On reduction of two degrees of freedom Hamiltonian systems by an S^1 action, and $SO_0(1,2)$ as a dynamical group

Toshihiro Iwai
Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Kyoto 606, Japan

(Received 14 September 1984; accepted for publication 23 November 1984)

Reduction by an S^1 action is a method of finding periodic solutions in Hamiltonian systems, which is known rather as the method of averaging. Such periodic solutions can be reconstructed as S^1 orbits by pulling back the critical points of an associated "reduced Hamiltonian" on a "reduced phase space" along the reduction. For Hamiltonian systems of two degrees of freedom, a geometric setting of the reduction is already accomplished in the case where the reduced phase space is a two-sphere in the Euclidean space $\mathbb{R}^3$, and the reduced Hamilton's equations of motion are Euler's equations. This article deals with the case where the reduced phase space will be a two-hyperboloid in the three-Minkowski space, and the reduced Hamilton's equations of motion will be Euler's equations with respect to the Lorentz metric. This reduction is associated with SU(1,1) symplectic action on the phase space $\mathbb{R}^4$. As a consequence of this association the reduced Hamiltonian system proves to admit a dynamical group $SO_0(1,2)$. A well-known reduction by an S^1 action occurs in the case of rotational-invariant Hamiltonian systems, which will be associated with SL(2,$\mathbb{R}$) symplectic action on $\mathbb{R}^4$. It is shown that the reduction associated with SU(1,1) and with SL(2,$\mathbb{R}$) are symplectically equivalent.

I. INTRODUCTION

Reduction by an S^1 action is a method of finding periodic solutions in Hamiltonian dynamical systems. Kummer[1] made an intensive use of the reduction method to find periodic solutions for the resonant Hamiltonians with two equal frequencies, i.e.,

$$H_{\pm} = \tfrac{1}{2}(x_1^2 + y_1^2) \pm \tfrac{1}{2}(x_2^2 + y_2^2) + O_3, \tag{1.1}$$

where x_j, y_j, $j = 1,2$, are the Cartesian coordinates in the phase space $\mathbb{R}^4$. The reduction method is known rather as the method of averaging. For a fair insight into this method, see Ref. 2. The level surface of the leading terms in the right-hand side of (1.1), denoted by $L_{\pm}$, are three-spheres or three-hyperboloids, according to whether the plus sign or minus sign is considered. The former will be referred to as the compact case, and the latter as the noncompact case. In the compact case, the reduction is closely related to the Hopf map $S^3 \to S^2$. Cushman and Rod[3] factored the Hopf map in terms of the momentum map associated with an SU(2) symplectic action on $\mathbb{R}^4$ to show that for polynomial Hamiltonians H_+ commutative with L_+ the reduced Hamilton's equations of motion are just Euler's equations restricted to S^2 in $\mathbb{R}^3$.

The purpose of this article is to accomplish a geometric setting for the reduction by an S^1 action in the noncompact case, which is linked with the "pseudo-Hopf" map of a three-hyperboloid to a two-hyperboloid. The pseudo-Hopf map will be factored in terms of the momentum map associated with an SU(1,1) symplectic action on $\mathbb{R}^4$. For Hamiltonians H_- commutative with L_- the reduced Hamilton's equations of motion are proved to be Euler's equations restricted to one sheet of a two-sheeted two-hyperboloid in the three-Minkowski space.

In both compact and noncompact cases, critical points for Euler's equations are pulled back to periodic solutions in the original Hamiltonian systems.

There is a widely known reduction of two degrees of freedom Hamiltonian systems by an S^1 action. The SO(2) action on $\mathbb{R}^2$, rotation, is symplectically lifted to that on $\mathbb{R}^4$, the phase space. The reduction by the SO(2) is just the elementary fact that rotational-invariant Hamiltonian systems of two degrees of freedom can be described in terms of r and p_r only, where original Hamiltonian systems are described in the polar coordinates and their conjugate momenta (r,θ,p_r,p_θ). The reduction will be shown to be associated with a symplectic action of SL(2,$\mathbb{R}$) on $\mathbb{R}^4$.

The fact that SU(1,1) is isomorphic with SL(2,$\mathbb{R}$) gives a symplectic equivalence between the reduction associated with SU(1,1) and that with SL(2,$\mathbb{R}$). To show this is another purpose of this article.

The third purpose is to show that the reduced Hamiltonian system admits $SO_0(1,2)$ as a dynamical group.

II. REDUCTION BY U(1)

A. The pseudo-Hopf map

Let $(\mathbb{R}^4,\omega)$ be a symplectic vector space with the Cartesian coordinates (x_j, y_j), $j = 1,2$, where ω is the standard symplectic form given by

$$\omega = dx_1 \wedge dy_1 + dx_2 \wedge dy_2. \tag{2.1}$$

Define H_2 by

$$H_2 = \tfrac{1}{2}(x_1^2 + y_1^2) - \tfrac{1}{2}(x_2^2 + y_2^2). \tag{2.2}$$

The flows of Hamilton's equation for the H_2 are generated by the Hamiltonian vector field X_2 determined by $i(X_2)\omega = dH_2$, $i(\ )$ denoting the interior product. These flows define a symplectic S^1 action on $\mathbb{R}^4$. To see this, it is convenient to introduce the complex vector space structure $\mathbb{C}^2$ in $\mathbb{R}^4$ by

$$z_1 = x_1 + iy_1, \quad z_2 = x_2 - iy_2. \tag{2.3}$$

We remark that z_2 is not set as $x_2 + iy_2$. Then ω and H_2 take the form

$$\omega = \frac{i}{2}\sum G_{jk}\, dz_j \wedge d\bar{z}_k, \quad H_2 = \frac{1}{2}\sum G_{jk} z_j \bar{z}_k,$$

with

$$G = (G_{jk}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.4}$$

respectively, where $\bar{z}_k$ are the complex conjugates of z_k. Further, the Hamiltonian vector field X_2 is expressed in the form

$$X_2 = -i \sum \left(z_j \frac{\partial}{\partial z_j} - \bar{z}_j \frac{\partial}{\partial \bar{z}_j} \right). \tag{2.5}$$

Integration of (2.5) yields a U(1) action on $\mathbb{C}^2$

$$\Phi_t : z \to e^{-it} z. \tag{2.6}$$

Here we have used the vector notation for points z of $\mathbb{C}^2$,

$$z = x + iGy, \quad \text{with } (x,y) \in \mathbb{R}^2 \times \mathbb{R}^2. \tag{2.7}$$

The action of Φ_t, which is easily known to be symplectic from (2.4) and (2.6), can be expressed by a 4×4 matrix acting on $\mathbb{R}^4$,

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} I_2 \cos t & G \sin t \\ -G \sin t & I_2 \cos t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{2.8}$$

where I_2 denotes the 2×2 unit matrix and G is the matrix defined in (2.4).

The level surface $H_2 = \frac{1}{2}(|z_1|^2 - |z_2|^2) = h$ is a three-hyperboloid which we denote by H^3, where we have assumed that h is a nonzero constant. It is clear that the matrix (2.8) acts on H^3.

