}$. In the common case of Eqs. (4.14) and (4.15) it is not so, generally speaking.

For the established motion the outflow condition of the species $s$ from $\Omega$ through $\Gamma$ has the form

$$\text{sgn}(n_j v^j_{(s)}|_{\Gamma_+}) = \text{sgn}(n_0 - \mathbf{n}\mathbf{v}_s|_{\Gamma_+}) < 0. \tag{5.18}$$

The condition of no outflow of the species $s$ from $\Omega$ through $\Gamma$ (i.e., inflow of $s$ in $\Omega$, or resting in $\Omega$) has the form

$$\text{sgn}(n_j v^j_{(s)}|_{\Gamma_+}) = \text{sgn}(n_0 - \mathbf{n}\mathbf{v}_s|_{\Gamma_+}) \geqslant 0, \tag{5.19}$$

For brevity one will say that $\Gamma$ is unstable with respect to the species $s$ if there is outflow of $s$ through $\Gamma$ [(5.18)], and that $\Gamma$ is stable with respect to species $s$, if there is no outflow of $s$ through $\Gamma$ [(5.19)]. The conditions (5.18) and (5.19) are invariant with respect to the gauge transformation (2.27).

Now let us consider the case where there is only one species (for instance, $s=-1$). Let $\rho_D$ and $\rho_R$ be, respectively, charge density DP and SP. Then one can see from (5.9) that in the coordinate system $K_0$,

$$\rho_D=\rho_{-1}-\rho_1\leqslant 0, \quad \rho_R=2\rho_1\leqslant 0. \tag{5.20}$$

Let us consider different cases.

(I) $(\mathbf{n},\mathbf{H})=0, \quad (\mathbf{E}\times\mathbf{H},\mathbf{n})=0.$ (5.21)

Then according to (2.15) and (2.16) the condition (5.10) is fulfilled and the surface four-current $j^i_{\text{sur}}$ on $\Gamma$ does not vanish, generally speaking. Also the components $\mathbf{H}_t$ and $E_n$ are usually discontinuous.

(II) $(\mathbf{n},\mathbf{H})\neq 0, \quad (\mathbf{E}\times\mathbf{H},\mathbf{n})=0.$

Then it follows from the last of Eqs. (5.14) and Eq. (5.20) that

$$[\rho_D]_\Gamma=[\rho_{-1}-\rho_1]_\Gamma=0. \tag{5.22}$$

By definition, $\rho_R=0$ outside the capture region $\Omega$; ;thus, from (5.14), (5.20), and (5.22),

$$\frac{1}{4\pi}\left[\frac{\partial E_n}{\partial n}\right]_\Gamma=[\rho_R]_\Gamma=2[\rho_{-1}]_\Gamma\geqslant 0, \quad s=-1. \tag{5.23}$$

(IIa) Let $\Gamma$ be unstable with respect to electrons ($s=-1$); then, as a result of (2.15) and (2.16), condition (5.18) takes the form

$$\left[\frac{\partial E_n}{\partial n}\right]_\Gamma\leqslant 0. \tag{5.24}$$

Conditions (5.23) and (5.24) are compatible in the only case where

$$[\rho_R]_\Gamma=0, \quad \rho_R|_\Gamma=0, \tag{5.25}$$

and the quantities $F^{ik}$, $\mathbf{E}^2-\mathbf{H}^2$, $(\mathbf{E},\mathbf{H})$, $\mu$, and $\psi$ with their first derivatives are continuous on $\Gamma$, and

$$F^{jl}_{\mathscr{E}}|_\Gamma=0, \quad \partial_k F^{jl}_{\mathscr{E}}|_\Gamma=0, \quad k,j,l=0,1,2,3.$$

(IIb) Let $\Gamma$ be stable with respect to the electrons ($s=-1$). Then, as a result of (2.15) and (2.16), condition (5.19) takes the form (5.23) and

$$\left[\frac{\partial \mathbf{H}_t}{\partial t}\right]_\Gamma=-\frac{4\pi}{H^2}[\rho_R]_\Gamma(\mathbf{E}H_n-\mathbf{H}E_n). \tag{5.26}$$

(III) $(\mathbf{n},\mathbf{H})=0, \quad (\mathbf{E}\times\mathbf{H},\mathbf{n})\neq 0.$

It follows from the last of Eqs. (5.14) that

$$\begin{aligned}&[\rho]_\Gamma=[\rho_1+\rho_{-1}]_\Gamma=0, \quad [\rho_D]_\Gamma=2[\rho_{-1}]_\Gamma,\\ &\left[\frac{\partial E_n}{\partial n}\right]_\Gamma=0, \quad \left[\frac{\partial \mathbf{H}_t}{\partial n}\right]_\Gamma=\frac{4\pi v s_1}{H^2}[\rho_D]_\Gamma(\mathbf{n}\times\mathbf{H}),\end{aligned} \tag{5.27}$$

$$\begin{aligned}&\left[\frac{\partial}{\partial n}(\mathbf{E},\mathbf{H})\right]_\Gamma=-\frac{4\pi[\rho_D]_\Gamma v s_1(\mathbf{n},\mathbf{E}\times\mathbf{H})}{H^2},\\ &\left[\frac{\partial\psi}{\partial n}\right]_\Gamma=-\frac{4\pi[\rho_D]_\Gamma v s_1(\mathbf{n},\mathbf{E}\times\mathbf{H})}{\mu H^2}.\end{aligned} \tag{5.28}$$

(IV) $(\mathbf{n},\mathbf{H})=0, \quad (\mathbf{E}\times\mathbf{H},\mathbf{n})\neq 0.$

Here one has the common case (5.14). It can be reduced to case (II), if $H_n^2<E_t^2$, or to case (III), if $H_n^2>E_t^2$, by means of a proper Lorentz transformation, which keeps the boundary at rest.

According to the capture region $\Omega$ definition the equation

$$F_{ik}j^k=0, \quad i=0,1,2,3, \tag{5.29}$$

is satisfied inside $\Omega$. As long as $j^i$ is a real vector, then $\lambda=0$, and the electric bivector $F^{ik}_{\mathscr{E}}$ vanishes. Let $\xi_2$, $\xi_3$ number the baselines $L_{(1)}$, $L_{(-1)}$, i.e.,

$$u^i_{(s)}\partial_i\xi_l=0, \quad s=\pm 1, \quad l=2,3. \tag{5.30}$$

Then the expression

$$\begin{aligned}F_{ik}&=a(\xi_2,\xi_3)\frac{\partial(\xi_2,\xi_3)}{\partial(x^i,x^k)}\\ &\equiv\frac{1}{2}a(\xi_2,\xi_3)\varepsilon_{iklm}\frac{\partial\xi_2}{\partial x_l}\frac{\partial\xi_3}{\partial x_m}, \quad i,k=0,1,2,3,\end{aligned} \tag{5.31}$$

where $a(\xi_2,\xi_3)$ is an arbitrary function, satisfies identically the last of Eqs. (3.49) and Eq. (5.29) with $j^i$ defined by (5.9). Equations (5.9), (5.29), and (5.30) are invariant with respect to the transformation $\xi_s\to\tilde{\xi}_s(\xi_2,\xi_3)$, $s=2,3$. For this reason one can set, without loss of generality,

$$a(\xi_2,\xi_3)=1, \quad F_{ik}=\frac{\partial(\xi_2,\xi_3)}{\partial(x^i,x^k)}, \quad i,k=0,1,2,3. \tag{5.32}$$

## VI. THE CAPTURE REGION OF THE MAGNETIZED ROTATING SPHERE

Let us consider a collisionless cold flow of massless electrons, which are ejected with zero energy from the surface of a conducting sphere. The sphere of the radius $R^*$ rotates with the angular velocity $\mathbf{\Omega}_0$ and has the magnetic dipole moment $\boldsymbol{\mu}_0$ [$\mathbf{\Omega}_0\|\boldsymbol{\mu}_0$, $(\mathbf{\Omega}_0,\boldsymbol{\mu}_0)>0$]. The sphere rotation generates a quadrupole electric field of the order $E\sim\beta H$, where

$$\beta=\Omega_0 R^*/c\ll 1 \tag{6.1}$$

and $\mathbf{H}$ is the magnetic field. The electrons move in this electromagnetic field. The electron flow is supposed to be steady state, established, and axially symmetric. The total electric charge $Q_{\text{tot}}$ is a parameter of the system.

Such a system is a simple model of the pulsar magnetosphere[1] with the sphere playing the role of the neutron star. The flow of collisionless, massless, cold electrons forms a capture region. Inside the capture region the electromagnetic field $F_{ik}$ has the form (5.32) and does not depend on $t$, $\varphi$ ($t,r,\vartheta,\varphi$ are spherical coordinates concentric to the sphere). Then at proper choice of $\xi_3$ one has

$$\xi_2=-\Omega(\xi_3)t+\varphi+\xi_2'(r,\vartheta), \quad \xi_3=\xi_3(r,\vartheta), \tag{6.2}$$

$$F_{ik} = \Omega(\xi_3)\frac{\partial(\xi_3,t)}{\partial(x^i,x^k)} - \frac{\partial(\xi_3,\varphi)}{\partial(x^i,x^k)} - \frac{\partial(\xi_3,\xi_2')}{\partial(x^i,x^k)},$$
$$i,k = 0,1,2,3, \tag{6.3}$$

where $\Omega$, $\xi_2'$, and $\xi_3$ are arbitrary functions of their arguments.

It is easy to verify that the first term of (6.3) describes the electric field, the second one the poloidal magnetic field, and the last one the toroidal magnetic field.

Let $V_0$, $V_1$, and $V_2$ be three-volumes of the configurational space, where $V_0$ is the volume of the sphere, $V_1$ is that of the capture region, and $V_2$ is that of the rest space. Let $\Gamma_{ik} = \Gamma_{ki}$ $(i,k = 0,1,2)$ be the two-dimensional boundaries between $V_i$ and $V_k$. Inside $V_0$ the condition (5.29) takes place, and therefore $V_0$ is essentially a capture region.

Let us use the following suppositions.

(a) There is only one species: electrons $s = \mathrm{sgn}(e) = -1$.

(b) The electron motion is collisionless and established.

(c) The capture region $V_1$ adjoins to the sphere region $V_0$.

(d) Each baseline $L_{(-1)}$ crosses the capture region $V_1$ not more than two times.

(e) All physical quantities do not depend on $t$, $\varphi$.

(f) The toroidal component $H_\varphi$ of the magnetic field is neglected:

$$\xi_2'(r,\vartheta) = 0. \tag{6.4}$$

The statement of the problem is described by the equations

$$A_i(\mathbf{x}) = A_i^*(\mathbf{x}) - \frac{1}{c}\int_{V_1\cup V_2}\frac{J_i(\mathbf{x}')d\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|} - \frac{1}{c}\int_{\Gamma_{02}}\frac{J_{(02)i}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}ds, \quad i = 0,1,2,3, \tag{6.5}$$

where $A_i$ is the total four-potential in the Cartesian coordinate system, $A_i^*(x)$ is the four-potential generated by the volume $V_0$, $J_i$ is the four-current of the electrons, and $J_{(02)i}$ is the surface four-current on the sphere surface $\Gamma_{02}$. The $J_{(02)i}$ is constituted by the positive charge that cannot be ejected from the sphere. Equations (6.5) are a corollary of the Maxwell equations (3.49) if $A_i$ has the Lorentz gauge $\partial_i A^i = 0$. In addition, relation (6.5) includes the absence of the surface charge on the boundary $\Gamma_{12}$ of the capture region $V_1$. The surface charge absence on $\Gamma_{12}$ permits one to determine the shape of $\Gamma_{12}$. The four-current $J^i$ is determined by (5.9), i.e., $J^i$ is expressed through $F^{ik}$ and $\rho_1$, $\rho_{-1}$. Thus (6.5) expresses $A_i$ through $A_i^*$, the quantities $\rho_1$, $\rho_{-1}$, and $J_{(02)i}$, and the shape of the boundary $\Gamma_{12}$.

Let $\Xi(r,\vartheta) = \mathrm{const}$ describe the shape of the line $l_{p(-1)}$, which is a projection of the base trajectory $l_{(-1)}$ onto the meridional plane $\varphi = \mathrm{const}$. Then $\xi$ satisfies the equation

$$(\mathbf{E}\sin\psi + \mathbf{H}\cos\psi, \nabla\Xi) = 0, \tag{6.6}$$

where $\psi$ is defined by (2.31).

Let $\mathbf{i}_r$, $\mathbf{i}_\vartheta$, and $\mathbf{i}_\varphi$ be unit vectors tangent to coordinate lines of the spherical coordinate system $r,\vartheta,\varphi$ with the polar axis along $\boldsymbol{\Omega}_0$. The physical components $\hat{J}_i$ of the vector $J^i$ can be represented in the form

$$\hat{J}_0 = \rho = \begin{cases} -\dfrac{\Omega_0}{4\pi}\nabla^2\Xi, & (r,\vartheta)\in V_1, \\ \dfrac{F(\Xi)|\nabla\Xi|}{r\sin\vartheta}\sqrt{\dfrac{H^2+\mu\sin^2\psi}{\mu}}, & (r,\vartheta)\in V_2, \end{cases} \tag{6.7}$$

$$\hat{J}_\varphi = \begin{cases} \dfrac{\Omega_0^2 r\sin\vartheta}{4\pi c}\nabla^2\Xi, & (r,\vartheta)\in V_1, \\ -\dfrac{F(\Xi)|\nabla\Xi|}{r\sin\vartheta}\sqrt{\dfrac{H^2-\mu\cos^2\psi}{\mu}}, & (r,\vartheta)\in V_2, \end{cases} \tag{6.8}$$

$$J_p = \frac{F(\Xi)}{r\sin\vartheta}\nabla\Xi\times\mathbf{i}_\varphi, \quad (r,\vartheta)\in V_1\cup V_2, \tag{6.9}$$

where $\mathbf{J}_p$ and $\mathbf{J}_\varphi$ are, respectively, the poloidal component of $J^i$ and the toroidal one, $\mu$ and $\psi$ are defined by (2.31), $F$ is a function determined from the continuity condition of (6.7) on the unstable part $\Gamma_{12}'$ of the boundary $\Gamma_{12}$, and $J^i$, determined by (6.7)–(6.9), satisfies the continuity equation identically.

Using the supposition (c) one can consider the volume $V_0\cup V_1$ as a united capture region. Comparing Eq. (6.3) with $F_{ik}$ obtained from (6.5) inside $V_0\cup V_1$, one concludes

$$\Omega(\xi_3) = \Omega_0, \quad \xi_3 = -A_\varphi = -\Xi,$$
$$A_0 = -(1/c)\Phi(\xi_3) = -(\Omega_0/c)(\Xi+\Phi_0). \tag{6.10}$$

Let us substitute (6.7)–(6.9) into (6.5), transform the obtained equations into the spherical coordinate system, and use dimensionless variables $\xi$, $\eta$, $a_0$, $a_\varphi$, $j_0$, and $j_\varphi$ instead of $\Xi$, $r$, $\hat{A}_0$, $A_\varphi$, $\hat{J}_0$, $\hat{J}_\varphi$:

$$\Xi = \mu_0\xi/R^*,$$
$$\eta = \frac{r}{R^*}, \quad \hat{A}_0 = \frac{\mu_0 c}{(R^*)^2}\beta a_0, \quad A_\varphi = \frac{\mu_0}{R^*}a_\varphi, \tag{6.11}$$
$$\hat{J}_0 = \frac{\mu_0 c}{(R^*)^4}\beta j_0, \quad \hat{J}_\varphi = \frac{\mu_0 c}{(R^*)^2}\beta^2 j_\varphi.$$

Let us suppose that the boundary $\Gamma_{12}$ can be described by

$$\Gamma_{12}: \quad \eta = \eta_0(\vartheta), \quad \vartheta\leqslant\vartheta_0, \quad \eta_0(\vartheta_0) = 1. \tag{6.12}$$

Then after calculation one obtains the following statement of the problem:

$$a_0(\eta,\vartheta) = a_0^*(\eta,\vartheta) - \int_0^{\pi/2}\sin\vartheta'\,d\vartheta' \times\int_1^\infty G_0(\eta,\vartheta;\eta',\vartheta')j_0(\eta',\vartheta')\eta'^2\,d\eta' - \int_{\vartheta_0}^{\pi} G_0(\eta,\vartheta;1,\vartheta')\sigma(\vartheta')\sin\vartheta'\,d\vartheta',$$
$$1<\eta<\infty, \quad 0\leqslant\vartheta\leqslant\pi, \tag{6.13}$$

$$a_\varphi(\eta,\vartheta) = a_\varphi^*(\eta,\vartheta) - \beta^2 \int_0^{\pi/2} d\vartheta' \times \int_1^\infty G_\varphi(\eta,\vartheta;\eta'\vartheta') j_\varphi(\eta',\vartheta')\eta'\, d\eta' - \beta^2 \int_{\vartheta_0}^{\pi/2} G_\varphi(\eta,\vartheta;1,\vartheta')\sigma(\vartheta')\sin\vartheta'\, d\vartheta', \quad 1<\eta<\infty, \quad 0\leqslant\vartheta\leqslant\pi, \tag{6.14}$$

$$a_\varphi(\eta,\vartheta) = \xi, \quad a_0(\eta,\vartheta) = -(\xi+\varphi_0), \quad 1\leqslant\eta\leqslant\eta_0(\vartheta), \quad \varphi_0 = \text{const}, \tag{6.15}$$

where $G_0$ and $G_\varphi$ are defined by

$$G_0(\eta,\vartheta;\eta',\vartheta') = \sum_{k=\pm 1} \int_0^{2\pi} [\eta^2+\eta'^2 - 2k\eta\eta'(\cos\vartheta\cos\vartheta' + \sin\vartheta\sin\vartheta'\cos\varphi')]^{-1/2} d\varphi', \tag{6.16}$$

$$G_\varphi(\eta,\vartheta;\eta',\vartheta') = \sum_{k=\pm 1} \int_0^{2\pi} \frac{\eta\eta'\sin\vartheta\sin\vartheta'\cos\varphi'\, d\varphi'}{\sqrt{\eta^2+\eta'^2 - 2k\eta\eta'(\cos\vartheta\cos\vartheta' + \sin\vartheta\sin\vartheta'\cos\varphi')}}, \tag{6.17}$$

and $j_0, j_\varphi$ are defined by the expressions

$$j_0(\eta,\vartheta) = -\frac{\nabla^2\xi}{4\pi}\Theta(\eta_0(\vartheta)-\eta) + \frac{F_1(\xi)|\nabla\xi|}{\eta\sin\vartheta} \times \sqrt{\frac{h^2+\mu_1\sin^2\psi}{\mu_1}}\Theta(\eta-\eta_0(\vartheta)), \tag{6.18}$$

$$j_\varphi(\eta,\vartheta) = \frac{\nabla^2\xi}{4\pi}\eta\sin\vartheta\,\Theta(\eta_0(\vartheta)-\eta) - \frac{F_1(\xi)|\nabla\xi|}{\eta\sin\vartheta} \times \sqrt{\frac{h^2-\mu_1\cos^2\psi}{\mu_1}}\Theta(\eta-\eta_0(\vartheta)), \tag{6.19}$$

$$\mathbf{h} = -\frac{\mathbf{i}_\varphi \times \nabla a_\varphi}{\eta\sin\vartheta}, \quad \mathbf{e} = \nabla a_0, \tag{6.20}$$

$$\nabla = \left\{\frac{\partial}{\partial\eta}, \frac{1}{\eta}\frac{\partial}{\partial\vartheta}, \frac{1}{\eta\sin\vartheta}\frac{\partial}{\partial\varphi}\right\}, \tag{6.21}$$

$$\mu_1 = \sqrt{(\mathbf{h}^2-\beta^2\mathbf{e}^2)^2 + 4\beta^2(\mathbf{e},\mathbf{h})^2},$$

$$\cos(2\psi) = (\mathbf{h}^2-\beta^2\mathbf{e}^2)/\mu_1, \quad \sin(2\psi) = 2\beta(\mathbf{e},\mathbf{h})/\mu_1. \tag{6.22}$$

The function $\Theta$ is defined by

$$\Theta(x) = \begin{cases} 1, & x>0, \\ 0, & x<0, \end{cases} \tag{6.23}$$

and $\xi$ is the solution of the equation

$$(\beta\mathbf{e}\sin\psi + \mathbf{h}\cos\psi, \nabla\xi) = 0 \tag{6.24}$$

determined by the boundary condition

$$\xi|_{\eta=1} = a_\varphi^*|_{\eta=1}. \tag{6.25}$$

The $\xi$ is defined inside the region $\eta>1$ of the plane $(\eta,\vartheta)$ with the cut along the stable part $\Gamma_{12}''$ of the boundary $\Gamma_{12}$. The $\Gamma_{12}''$ is determined by

$$\eta = \eta_0(\vartheta), \quad [(\mathbf{n},\nabla)(\mathbf{n},\mathbf{e})]_{\Gamma_{12}''} > 0, \quad \mathbf{n} = C\left\{1, -\frac{1}{\eta}\frac{\partial\eta_0}{\partial\vartheta}, 0\right\}, \quad C = \left(1+\frac{1}{\eta^2}\left(\frac{\partial\eta_0}{\partial\vartheta}\right)^2\right)^{-1/2}, \tag{6.26}$$

where $F_1(\xi)$ is determined from continuity of $j_0$ on the unstable part $\Gamma_{12}'$ of $\Gamma_{12}$. If $\Gamma_{12}'$ vanishes, then $F_1 = 0$. The $\sigma$ is the dimensionless charge density on $\Gamma_{02}$. The functions $\sigma(\vartheta)\geqslant 0$, $\eta_0(\vartheta)$, $\xi(\eta,\vartheta)$, $a_0(\eta,\vartheta)$, $a_\varphi(\eta,\vartheta)$, and the constant $\vartheta_0$ are to be determined. The $\vartheta_0$ is determined from the condition

$$\eta_0(\vartheta_0) = 1. \tag{6.27}$$

The constant $\varphi_0$ is to be given. The $\varphi_0$ determines the total charge $Q_{\text{tot}}$ of the system.

The $a_0^*$ and $a_\varphi^*$ are determined by the expressions

$$a_0^* = -\frac{1}{\eta}\left(\frac{2}{3} - \frac{1}{5\eta^2}(2-3\sin^2\vartheta)\right), \quad a_\varphi^* = (\sin^2\vartheta)/\eta, \tag{6.28}$$

From (6.1),

$$a_\varphi \simeq a_\varphi^*(\eta,\vartheta) = (\sin^2\vartheta)/\eta \tag{6.29}$$

is an approximate solution of Eq. (6.14). This approximation is valid in the region $\eta\sin\vartheta \ll 1/\beta$. If $\varphi_0>0$ and $\varphi_0$ is large enough ($Q_{\text{tot}}>0$ and $Q_{\text{tot}}$ is large enough), then the whole $\Gamma_{12}$ is stable with respect to electrons, $F_1=0$, and (6.29) is a good approximation. This case has been investigated in Ref. 9 for electrons, and in Ref. 10 for two species: electrons and positrons. In Refs. 3 and 11, the case $Q_{\text{tot}}=0$ ($\varphi_0=0$) has been investigated with approximation (6.29) and with an additional supposition about stability of the whole $\Gamma_{12}$ [$F_1(\xi)=0$]. All investigations show that the electric charge $Q_2$ of region $V_2$ is much smaller than the charge $Q_1$ of region $V_1$. The influence of $Q_2$ upon the total magnetosphere structure is small. The region $V_1'$ filled with trajectories $l_{(-1)}$ crossing unstable $\Gamma_{12}'$ is small relative to the whole $V_1$. But the $V_1'$ (the electron outflow channel) is the active part of the magnetosphere. The $V_1'$ is interesting from this standpoint.

The full solution of the system (6.13)–(6.25) is a rather complicated problem. Here only some qualitative features of $V_1$ and $\Gamma_{12}$, which can be obtained without solving the system (6.13)–(6.25), will be considered. The expected shape of $\Gamma_{12}$ is shown schematically in Fig. 2 in the case of rather small $Q_{\text{tot}}>0$. One can see a cusp of $\Gamma_{12}$ at $\xi=\xi_P$, where $P$ is the branch point of the function $\xi=\xi(\eta,\vartheta)$. At least one

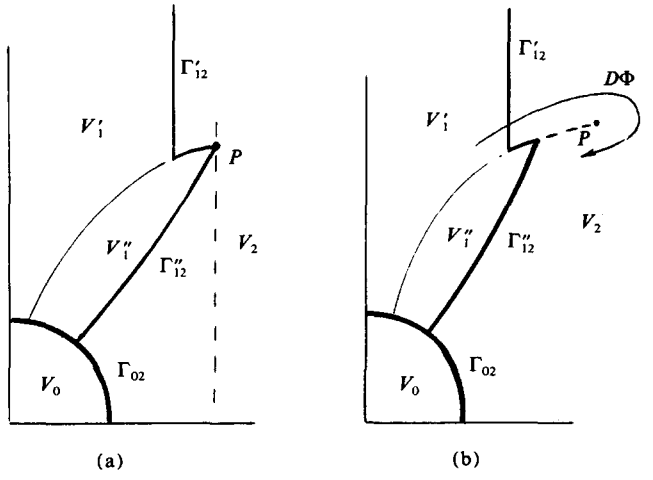



FIG. 2. The magnetosphere scheme of a rotating magnetized sphere in the massless approximation. (a) The branch point $P$ lies on the capture region $V_1$ boundary. (b) The branch point $P$ lies in the DP region $V_2$. Here $V_0$ is the sphere, $V'_1$ is the leaky capture region with the unstable boundary $\Gamma'_{12}$, and $V''_1$ is the leaky capture region with the stable boundary $\Gamma''_{12}$.

branch point must exist if there is an unstable part $\Gamma'_{12}$ of $\Gamma_{12}$. At the value $\xi = \xi_P$ the $\Gamma'_{12}$ passes into the stable part $\Gamma''_{12}$ of $\Gamma_{12}$. Indeed, due to the charge conservation law in the steady-state established case, each base trajectory, crossing $\Gamma'_{12}$ ($\xi < \xi_P$), must return and cross $\Gamma''_{12}$ ($\xi < \xi_P$ on $\Gamma''_{12}$). On the other hand, according to (6.15) and (6.29), the $\xi$ depends on $\vartheta$ monotonically inside $V_0$ (and $\xi > \xi_P$ on $\Gamma''_{12}$). Thus the function $\xi$ is multivalued. Thus to make the $\xi$ single-valued the cut along $\Gamma''_{12}$ and along the base trajectory intercept ($\xi = \xi_P$) up to where the point $P$ is used.

The cusp of $\Gamma_{12}$ at $\xi = \xi_P$ arises, because one has inside the capture region $V_1$,

$$|\rho| \geqslant |\rho_D| = \left| \frac{HF(\xi)}{\sqrt{1 - (\Omega_0^2 r^2 \sin^2 \vartheta)/c^2}} \right|, \quad \xi < \xi_P,$$

$$|\rho| = \left| \frac{(\boldsymbol{\Omega}_0, \mathbf{H})}{2\pi c(1 - \Omega_0^2 r^2 \sin^2 \vartheta / c^2)} \right|, \quad (\eta, \vartheta) \in V_1, \quad (6.30)$$

where $F(\xi)$ is the same as in (6.7). An investigation of relations (6.30) shows that, in the case $\xi > \xi_P$, $r \sin \vartheta$ can be indefinitely close to $R_L = c/\Omega_0$, whereas, in the case $\xi < \xi_P$, $r \sin \vartheta$ cannot be larger than $0.8R_L$–$0.9R_L$. In other words, the $\Gamma'_{12}$ cannot approach closely to $r \sin \vartheta = R_L$ due to the instability of $\Gamma'_{12}$.


[1]F. C. Michel, "Theory of pulsar magnetosphere," Rev. Mod. Phys. **54**, 1 (1982).

[2]B. Bertotti, "Relativistic plasma in a strong magnetic field," in *Colloques Internationaux du C.N.R.S.*, Vol. 184, *La Magnetohydrodynamique Classique et Relativiste* (CNRS, Paris, 1970), p. 251.

[3]Yu. A. Rylov, "On the electron cap shape of a rotating neutron star with a strong magnetic field," Astrophys. Space Sci. **51**, 59 (1977).

[4]D. M. Fradkin, "Covariant electromagnetic projection operators and covariant description of charged particles guiding centre motion," J. Phys. A: Math. Gen. **11**, 1069 (1978).

[5]P. O. Vandervoort, "The relativistic motion of a charged particle in an inhomogeneous electromagnetic field," Ann. Phys. (NY) **10**, 401 (1968).

[6]R. G. Littlejohn, "Hamilton formulation of guiding center motion," Phys. Fluids **24**, 1730 (1981).

[7]T. G. Northrop, *The Adiabatic Motion of Charged Particles* (Wiley, New York, 1963).

[8]N. N. Moiseyev, *Asimptoticheskie metody nelineyinoyi mechaniki* (Nauka, Moscow, 1981), p. 183.

[9]Yu. A. Rylov, "The magnetospheric electron and proton region of the charged rotating neutron star with the strong magnetic field," Astron. J. USSR **53**, 44 (1976) [Sov. Astron. J. **20**, 23 (1976)].

[10]J. Krause-Polstorff and F. C. Michel, "Pulsar space charging," Astron. Astrophys. **144**, 72 (1985); "Electrosphere of an aligned magnetized neutron star," Mon. Not. R. Astron. Soc. **213**, 43 (1985).

[11]Yu. A. Rylov, "The structure of polar caps and the equatorial belt of pulsar," Astrophys. Space Sci. **117**, 5 (1985).

# An electron trapped in a rotating magnetic field

Bogdan Mielnik
*Departamento de Física, CINVESTAV, Apdo postal 14-740, 07000 México, DF, México and Institute of Theoretical Physics, Warsaw University, Hoża 69, Warsaw, Poland*

David J. Fernández C.
*Departamento de Física, CINVESTAV, Apdo postal 14-740, 07000 México, DF, México*

(Received 23 September 1987; accepted for publication 3 August 1988)

The motion of the Schrödinger charged particle in a homogeneous, rotating magnetic field represents an exactly soluble model of the Schrödinger quantum dynamics. The Heisenberg motion trajectories are found and classified. It turns out that the rotating magnetic field of not too high intensity/frequency ratio forms a kind of trap, confining the nonrelativistic charged particle into a quasiperiodic motion. As the intensity/frequency ratio crosses a stability threshold, a sudden qualitative change of the Heisenberg trajectory occurs and the particle is expulsed out of the trap. It is shown that for some values of physical parameters our exact model simulates the behavior of the Schrödinger electron in the laser beam traps.

## I. INTRODUCTION

In the mathematical theory of dynamical systems[1] the time evolution is viewed as a sequence of simple evolution acts represented by the iterations of a certain given mapping $g$: $\Phi \to \Phi$ of an abstract space of states $\Phi$. This simple image has exhibited a number of unexpected features (such as cyclic orbits, instability thresholds, strange attractors, etc.), shedding some new light onto the mechanism of the dynamical processes.