Linear combinations of $z_j \bar{z}_k$ are clearly invariant under Φ_t. Among them, we call the quadratic polynomials defined below the "pseudo-Hopf" variables

$$V_1 = \tfrac{1}{2}(-x_1x_2 + y_1y_2) = -\tfrac{1}{2}\,\mathrm{Re}(z_1\bar{z}_2),$$
$$V_2 = \tfrac{1}{2}(x_1y_2 + y_1x_2) = \tfrac{1}{2}\,\mathrm{Im}(z_1\bar{z}_2),$$
$$V_3 = \tfrac{1}{4}(x_1^2 + y_1^2 + x_2^2 + y_2^2) = \tfrac{1}{4}(|z_1|^2 + |z_2|^2),$$
$$V_0 = \tfrac{1}{4}(x_1^2 + y_1^2 - x_2^2 - y_2^2) = \tfrac{1}{4}(|z_1|^2 - |z_2|^2), \tag{2.9}$$

where Re and Im indicate the real and the imaginary parts, respectively. We note here that $V_j, j = 1,2,3$, span a Lie algebra su(1,1) under the Poisson bracket

$$\{V_1,V_2\} = -V_3, \quad \{V_2,V_3\} = V_1, \quad \{V_3,V_1\} = V_2. \tag{2.10}$$

The V_0 is commutative with all the V_j. Furthermore, the square of V_0 equals the Casimir invariant for su(1,1), that is, one has

$$-V_1^2 - V_2^2 + V_3^2 = V_0^2. \tag{2.11}$$

Since $V_0 = \frac{1}{2}H_2$, the level surface H^3 is given by $V_0 = \frac{1}{2}h$ in $\mathbb{R}^4$. Equation (2.11) then defines a two-sheeted two-hyperboloid $-V_1^2 - V_2^2 + V_3^2 = (h/2)^2$ in $\mathbb{R}^3$ when (x, y) are restricted to H^3. By H^2_+ and H^2_- we mean the upper sheet and the lower sheet of the two-hyperboloid, respectively. We thus define "pseudo-Hopf" maps $H^3 \to H^2_+$ and $H^3 \to H^2_-$ by

$$\sigma_+ : (z_1,z_2) \mapsto (V_1,V_2,V_3), \tag{2.12a}$$

$$\sigma_- : (z_1,z_2) \mapsto (V_1,V_2,-V_3), \tag{2.12b}$$

respectively. Then H^3 can be shown to be a fiber space over $H^2_\pm$ with fiber S^1. In fact, since $V_j, j = 1,2,3$, are invariant under the U(1) action (2.6), each fiber $\sigma_\pm^{-1}(v)$ for $v \in H^2_\pm$ is an integral manifold of X_2 which is diffeomorphic to S^1.

The fiber space structure for $h > 0$ is easy to see when we introduce the following coordinates in $\mathbb{C}^2$:

$$z_1 = Re^{i[(\psi+\phi)/2]} \cosh(\tau/2), \quad z_2 = Re^{[i(\psi-\phi)/2]} \sinh(\tau/2), \tag{2.13}$$

where

$$0 \leqslant (\psi + \phi)/2 < 2\pi, \quad -\pi \leqslant (\psi - \phi)/2 < \pi, \quad R > 0, \quad \tau \geqslant 0.$$

H^3 is given by $R = (2h)^{1/2}$. It is also known from (2.6) and (2.13) that these coordinates are subject to the transformation $\psi \mapsto \psi - t$ and the others fixed. This means that ψ is the fiber coordinate.

B. Reduction of the Hamiltonian system $(\mathbb{R}^4, \omega, H)$

Following Cushman and Rod,[3] we consider polynomial Hamiltonian functions of the form $H = \Sigma H_{2k}$, where H_{2k} are homogeneous polynomials of degree $2k$, $1 \leqslant k \leqslant n$, in the variables $(x_j, y_j), j = 1,2$. The H is assumed, in addition, to be in normal form with respect to H_2; that is, H and H_2 commute under the Poisson bracket $\{H,H_2\} = 0$. As will be shown in Sec. II F, the commutativity means that H is a polynomial in the pseudo-Hopf variables $V_j, j = 0,...,3$. The commutativity implies, moreover, that the flows of the Hamiltonian vector field X_H take place on H^3 for any fixed $h \neq 0$.

As was already shown in (2.6)–(2.8), the flow of X_2 defines the symplectic U(1) action Φ_t on $\mathbb{R}^4$. The momentum map associated with Φ_t is H_2 itself, which is manifestly Ad* equivariant because $H_2 \circ \Phi_t = H_2$. This will be touched on again in Sec. II C.

According to Ref. 4, reduction of $(\mathbb{R}^4, \omega, H)$ by the U(1) action is carried out as follows. Since Φ_t acts freely and properly on the momentum manifold $M = H_2^{-1}(h)$, which is nothing but the level surface H^3, the reduction process yields a smooth orbit manifold $M_R = M/S^1$ [$\mathrm{U}(1) \cong S^1$], which is diffeomorphic to $H^2_+ \cong H^2_-$ because H^3 is a fiber space over $H^2_\pm$ with fiber S^1. The natural projection $\pi: M \to M_R$ together with ω determines a unique symplectic form ω_R on M_R by $\pi^*\omega_R = i^*\omega$, where the superscript asterisk indicates the pullback and $i: M \to \mathbb{R}^4$ is the inclusion map. Since the Hamiltonian H is invariant under the action of Φ_t, the reduced Hamiltonian H_R is induced on M_R in the manner such that $H_R \circ \pi = H \circ i$. Thus we obtain the reduced system (M_R, ω_R, H_R). The reduced Hamiltonian vector field X_{H_R} is given by $\pi_* X_H(p) = X_{H_R}(\pi(p))$ for $p \in M$, where the subscript asterisk indicates the tangent map.

In what follows, we will perform the reduction in the coordinates introduced in (2.13). For $h > 0$, the momentum manifold M is given by $R = (2h)^{1/2}$, and the pseudo-Hopf variables restricted on M are written as

$$V_1 \circ i = -(h/2)\cos\phi \sinh\tau, \quad V_2 \circ i = (h/2)\sin\phi \sinh\tau, \tag{2.14}$$
$$V_3 \circ i = (h/2)\cosh\tau, \quad V_0 \circ i = h/2,$$

where i remains to be the inclusion map. These are independent of the fiber coordinate ψ. The two-hyperboloid H^2 defined by (2.11) with V_j restricted to $M = H^3$ then has coordinates ϕ and τ with $0 \leqslant \phi < 2\pi$ and $\tau \geqslant 0$.

Since the U(1)-invariant Hamiltonian H should be a function of ϕ and τ only, it is regarded as a function H_R on H^2. This is the meaning of $H_R \circ \pi = H \circ i$. Now the symplectic form ω can be expressed, after calculation, in the form

$$\omega = \tfrac{1}{2}R\, dR \wedge (d\psi + \cosh\tau\, d\tau) + \tfrac{1}{4}R^2 \sinh\tau\, d\tau \wedge d\phi. \tag{2.15}$$

Restricted to M, ω gets the form

$$i^*\omega = (h/2)\sinh \tau \, d\tau \wedge d\phi. \tag{2.16}$$

The right-hand side of (2.16) is thought of as a two-form ω_R on H^2. This is the meaning of $\pi^*\omega_R = i^*\omega$.

For $h<0$, it is convenient to set

$$z_1 = Re^{i[(\psi-\phi)/2]}\sinh(\tau/2), \quad z_2 = Re^{i[(\psi+\phi)/2]}\cosh(\tau/2), \tag{2.17}$$

in place of (2.13). Then the momentum manifold M is given by $R=(2|h|)^{1/2}$, and V_j by similar equations to (2.14).

C. The momentum map associated with U(1,1)

Consider U(1,1) acting on $\mathbb{C}^2$. Elements g in U(1,1) satisfy the matrix relation $g^*Gg = G$, where g^* is the Hermitian conjugate of g. From the expression (2.4), we see that ω is invariant under the action of U(1,1); that is, U(1,1) acts symplectically on $\mathbb{C}^2$. By u(1,1) we mean the Lie algebra of U(1,1) consisting of α satisfying $\alpha^*G + G\alpha = 0$. The α's can be written as

$$\frac{1}{2}\begin{pmatrix} -ic_3 - ic_0 & c_2 + ic_1 \\ c_2 - ic_1 & ic_3 - ic_0 \end{pmatrix}, \quad \text{with } c_j \in \mathbb{R}. \tag{2.18}$$

The U(1,1) and u(1,1) have realizations in 4×4 real matrix form. Let $A+iB$ and $a+ib$ be elements of U(1,1) and of u(1,1), respectively. Then pairs (A,B) and (a,b) should be subject to

$$A^TGA + B^TGB = G, \quad A^TGB = B^TGA, \tag{2.19}$$

$$Ga + a^TG = 0, \quad Gb = b^TG, \tag{2.20}$$

respectively, where the superscript T indicates the transpose. Now the actions of U(1,1) and u(1,1) on $\mathbb{C}^2$, $(A+iB)(x+iGy)$ and $(a+ib)(x+iGy)$, yield the desired matrices

$$\begin{pmatrix} A & -BG \\ GB & GAG \end{pmatrix} \tag{2.21a}$$

and

$$\begin{pmatrix} a & -bG \\ Gb & GaG \end{pmatrix}, \tag{2.21b}$$

respectively.

Let the element (2.18) be expressed as $a+ib$. Then the matrix (2.21b) takes the form

$$\frac{1}{2}\begin{pmatrix} 0 & c_2 & c_3+c_0 & c_1 \\ c_2 & 0 & c_1 & c_3-c_0 \\ -c_3-c_0 & c_1 & 0 & -c_2 \\ c_1 & -c_3+c_0 & -c_2 & 0 \end{pmatrix}. \tag{2.22}$$

We take a basis $\{e_j\}$, $j=0,\dots,3$, for u(1,1) in the manner such that the matrix (2.22) is a linear combination $\Sigma c_j e_j$. The commutation relations among e_j are given by

$$[e_1,e_2] = -e_3, \quad [e_2,e_3] = e_1, \quad [e_3,e_1] = e_2,$$
$$[e_0,e_j] = 0, \quad j=1,2,3. \tag{2.23}$$

Let ζ be any element of u(1,1) with the matrix expression (2.21b). Then the infinitesimal symplectic action on $P=\mathbb{R}^4$ associated with ζ, denoted by ζ_P, is the vector field

$$\zeta_P = (ax - bGy)\cdot\frac{\partial}{\partial x} + (Gbx + GaGy)\cdot\frac{\partial}{\partial y}, \tag{2.24}$$

where $\partial/\partial x = (\partial/\partial x_1, \partial/\partial x_2)^T$ and $\partial/\partial y = (\partial/\partial y_1, \partial/\partial y_2)^T$, and the dot here stands for the standard inner product. The ζ_P has a generating function F such that $i(\zeta_P)\omega = dF$; that is,

$$\frac{\partial F}{\partial x} = -(Gbx + GaGy), \quad \frac{\partial F}{\partial y} = ax - bGy. \tag{2.25}$$

F is then found out to be

$$\begin{aligned} F(p) &= ax\cdot y - \tfrac{1}{2}bGy\cdot y - \tfrac{1}{2}Gbx\cdot x \\ &= \tfrac{1}{2}\omega(\zeta p, p), \quad p = \binom{x}{y} \in \mathbb{R}^4. \end{aligned} \tag{2.26}$$

For a general case, see Ref. 4, p. 190. Expanding Eq. (2.26), for $\zeta = \Sigma c_j e_j$, with respect to e_j, we obtain

$$\tfrac{1}{2}\omega(\zeta p, p) = c_1V_1 + c_2V_2 + c_3V_3 + c_0V_0, \tag{2.27}$$

where V_j are the pseudo-Hopf variables defined by (2.9).

The momentum map $L:\mathbb{R}^4 \to \mathrm{u}(1,1)^*$ is then given (see Ref. 4, p. 288) by

$$L(p)\cdot\zeta = \tfrac{1}{2}\omega(\zeta p, p), \tag{2.28}$$

where the dot means the paring of covectors and vectors. The L itself has the form

$$L(p) = V_1e_1^* + V_2e_2^* + V_3e_3^* + V_0e_0^*, \tag{2.29}$$

where $\{e_j^*\}$ is the basis of u(1,1)* dual to $\{e_j\}$.

The momentum map L is Ad* equivariant. In fact, for $g\in U(1,1)$ and $\zeta \in \mathrm{u}(1,1)$ we have, from (2.28),

$$\begin{aligned} L(gp)\cdot\zeta &= \tfrac{1}{2}\omega(\zeta gp, gp) = \tfrac{1}{2}\omega(g^{-1}\zeta gp, p) \\ &= L(p)\cdot \mathrm{Ad}_{g^{-1}}\zeta, \end{aligned}$$

or

$$L(gp) = \mathrm{Ad}^*_{g^{-1}} \circ L(p), \quad \text{for } p\in\mathbb{R}^4. \tag{2.30}$$

In what follows we specialize U(1,1) to subgroups. The matrix (2.8) gives a one-parameter subgroup U(1) of U(1,1). In fact, the matrix (2.21a) with $A = I_2\cos t$ and $B = -I_2\sin t$ is just (2.8). Further, the Lie algebra u(1) of U(1) has a basis e_0, because $\exp 2te_0$ equals (2.8). The momentum map associated with U(1) is then a restriction of L. In effect, setting $\zeta = e_0$ in (2.28), we get $L(p)\cdot e_0 = V_0 = \tfrac{1}{2}H_2$, so that $L(p)|_{\mathrm{u}(1)^*} = \tfrac{1}{2}H_2e_0^*$. Accordingly, the Ad* equivariance of H_2 under the U(1) action is a special case of that of L.

Here, SU(1,1) is a subgroup of U(1,1) with $\det(A+iB)=1$. The Lie algebra su(1,1) has a basis $\{e_j\}$, $j=1,2,3$, given in (2.23). We denote by J the momentum map associated with SU(1,1); $J:\mathbb{R}^4\to \mathrm{su}(1,1)^*$, which is a restriction of L to su(1,1)*. Let $\xi\in \mathrm{su}(1,1)$. Then Eqs. (2.28) and (2.29) reduce to

$$J(p)\cdot\xi = \tfrac{1}{2}\omega(\xi p, p), \tag{2.31}$$

$$J(p) = V_1e_1^* + V_2e_2^* + V_3e_3^*, \tag{2.32}$$

respectively, where $\{e_j^*\}$, $j=1,2,3$, is the basis dual to $\{e_j\}$. The Ad* equivariance of J is now transparent:

$$J(gp) = \mathrm{Ad}^*_{g^{-1}} \circ J(p). \tag{2.33}$$

We note that SU(1,1) and U(1) commute and that the Lie algebra su(1,1) + u(1) of SU(1,1)×U(1) coincides with u(1,1), so that the momentum map L can be thought of as associated with SU(1,1)×U(1). We will hereafter give our concern to SU(1,1)×U(1) rather than U(1,1). We remark in conclusion that SU(1,1)×U(1) is a covering group of U(1,1).