In quantum theories the space of states is an infinite-dimensional projective sphere in a Hilbert space. A typical scenario where the iterations of a mapping might play a similar role is that of an evolution process induced by periodically varying external fields. If the variation period of the external forces is $\tau$, then the dynamical operation of special interest is $U(\tau) = U(0,\tau)$, its repetitions permitting detection of the basic tendencies of the system. In the literature the general problem of evolution in time-dependent fields is seldom visited because of the complication of the continuous Baker–Campbell–Hausdorff composition formula.[2–6] However, if the external fields are quadratic electric potentials and homogeneous magnetic fields (the Hamiltonians quadratic in $q,p$), then it is known from Wigner theory[7] that the quantum mechanical evolution in the Heisenberg frame is exactly reducible to the corresponding classical evolution. Some deep consequences of this fact have been derived by Lewis and Riesenfield[8] and further generalized by Malkin, Manko, and Trifonow.[9,10] They all deal with the Heisenberg motion equations in the form of finite-dimensional matrix equations, some of which could have been effectively solved. On the other hand, a more complete study of the quadratic Hamiltonians was carried out by Moshinsky and Winternitz.[11] Recently, details of the behavior of the Schrödinger particle in time-dependent fields were investigated by one of us,[12,13] yielding special solutions in the form of closed trajectories ("evolution loops"), which might be essential for the techniques of generating arbitrary unitary operations.[12] In particular, it has been found that a sequence of space-homogeneous magnetic pulses in three mutually perpendicular directions can confine a charged particle into a periodic motion, in fact providing a nonspin analog of the spin echo.[13] This has directed our attention toward the general trapping and focusing possibilities of the time-dependent magnetic fields. Our purpose below is to examine the behavior of the Schrödinger electron in three sequencies of sinusoidal, mutually orthogonal magnetic pulses arriving with a phase difference of $2\pi/3$. Such a field is a kind of continuous analog of the pulse pattern of Ref. 13, and simultaneously a not too distant relative of the magnetic traps[14,15] and "standing wave traps."[16–18] As the magnetic field in question we take

$$\bar{B}(t) = b \sin \omega \bar{t} + b \sin(\omega t + 2\pi/3)\bar{\kappa} + b \sin(\omega t + 4\pi/3)\bar{l}, \tag{1.1}$$

where $\bar{j}, \bar{\kappa}, \bar{l}$ are three fixed orthogonal unit vectors in $\mathbb{R}^3$.

The field (1.1) can be approximately created by an uncomplicated experimental arrangement consisting of three electromagnets fed by the ordinary three-phase electric current (Fig. 1). The magnetic vector (1.1) has to be accompanied by a certain circular electric field. In nonrelativistic approximation, and assuming that the electromagnets possess a cylindrical symmetry, both can be represented by a vector potential

$$\bar{A}(\bar{x},t) = -\tfrac{1}{2}\bar{r}\times\bar{B}(t), \quad \bar{r} = \begin{Vmatrix} x \\ y \\ z \end{Vmatrix}. \tag{1.2}$$

At first sight the time dependence of $\bar{B}(t)$ in (1.1) might seem involved, but one can easily see that in fact $|\bar{B}(t)| = \text{const}$, and that $\bar{B}(t)$ is just rotating in a fixed plane with a constant angular velocity $\omega$. After choosing adequately three mutually orthogonal unit vectors $\bar{m}, \bar{n}, \bar{s}$ (Fig. 2), the time-dependent $\bar{B}(t)$ can be represented as

$$\bar{B}(t) = B\bar{n}(t), \qquad \bar{n}(t) = \cos \omega t \bar{n} + \sin \omega t \bar{m}. \tag{1.3}$$

So our arrangement is a magnetic alternative of the widely used "rotating field model."[19,20] Yet we are going to show that it possesses quite different physical properties.

 0022-2488/89/020537-13$02.50 © 1989 American Institute of Physics 

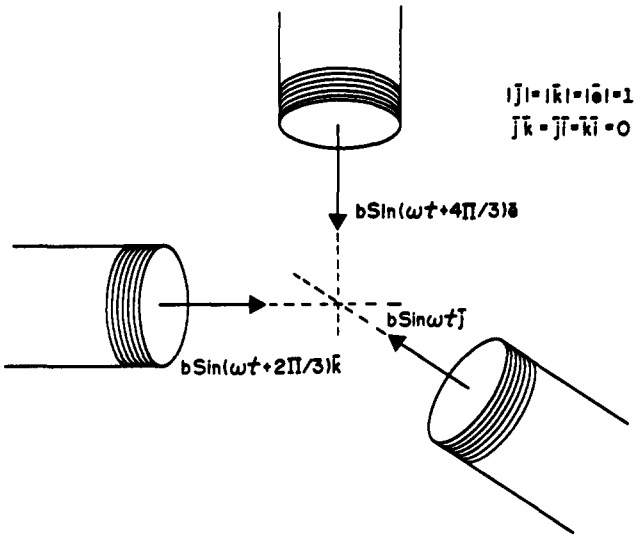



FIG. 1. The sinusoidal pulses of the magnetic field in three mutually orthogonal directions. (The cylindrical symmetry of the electromagnets is essential.)

## II. THE ROTATING FRAME

The evolution of the charged Schrödinger particle in the field (1.2) is given by the unitary operator $U(t)$,

$$\frac{dU(t)}{dt} = -iH(t)U(t), \quad U(0)=1, \tag{2.1}$$

where $H(t)$ is the time-dependent Hamiltonian

$$H(t) = \tfrac{1}{2}[\bar{p} - \bar{A}]^2 = \tfrac{1}{2}[\bar{p} + \tfrac{1}{2}\bar{r}\times\bar{B}(t)]^2. \tag{2.2}$$

(We put for simplicity $m = \hbar = e/c = 1$. The return to the general system of units will be carried out in Sec. III A.) Executing squares and putting $a = B/2$ one obtains

$$H(t) = \tfrac{1}{2}\bar{p}^2 - a\bar{n}(t)\bar{M} + \tfrac{1}{2}a^2\bar{r}_\perp(t)^2, \tag{2.3}$$

where $\bar{M}$ is the angular momentum, $\bar{M} = \bar{r}\times\bar{p}$, and $\bar{r}_\perp(t)$ is the projection of $\bar{r}$ onto the subspace orthogonal to $\bar{n}(t)$:

$$\bar{r}_\perp(t) = \bar{r} - [\bar{r}\bar{n}(t)]\bar{n}(t). \tag{2.4}$$

As is easily noticed, the time dependence of the Hamiltonian (2.3) can be generated with the help of the rotation operator $\exp i\omega\bar{M}\bar{s}$. Indeed

$$\bar{n}(t)\bar{M} = e^{i\omega t\bar{M}\bar{s}}\bar{n}\bar{M}e^{-i\omega\bar{M}\bar{s}}, \tag{2.5}$$

$$\bar{n}(t)\bar{r} = e^{i\omega t\bar{M}\bar{s}}\bar{n}\bar{r}e^{-i\omega t\bar{M}\bar{s}}. \tag{2.6}$$

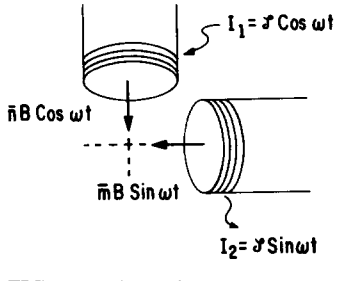



FIG. 2. An alternative arrangement to produce the rotating electromagnetic field (1.2).

Hence

$$H(t) = e^{i\omega t\bar{M}\bar{s}}[\tfrac{1}{2}\bar{p}^2 - a\bar{n}\bar{M} + \tfrac{1}{2}a^2\bar{r}_\perp^2]e^{-i\omega t\bar{M}\bar{s}}, \tag{2.7}$$

$$\bar{r}_\perp = \bar{r}_\perp(0) = \bar{r} - (\bar{n}\bar{r})\bar{n}. \tag{2.8}$$

This suggests that the evolution problem (2.1), (2.2) can be simplified by making the transition to a "rotating frame." Indeed, write down (2.1) as

$$\frac{dU(t)}{dt} = -ie^{i\omega t\bar{M}\bar{s}}\left[\frac{1}{2}\bar{p}^2 - a\bar{n}\bar{M} + \frac{1}{2}a^2\bar{r}_\perp^2\right]e^{-i\omega t\bar{M}\bar{s}}U(t). \tag{2.9}$$

Then put

$$U(t) = e^{i\omega t\bar{M}\bar{s}}W(t). \tag{2.10}$$

The operator $W(t)$ represent the evolution "as seen from the rotating frame." Because of (2.9),

$$\frac{dW(t)}{dt} = -iGW(t), \tag{2.11}$$

where $G$ turns out to be a time-independent generator:

$$G = \bar{p}^2/2 + \tfrac{1}{2}a^2\bar{r}_\perp^2 - a\bar{n}\bar{M} + \omega\bar{s}\bar{M}. \tag{2.12}$$

Choosing—for simplicity—the $x$, $y$, and $z$ axes in the directions $\bar{n}$, $\bar{m}$, and $\bar{s}$, respectively, one can represent $G$ as

$$G = \bar{p}^2/2 + a^2(y^2+z^2)/2 - aM_x + \omega M_z. \tag{2.13}$$

Since $G$ is a quadratic form in $\bar{r},\bar{p}$, (2.11) can be reduced to a $6\times6$-matrix equation and solved on computers with any desired accuracy. Since, however, $G$ is time independent, (2.11) admits also an exact solution. Its form depends on the parameters $a,\omega$ involved in (2.13).

Since the operator $W(t)$ yields only a part of the evolution, the true evolution trajectories in the initial frame will be obtained by superposing the trajectories of $W(t)$ with a continuous circular motion around $\bar{s}$. Note, however, that in the infinite sequence of time moments $t = \pm n\tau$, $\tau = 2\pi/\omega$ ($n = 0,1,2,\ldots$) the rotations reduce to the identity, and $W(t)$ coincides with the full evolution operator. Hence in the corresponding sequence of time intervals $[n\tau,(n+1)\tau]$ $G$ becomes an "effective Hamiltonian," and the repeated application of $W(\tau) = e^{-i\tau G}$ defines the asymptotic behavior of the Heisenberg trajectories, visible equally well for the rotating and for the initial observer.

## III. EVOLUTION MATRIX AND EVOLUTION TRAJECTORIES

The structure of the evolution operator $\exp itG$ might now be determined in agreement with the formalism of Ref. 11 by examining the quadratic form (2.13). The operator $iG$ is defined (with accuracy to an additive constant) by the real $6\times6$ matrix $\Lambda$ representing the commutation operation $\mathrm{Ad}_{iG}$ applied to the six canonical observables $q = \{q_j\}_{j=1,\ldots,6} = \{x,y,z,p_x,p_y,p_z\}$:

$$\left[iG, \begin{vmatrix}\bar{x}\\ \bar{p}\end{vmatrix}\right] = \Lambda\begin{vmatrix}\bar{x}\\ \bar{p}\end{vmatrix}. \tag{3.1}$$

The corresponding operator $W(t)$ is therefore represented by the following transformation of $\bar{x}$ and $\bar{p}$ variables (interpretable as the Heisenberg trajectory in the rotating frame):

$$e^{itG}\begin{Vmatrix}\bar{x}\\ \bar{p}\end{Vmatrix}e^{-itG}=e^{t\Lambda}\begin{Vmatrix}\bar{x}\\ \bar{p}\end{Vmatrix}. \tag{3.2}$$

Using the expression (2.13) for $iG$ and executing the commutations in (3.1) one finds

$$\Lambda=\left\|\begin{array}{cccc|cc} 0 & -\omega & 0 & 1 & & \\ \omega & 0 & a & & 1 & \\ 0 & -a & 0 & & & 1 \\ \hline 0 & & & 0 & -\omega & 0 \\ & -a^2 & & \omega & 0 & a \\ & & -a^2 & 0 & -a & 0 \end{array}\right\|. \tag{3.3}$$

(The unoccupied places mean zeros.) The motions generated by (3.2) depend on the spectral decomposition of the matrix $\Lambda$, which in turn depends on the characteristic polynomial

$$D(\lambda)=\mathrm{Det}(\lambda-\Lambda)=\lambda^6+2(\omega^2+2a^2)\lambda^4+(\omega^4+3\omega^2a^2)\lambda^2+a^2\omega^4. \tag{3.4}$$

Introducing the new variable $\sigma=\lambda^2/\omega^2$ one can write it as

$$\Delta_\alpha(\sigma)=D(\lambda)\omega^{-6}=\sigma^3+2(1+2\alpha^2)\sigma^2+(1+3\alpha^2)\sigma+\alpha^2, \tag{3.5}$$

where $\alpha=a/\omega=a\tau/2\pi$. Hence the algebraic types of $\Lambda$ (and the corresponding types of the motion trajectories) can be classified according to the roots of the third-order polynomial $\Delta_\alpha(\sigma)$ (see Fig. 3). They reveal three different behaviors with an instability threshold at $|\alpha|=\alpha_{\mathrm{cr}}$, where the Cardan discriminant changes the sign:

$$\alpha_{\mathrm{cr}}\cong 0.579\,982\,655\,598\ldots. \tag{3.6}$$

[Curiously, the value of $\alpha_{\mathrm{cr}}$ is within 0.5% of Euler's constant. (We owe this observation to the referee of this paper.)]

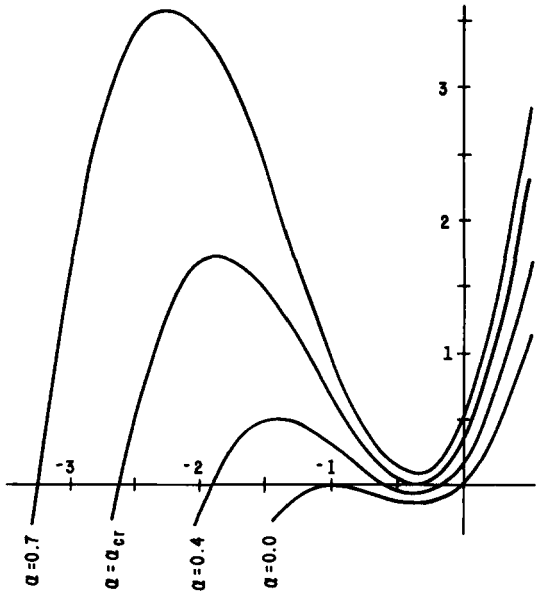



FIG. 3. The shapes of $\Delta_\alpha(\sigma)$.

If $0\neq|\alpha|<\alpha_{\mathrm{cr}}$, the function $\Delta_\alpha(\sigma)$ has three negative real roots $\Rightarrow$ the characteristic polynomial $D(\lambda)$ has six different purely imaginary roots $\Rightarrow$ the matrix $\exp\Lambda$ defines circulating (confined) Heisenberg trajectories in the space of the canonical variables $\bar{x},\bar{p}$. If $|\alpha|=\alpha_{\mathrm{cr}}$, two roots of $\Delta_\alpha(\sigma)$ coincide ($\Delta_\alpha$ degenerate). If $|\alpha|>\alpha_{\mathrm{cr}}$, the function $\Delta_\alpha(\sigma)$ has one real and two conjugated complex roots $\Rightarrow$ $D(\lambda)$ has two imaginary and four complex roots $\zeta$, $-\zeta,\zeta^*,-\zeta^*$, with $\mathrm{Re}\,\zeta\neq 0$ causing an exponential explosion and implosion in the dynamical problem (2.11). Hence the resulting motion trajectories fall into three different classes A, B, C, which we shall discuss below.

### A. $0<|\alpha|<\alpha_{\mathrm{cr}}$: Weak fields and quick rotations

The polynomial $\Delta_\alpha(\sigma)$ has three different real roots $\sigma_0<\sigma_1<\sigma_2<0$ given by Cardan formulas[21]:

$$\sigma_k=2\rho^{1/3}\cos[(\theta+2k\pi)/3]-\tfrac{4}{3}\alpha^2-\tfrac{2}{3}\quad(k=0,1,2), \tag{3.7}$$

$$\rho=2^{-1}3^{-3}[\kappa(\alpha)^2+Q(\alpha)]^{1/2},\qquad \theta=\tan^{-1}\{Q(\alpha)^{1/2}/\kappa(\alpha)\}, \tag{3.8}$$

$$\kappa(\alpha)=2-33\alpha^2-84\alpha^4-2^7\alpha^6,\qquad Q(\alpha)=3^3\alpha^2(8+\alpha^2-36\alpha^4-112\alpha^6). \tag{3.9}$$

The characteristic polynomial $D(\lambda)$ thus yields six different imaginary eigenvalues:

$$\mp i\omega_0,\quad \mp i\omega_1,\quad \mp i\omega_2;\qquad \omega_k=\omega\sqrt{|\sigma_k|}. \tag{3.10}$$

Consistently, the evolution matrix $\exp t\Lambda$ [generating the Heisenberg trajectory (3.2)] has bounded time-dependent entries and defines "circular phenomena." On the classical level it means the bounded phase-space trajectories. In the original rest frame these trajectories are affected by an extra rotational drift around $\bar{s}$ but always remain limited. To find out explicitly the time-dependent operators $\bar{x}(t),\bar{p}(t)$, the spectral decomposition of $\Lambda$ is relevant. Since $\Lambda$ is non-Hermitian, its spectral form can be most simply obtained in terms of the right and left eigenvectors (*eigenvectors* and *eigenforms*) $u_j,u_j^*,e_j,e_j^*$ $(j=0,1,2)$,

$$\Lambda u_j=-i\omega_j u_j,\quad e_j\Lambda=-i\omega_j e_j,\qquad \Lambda u_j^*=i\omega_j u_j^*,\quad e_j^*\Lambda=i\omega_j e_j^*, \tag{3.11}$$

normalized to be a dual system:

$$(e_j u_k)=(e_j^* u_k^*)=\delta_{ik},\qquad (e_j u_k^*)=(e_j^* u_k)=0. \tag{3.12}$$

[Here $(eu)$ means the number output of the matrix multiplication of the row vector $e$ by the column vector $u$.] After calculation one arrives at the following dual bases of forms $e_j=e(\sigma_j)$, $e_j^*=e(\sigma_j)^*$ $(j=0,1,2)$ and vectors $u_j=u(\sigma_j)$, $u_j^*=u(\sigma_j)^*$ $(j=0,1,2)$:

$$e(\sigma)=K(\sigma)\left\|1,\ -i\sqrt{|\sigma|},\ \frac{\alpha(2\sigma+1)}{\sigma}\right.,$$

$$\frac{i(1+3\sigma)}{\omega\sqrt{|\sigma|}(\sigma+1)}, \frac{2\sigma}{\omega(\sigma+1)}, \frac{-i(\sigma+1)}{\omega\alpha\sqrt{|\sigma|}}\Bigg|\Bigg|, \quad \text{and c.c.,} \tag{3.13}$$

$$u(\sigma) = N(\sigma)\left|\left|\begin{array}{c} (1+3\sigma)/2\sigma\sqrt{|\sigma|} \\ i \\ -(\sigma+1)^2/2\alpha\sigma\sqrt{|\sigma|} \\ -i\omega(\sigma+1)/2\sigma \\ -\omega(\sigma+1)/2\sqrt{|\sigma|} \\ -i\omega\alpha(\sigma+1)(2\sigma+1)/2\sigma^2 \end{array}\right|\right|, \quad \text{and c.c.} \tag{3.14}$$

Here $K(\sigma)$ and $N(\sigma)$ are numerical factors, bound only by the normalization condition

$$(e(\sigma)u(\sigma)) = 1 \Rightarrow K(\sigma)N(\sigma) = -\sigma^2\sqrt{|\sigma|}/(4\sigma^3+2\sigma^2+3\sigma+1),$$

and the desired forms $e_j$ and vectors $u_j$ are obtained just by feeding the triple of Cardan's roots $\sigma = \sigma_0, \sigma_1, \sigma_2$ into the functions (3.13), (3.14). [In Sec. IV we shall additionally fix the normalization function $K(\sigma)$ so as to obtain the most convenient description of the "effective Hamiltonian" $G$ and Heisenberg time-dependent operators.]

On account of (3.12), the tensor products $u_j \otimes e_j$, $u_j^* \otimes e_j^*$ (the $u \otimes e$ generalizing Dirac's $|u\times e|$) are projectors, and they yield the spectral decomposition of the unit matrix 1:

$$\mathbf{1} = \sum_{j=0}^{2} (u_j \otimes e_j + u_j^* \otimes e_j^*). \tag{3.15}$$

The corresponding decomposition of $\Lambda$ is

$$\Lambda = \Lambda\mathbf{1} = \sum_{j=0}^{2} (-i\omega_j u_j \otimes e_j + i\omega_j u_j^* \otimes e_j^*). \tag{3.16}$$

Consistently,

$$e^{t\Lambda} = \sum_{j=0}^{2} [e^{-it\omega_j} u_j \otimes e_j + e^{+it\omega_j} u_j^* \otimes e_j^*]. \tag{3.17}$$

The spectral representation (3.17) provides immediately the evolution trajectories in either quantum or classical cases. In fact, let $q$ denote a column of the six (classical and/or quantum) canonical variables:

$$q = \left|\left|\begin{array}{c} q_1 \\ \vdots \\ q_6 \end{array}\right|\right| = \left|\left|\begin{array}{c} \bar{x} \\ \bar{p} \end{array}\right|\right|. \tag{3.18}$$

Introducing (3.18) to (3.2) one has

$$q(t) = e^{itG} q e^{-itG} = e^{t\Lambda} q = \sum_{j=0}^{2} [e^{-i\omega_j t} u_j A_j + e^{+i\omega_j t} u_j^* A_j^*], \tag{3.19}$$

where $A_j$ and $A_j^*$ are scalar coefficients defined with the help of the eigenforms

$$A_j = (e_j q), \quad A_j^* = (e_j^* q). \tag{3.20}$$

The formula (3.19) decomposes the time dependence of the Heisenberg variables $\bar{x}(t), \bar{p}(t)$ into the simplest "circular modes" along three pairs of distinguished six-vectors $u_k$, $u_k^*$ $(k = 0,1,2)$. This decomposition is valid in classical, as well as in quantum, cases, but to see the physical sense of (3.19) it is worthwhile to consider the picture of the classical phase-space trajectories.

In classical mechanics the scalars (3.20) are just numbers and represent the initial conditions defining the motion. In order to express (3.12) in terms of real coordinates, put $A_j = \frac{1}{2}C_j e^{i\phi_j}$ and introduce in each two-dimensional subspace of $u_j, u_j^*$ as an alternative basis consisting of two real vectors,

$$a_j = (1/2i)(u_j - u_j^*), \quad b_j = \tfrac{1}{2}(u_j + u_j^*). \tag{3.21}$$

Then (3.19) becomes

$$q(t) = \sum_{j=0}^{2} C_j [\sin(\omega_j t - \phi_j) a_j + \cos(\omega_j t - \phi_j) b_j]. \tag{3.22}$$

Henceforth (3.19) splits into three elementary circular motions taking place in three distinguished two-dimensional subspaces of the real six-dimensional phase space, with frequencies $\omega_j$ defined by the Cardan roots (3.10). In order to obtain the trajectories in the three-dimensional configuration space, it is enough to make the transition from the six-vectors $a_j$, $b_j$ $(j = 0,1,2)$ to the three-vectors proportional to their upper triples of coordinates:

$$\bar{a}_j = \left|\left|\begin{array}{c} 0 \\ 1 \\ 0 \end{array}\right|\right|, \quad \bar{b}_j = \left|\left|\begin{array}{c} (1+3\sigma)/2\sigma\sqrt{|\sigma|} \\ 0 \\ -(\sigma+1)^2/2\alpha\sigma\sqrt{|\sigma|} \end{array}\right|\right|_{\sigma=\sigma_j} \tag{3.23}$$

One thus obtains the family of the configuration space trajectories labeled by six integration constants $D_j$, $\phi_j$ $(D_j = C_j N_j)$, $(j = 0,1,2)$:

$$\bar{x}(t) = \sum_{j=0}^{2} D_j [\sin(\omega_j t - \phi_j)\bar{a}_j + \cos(\omega_j t - \phi_j)\bar{b}_j]. \tag{3.24}$$

As one can see, for special initial conditions $(D_j = D\delta_{jj_0})$ they degenerate to three simple "circular modes" represented by families of concentric elliptic trajectories on the exceptional two-dimensional planes $P[\sigma_j]$ spanned by the vectors $\bar{a}_j$, $\bar{b}_j$, (see Fig. 4). These planes are defined by the physical parameters of the system; they possess a common intersection along the vector $\bar{m}$ (like three pages of an open book), and their angles $\beta_j$ with the rotation axis $\bar{s}$ are determined by the three Cardan roots (3.7):

$$\tan\beta_j = -(1+3\sigma_j)\alpha/(1+\sigma_j)^2. \tag{3.25}$$

The semiaxes of the corresponding elliptical motions are proportional to the norms of the vectors $\bar{a}_j$, $\bar{b}_j$:

$$\bar{a}_j = |\bar{a}_j| = 1, \quad b_j = |\bar{b}_j| = \sqrt{1 + \frac{1}{2\sigma} + \frac{5}{4\sigma^2} + \frac{1}{\sigma^3} + \frac{1}{4\sigma^4}}\Bigg|_{\sigma=\sigma_j}. \tag{3.26}$$

The positions of the exceptional planes $P[\sigma_j]$ are listed on Fig. 5. (The three angles $\beta_0, \beta_1, \beta_2$ are plotted against the field parameter $\alpha = a/\omega$.) As one can see, for small $\alpha > 0$

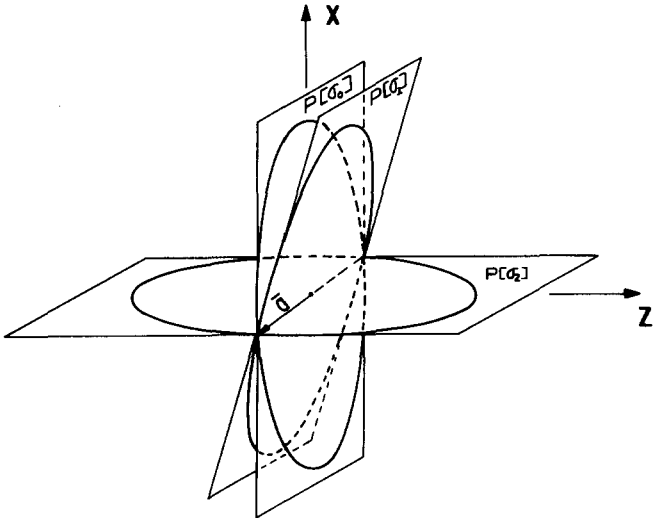



FIG. 4. Three exceptional planes $P[\sigma_j]$ (case A) associated with the Cardan roots $\sigma_0$, $\sigma_1$, $\sigma_2$ and hosting the elliptic trajectories (view from the rotating frame).

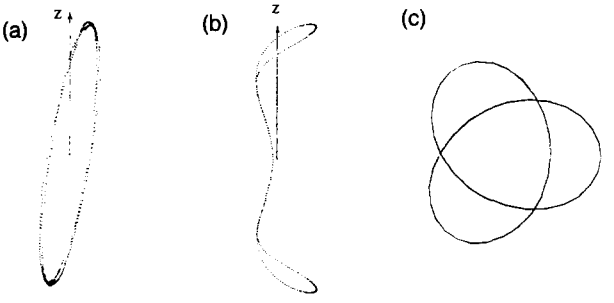



FIG. 6. Three exceptional trapped trajectories in the original frame (computer pictures). (a) The case of $\alpha \cong 0.530\ 330\ 085\ 89$ and $\sigma_2 = -\frac{1}{4}$. The orbits associated with the root $\sigma_2$ are ellipses in the rotating frame, and they remain ellipses in the original one. (The computer was asked to show a slightly deformed and precessing variant of one of these orbits seen in perspective.) (b) An exceptional trajectory that occurs for $\alpha \cong 0.350\ 150\ 487\ 262$ and is associated with the Cardan root $\sigma_2 = -\frac{1}{9}$. (c) One of closed orbits that occur out of the proper confinement domain ($\alpha \cong 1.390\ 282\ 671\ 27$, $\sigma_0 = -9$).

there is one special plane $P[\sigma_2]$, nearly orthogonal to $\bar{n}$, on which the circulation is relatively slow ($\omega_2 = \omega\sqrt{|\sigma_2|} \cong \omega\alpha$) and the motion trajectories have the form of extreme elongated ellipses ($b_2 \cong 1/2\alpha^2$), whereas two other planes $P[\sigma_1]$, $P[\sigma_0]$ stick almost to $\bar{n}$ and host almost circular orbits. As $\alpha$ grows, $P[\sigma_2]$ and $P[\sigma_1]$ tend to coincide for the critical value $\alpha = \alpha_{cr}$ (at the angle $\tilde{\beta}_1 = 0.032\ 472\ 67\pi$), whereas $P[\sigma_0]$ conserves its character even for $\alpha > \alpha_{cr}$, bearing approximately circular trajectories and adopting a position orthogonal to the magnetic field as $\alpha \to +\infty$. The orbits on the root planes $P[\sigma_0]$, $P[\sigma_1]$, $P[\sigma_2]$ are shown on Fig. 4, whereas Fig. 6 represents some exceptional trajectories seen from the original (nonrotating) frame. The general trajectory in $\mathbb{R}^3$ is the superposition of the three elliptic circulations (3.24) (of frequencies $\omega_j = \omega\sqrt{|\sigma_j|}$) and the standard rotation with frequency $\omega$. An intriguing question is whether for some values of $\alpha$ the four frequencies $\omega$ and $\omega_j$ can become commensurable ($\sqrt{|\sigma_j|}$ rational for $j = 0,1,2$), thus producing the phenomenon of an evolution loop (all the phase-space trajectories simultaneously closed). To check this we have assumed $\sigma_0 = -r^2 = -(m/n)^2$ to be a rational root of the equation $\Delta_\alpha(\sigma_0) = 0$, and determined two other roots using the quadratic equation derived from Vieta formulas:

$$\sigma^2 + \frac{3r^4 - 3r^2 + 2}{4r^4 - 3r^2 + 1}\sigma + \frac{(r^2-1)^2}{4r^4 - 3r^2 + 1} = 0. \qquad (3.27)$$

The computer was asked to insert a rational $r$ into (3.27) and to check whether its two roots have the form $\sigma_1 = -r_1^2$, $\sigma_2 = -r_2^2$, where $r_1$, $r_2$ are rational. For $r = m/n \in [0,1.621\ 8182]$ with denominators $n$ running from 1,...,35, the answer turns out negative. Thus we did not detect the loop phenomenon. [Perhaps the special motions (3.22) provide yet another "physical construction" of noncommensurable quantities.] The absence of rational relations determines the character of the motion (3.24). The resulting trajectory, in general, is dense in a certain domain in $\mathbb{R}^3$, and so is the sequence of points obtained by iterations of $U(\tau) = W(\tau)$. In Fig. 7 we have plotted the computer picture of a typical motion trajectory departing from $\bar{x} = 0$ and with $\bar{p}_0$ chosen at random. The resulting cloud illustrates a curious phenomenon of trapping: a charged particle unable to leave a limited space domain because of the rotating magnetic field (an effect that has not so simple an analog in magnetostatics). The space volume that comprises the trajectory may be estimated by remembering that the link between the initial conditions and the coefficients $D_j$ of (3.20) [$D_j = D(\sigma)|_{\sigma=\sigma_j}$] involves the products of the normalization constants $N_j K_j = N(\sigma)K(\sigma)|_{\sigma=\sigma_j}$. One has

$$|\bar{x}(t)| < \sum_{\sigma_0,\sigma_1,\sigma_2} |D(\sigma)|\,|\bar{\mu}(\sigma)|, \qquad (3.28)$$

where $\mu(\sigma) = \max\{1, |\bar{b}(\sigma)|\}$, and

$$|D(\sigma)| = |N(\sigma)C(\sigma)| = |N(\sigma)|\,2|(e(\sigma)q_0)|. \qquad (3.29)$$

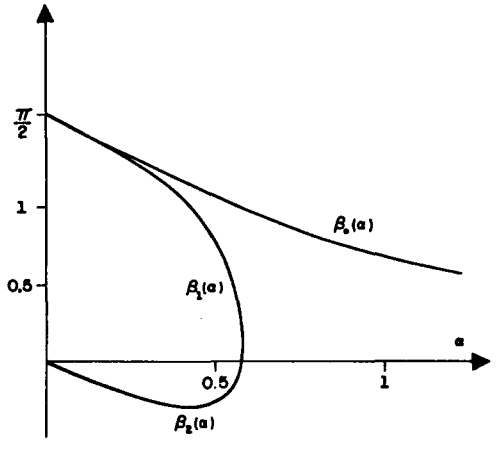



FIG. 5. The angles $\beta_0(\alpha)$, $\beta_1(\alpha)$, $\beta_2(\alpha)$ between the root planes $P[\sigma_j]$ and the rotation axis $z$.