D. Factoring the canonical projection $\pi: M \to M/S^1$

In this section, we are going to break up the natural projection $\pi: M \to M/S^1$ through the following diagram:

$$\begin{array}{ccc} M = H^3 & \xrightarrow{J} & \mathrm{su}(1,1)^* \\ \pi \downarrow & & \downarrow \\ M_R \simeq H^2_\pm & \longleftarrow & \mathrm{su}(1,1) \end{array}, \tag{2.34}$$

where the maps other than J and π will be defined in the sequel. The reasoning to be done in this section is a translation of that in Ref. 3 into a noncompact case.

Let M continue to denote the three-hyperboloid H^3 determined by $H_2^{-1}(h)$. The action of SU(1,1) is transitive on M, because for any point (z_1,z_2) with $|z_1|^2 - |z_2|^2 = 2h > 0$ and a fixed point $p_0 = ((2h)^{1/2},0)$ of $M \subset \mathbb{C}^2$, we have

$$\begin{pmatrix} z_1(2h)^{-1/2} & \bar{z}_2(2h)^{-1/2} \\ z_2(2h)^{-1/2} & \bar{z}_1(2h)^{-1/2} \end{pmatrix} \begin{pmatrix} (2h)^{1/2} \\ 0 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \tag{2.35}$$

and for (z_1,z_2) with $|z_1|^2 - |z_2|^2 = 2h < 0$ and $p_0 = (0,(2|h|)^{1/2})$,

$$\begin{pmatrix} \bar{z}_2(2|h|)^{-1/2} & z_1(2|h|)^{-1/2} \\ \bar{z}_1(2|h|)^{-1/2} & z_2(2|h|)^{-1/2} \end{pmatrix} \begin{pmatrix} 0 \\ (2|h|)^{1/2} \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}. \tag{2.36}$$

Therefore, the Ad* equivariance (2.33) together with $M = \mathrm{SU}(1,1)\cdot p_0$ yields

$$J(M) = \{\mathrm{Ad}^*_{g^{-1}} \circ J(p_0) | g \in \mathrm{SU}(1,1)\} \subset \mathrm{su}(1,1)^*, \tag{2.37}$$

where $J(p_0) = (|h|/2)e_3^*$. Thus $J(M)$ is a coadjoint orbit, so that it is even dimensional and has a standard symplectic form (the Kirillov–Kostant–Souriau theorem).

We now determine a symplectic form Ω on $J(M)$ such that

$$J^*\Omega = i^*\omega, \tag{2.38}$$

where $i: M \to \mathbb{R}^4$ is the inclusion map. Let $\mathbb{R}^4 = P$ and $\mathrm{su}(1,1)^* = Q$. For $\xi \in \mathrm{su}(1,1)$, we mean by $\xi_P(p)$ and $\xi_Q(\mu)$ the infinitesimal generator of $\exp t\xi$ action at $p \in P$ and of $\mathrm{Ad}^*_{\exp(-t\xi)}$ action at $\mu \in Q$, respectively. Then differentiated, the Ad* equivariance (2.33) of J with $g = \exp t\xi$ gives $J_* \xi_P(p) = \xi_Q(\mu)$ with $\mu = J(p)$, where J_* is the tangent map of J. Moreover, we note that $\xi_P(p)$ has a generating function $\hat{J}(\xi)$ defined by $\hat{J}(\xi)(p) = J(p)\cdot\xi$; that is, $\xi_P(p) = X_{\hat{J}(\xi)}$, a Hamiltonian vector field (see Ref. 4, p. 276). The Ad* equivariance also implies that $\{\hat{J}(\xi),\hat{J}(\eta)\} = \hat{J}([\xi,\eta])$ for $\xi,\eta \in \mathrm{su}(1,1)$ (see Ref. 4, p. 281). We are now in a position to write down Ω by the use of the above facts. Since the coadjoint action of SU(1,1) is clearly transitive on $J(M)$, all $\xi_Q(\mu)$ cover the tangent space to $J(M)$ at μ. We then obtain for $\xi,\eta \in \mathrm{su}(1,1)$ and $\mu = J(p)$

$$\begin{aligned} \Omega(\xi_Q,\eta_Q)(\mu) &= \Omega(J_* \xi_P, J_* \eta_P)(J(p)) \\ &= (J^*\Omega)(\xi_P,\eta_P)(p) \\ &= \omega(\xi_P,\eta_P)(p) = \omega(X_{\hat{J}(\xi)}, X_{\hat{J}(\eta)})(p) \\ &= \{\hat{J}(\xi),\hat{J}(\eta)\}(p) = \hat{J}([\xi,\eta])(p) \\ &= J(p)\cdot[\xi,\eta] = \mu\cdot[\xi,\eta]. \end{aligned} \tag{2.39}$$

This shows that Ω is the Kirillov–Kostant–Souriau form (see Ref. 4, p. 281 and Ref. 5, p. 230). The deduction (2.39) is quite the same as that in Ref. 3, but we reproduced it for the sake of consistency.

According to the diagram (2.34), we now pass from su(1,1)* to su(1,1). Let $\xi = \Sigma_{j=1}^{j=3} c_j e_j$ and $\xi' = \Sigma_{j=1}^{j=3} c_j' e_j$ be in su(1,1). We define an indefinite inner product γ on su(1,1) by

$$\gamma(\xi,\xi') = -\mathrm{tr}(\xi\xi') = -c_1c_1' - c_2c_2' + c_3c_3'. \tag{2.40}$$

Thus su(1,1) is identified with the Minkowski space of dimension 3. The vector space isomorphism γ^b of su(1,1) with su(1,1)* is induced by

$$\gamma^b(\xi)\cdot\xi' = \gamma(\xi,\xi'). \tag{2.41}$$

For the basis $\{e_j\}$ and $\{e_j^*\}$ we have

$$\gamma^b(e_1) = -e_1^*, \quad \gamma^b(e_2) = -e_2^*, \quad \gamma^b(e_3) = e_3^*. \tag{2.42}$$

A simple calculation shows that, on setting $\gamma^\# = (\gamma^b)^{-1}$,

$$\gamma^b \circ \mathrm{Ad}_g = \mathrm{Ad}^*_{g^{-1}} \circ \gamma^b, \quad \text{or} \quad \mathrm{Ad}_g \circ \gamma^\# = \gamma^\# \circ \mathrm{Ad}^*_{g^{-1}}. \tag{2.43}$$

We know from (2.43) that $(\gamma^\# \circ J)(M)$ is an orbit of SU(1,1) in the adjoint representation.

Let $T = \mathrm{su}(1,1)$ and $Q = \mathrm{su}(1,1)^*$. Differentiated, Eq. (2.43) with $g = \exp t\xi$ yields

$$\gamma^b(\xi_T) = \xi_Q \circ \gamma^b, \tag{2.44}$$

where ξ_T and ξ_Q are the infinitesimal generators of the adjoint action and of the coadjoint action of $\exp t\xi \in \mathrm{SU}(1,1)$. Here we have used the fact that $(\gamma^b)_* = \gamma_b$, as γ is linear. We note that $\xi_T(\nu) = [\xi,\nu]$ for $\nu \in \mathrm{su}(1,1)$. Now the symplectic form $\tilde{\Omega}$ defined by $\tilde{\Omega} = (\gamma^b)^*\Omega$ on $(\gamma^\# \circ J)(M)$ is expressed, for $\xi,\eta,\nu \in \mathrm{su}(1,1)$, in the form

$$\tilde{\Omega}(\xi_T,\eta_T)(\nu) = \gamma(\nu,[\xi,\eta]), \tag{2.45}$$

which can be easily proved on account of (2.39) and (2.44).