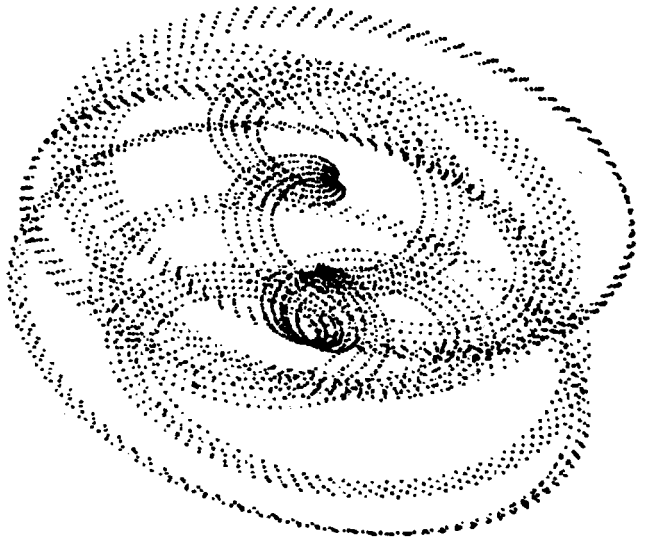


FIG. 7. One of the evolution trajectories for $\alpha = 0.25$, $\omega = 2\pi$ (case A) in the rotating frame, departing from the initial condition $\bar{x}_0 = 0$, $p_{0x} = p_{0y} = p_{0z} = 1$. The graphic, computed in dimensionless coordinates, corresponds to 20 rotation periods of $\bar{B}(t)$ and is simultaneously confined in three coordinate intervals $[-1.42, +1.49] \times [-1.39, 1.38] \times [-0.81, 0.81]$.

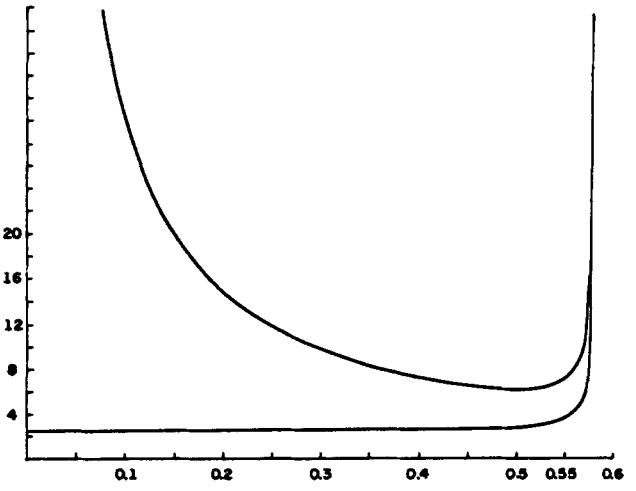



FIG. 8. The functions $F(\alpha)$ and $\Phi(\alpha)$ providing an estimate of the trapped trajectories.

A short calculation henceforth shows that the configuration space radius of a general trajectory admits the limit

$$|\bar{x}(t)| < F(\alpha)|\bar{x}_0| + (1/\omega)\Phi(\alpha)|\bar{p}_0|, \tag{3.30}$$

where

$$F(\alpha) = \sum_{\sigma_0,\sigma_1,\sigma_2} 2\sigma^2 |\bar{b}(\sigma)|\,|4\sigma^3 + 2\sigma^2 + 3\sigma + 1|^{-1} \times \sqrt{-\sigma + \sigma^2 + \frac{(\sigma+1)^2(2\sigma+1)^2}{(4\sigma^2+3\sigma+1)}}, \tag{3.31}$$

$$\Phi(\alpha) = \sum_{\sigma_0,\sigma_1,\sigma_2} 2|\sigma|^{3/2}\,|\sigma+1|^{-1}\,|\bar{b}(\sigma)| \cdot |4\sigma^3 + 2\sigma^2 + 3\sigma + 1|^{-1} \times \sqrt{4\sigma^4 + (\sigma+1)^2(4\sigma^2+3\sigma+1) - \sigma(1+3\sigma)^2}. \tag{3.32}$$

The computer pictures of $F(\alpha)$ and $\Phi(\alpha)$ are plotted in Fig. 8. As can be noticed, the limitation (3.30) works differently for the orbits with $\bar{p}_0 = 0$ (the "$x$ orbits") and the orbits with $\bar{x}_0 = 0$ (the "$p$ orbits"). For the $x$ orbits the trap is most efficient for $\alpha < 0.56$ [$\Rightarrow F(\alpha) < 4$], while for the $p$ orbit it is the best for $0.36 < \alpha < 0.56$ [$\Rightarrow \Phi(\alpha) < 8$]. For an unspecified trajectory the interval $0.36 < \alpha < 0.56$ represents the "optimal efficiency region" in which

$$|\bar{x}(t)| < 4|\bar{x}_0| + (8/\omega)|\bar{p}_0|. \tag{3.33}$$

As can be seen, the rotating confinement occurs for not too strong fields and has a certain natural efficiency limit. If $\omega = \text{const}$, the ratio of the confinement radius to the initial moment $|\bar{p}_0|$ for a typical trajectory cannot be reduced more than to $(1/\omega)\Phi\,\min \cong 6.2/\omega$, no matter the value of $\bar{B}$ employed. If $\bar{B}$ is too strong, the confining effect is "overdone" and the trajectories explode (see the discussion in Secs. III B and III C). The physical orders of magnitude in cgs units can be recovered by putting

$$\omega = 2\pi/T, \quad a = eB/2mc, \tag{3.34}$$

where $T$ is the rotation period of $\bar{B}(t)$ in seconds and $e/m$ is the charge/mass ratio of the trapped particle. Out of curiosity we have collected some optimal confinement radii achievable for $\bar{B}$ rotating with typical electromagnetic frequencies (Table I).

As can be noticed, the magnetic fields rotating with the standard frequency of the variable electric current ($\nu = 1/T = 60/\text{sec}$) can sustain the electrons only on very large trajectories, which might suggest an image of a relatively cold and dispersed cloud of charged particles kept in vacuum by very weak rotating fields ($|\bar{B}| < 4.14 \times 10^{-7}$ G). For the frequencies $\cong 10^6$/sec, the "cold" 3 °K electron is kept on the orbit of the radius $\cong 6$ mm by $|\bar{B}(t)|$ of the order of magnitude $\cong 10^{-1}$ G, creating a shallow alternative of the ion traps.[14,15] The confinement becomes tighter above $\nu > 10^6$/sec, the frequencies not difficult to achieve for the rotating fields produced by systems of core-free electromagnets. Moreover, there are also different physical mechanisms producing the same fields.

We refer to the behavior of the fields in some points of standing waves. Consider the electromagnetic standing wave described by the vector potential $\bar{A}_{\bar{m}}(\bar{s}\bar{x},\omega t)$ consistently with both Coulomb and Lorentz gauges:

$$\bar{A}_{\bar{m}}(\bar{s}\bar{x},\omega t) = A\bar{m}\sin(\omega\bar{s}\bar{x}/c)\sin\omega t \quad (A \in \bar{\mathbb{R}}), \tag{3.35}$$

where $\bar{n},\bar{m},\bar{s}$ are three arbitrary orthogonal unit vectors. By crossing two perpendicular standing waves of the form (3.35) one defines the field

$$A_{\bar{m},\bar{s}}(\bar{x},\omega t) = \tfrac{1}{2}[\bar{A}_{\bar{m}}(\bar{s}\bar{x},\omega t) - \bar{A}_{\bar{s}}(\bar{m}\bar{x},\omega t)] = \tfrac{1}{2}A\,[\bar{m}\sin(\omega\bar{s}\bar{x}/c) - \bar{s}\sin(\omega\bar{m}x/c)]\sin\omega t, \tag{3.36}$$

TABLE I. The size of classical trapped trajectories.

| Frequency (in rotation/sec) | Intensity of the optimally confining field $\bar{B}(t)$ (in G) | The optimal confinement radius for 3 °K electron departing from the center | The threshold value of the magnetic field marking the instability domain (in G) |
|---|---|---|---|
| 1/sec | $3.61\times10^{-7}$ | 6 km | $4.14\times10^{-7}$ |
| 60/sec | $2.16\times10^{-5}$ | 100 m | $2.48\times10^{-5}$ |
| $10^3$/sec | $3.61\times10^{-4}$ | 6 m | $4.14\times10^{-4}$ |
| $10^6$/sec | $3.61\times10^{-1}$ | 6 mm | $4.14\times10^{-1}$ |
| $10^7$/sec | 3.61 | 0.6 mm | 4.14 |

which in the vicinity of $\bar{x}=0$ (i.e., for $|\bar{x}|\ll\lambda=2\pi c/\omega$) behaves approximately like one of the oscillation modes of our rotating potential (1.2):

$$\bar{A}_{\bar{m},\bar{s}}(\bar{x},\omega t)\cong\tfrac{1}{2}(A\omega/c)[\bar{m}(\bar{s}\bar{x})-\bar{s}(\bar{m}\bar{x})]\sin\omega t$$
$$=\tfrac{1}{2}\bar{r}(A\omega/c)\times\bar{n}\sin\omega t. \tag{3.37}$$

Hence by superposing two fields $\bar{A}_{\bar{m},\bar{s}}$ and $\bar{A}_{\bar{s},\bar{n}}$ with the phase shift $\pi/2$ (equivalent to four standing waves) one obtains the time-dependent periodic field

$$\bar{A}(\bar{x},t)=\bar{A}_{\bar{n},\bar{s}}(\bar{x},\omega t)+\bar{A}_{\bar{m},\bar{s}}(\bar{x},\omega t-\pi/2)$$
$$=-\tfrac{1}{2}A\left[\left(\bar{m}\sin\frac{\omega\bar{s}\,\bar{x}}{c}-\bar{s}\sin\frac{\omega\bar{m}\bar{x}}{c}\right)\cos\omega t\right.$$
$$\left.-\left(\bar{s}\sin\frac{\omega\bar{n}\bar{x}}{c}-\bar{n}\sin\frac{\omega\bar{s}\,\bar{x}}{c}\right)\sin\omega t\right], \tag{3.38}$$

which near to $\bar{x}=0$ imitates our rotating trap (1.2):

$$\bar{A}(\bar{x},t)\cong-\tfrac{1}{2}\bar{r}(A\omega/c)\times[\bar{n}\cos\omega t+\bar{m}\sin\omega t]$$
$$=-\tfrac{1}{2}\bar{r}\times\bar{B}(t)\quad(B=A\omega/c). \tag{3.39}$$

It is furthermore essential that the point $\bar{x}=0$ is not unique; the same situation occurs around the lattice of the nodal points

$$\mathcal{N}=\{\lambda(l_1\bar{n}+l_2\bar{m}+l_3\bar{s}):l_1,l_2,l_3=0,\pm1,\pm2,\dots\},$$

where the standing wave (3.38) approximates the field (1.2). [Besides $\mathcal{N}$, the field (3.38) has also a subset of nodal points of different structure, to be treated elsewhere.] Since the oscillating forces in between the nodal points create a repulsive effect,[14–17] the whole standing wave (3.38) behaves like a "radiation-made crystal," with the lattice of nodal centers (some of which are "little copies" of our rotating trap) is encrusted into a repulsive field medium. This analogy resembles the approach of Letokhov and Minogin[16] with one slight diference.

In the standard description of the electron in the standing waves, the attention is focused on the oscillating electric field $\bar{E}(\bar{x},t)=\bar{E}(\bar{x})\sin\omega t$ [which is assumed to create an "effective potential" $(e/2\omega)^2|E|^2$], while the role of the magnetic component is neglected.[14] However, once the electron is caught near a point in $\mathcal{N}$, the exact sine functions of space coordinates in (3.38) become of secondary importance, whereas there is no good reason to neglect the fluctuating magnetic field. Henceforth it seems that our exact model might give a better insight into what truly happens in the vicinity of each attraction center. As far as the classical $p$ trajectories are concerned, this requires that the confinement radius (3.30) be a small fraction of the wavelength $\lambda/2\pi=c/\omega$. After introducing the cgs units this yields

$$(1/m\omega)|\bar{p}_0|\,\Phi(\alpha)\ll c/\omega,$$

i.e., the following dimensionless parameter ought to be small:

$$\varepsilon=(|\bar{v}_0|/c)\Phi(\alpha)\ll1, \tag{3.40}$$

where $|\bar{v}_0|$ is the initial velocity of the $p$ trajectory, and

$$\alpha=\frac{e}{mc}\frac{B}{2\omega}=\frac{1}{2}\frac{eA}{mc^2}.$$

Thus for $\varepsilon<10^{-4}$ one can replace (3.38) by (3.39) with the relative error $<(\epsilon-\sin\epsilon)/\epsilon\cong6^{-1}\times10^{-8}$. If the parameter $\alpha$ of the standing wave (3.39) is near to the optimal value 0.505 182 51 $\Rightarrow\Phi(\alpha)\cong6.2$, this still imposes a fairly tolerant condition on the particle velocity: $|\bar{v}_0|<(c/6.2)10^{-4}\cong4$ km/sec. Provided that this holds, all our graphics in Figs. 4, 6, and 7 become interpretable as the electron trajectories near to one of the lattice centers. [For the trajectories which start *near*, but not exactly *in*, one of the attraction centers, the evaluations are not essentially worse; see the function $F(\alpha)$ in Fig. 8.] It thus seems that the confinement of a relatively cold cloud of electrons in the lattice of zero points of a standing radio wave is not unreal, and that it might have to do with an ensemble of our rotating systems. (An interrelation with the laser beam traps is also imminent.)

### B. ($|\alpha|=\alpha_{cr}$): The threshold

We return to our principal subject.

If $\alpha\to\pm\alpha_{cr}$, the two roots of $\Delta_\alpha$ coincide; the resulting $\Delta_{\alpha_{cr}}$ has one single root at $\tilde{\sigma}_0=-2.630\,294$ and a double root at $\tilde{\sigma}_1=-0.357\,6126$. The matrix $\Lambda$ has four linearly independent eigenvectors only and is nondiagonalizable. Its canonical form is of Jordan type and can be best expressed with the help of the eigenvectors and eigenforms $u_0, u_0^*, u_1, u_1^*, e_0, e_0^*, e_1, e_1^*$,

$$(\Lambda+i\omega_j)u_j=0=e_j(\Lambda+i\omega_j), \tag{3.41}$$
$$(\Lambda-i\omega_j)u_j^*=0=e_j^*(\Lambda-i\omega_j)\quad(j=0,1),$$

as well as the generalized eigenvectors (eigenforms) $v_1, v_1^*$, $f_1, f_1^*$ representing the nontrivial Jordan structure:

$$(\Lambda+i\omega_1)v_1=u_1,\quad f_1(\Lambda+i\omega_1)=e_1,$$
$$(\Lambda-i\omega_1)v_1^*=u_1^*,\quad f_1^*(\Lambda-i\omega_1)=e_1^*, \tag{3.42}$$

where $\omega_0 = \omega\sqrt{|\sigma_0|}, \omega_1 = \omega\sqrt{|\sigma_1|}$. Note an immediate consequence of (3.41), (3.42):

$$(\Lambda + i\omega_1)^2 v_1 = 0, \quad f_1(\Lambda + i\omega_1)^2 = 0, \quad \text{and c.c.} \tag{3.43}$$

The Jordan vectors and forms of a nondiagonalizable matrix, in general, are not dual. However, $u_j, v_1$ and $e_j f_1$ admit now the freedom of the gauge transformations,

$$\begin{aligned} u_j' &= N_j u_j, \quad e_j' = K_j e_j, \\ v_1' &= N v_1 + \mu u_1, \quad f_1' = K f_1 + \eta e_1, \end{aligned} \tag{3.44}$$

and by fixing $u_j, v_1, e_j, f_1$ so that

$$(e_0 u_0) = (e_1 v_1) = (f_1 u_1) = 1, \quad (f_1 v_1) = 0, \tag{3.45}$$

one achieves the duality of the systems, with $e_1$ dual to $v_1$ and $f_1$ dual to $u_1$. The unit matrix then becomes

$$\begin{aligned} \mathbb{1} = {} & u_0 \otimes e_0 + u_0^* \otimes e_0^* + u_1 \otimes f_1 \\ & + u_1^* \otimes f_1^* + v_1 \otimes e_1 + v_1^* \otimes e_1^*, \end{aligned} \tag{3.46}$$

and owing to (3.41)–(3.43),

$$\begin{aligned} e^{t\Lambda} &= e^{t\Lambda}\mathbb{1} \\ &= e^{-i\omega_0 t} u_0 \otimes e_0 + e^{i\omega_0 t} u_0^* \otimes e_0^* \\ &\quad + e^{-i\omega_1 t} v_1 \otimes e_1 + e^{i\omega_1 t} v_1^* \otimes e_1^* \\ &\quad + e^{-i\omega_1 t} u_1 \otimes (f_1 + te_1) + e^{i\omega_1 t} u_1^* \otimes (f_1^* + te_1^*). \end{aligned} \tag{3.47}$$

The Heisenberg trajectory for the canonical variables $q = \|q_j\|_{j=1,\ldots,6}$ can now be written as

$$\begin{aligned} q(t) = e^{t\Lambda} q = {} & (e^{-i\omega_0 t} A_0 u_0 + e^{i\omega_0 t} A_0^* u_0^*) \\ & + (e^{-i\omega_1 t} B v_1 + e^{i\omega_1 t} B^* v_1^*) \\ & + [e^{-i\omega_1 t}(A_1 + tB) u_1 + \text{c.c.}], \end{aligned} \tag{3.48}$$

where $A_0, A_1, B$ are the operators and/or $c$-number scalars

$$A_0 = (e_0 q), \quad B = (e_1 q), \quad A_1 = (f_1 q). \tag{3.49}$$

As one can observe, the trajectory (3.48) has a spirally expanding part as $t \to +\infty$, associated with the scalar $B$. Introducing the basis of six real vectors,

$$\begin{aligned} a_j &= \text{Im}\, u_j, \quad b_j = \text{Re}\, u_j \quad (j = 0,1), \\ a &= \text{Im}\, v_1, \quad b = \text{Re}\, v_1, \end{aligned} \tag{3.50}$$

one can decompose the (six-dimensional) real phase space $\mathscr{P}$ into three two-dimensional subspaces $\mathscr{P}_0$, $\mathscr{P}_1$, and $\mathscr{V}$ spanned by $\{a_0, b_0\}$, $\{a_1, b_1\}$, and $\{a, b\}$, and hosting the simplest forms of the motion. These turn out to be elliptic circulations (with frequencies $\omega_0$ and $\omega_1$) in $\mathscr{P}_0$ and $\mathscr{V}$, and a family of diverging Archimedes spirals on $\mathscr{P}_1$. The subspaces $\mathscr{P}_0, \mathscr{P}_1, \mathscr{V}$ correspond to three special planes $P_0, P_1, V$ in the configuration space $\mathbb{R}^3$ that host the "configuration space shadows" of the above trajectories. These planes intersect again along the $y$ axis of the rotating frame. The plane $P_0$ is the limiting case of $P[\sigma_0]$ for $\sigma_0 = \tilde{\sigma}_0$, and its angle with $\bar{s}$ is $\tilde{\beta}_0 \cong 0.313\ 1936\pi$. The $P_1$ plane is the result of the coincidence of $P[\sigma_1]$ and $P[\sigma_2]$ for the critical values $\sigma_1 = \sigma_2 = \tilde{\sigma}_1$ and lies at the angular distance $\tilde{\beta}_1 \cong 0.032\ 472\ 67\pi$ from $\bar{s}$. The $V$ plane, in turn, is an extra plane appearing in the critical case alone, forming with $\bar{s}$ the angle $\tilde{\beta} \cong 0.480\ 3933\pi$. Note that the motions on $P_0$, $P_1$, $V$ are linked. In the absence of the component in $V$ the motion in $P_1$ is circular; however, if there is a nontrivial component in $V$, the motion in $P_1$ starts to diverge (as if the elliptic circulation in $V$ were a kind of clock mechanism measuring the spiral development in $P_1$). Since the components in $V$ or in $P_1$ are missing only for very special initial conditions (of measure zero), one thus sees that for $\alpha$ attaining $\alpha_{cr}$ the global picture of motions is radically changed: the confinement is broken and almost all trajectories escape to infinity (see Fig. 9). Before discussing the consequences of this fact, the next case is worth examining.

### C. ($|\alpha| > \alpha_{cr}$): Strong fields and slow rotations

As $|\alpha| > \alpha_{cr}$ the algebraic type of the matrix $\Lambda$ once more changes. The polynomial $\Delta_\alpha(\sigma)$ now has one real and two complex conjugate roots: $\sigma_0$ and $\sigma_1 = \sigma_2^*$, where $\sigma_{1,2}$ are determined in terms of $\sigma_0$ as the roots of the quadratic equation

$$(4\sigma_0^2 + 3\sigma_0 + 1)\sigma^2 + (3\sigma_0^2 + 3\sigma_0 + 2)\sigma + (\sigma_0 + 1)^2 = 0. \tag{3.51}$$

The characteristic polynomial $D(\lambda)$ has therefore two purely imaginary and four different complex roots,

$$\pm i\omega\sqrt{|\sigma_0|}, \quad \pm\omega\zeta, \quad \pm\omega\zeta^*, \tag{3.52}$$

where $\zeta = [\sigma_1]^{1/2} = \eta + i\kappa$, with $\eta = \text{Re}\,\zeta > 0, \kappa = \text{Im}\,\zeta \neq 0$. The matrix $\Lambda$ is diagonalizable and defines the dual systems of the eigenvectors and eigenforms,

$$\begin{aligned} &\Lambda u_0 = -i\omega_0 u_0, \quad \Lambda u_\pm = \pm\omega\zeta u_\pm, \\ &e_0\Lambda = -i\omega_0 e_0, \quad e_\pm \Lambda = \pm\omega\zeta e_\pm, \quad \text{and c.c.}, \end{aligned} \tag{3.53}$$

which turn out to be $u_0 =$ as in formula (3.14),

$$u_- = -C(\zeta) \left\| \begin{array}{c} (1+3\zeta^2)/2\zeta^3 \\ 1 \\ -(1+\zeta^2)^2/2\alpha\zeta^3 \\ -\omega(1+\zeta^2)2\zeta^2 \\ -\omega(1+\zeta^2)/2\zeta \\ -\omega\alpha(1+\zeta^2)(1+2\zeta^2)/2\zeta^4 \end{array} \right\|, \tag{3.54}$$

$$u_+ = C(\zeta) \left\| \begin{array}{c} -(1+3\zeta^2)/2\zeta^3 \\ 1 \\ (1+\zeta^2)^2/2\alpha\zeta^3 \\ -\omega(1+\zeta^2)/2\zeta^2 \\ \omega(1+\zeta^2)/2\zeta \\ -\omega\alpha(1+\zeta^2)(1+2\zeta^2)/2\zeta^4 \end{array} \right\|,$$

with $C(\zeta) = \zeta^5/(4\zeta^6 + 2\zeta^4 + 3\zeta^2 + 1)$; and $e_0 =$ as in formula (3.13),

$$\begin{aligned} e_- = {} & \left\| 1, -\zeta, \frac{\alpha(1+2\zeta^2)}{\zeta^2}, -\frac{(1+3\zeta^2)}{\omega\zeta(1+\zeta^2)}, \right. \\ & \left. \frac{2\zeta^2}{\omega(1+\zeta^2)}, \frac{(1+\zeta^2)}{\omega\alpha\zeta} \right\|, \\ e_+ = {} & \left\| 1, \zeta, \frac{\alpha(1+2\zeta^2)}{\zeta^2}, \frac{(1+3\zeta^2)}{\omega\zeta(1+\zeta^2)}, \right. \\ & \left. \frac{2\zeta^2}{\omega(1+\zeta^2)}, -\frac{(1+\zeta^2)}{\omega\alpha\zeta} \right\|. \end{aligned} \tag{3.55}$$

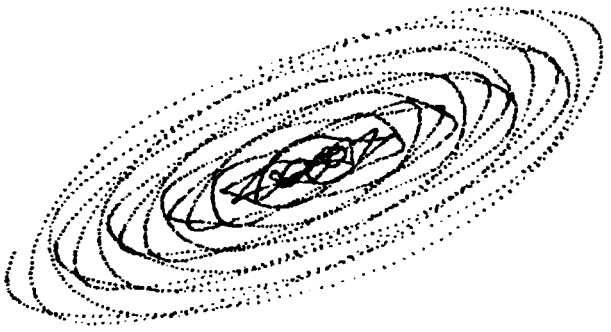


FIG. 9. A slowly expanding trajectory of the critical case $\alpha = \alpha_{cr} \cong 0.579\,982\,655\,598$ seen in the rotating frame [computer picture corresponding to the 30 rotations of $\bar{B}(t)$].

The spectral form of $\Lambda$ becomes

$$\Lambda = -i\omega_0(u_0 \otimes e_0 - u_0^* \otimes e_0^*) - \omega(\zeta u_- \otimes e_- + \zeta^* u_-^* \otimes e_-^*) + \omega(\zeta u_+ \otimes e_+ + \zeta^* u_+^* \otimes e_+^*), \quad (3.56)$$

and so

$$e^{t\Lambda} = e^{-i\omega_0 t} u_0 \otimes e_0 + e^{i\omega_0 t} u_0^* \otimes e_0^* + e^{-\lambda t}(e^{-i\rho t} u_- \otimes e_- + e^{i\rho t} u_-^* \otimes e_-^*) + e^{\lambda t}(e^{i\rho t} u_+ \otimes e_+ + e^{-i\rho t} u_+^* \otimes e_+^*), \quad (3.57)$$

where $\lambda = \omega\eta$ and $\rho = \omega\kappa$. By introducing $\mathscr{P}_0$, $\mathscr{P}_-$, $\mathscr{P}_+$ as the real sections of the two-dimensional complex subspaces of $\mathbb{R}^6$ spanned by the pairs of vectors $\{e_0, e_0^*\}$, $\{e_-, e_-^*\}$, and $\{e_+, e_+^*\}$, one sees that the evolution matrix (3.57) generates the elliptic circulation in $\mathscr{P}_0$, exponentially shrinking spirals in $\mathscr{P}_-$, and exponentially expanding spirals in $\mathscr{P}_+$. In the configuration space $\mathbb{R}^3$ the faithful shadows of these motions will take place on the three "root planes" $P_0$, $P_-$, $P_+$ defined as the configuration space projections of $\mathscr{P}_0$, $\mathscr{P}_-$, $\mathscr{P}_+$. It turns out that the $P_-$ and $P_+$ are transformed into each other by the reflection $y \to -y$, intersect along the vector

$$\bar{a} = \mathrm{Im}\{[(1+3\zeta^2)/2\zeta^3]\bar{n} - [(1+\zeta^2)^2/2\zeta^2]\bar{s}\},$$

and form the angle $\theta(-,+)$, which varies between 0 and $\pi$ as $\alpha \in (-\infty, \alpha_{cr})$. (Thus $P_-$ and $P_+$ tend to coincide twice, for $\alpha \to -\infty$ and $\alpha \to \alpha_{cr}$.)

Since almost all initial conditions possess a nonvanishing part in $\mathscr{P}_+$, all but exceptional trajectories "explode" as $t \to +\infty$ (see Fig. 10). A question about the physical sense of this effect arises: Does it mean that the particle passively "evaporates" to infinity [just since $\bar{B}(t)$ does not rotate quickly enough to block all the escape routes], or is the particle actively expulsed? Since the particle momentum exponentially diverges, it seems that the "deconfinement" represents a new kind of "semiclassical resonance" depending not on the frequency alone but on the ratio $a/\omega$, and affecting otherwise free particles in pulsating coherent fields (tentatively interpretable as the "cyclotronic resonance without cyclotron.") If our argument is right, this phenomenon might be no less important than the "confinement." Thus one can imagine physical situations where a rotating magnetic field $\bar{B}(t)$ works as a particle filter (if two kinds of particles with different charge/mass ratios define the parameter $\alpha$ below and above the threshold $\alpha_{cr}$). Interrelation with the models of quantum friction[22] is imminent. Finally, some critical phenomena concerning the standing wave traps of Sec. III A may be expected.

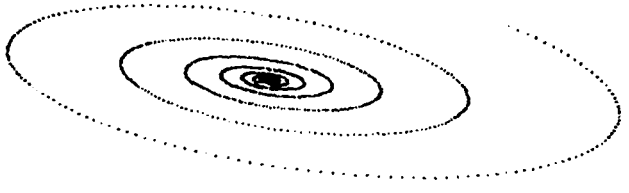


FIG. 10. An exponentially expanding trajectory for $\alpha = 0.6$, $\omega = 2\pi$ (case C) integrated within 20 rotation periods, in the rotating frame. It is seen that one of the evolution modes has quickly disappeared (the dark center of the graphic), giving place to an almost spiral expansion.

Consider again the "wave crystal" (3.38) composed of a lattice of "vibrating traps" (3.39). When this structure is described in agreement with the traditional high-frequency approximation[14,16] the existence of a critical parameter is not clear. The situation is different within our rotating model. In fact, when the parameter $\alpha = eB/2m\omega c = \frac{1}{2} eA/mc^2$ crosses the threshold value (3.6), i.e., for

$$A > 1.159\,965\,3112(mc^2/e), \quad (3.58)$$

all the attractive nodal points of the standing wave (3.38) should undergo a metamorphosis into repulsive centers, thus affecting the global properties of the whole standing wave trap. To obtain a more detailed picture of this phenomenon the electron wave functions are essential.

## IV. STRUCTURE OF THE GENERATOR: "DRIFTING STATES"

The solution of the evolution problem (2.1), (2.2) in terms of the Heisenberg trajectory $q(t), p(t)$ is most natural; yet it leaves implicit the behavior of the Schrödinger wave functions. To recover this last aspect it is useful to split the generator (2.13) into commuting "irreducible parts" related to the simplest modes of behavior of the Schrödinger wave packets. It turns out that the desired decomposition is most easily derived from the matrix $\Lambda$. Below, we shall limit our considerations to the case A ($0 < |\alpha| < \alpha_{cr}$). Consider again the operators $A_j$, $A_j^*$ defined by the eigenforms $e_j$, $e_j^*$ [see (3.13)–(3.20)]. By definition they obey a special commutation relation with $G$:

$$[G, A_j] = -i[iG, (e_j q)] = -i(e_j [iG, q]) = -i(e_j \Lambda q) = -\omega_j (e_j q) = -\omega_j A_j, \quad (4.1)$$

$$[G, A_j^*] = \omega_j A_j^*.$$

Simultaneously, $A_j$, $A_k^*$ commute to

$$[A_k, A_j] = [A_k^*, A_j^*] = 0, \quad (4.2)$$

$$[A_k, A_j^*] = \gamma_k \delta_{kj}, \quad \gamma_k \in \mathbb{R}. \quad (4.3)$$

The commutation rules (4.2), (4.3) may be straightfor-

wardly checked using (3.20), but they are in fact a consequence of a certain general lemma.

*Lemma 1:* If $A$, $B$, $G$ are three operators in a Hilbert space $\mathscr{H}$, and

$$[A,B] = c \text{ number}, \quad [G,A] = \lambda A, \quad [G,B] = \mu B, \quad \lambda,\mu \in \mathbb{C}, \tag{4.4}$$

then $(\lambda + \mu) \neq 0 \Rightarrow [A,B] = 0$.

*Proof:*

$$\begin{aligned}(\lambda+\mu)[A,B] &= [\lambda A,B] + [A,\mu B][[G,A],B] + [A,[G,B]] \\ &= [B,[A,G]] + [A,[G,B]] \\ &= -[G,[A,B]] = -[G,c \text{ number}] = 0.\end{aligned}$$

(*Note:* A refinement of this lemma to the case when $A$, $B$, $G$ and their products are defined not on the whole of $\mathscr{H}$ but on a certain common dense domain $\mathscr{D} \subset \mathscr{H}$ is easily obtained.)