The last stage in tracing the diagram (2.34) is to identify su(1,1) with the Minkowski space of dimension 3. Let $(\mathbb{R}^3,\Gamma)$ be the Minkowski space endowed with the Lorentz metric Γ such that $(\Gamma(f_j,f_k)) = \mathrm{diag}(-1,-1,+1)$ for the standard basis $\{f_j\}$, $j = 1,2,3$, of $\mathbb{R}^3$. The Hodge star operator $*$ with respect to the Lorentz metric Γ is given by

$$*(f_1 \wedge f_2) = f_3, \quad *(f_2 \wedge f_3) = -f_1, \quad *(f_3 \wedge f_1) = -f_2. \tag{2.46}$$

Define a linear map $\lambda: \mathbb{R}^3 \to \mathrm{su}(1,1)$ by

$$\lambda(f_j) = -e_j, \quad j = 1,2,3. \tag{2.47}$$

Then λ is an isometry of $(\mathbb{R}^3,\Gamma)$ with $(\mathrm{su}(1,1),\gamma)$; that is, $\lambda^*\gamma = \Gamma$.

The commutation relations (2.22) and the star operators (2.46) now give, for $j,k = 1,2,3$,

$$[\lambda(f_j),\lambda(f_k)] = \lambda(*(f_j \wedge f_k)), \tag{2.48}$$

which implies that λ is a Lie algebra isomorphism of $(\mathbb{R}^{2,1}, *\circ\wedge)$ with su(1,1), where $\mathbb{R}^{2,1}$ denotes the Minkowski space $(\mathbb{R}^3,\Gamma)$.

Letting $\tilde{J} = \lambda^{-1} \circ \gamma^\# \circ J$, we obtain from (2.32), (2.42), and (2.47)

$$\tilde{J}(p) = V_1 f_1 + V_2 f_2 - V_3 f_3 = (V_1, V_2, -V_3) \in \mathbb{R}^3, \tag{2.49}$$

which shows that $\tilde{J}$ restricted to M is just the pseudo-Hopf map defined by (2.21b). We have $\tilde{J}(M) = H^2_-$, indeed. Since $M_R = M/S^1 \simeq H^2_-$, we can choose $\tilde{J}|_M$ for $\pi: M \to M_R$.

If we choose another basis $\{e_j'\}$ with $e_1' = -e_1$, $e_2' = e_2$, and $e_3' = -e_3$, which satisfies the same commutation relations as $\{e_j\}$ does, we have, in place of (2.32),

$$J(p) = -V_1e_1^* + V_2e_2^* + V_3e_3^* = V_1e_1'^* + V_2e_2'^* - V_3e_3'^*, \tag{2.50}$$

where $\{e_j'^*\}$ is the dual basis to $\{e_j'\}$. Equation (2.49) then becomes

$$J(p) = -V_1f_1 + V_2f_2 + V_3f_3 = V_1f_1' + V_2f_2' - V_3f_3', \tag{2.51}$$

where $\{f_j'\}$ is a basis of $\mathbb{R}^3$ defined by $f_1' = -f_1, f_2' = f_2$, and $f_3' = -f_3$ such that $\lambda(f_j') = -e_j'$, $j = 1,2,3$. In this case, we have $\widetilde{J}(M) = H^2_+$ with respect to the basis $\{f_j\}$. However, $\widetilde{J}(M)$ may be considered as H^2_- with respect to the basis $\{f_j'\}$ on account of (2.51). Hence we do not need to distinguish H^2_+ and H^2_-, and denote one of them by H^2 for short. Thus we have realized M_R in $\mathbb{R}^3$ as one sheet of a two-sheeted two-hyperboloid and have factored the projection $\pi: M \rightarrow M_R$ into

$$\pi = \lambda^{-1} \circ \gamma^{\#} \circ J|_M. \tag{2.52}$$

We here point out that because of (2.11) the one-sheeted two-hyperboloid is not admitted as a realization of M_R.

Accompanying the realization of M_R, the symplectic form ω_R is also realized on H^2. Consider the form $\lambda^* \widetilde{\Omega}$ induced on H^2. Then from the factorization (2.52), we have

$$\pi^*(\lambda^*\widetilde{\Omega}) = J^*(\gamma^{\#})^*(\lambda^{-1})^*\lambda^*\widetilde{\Omega} = J^*\Omega = i^*\omega, \tag{2.53}$$

so that $\omega_R = \lambda^*\widetilde{\Omega}$. Thus we can take $(H^2, \lambda^*\widetilde{\Omega})$ as a model for the reduced phase space (M_R, ω_R).

We now write out ω_R in an explicit form. Let $\lambda(\tilde{\xi}) = \xi$, $\lambda(\tilde{\eta}) = \eta$, and $\lambda(\tilde{\nu}) = \nu$ for ξ, η, ν in su(1,1). A simple computation which will be written in the next paragraph shows that tangent vectors to H^2 at $\tilde{\nu}$ have the form $*(\tilde{\xi} \wedge \tilde{\eta})$, which corresponds to the ordinary vector product in $\mathbb{R}^3$. We note also that $\lambda(*(\tilde{\xi} \wedge \tilde{\eta})) = [\xi, \eta]$, which is a consequence of (2.48). Then from the facts stated above, we have

$$\begin{aligned}
&\omega_R(*(\tilde{\xi} \wedge \tilde{\nu}), *(\tilde{\eta} \wedge \tilde{\nu}))(\tilde{\nu}) \\
&\quad = (\lambda^*\widetilde{\Omega})(*(\tilde{\xi} \wedge \tilde{\nu}), *(\tilde{\eta} \wedge \tilde{\nu}))(\tilde{\nu}) \\
&\quad = \widetilde{\Omega}([\xi,\nu],[\eta,\nu])(\nu) \\
&\quad = \gamma(\nu,[\xi,\eta]) \\
&\quad = \gamma[\lambda(\tilde{\nu}), \lambda(*(\tilde{\xi} \wedge \tilde{\eta}))] \\
&\quad = (\lambda^*\gamma)(\tilde{\nu}, *(\tilde{\xi} \wedge \tilde{\eta})) \\
&\quad = \Gamma(\tilde{\nu}, *(\tilde{\xi} \wedge \tilde{\eta})).
\end{aligned} \tag{2.54}$$

It is convenient for us to get accustomed to vector calculus in the Minkowski space $(\mathbb{R}^3, \Gamma)$. We show that an arbitrary tangent vector to H^2 at $v \in H^2$ is expressed in the form $*(w \wedge v)$ with $w \in \mathbb{R}^3$. Let

$$q(v) = -v_1^2 - v_2^2 + v_3^2. \tag{2.55}$$

Then H^2 is defined by $q(x) = (h/2)^2$. For $w = \Sigma w_j f_j$ and $v = \Sigma v_j f_j$, one has, by definition,

$$*(w \wedge v) = -\begin{vmatrix} w_2 & w_3 \\ v_2 & v_3 \end{vmatrix} f_1 - \begin{vmatrix} w_3 & w_1 \\ v_3 & v_1 \end{vmatrix} f_2 + \begin{vmatrix} w_1 & w_2 \\ v_1 & v_2 \end{vmatrix} f_3. \tag{2.56}$$

It is now clear that grad q and $*(w \wedge v)$ are orthogonal as vectors in the Euclidean space. Thus $*(w \wedge v)$ is tangent to H^2. Further, we give a formula useful in the next section; for $A,B,C \in \mathbb{R}^3$, one has

$$\Gamma(*(A \wedge B), C) = \Gamma(A, *(B \wedge C)), \tag{2.57}$$

which can be proved by a simple calculation.