Notice that none of the $\gamma_k$ in (4.3) may vanish. Indeed, if any of the $\gamma_k$'s were zero, the corresponding $A_k$ would commute with every $A_j$ and $A_j^*$ $(j = 0,1,2)$. As the eigenforms $e_j$, $e_j^*$ are a basis among the forms $\xi = \|\xi_1,\ldots,\xi_6\|$, $A_k$ would commute with any linear combination $(\xi q) = \xi_1 q_1 + \cdots + \xi_6 q_6$. Hence it should commute with every canonical observable $q_j, j = 1,\ldots,6$, which is impossible since $e_k \neq 0$.

We shall show that $A_k^*$, $A_k$ $(k = 0,1,2)$ play the role of generalized creation and annihilation operators, and provide the decomposition of $G$ into three commuting canonical parts.

*Lemma 2:* $G$ is representable as

$$G = \sum_{k=0}^{2} \frac{\omega_k}{\gamma_k} A_k^* A_k + g_0, \tag{4.5}$$

where $g_0 \in \mathbb{R}$.

*Proof:* Denote

$$g_0 = G - \sum_{k=0}^{2} \left(\frac{\omega_k}{\gamma_k}\right) A_k^* A_k. \tag{4.6}$$

Because of (4.1), (4.2), and (4.5), $g_0$ commutes with every $A_j$ and $A_j^*$:

$$\begin{aligned}[g_0,A_j] &= [G,A_j] - (\omega_j/\gamma_j)[A_j^*,A_j]A_j \\ &= -\omega_j A_j - (\omega_j/\gamma_j)(-\gamma_j)A_j = 0, \qquad (4.7) \\ [g_0,A_j^*] &= 0, \quad j = 0,1,2.\end{aligned}$$

Thus $g_0$ commutes with all $q_j$'s, and $g_0 \in \mathbb{R}$.

*Observation 1:* A more detailed argument shows that $g_0 = \frac{1}{2}(\omega_0 + \omega_1 + \omega_2)$ (cf. also Refs. 9 and 10).

*Observation 2:* In the expressions (4.3), (4.6) the absolute values of the $\gamma_k$'s are, in fact, superfluous. By redefining $e_j' = e_j/\sqrt{|\gamma_j|}$ and $u_j = \sqrt{|\gamma_j|}\, u_j$ one might easily achieve $\gamma_k = \pm 1$, $k = 0,1,2$. In contrast, the signs of $\gamma_k$ contain essential information about the structure of $G$ and cannot be affected by any redefinition of constants. In our case, taking $A(\sigma) = (e(\sigma)q)$ one finds

$$\gamma(\sigma) = [A(\sigma),A(\sigma)^*] = |K(\sigma)|^2(2/\omega)|\sigma|^{-3/2}(\sigma+1)^{-1} \times [4\sigma^3 + 2\sigma^2 + 3\sigma + 1]. \tag{4.8}$$

The function on the right-hand side changes the sign at $\sigma = -1$ and at $\sigma = \tilde{\sigma}_1$: it is positive in $(\tilde{\sigma}_0, -1)$, negative in $(-1,\tilde{\sigma}_1)$, and positive again in $(\tilde{\sigma}_1,0)$. (The critical values $\sigma = \tilde{\sigma}_0, -1, \tilde{\sigma}_1, 0$ do not belong to case A.) The required normalization of the $A_j$'s can therefore be obtained by fixing

$$|K(\sigma)| = (\omega/2)^{1/2}|\sigma|^{3/4}\,|\sigma+1|^{1/2}/\sqrt{4\sigma^3 + 2\sigma^2 + 3\sigma + 1}, \tag{4.9}$$

which yields $\gamma_0 = \gamma_2 = +1$, $\gamma_1 = -1$, and brings the formulas (4.3) and (4.5) into

$$[A_0,A_0^*] = 1, \quad [A_1,A_1^*] = -1, \quad [A_2,A_2^*] = 1. \tag{4.10}$$

Henceforth

$$G = \omega_0 A_0^* A_0 - \omega_1 A_1^* A_1 + \omega_2 A_2^* A_2 + g_0. \tag{4.11}$$

The representation (4.11) reveals a curious fact: the generator $G$ is not positively defined. Though the parts of $G$ are equivalent to the Hamiltonians of linear oscillators, the signs are alternating. Since we are on the ground of the traditional quantum mechanics, the attempts to get rid of this property by maneuvers similar to the Gupta formalism would be pointless. Departing from (4.11) one can construct the stationary wave packets in the rotating frame. The simplest one (the "ground state"), $\psi_0$, can be defined by

$$A_0\psi_0 = A_1^*\psi_0 = A_2\psi_0 = 0. \tag{4.12}$$

The integrability of Eqs. (4.12) and the existence of a square integrable solution $\psi_0$ is a consequence of some general facts. Let $C^2(\mathbb{R}^n)$ be the space of twice differentiable complex functions $\phi = \{\phi(\mathbf{x})\}$ of the argument $\mathbf{x} = \|x_j\|_{j=1,\ldots,n} \in \mathbb{R}^n$. Let now $B_1,\ldots,B_n$ be a system of $n$ differential operators on $C^2(\mathbb{R}^n)$, each $B_k$ $(k = 1,\ldots,n)$ being a linear combination of the $2n$ canonical operators $x_j$, $p_j = (1/i)(\partial/\partial x_j)$, and let $B_1^*,\ldots,B_n^*$ denote their Hermitian conjugates.

*Lemma 3:* If the operators $B_1,\ldots,B_n$ obey the commutation relations in $C^2(\mathbb{R}^n)$,

$$[B_j,B_k] = 0, \tag{4.13}$$

$$[B_j,B_k^*] = \delta_{jk} \quad (j,k = 1,\ldots,n), \tag{4.14}$$

then the system of $n$ differential equations

$$B_j\psi_0(\mathbf{x}) = 0, \quad j = 1,\ldots,n, \tag{4.15}$$

has a square-integrable solution

$$\psi_0(\mathbf{x}) = ce^{-(1/2)\phi(\mathbf{x})}, \tag{4.16}$$

where $\phi(\mathbf{x})$ is a complex quadratic form of $\mathbf{x}$ with a positively defined real part.

*Note:* This lemma is almost immediate in the Hilbert space formalism of quantum mechanics. [Owing to the commutation relations (4.13), (4.14), the $B_1,\ldots,B_n$ are unitarily equivalent to the annihilation operators of the $n$-dimensional oscillator; hence the solution $\psi_0(\mathbf{x})$ arises as a unitary image of the oscillator ground state.] However, for the sake of an explicit expression, an independent, "constructive proof" might be of interest.

*Constructive proof of Lemma 3:* According to our assumption, the operators $B_j$ and $B_j^*$ are of the form

$$B_j = \alpha_j\nabla + \beta_j\mathbf{x}, \quad B_j^* = -\alpha_j^*\nabla + \beta_j^*\mathbf{x} \quad (j = 1,\ldots,n), \tag{4.17}$$

where $\alpha_j$, $\beta_j$ ($j=1,...,n$) are certain $n$-component complex vectors, $\nabla$ is the gradient operator, and the "products" $\alpha\nabla$, $\beta x$ have the usual Euclidean geometry sense. Note that in view of (4.13) the vectors $\alpha_j$ must be linearly independent. Indeed, assume $c_1,...,c_n$ are complex numbers such that $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$; then the operators $B = c_1B_1 + \cdots + c_nB_n$ and $B^* = c_1^*B_1^* + ... + c_n^*B_n^*$ do not contain differentiations $\Rightarrow$

$$[B,B^*] = 0 \Rightarrow \sum_{i,k=1}^{n} c_j c_k^* [B_j, B_k^*] = \sum_{j=1}^{n} |c_j|^2 = 0 \Rightarrow c_1 = \cdots = c_n = 0.$$

Quite similarly, $\beta_1,...,\beta_n$ are independent.

In view of the commutativity (4.13) the system of equations (4.15) is integrable, and since $B_j$ are linear in $x$, the solution $\psi_0$ has to be of the form

$$\psi_0(x) = ce^{-(1/2)a_{ik}x_ix_k} = ce^{-(1/2)(xax)}, \tag{4.18}$$

where $a = \|a_{ik}\|$ is a symmetric complex coefficient matrix. By feeding (4.18) into (4.15) one obtains the following necessary and sufficient set of conditions:

$$a\alpha_j = \beta_j \quad (j=1,...,n). \tag{4.19}$$

In view of the linear independence of the vector systems $\{\alpha_j\}$ and $\{\beta_j\}$, (4.19) provides a unique and noncontradictory definition of a nonsingular matrix $a$. Moreover, (4.13) $\Rightarrow \alpha_j\beta_k = \alpha_k\beta_j$; hence the symmetry of the matrix elements $(\alpha_k a\alpha_j)$ is granted.

To obtain an explicit expression for $a$, define a basis $s_1,...,s_n$ dual to $\alpha_1,...,\alpha_n$:

$$s_k\alpha_j = \delta_{kj} = s_k^*\alpha_j^*. \tag{4.20}$$

Since the tensor products $s_k^* \otimes s_j$ span the whole $n\times n$ matrix space, one may seek $a$ in the form

$$a = \alpha_{kj}s_k^* \otimes s_j \quad \text{(summation convention)}, \tag{4.21}$$

where $\alpha_{kj}$ is another matrix of coefficients. By forming now the scalar products of (4.21) with $\alpha_k^*$ and $\alpha_j$ one finds

$$\alpha_{kj} = (\alpha_k^* a\alpha_j) = \alpha_k^*\beta_j, \tag{4.22}$$

and by using (4.14) one sees that

$$\alpha_{kj} + \alpha_{jk}^* = \alpha_k^*\beta_j + \alpha_j\beta_k^* = \delta_{kj}. \tag{4.23}$$

Hence the decomposition of (4.21) into the Hermitian and anti-Hermitian parts yields

$$a = S + i\Omega, \tag{4.24}$$

where

$$S = \tfrac{1}{2}s_j^* \otimes s_j, \tag{4.25}$$

$$\Omega = [(\alpha_k^*\beta_j - \alpha_j\beta_k^*)/2i]s_k^* \otimes s_j. \tag{4.26}$$

The expression (4.18) henceforth becomes

$$\psi_0(x) = c\exp\left[-\frac{1}{4}\sum_{j=1}^{n}|s_jx|^2 - \frac{i}{2}(x\Omega x)\right], \tag{4.27}$$

where $(x\Omega x)$ is real.

*Corollary:* As a by-product the following expressions for $B_j$ and $B_j^*$ are obtained:

$$B_j = e^{-(1/2)\Phi}\alpha_j\nabla e^{(1/2)\Phi},$$
$$B_j^* = -e^{(1/2)\Phi^*}\alpha_j^*\nabla e^{-(1/2)\Phi^*}, \tag{4.28}$$

where $-\frac{1}{2}\Phi$ is the exponent in (4.18).

In our case $B_1 = A_0$, $B_2 = A_1^*$, $B_3 = A_2$, and Lemma 3 yields the following expression for the "basic" Schrödinger packet (where the quadratic form $\frac{1}{2}\Sigma_{j=0}^{2}|\bar{s}_j\bar{x}|^2$ is already reduced to the three orthogonal main axes along the unit vectors $m_0$, $m_1$, $m_2$):

$$\psi_0(x) = c\exp\left[-\frac{1}{2}\omega\sum_{j=0}^{2}\lambda_j(\alpha)\,|m_jx|^2 - \frac{i}{2}\Omega(x)\right]. \tag{4.29}$$

Here $m_0$, $m_1$, $m_2$ are defined in terms of $s_0$, $s_1$, $s_2$ as the eigenvectors of the $3\times3$ matrix (4.25). The $m_1$ coincides with $\bar{m}$ and does not depend on $\alpha$. The remaining two vectors $m_0$ and $m_2$ lie in the plane spanned by the magnetic field $B$ and by the rotation axis $s$ (viewed in the rotating frame), forming with $s$ the $\alpha$-dependent angles $\phi_1(\alpha)$ and $\phi_2(\alpha) = (\pi/2) + \phi_1(\alpha)$ [see Fig. 11(a)]. In turn, the numbers $\lambda_0(\alpha), \lambda_1(\alpha), \lambda_2(\alpha) > 0$ are the roots of the characteristic equation of the matrix (4.25), their dependence on $\sigma$ being plotted in Fig. 11(b). The asymptotic values for $\alpha\to0$ and $\alpha\to\alpha_{cr}$ are of some interest. As $\alpha\to0$, the $\lambda_j(\alpha)$ behave as

$$\lambda_j(\alpha) \underset{\alpha\to0}{\sim} \alpha\lambda_j, \quad \lambda_0 = \lambda_1 = 1/\sqrt{2}, \quad \lambda_2 = 1. \tag{4.30}$$

Note that the limit $\alpha\to0$ can be approached in two ways: as $a\to0$ and/or $\omega\to+\infty$. If $a\to0$, the confinement disappears and the Gaussian packet (4.29) dissolves. A more in-

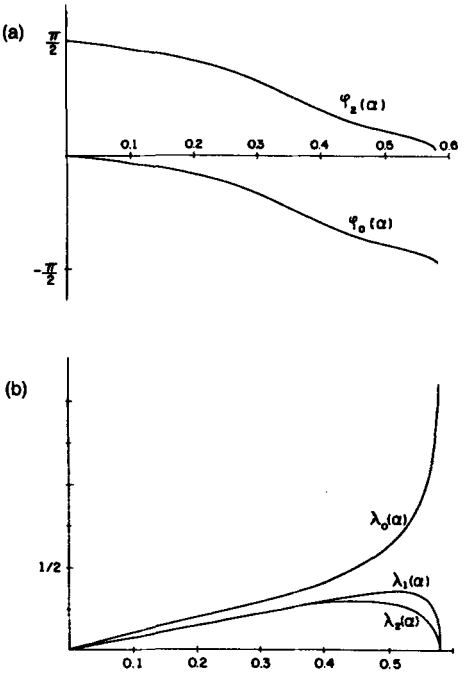



FIG. 11. The parameters characterizing the ground state (4.29) plotted versus the variable $\alpha$: (a) the angles $\rho_0(\alpha)$, $\rho_2(\alpha)$ between the vectors $\bar{m}_0(\alpha)$, $\bar{m}_2(\alpha)$ and the rotation axis $z$; (b) the eigenvalues of the real quadratic form $(\bar{x}S\bar{x})$.

teresting phenomenon occurs if $a = \text{const}$ and $\omega \to +\infty$ (the "high-frequency confinement"). In that case

$$\omega\lambda_j(\alpha) = \omega\lambda_j(a/\omega) \underset{\omega\to+\infty}{\sim} \omega(a/\omega)\lambda_j = a\lambda_j,$$

and the ground state (4.29) acquires the asymptotic shape checking the high-frequency approximation[14,15,16]:

$$\psi_0(\mathbf{x}) \underset{\omega\to+\infty}{\to} ce^{-(1/2)a(z^2 + x^2/\sqrt{2} + y^2/\sqrt{2})}. \tag{4.31}$$

On the other extreme, for $|\alpha| \to \alpha_{cr}$ the main axes of the quadratic exponent in (4.29) stick to $\tilde{m}_0$, $\tilde{m}_1$, $\tilde{m}_2$, where $\tilde{m}_1 = \bar{m}$, and $\tilde{m}_0$, $\tilde{m}_2$ are two orthogonal unit vectors in the $xz$ plane forming with $\bar{s}$ the angles $\phi_0 \cong -0.467\,527\,33\pi$ and $\tilde{\phi}_2 = \pi/2 + \tilde{\phi}_0 \cong 0.032\,472\,67\pi$, respectively. Thus $\bar{m}$ and $\tilde{m}_2$ define the critical plane already described in Sec. III B as $P_1 = P[\tilde{\sigma}_1]$. In turn, the coefficients $\lambda_1(\alpha)$ and $\lambda_2(\alpha)$ vanish as $|\alpha| \to \alpha_{cr}$ and $\lambda_2(\alpha)/\lambda_1(\alpha) \to 0.357\,612\,68$, whereas

$$\lambda_0(\alpha) \underset{\alpha\to\alpha_{cr}}{\to} 3.243\,619\ldots.$$

Henceforth, as $\alpha \to \alpha_{cr}$, the Gaussian wave packet (4.29) spills anisotropically all over the critical plane $P_1 = P[\tilde{\sigma}_1]$ in a good accordance with the classical case.