E. Euler's equations

So far we have obtained $(H^2, \lambda^*\widetilde{\Omega})$ as a model for the reduced phase space (M_R, ω_R). Given a Hamiltonian function H which is invariant under the U(1) action (2.6), one can determine a unique function H_R on M_R such that $H_R \circ \pi = H \circ i$ to obtain the Hamiltonian vector field X_{H_R} on M_R through $i(X_{H_R})\omega_R = dH_R$. Our purpose in this section is to get an explicit form of X_{H_R}.

As was anticipated in Sec. II B and will be proved in Sec. II F, polynomial Hamiltonian functions H_{2k} invariant under U(1) are polynomials in the pseudo-Hopf variables

$$H_{2k}(x_1,x_2,y_1,y_2) = F_k(V_1,V_2,V_3,V_0), \tag{2.58}$$

where F_k are polynomials of degree k in the prescribed variables. Letting (v_j) be the Cartesian coordinates in $\mathbb{R}^3$, we define on $\mathbb{R}^3$

$$F(v) = \sum_{k=1}^{k=n} F_k(v_1,v_2,-v_3,h/2). \tag{2.59}$$

Then the function $H_R = F|_{M_R}$, a restriction of F to $M_R \simeq H^2$, clearly satisfies $H_R \circ \pi = H \circ i$ with $H = \Sigma H_{2k}$.

We now derive the Hamiltonian vector field X_{H_R} by using (2.54) together with $i(X_{H_R})\omega_R = dH_R$. Set $X_{H_R}(v) = *(A \wedge v)$ for $v \in H^2$ and $A \in \mathbb{R}^3$, since $X_{H_R}(v)$ is tangent to H^2. Then for an arbitrary tangent vector $*(B \wedge v)$ to H^2 at v, we have

$$\begin{aligned}
\omega_R(X_{H_R}(v), *(B \wedge v))(v) &= \omega_R(*(A \wedge v), *(B \wedge v))(v) \\
&= \Gamma(v, *(A \wedge B)) \\
&= -\Gamma(*(A \wedge v), B) \\
&= -\Gamma(X_{H_R}(v), B).
\end{aligned} \tag{2.60}$$

Here we have used the formula (2.57). We next obtain, on account of $H_R = F|_{M_R}$,

$$\begin{aligned}
dH_R(v)(*(B \wedge v)) &= \Gamma(\nabla F, *(B \wedge v)) \\
&= -\Gamma(*(\nabla F \wedge v), B),
\end{aligned} \tag{2.61}$$

where ∇F is the gradient of F with respect of the Lorentz metric Γ. From (2.60) and (2.61) we conclude that $X_{H_R}(v) = *(\nabla F \wedge v)$, so that the equation of motion is expressed in the form

$$\frac{dv}{dt} = X_{H_R}(v) = *(\nabla F \wedge v), \quad \text{for } v \in H^2, \tag{2.62}$$

which is Euler's equation in the Minkowski space $\mathbb{R}^{2,1}$. In the ordinary notation of vector calculus in $\mathbb{R}^3$, Eq. (2.62) is put into

$$\frac{dv}{dt} = \frac{1}{2}(\operatorname{grad} F \times \operatorname{grad} q),$$

where grad indicates the gradient in the Euclidean sense. This equation was found by Kummer.[1]

F. Normal form

We now proceed to prove the relation (2.58). Though the proof is the same as that in Ref. 3, we reproduce it for consistency. Recall that $L_{X_2} V_j = 0, j = 0,...,3$, and the relation (2.11) among the pseudo-Hopf variables. Let $P_n(x_1,x_2,y_1,y_2)$ and $Q_m(V_1,V_2,V_3,V_0)$ be the vector spaces of homogeneous polynomials of degree n and m in the specified variables, respectively. Since L_{X_2} is a derivation, we have

$$Q_m(V_1,V_2,V_3,V_0) \subseteq \ker(L_{X_2}|P_{2m}(x_1,x_2,y_1,y_2)). \tag{2.63}$$

We make an attempt to show that equality holds in (2.63). To this end, we compute the dimension of the vector spaces on both sides of (2.63). From the relation (2.11) it follows that

$$Q_m(V_1,V_2,V_3,V_0) = Q_m(V_1,V_2,V_3) + V_0 Q_{m-1}(V_1,V_2,V_3). \tag{2.64}$$

The dimension of the right-hand side of (2.64) is $(m+1)^2$.

The dimension of $\ker(L_{X_2}|P_n)$ is easily computed when the polynomials in P_n are written in the complex variables introduced in (2.3). In fact, since polynomials in P_n then get the form

$$\sum c_{j_1 j_2 k_1 k_2} z_1^{j_1} z_2^{j_2} \bar{z}_1^{k_1} \bar{z}_2^{k_2}, \tag{2.65}$$

with $j_1 + j_2 + k_1 + k_2 = n$ and $c_{j_1 j_2 k_1 k_2} = \bar{c}_{k_1 k_2 j_1 j_2}$, and since X_2 is expressed in the form (2.5), the Lie derivatives of the basis monomials are calculated to be

$$L_{x_2} z_1^{j_1} z_2^{j_2} \bar{z}_1^{k_1} \bar{z}_2^{k_2} = -i(j_1 + j_2 - k_1 - k_2) z_1^{j_1} z_2^{j_2} \bar{z}_1^{k_1} \bar{z}_2^{k_2}, \tag{2.66}$$

so that the dimension of $\ker(L_{X_2}|P_n)$ equals the number of non-negative integer solutions to

$$\begin{aligned} &j_1 + j_2 - k_1 - k_2 = 0, \\ &j_1 + j_2 + k_1 + k_2 = n. \end{aligned} \tag{2.67}$$

For $n = 2m$, the number of non-negative integer solutions to (2.67) is $(m+1)^2$, and hence equality holds in (2.63). An analogous equation to (2.66) and the same equations as (2.67) appeared in Ref. 6 in reducing the quantum harmonic oscillator.

G. $SO_0(1,2)$ as a dynamical group

By a dynamical group we mean a Lie group which acts on the phase space symplectically, and whose Lie algebra has a realization in functions on the phase space under the Poisson bracket such that the Hamiltonian is a function of the generators in the realization Lie algebra. Our aim in this section is to show that $SO_0(1,2)$ is a dynamical group for the Hamiltonian system (M_R, ω_R, H_R).

We start by looking for what group is acting on M_R. So far we have reduced the dynamical system $(\mathbb{R}^4,\omega,H)$ to (M_R,ω_R,H_R). We recall here that SU(1,1) acts on $\mathbb{R}^4$ symplectically, leaving $M \subset \mathbb{R}^4$ invariant, and that SU(1,1) and U(1) commute. Since $M_R = M/S^1$ [$S^1 \simeq U(1)$], SU(1,1) can act on M_R; in fact, the action of SU(1,1) is well defined by $\Phi_g(\pi(x)) = \pi(gx)$ for $g \in SU(1,1)$ and $x \in M$. The Φ_g are symplectic, that is, $\Phi_g^* \omega_R = \omega_R$, because the SU(1,1) leaves ω invariant and because $\pi^* \omega_R = i^* \omega$.

From (2.33), (2.43), and (2.52), we get the Φ_g in the form

$$\Phi_g(\pi(x)) = \lambda^{-1} \circ \mathrm{Ad}_g(\gamma^{\#} \circ J(x)), \tag{2.68}$$

which shows that the action of SU(1,1) on M_R is identified with the adjoint representation of SU(1,1) in its Lie algebra su(1,1). The M_R is, of course, identified with an adjoint orbit of SU(1,1). It follows therefore that $SU(1,1)/\mathbb{Z}_2 \simeq SO_0(1,2)$ acts on M_R, where $\mathbb{Z}_2 = \{I_2, -I_2\}$ and $SO_0(1,2)$ denotes the identity component of the Lorentz group O(1,2). It is also clear from the above that the action of $SO_0(1,2)$ on M_R is symplectic. The action of $SO_0(1,2)$ was pointed out in Ref. 1 without mentioning its symplecticity.

We proceed to look into the $SO_0(1,2)$ action. The Φ_g given by (2.68) can be expressed in the 3×3 matrix form, because M_R is realized as a surface H^2 in $\mathbb{R}^3$ by $(v_j)|_{H^2} = (V_1, V_2, -V_3)$ in the Cartesian coordinates (v_j). For $g = \exp te_j, j = 1,2,3$, we obtain the matrices

$$\Phi_{\exp te_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cosh t & -\sinh t \\ 0 & -\sinh t & \cosh t \end{pmatrix},$$

$$\Phi_{\exp te_2} = \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & 1 & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix},$$

$$\Phi_{\exp te_3} = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.69}$$

Infinitesimal generators of the flows $v \to \Phi_{\exp te_j}(v), j = 1,2,3$, are then given by

$$\begin{aligned} L_1 &= -v_3 \frac{\partial}{\partial v_2} - v_2 \frac{\partial}{\partial v_3}, \\ L_2 &= v_3 \frac{\partial}{\partial v_1} + v_1 \frac{\partial}{\partial v_3}, \\ L_3 &= -v_2 \frac{\partial}{\partial v_1} + v_1 \frac{\partial}{\partial v_2}. \end{aligned} \tag{2.70}$$

The generators L_j, $j = 1,2,3$, are Hamiltonian vector fields, because the $\Phi_{\exp te_j}$ are symplectic. We wish to obtain generating functions (or Hamiltonians) for L_j. To this end, it is of practical use to define functions $(V_j)_R$ on $M_R \simeq H^2$ by $v_j|_{M_R} = (V_j)_R$ for $j = 1,2$, and by $v_3|_{M_R} = -(V_3)_R$. These functions are special cases of the Hamiltonian H_R with $H = V_j$, $j = 1,2,3$ [see (2.58) and (2.59)]. Then we have $(V_j)_R \circ \pi = V_j \circ i$, i being the inclusion map $M \to \mathbb{R}^4$. According to (2.62), the Hamiltonian vector fields $X_{(V_j)_R}, j = 1,2,3$, associated with $(V_j)_R$ are given at $v \in H^2$ by $*(\nabla v_j \wedge v)$ for $j = 1,2$, and by $*(-\nabla v_3 \wedge v)$. After short calculation, we obtain

$$\begin{aligned} X_{(V_1)_R}(v) &= -v_3 f_2 - v_2 f_3, \\ X_{(V_2)_R}(v) &= v_3 f_1 + v_1 f_3, \\ X_{(V_3)_R}(v) &= -v_2 f_1 + v_1 f_2. \end{aligned} \tag{2.71}$$

Equations (2.70) and (2.71) imply that $L_j = X_{(V_j)_R}$, so that $(V_j)_R$ are Hamiltonians for L_j. It deserves mention here that the mapping $e_j \to -L_j$ is a Lie algebra homomorphism.

We are now in a position to prove $SO_0(1,2)$ to be a dynamical group. We are left with a task to show that $(V_j)_R$, $j=1,2,3$, span the Lie algebra so(1,2), for the Hamiltonian H_R is known to be a function of $(V_j)_R$ (Sec. II F). We recall here that the functions V_j, $j=1,2,3$, span the Lie algebra su(1,1)$\simeq$so(1,2) under the Poisson bracket [see (2.10)]. Further we note that $\pi_* X_{V_j}(x) = X_{(V_j)_R}(\pi(x))$. Then by using $i^*\omega = \pi^*\omega_R$, we obtain

$$\begin{aligned}\{V_j,V_k\}\circ i(x) &= \omega(X_{V_j},X_{V_k})\circ i(x)\\ &= (i^*\omega)(X_{V_j},X_{V_k})(i(x))\\ &= (\pi^*\omega_R)(X_{V_j},X_{V_k})(x)\\ &= \omega_R(\pi_*X_{V_j},\pi_*X_{V_k})(\pi(x))\\ &= \omega_R(X_{(V_j)_R},X_{(V_k)_R})(\pi(x))\\ &= \{(V_j)_R,(V_k)_R\}(\pi(x)), \end{aligned} \tag{2.72}$$

so that $(V_j)_R$, $j=1,2,3$, satisfy the so(1,2) commutation relations.

III. REDUCTION BY SO(2)

Rotational-invariant Hamiltonian systems with two degrees of freedom are reduced to Hamiltonian systems with one degree of freedom, which give systems on the positive real line ($r>0$). This is an application of Routh's procedure for cyclic coordinates.[7] We will shed new light on this well-known procedure.

Let (x_1,x_2,y_1,y_2) be the Cartesian coordinates of the phase space $\mathbb{R}^4$ and ω the standard symplectic form given by (2.1). Rotations in $\mathbb{R}^2$, the x space, lift to $\mathbb{R}^4$; that is, SO(2) acts on $\mathbb{R}^4$ through

$$\rho(t) = \begin{pmatrix} \exp tN & 0 \\ 0 & \exp tN \end{pmatrix}, \quad \text{with } N = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.1}$$

The action of SO(2) is clearly symplectic. The angular momentum is the very momentum map associated with the SO(2). We denote it by $2W_0$:

$$W_0 = \tfrac{1}{2}\langle y, Nx\rangle, \tag{3.2}$$

where $\langle\ ,\ \rangle$ is the standard inner product on $\mathbb{R}^2$, and x and y are vectors in $\mathbb{R}^2$.

Let us write SL(2,$\mathbb{R}$) in the 4×4 matrix form

$$\begin{pmatrix} aI_2 & bI_2 \\ cI_2 & dI_2 \end{pmatrix}, \quad \text{with } ad-bc=1. \tag{3.3}$$

The SL(2,$\mathbb{R}$) is commutative with the SO(2) given in (3.1), and acts on $\mathbb{R}^4$ symplectically. We will compute the momentum map associated with the SL(2,$\mathbb{R}$). We take a basis $\{e_j''\}$, $j=1,2,3$, of the Lie algebra sl(2,$\mathbb{R}$) so that any element $\xi = \Sigma a_j e_j''$ may be expressed as

$$\frac{1}{2}\begin{pmatrix} a_2 & 0 & a_1+a_3 & 0 \\ 0 & a_2 & 0 & a_1+a_3 \\ a_1-a_3 & 0 & -a_2 & 0 \\ 0 & a_1-a_3 & 0 & -a_2 \end{pmatrix}. \tag{3.4}$$

Then the commutation relations among the e_j'' are

$$[e_1'',e_2''] = -e_3'', \quad [e_2'',e_3''] = e_1'', \quad [e_3'',e_1''] = e_2''. \tag{3.5}$$

In a similar manner to that in Sec. II C, we can find a generating function of the infinitesimal symplectic transformation corresponding to (3.4). In fact, for $\xi = \Sigma a_j e_j''$ and $p = \binom{x}{y} \in \mathbb{R}^4$ we have a generating function

$$\tfrac{1}{2}\omega(\xi p, p) = a_1W_1 + a_2W_2 + a_3W_3, \tag{3.6}$$

where

$$\begin{aligned} W_1 &= \tfrac{1}{4}(\langle y,y\rangle - \langle x,x\rangle),\\ W_2 &= \tfrac{1}{2}\langle x,y\rangle,\\ W_3 &= \tfrac{1}{4}(\langle y,y\rangle + \langle x,x\rangle). \end{aligned} \tag{3.7}$$