The aforementioned ground state (4.29) is an analog of the classical "rest state" (the point particle at rest in the coordinate center) and, simultaneously, one of the familiar "squeezed states."[23,24] The analogs of other special orbits described in Sec. III can now be easily constructed by applying powers of $A_0^*$, $A_1$, and $A_2^*$ to the basic state (4.29). This leads to the triple sequence of eigenvectors

$$\psi_{n,m,r} = A_0^{*n} A_1^m A_2^{*r} \psi_0 \quad (n,m,r = 0,1,2,\ldots), \tag{4.32}$$

expressible as

$$\psi_{n,m,r}(\mathbf{x}) = C_{nmr}\psi_0(\mathbf{x})\left[e^{S(\mathbf{x})}(\partial_0^n \partial_1^m \partial_2^r) e^{-S(\mathbf{x})}\right], \tag{4.33}$$

where

$$S(\mathbf{x}) = (\mathbf{x}S\mathbf{x}), \tag{4.34}$$

$$\begin{aligned}
\partial_0 &= (3\sigma_0 + 1)\frac{\partial}{\partial x} - 2i|\sigma_0|^{3/2}\frac{\partial}{\partial y} - (\sigma_0 + 1)^2\alpha^{-1}\frac{\partial}{\partial z},\\
\partial_1 &= -(3\sigma_1 + 1)\frac{\partial}{\partial x} - 2i|\sigma_1|^{3/2}\frac{\partial}{\partial y} + (\sigma_1 + 1)^2\alpha^{-1}\frac{\partial}{\partial z},\\
\partial_2 &= (3\sigma_2 + 1)\frac{\partial}{\partial x} - 2i|\sigma_2|^{3/2}\frac{\partial}{\partial y} - (\sigma_2 + 1)^2\alpha^{-1}\frac{\partial}{\partial z},
\end{aligned} \tag{4.35}$$

and, corresponding to the triple sequence of eigenvalues,

$$\lambda_{n,m,r} = n\omega_0 - m\omega_1 + r\omega_2 \quad (n,m,r = 0,1,2,\ldots). \tag{4.36}$$

Since the frequencies $\omega_0$, $\omega_1$, $\omega_2$ (as far as we have checked) are not commensurable, the resulting set of eigenvalues is dense on the real axis $\mathbb{R}$. Thus the spectrum of $G$ coincides with $\mathbb{R}$, though it is not continuous.

### A. Absence of energy interpretation

The eigenstates (4.32) are worth separate discussion. While stationary in the rotating frame, in the original one they are afflicted by an extra rotational drift with a constant angular velocity $\omega$. This is by no means unexpected. Since the external field $\bar{B}(t)$ is time dependent, the motion problem (2.1) in the original inertial frame has no stationary solutions. Our "drifting waves" are precisely "as stationary as possible," which in this case means periodic behavior. An intriguing question, however, arises. What happens to the above periodic patterns of motion if they are allowed to interact with an external radiation field? In the case of the conventional oscillator the answer is simple: after radiating some quanta the system has a tendency to settle down in its ground state. An analogous answer in the case of our drifting waves might remain true, though the argument is no longer so simple. Here, in contrast to the harmonic oscillator, it is not granted that the observable $G$ serving as the evolution generator (in the rotating frame) is involved in any conservative balance (so that if the system changes the $G$ eigenstate, it must simultaneously omit or receive some "$G$ quanta.") In fact, such an interpretation would be wrong, since then the system could become an infinite "$G$ source" at the cost of pushing the electron down into the $G$ negative states. Henceforth our rotating arrangement is an example of a new physical situation, where the (effective) evolution generator has no energy interpretation and the interaction with the outside world has to be separately treated. This might be relevant for the general problem of the quantum Hamiltonians in the noninertial frames (see, e.g., Ref. 25). (While permitting one to find stationary states of motion, these "Hamiltonians" do not, however, enter into the universal conservative balance, and so their spectrum is not interpretable as a radiation spectrum "in eyes of a noninertial observer.")

### B. Corrections to the high-frequency approximation

In spite of lacking the sense of energy, the generator $G$ permits one to improve the high-frequency approximation commonly used to determine the effects of quickly oscillating fields.[14–17] Given an electromagnetic field

$$\bar{E} = \omega\bar{E}(\bar{x})\sin(\omega t + \delta_1), \quad \bar{H} = \omega\bar{H}(\bar{x})\sin(\omega t + \delta_2),$$

the average motion of an electric charge $e$ within this approximation is described in terms of a noncovariant effective potential $V_{ef}(\bar{x}) = (e/2m)^2[\bar{E}(\bar{x})]^2$. For our rotating field (1.2) it leads to the effective Hamiltonian:

$$H_{ef} = \frac{\bar{P}^2}{2} + \frac{a^2}{2}\left[z^2 + \frac{x^2}{2} + \frac{y^2}{2}\right]. \tag{4.37}$$

As can be seen, our exact solution (4.29) for $\omega \to +\infty$ tends indeed to the ground state of the oscillator Hamiltonian $H_{ef}$, thus confirming the high-frequency approximation for the Schrödinger wave packets. However, if $\omega$ is large but not infinite (or if both $\omega$ and $a$ are large), our $G$ theory yields different results. In fact, for $\alpha \neq 0$, the ground state $\psi_0(\bar{x},t)$ (seen from the original frame) leads to the variable probability density in $\mathbb{R}^3$. Its time average is

$$\rho(\bar{x}) = \frac{1}{T}\int_0^T |\psi_0(\bar{x},t)|^2\,dt, \tag{4.38}$$

where $\psi_0(\bar{x},t)$ is the function (4.29) with $\mathbf{m}_0$, $\mathbf{m}_1$, $\mathbf{m}_2$ substituted by their rotating variants $\mathbf{m}_0(t)$, $\mathbf{m}_1(t)$, $\mathbf{m}_2(t)$. The function (4.38) is stationary but not Gaussian. If $|\alpha| > 0$,

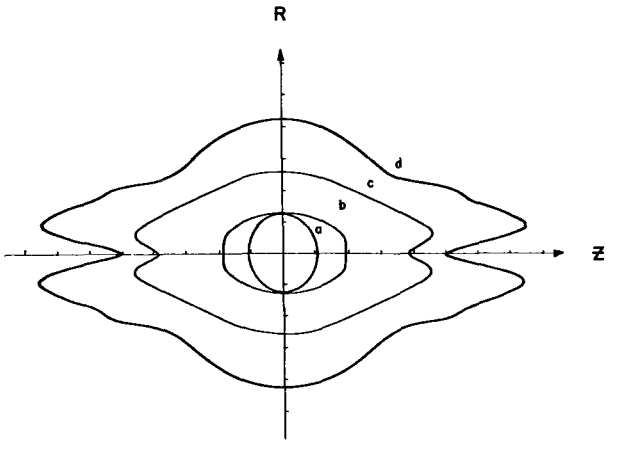



FIG. 12. The surfaces $\rho(\bar{x})/\rho(0) = \text{const}$ for $a = \frac{1}{10}$ and varying $\omega$: (a) $\omega \cong 316.07 \Rightarrow \alpha \cong 0.316\,38$ (the shape predicted by the high-frequency approximation); (b) $\omega \cong 17.69$, $\alpha \cong 0.5653$ (the dissention from the high-frequency approximation predicted by our exact model); (c) $\omega \cong 17.2495$; (d) $\omega \cong 17.2428$ (the anisotropic expansion of the probability density in the subcritical cases).

$\rho(\bar{x})$ starts to deviate from the ground state density of the effective Hamiltonian $H_{\text{ef}}$, and as $\alpha \to \pm \alpha_{\text{cr}}$ it extends tentacles along the $z$ axis [see the surfaces $\rho(\bar{x}) = \text{const}$ in Fig. 12]. This deformation might prove decisive for the properties of the standing wave traps of the particular structure (3.38).

### C. The structure of the "radiation-made crystals"

The physical conditions met by the Schrödinger electron near the nodal points of the standing wave (3.38) are determined by two parameters: the wavelength $\lambda = 2\pi c/\omega$, and the "average radius" $p_0$ of the wave packet (4.29) (defined as the quadratic deviation of the Gaussian probability distribution $\rho_0 = \langle \bar{x}^2 \rangle^{1/2}$). In the traditional cgs units,

$$\rho_0 = \frac{1}{\sqrt{2\omega}} \left[ \frac{1}{\lambda_0(\alpha)} + \frac{1}{\lambda_1(\alpha)} + \frac{1}{\lambda_2(\alpha)} \right]^{1/2} \left( \frac{\hbar}{m_e} \right)^{1/2}$$
$$\underset{\alpha \to 0}{\sim} \frac{1}{\sqrt{2\alpha\omega}} \left( \frac{1+\sqrt{2}}{2} \right)^{1/2} \left( \frac{\hbar}{m_e} \right)^{1/2}. \tag{4.39}$$

Depending on the relation between these two quantities, four different physical situations are predictable.

(1°) If $\alpha$ is small ($\alpha \lesssim 10^{-19} \nu$ sec), the confinement is weak ($\rho_0 > \lambda/100$) and the localization of the ground state $\psi_0$ around the nodal point is insufficient to justify the use of our rotating model. (The high-frequency approximation is most exact in this area.)

(2°) As $\alpha$ grows ($10^{-19} \nu \sec \lesssim \alpha \lesssim \alpha_{\text{cr}} - 2 \times 10^{-37} \nu^2 \sec^2$) the confinement becomes tighter ($\rho_0 < \lambda/100$). [In fact, we have checked that if $\nu < 10^{14}$/sec, then for most of the $\alpha$ values in this interval 99% of the Gaussian packet (4.29) is compressed within the sphere of the radius of $10^{-3}$ of the standing wavelength.] The electrons in the fundamental state of motion (represented by the wave packet $\psi_0$) are therefore trapped around the single nodal points, and our rotating model is good for describing the nature of this trap.

(3°) If $\alpha_{\text{cr}} - 2 \times 10^{-37} \nu^2 \sec^2 \lesssim \alpha \lesssim \alpha_{\text{cr}}$, the situation changes once more. For $|\alpha| \to \alpha_{\text{cr}}$ the rotating packet $\psi_0(\bar{x}, t)$ expands, and our rotating model is no longer globally applicable; yet it gives a hint about the phenomenon of anisotropic subcritical conductivity, which should follow the design of the surfaces in Fig. 12.

(4°) Finally, as $|\alpha|$ crosses $\alpha_{\text{cr}}$, the attractive nodal points of the standing wave undergo a sudden phase transition into repulsive centers, and the whole nodal lattice $\mathcal{N}$ tends to expel the electrons anisotropically with a preference for a certain cone of directions around the axis $z$.

### D. The question of positivity

The indefinite character of $G$ might also mark some natural applicability limits of the positivity axiom in quantum theories. This axiom seems firmly established whenever the evolution generator describes a continuous flux of evolution and has an interpretation of a conservative quantity. However, if $G$ is just an average or effective generator, describing a summary effect of some unperceived microfluctuations and concerning only a discrete sequence of time moments, then our example shows that the positivity cannot be demanded.

## ACKNOWLEDGMENTS

The authors are grateful to colleagues in the Departamento de Física, CINVESTAV, México, and at the Institute of Theoretical Physics, Warsaw University, Warsaw, Poland, for their interest in this paper and helpful discussion.

One of us (B.M.) wishes also to acknowledge support from Polish Interdisciplinary Grant CPBP 01.03.

[1] P. Collet and J. P. Eckman, *Iterated Maps on the Interval as Dynamical Systems, Progress in Physics*, No. 1, edited by A. Jaffe and D. Ruelle (Birkhäuser, Cambridge, MA, 1980).
[2] W. Magnus, Commun. Pure Appl. Math. 7, 649 (1954).
[3] J. Wilcox, J. Math. Phys. 8, 962 (1967).
[4] I. Bialynicki-Birula, B. Mielnik, and J. Plebański, Ann. Phys. (NY) **51**, 187 (1969).
[5] B. Mielnik and J. Plebański, Ann. Inst. H. Poincaré **11**, 215 (1970).
[6] M. Suzuki, J. Math. Phys. **26**, 61 (1985).
[7] E. Wigner, Phys. Rev. **40**, 749 (1932).
[8] H. R. Lewis and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[9] A. Malkin, V. I. Man'ko, and D. A. Trifonow, Phys. Rev. D **2**, 1371 (1970).
[10] A. Malkin, V. I. Man'ko, and D. A. Trifonow, J. Math. Phys. **14**, 576 (1973).
[11] M. Moshinsky and P. Winternitz, J. Math. Phys. **21**, 1667 (1980).
[12] B. Mielnik, J. Math. Phys. **27**, 2290 (1986).
[13] B. Mielnik, Lett. Math. Phys. **12**, 49 (1986); **13**, 259 (1987).
[14] D. E. Toschek, "Atomic particles in traps," in *New Trends in Atomic Physics*, Les Houches, 1982, edited by G. Grynberg and R. Stora (North-Holland, Amsterdam, 1984).
[15] L. S. Brown and G. Gabrielse, Rev. Mod. Phys. **58**, 233 (1986).
[16] V. S. Letokhov and V. G. Minogin, Sov. Phys. JETP **74**, 1318 (1978).
[17] S. Stenholm, Rev. Mod. Phys. **48**, 699 (1986).
[18] W. D. Phillips and H. J. Metcalf, Sci. Am. **256** (3), 36 (1987).
[19] F. V. Bunkin and A. M. Prokhorov, Sov. Phys. JETP **46**, 1090 (1964).
[20] A. J. Tip, J. Phys. A: Math. Gen. **16**, 3237 (1983).
[21] J. Plebański, "On the generators of the $N$-dimensional pseudo-unitary and pseudo-orthogonal groups," CINVESTAV preprint, Mexico, 1966.
[22] H. Dekkner, Phys. Rep. **80**, 1 (1981); also J. Plebański, "On the linear unitary transformations of two canonical variables," 1987.
[23] B. R. Mollow, Phys. Rev. **162**, 1256 (1967).
[24] D. F. Walls, Nature (London) **306**, 141 (1983).
[25] E. Schmutzer and J. Plebański, Fortschr. Phys. **25**, 37 (1977).

# Erratum: Evolution loops [J. Math. Phys. 27, 2290 (1986)]

Bogdan Mielnik
*Departamento de Fisica, Centro de Investigacion y de Estudios Avanzados del Instituto Politécnico Nacional, Apdo. Postal 14-740, Mexico, D.F., Mexico*

(Received 24 August 1988; accepted for publication 21 September 1988)

The values of the parameter $\alpha$ defining the confined and/or loop motions [p. 2296, below (4.19)] are incorrect. The correct values are for the first confinement interval, $|\alpha| \lesssim 17.92$; for the three simplest loop generating values,

$$S(\alpha) = -1 \Rightarrow \alpha \cong \pm 15.489,$$

$$S(\alpha) = 0 \Rightarrow \alpha \cong \pm 12.619,$$

$$S(\alpha) = +1 \Rightarrow \alpha \cong \pm 8.904;$$

and for the second confinement interval, $148.413 \lesssim |\alpha| \lesssim 149.617$.

In formula (6.29), the second matrix should be squared. Consistently, (6.31) should read

$$r(\alpha) = 2\cos 2\alpha - \alpha \sin 2\alpha = 2\cos(2\pi l/n). \quad (6.31)$$

When writing this paper the author was not aware of the theory of *ion traps*; hence a part of literature is missing. In particular, the following references should be added.

[1]E. D. Toschek, "Atomic particles in traps," in *New Trends in Atomic Physics*, edited by G. Grynberg and R. Stora (North-Holland, Amsterdam, 1984).
[2]L. S. Brown and G. Gabrielse, Rev. Mod. Phys. **58**, 233 (1986).

 0022-2488/89/020550-01$02.50 © 1989 American Institute of Physics 