We note here that W_j, $j=1,2,3$, are well known to satisfy the commutation relations

$$\{W_1,W_2\} = -W_3, \quad \{W_2,W_3\} = W_1, \quad \{W_3,W_1\} = W_2, \tag{3.8}$$

and that W_j's are related by

$$-W_1^2 - W_2^2 + W_3^2 = W_0^2, \tag{3.9}$$

which are similar to (2.10) and (2.11). We also remark that W_j, $j=0,\ldots,3$, are all invariant under the SO(2) action. The momentum map $J:\mathbb{R}^4 \to \mathrm{sl}(2,\mathbb{R})^*$ is then defined by

$$J(p)\cdot\xi = \tfrac{1}{2}\omega(\xi p, p). \tag{3.10}$$

The Ad* equivariance of J can be proved in the same manner as in (2.30).

We now consider the momentum manifold M determined by $2W_0 = \langle y, Nx\rangle = l$, l being a nonzero constant. Since for any x in $\mathbb{R}^2 - \{0\}$, the equation $2W_0 = l$ determines a linear equation in $y = \binom{y_1}{y_2}$, M is diffeomorphic to $(\mathbb{R}^2 - \{0\})\times\mathbb{R}$. We will get a concrete idea of the topology of M in the next section. Now M admits the action of SO(2) because W_0 is invariant under the SO(2) action (3.1). Therefore, one can get the orbit manifold M/S^1, which we denote by M_R, the same notation as in Sec. II.

We wish to realize the M_R to be a surface in $\mathbb{R}^3$. For this purpose, like (2.12b), we consider the map of M to $\mathbb{R}^3$

$$\kappa:(x,y) \mapsto (W_1, W_2, -W_3). \tag{3.11}$$

For the same reason as in Sec. II D [see (2.51)] we do not need to consider the map corresponding to (2.12a). From Eq. (3.9) with $W_0 = l/2$ it follows that $\kappa(M)$ is one sheet of a two-sheeted two-hyperboloid in $\mathbb{R}^3$, which we denote by H^2, the same notation as in Sec. II [$W_0 = l/2$ will be seen to be equivalent to $H_2 = h = l$ in (4.10)]. Since W_j, $j=1,2,3$, are invariant under SO(2), the inverse image $\kappa^{-1}(w)$ of $w \in H^2$ must be an invariant manifold for SO(2) which is diffeomorphic to S^1. Therefore we conclude that M/S^1 is realized as a surface H^2 in $\mathbb{R}^3$, and that the map κ is identified with the natural projection $\pi:M\to M_R$. The projection π determines a unique symplectic form ω_R on M_R by $\pi^*\omega_R = i^*\omega$, where $i:M\to\mathbb{R}^4$ is the inclusion map. Thus we have a reduced symplectic manifold (M_R,ω_R) as a surface in $\mathbb{R}^3$.

We turn to describing the reduction in the polar coordinates (r,θ) and their conjugate momentum variables (p_r, p_θ). We have, as is well known,

$$x_1 = r\cos\theta,\quad y_1 = p_r\cos\theta - [\sin\theta/r]\,p_\theta,$$
$$x_2 = r\sin\theta,\quad y_2 = p_r\sin\theta + [\cos\theta/r]\,p_\theta, \tag{3.12}$$

and

$$rp_r = \langle y,x\rangle,\quad p_\theta = \langle y,Nx\rangle. \tag{3.13}$$

The momentum manifold M is then given by $p_\theta = l$, so that (r,θ,p_r) is a local coordinate system for M. The action of SO(2) is described in terms of (r,θ,p_r) as $\theta\to\theta+t$ with r and p_r fixed. The orbit manifold $M/S^1\simeq H^2$ then has the coordinates (r,p_r), and can be described by the equations

$$\begin{aligned} w_1 &= \tfrac14(p_r^2 + l^2/r^2 - r^2),\\ w_2 &= \tfrac12 rp_r,\\ -w_3 &= \tfrac14(p_r^2 + l^2/r^2 + r^2),\end{aligned} \tag{3.14}$$

where (w_j) are the Cartesian coordinates in $\mathbb{R}^3$. Equations (3.14) result from (3.7) and (3.11)–(3.13), as is easily verified.

The symplectic forms ω and ω_R are expressed as

$$\omega = dr\wedge dp_r + d\theta\wedge dp_\theta,\quad \omega_R = dr\wedge dp_r, \tag{3.15}$$

respectively. It is clear that the relation $\pi^*\omega_R = i^*\omega$ holds for the above forms.

In the same manner as in Sec. II D, we can factor the projection $\pi:M\to M/S^1$. We first prove that $J(M)$ is a coadjoint orbit of SL(2,$\mathbb{R}$) in sl(2,$\mathbb{R}$)*. To this end, we show that SL(2,$\mathbb{R}$) acts transitively on M. Our discussion is broken up into two parts, according to whether l is positive or negative. First we take the case where l is positive. Then for a fixed point $(\sqrt{l},0,0,\sqrt{l})^T$ and an arbitrary point $(x_1,x_2,y_1,y_2)^T$ of M we have

$$\frac{1}{\sqrt{l}}\begin{pmatrix} x_1 & 0 & x_2 & 0\\ 0 & x_1 & 0 & x_2\\ y_1 & 0 & y_2 & 0\\ 0 & y_1 & 0 & y_2\end{pmatrix}\begin{pmatrix}\sqrt{l}\\0\\0\\\sqrt{l}\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\y_1\\y_2\end{pmatrix}. \tag{3.16}$$

When l is negative, we have for $(\sqrt{|l|},0,0,-\sqrt{|l|})^T$

$$\frac{1}{\sqrt{|l|}}\begin{pmatrix} x_1 & 0 & -x_2 & 0\\ 0 & x_1 & 0 & -x_2\\ y_1 & 0 & -y_2 & 0\\ 0 & y_1 & 0 & -y_2\end{pmatrix}\begin{pmatrix}\sqrt{|l|}\\0\\0\\-\sqrt{|l|}\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\y_1\\y_2\end{pmatrix}. \tag{3.17}$$

Equations (3.16) and (3.17) mean that SL(2,$\mathbb{R}$) is transitive on M. Therefore we obtain, by using the Ad* equivariance of J,

$$\begin{aligned} J(M) &= \{J(gp_0)\,|\,g\in \mathrm{SL}(2,\mathbb{R})\}\\ &= \{\mathrm{Ad}^*_{g^{-1}}\circ J(p_0)\,|\,g\in\mathrm{SL}(2,\mathbb{R})\},\end{aligned} \tag{3.18}$$

where $p_0 = (\sqrt{l},0,0,\sqrt{l})^T$ for l positive or $p_0 = (\sqrt{|l|},0,0,-\sqrt{|l|})^T$ for l negative. Thus $J(M)$ turns out to be a coadjoint orbit in sl(2,$\mathbb{R}$)*.

Since sl(2,$\mathbb{R}$) and su(1,1) are isomorphic Lie algebras [see (2.23) and (3.5)], the same reasoning as that done in Sec. II D can go through to put $J(M)$ into an adjoint orbit in sl(2,$\mathbb{R}$). Further, sl(2,$\mathbb{R}$) can be identified with $\mathbb{R}^3$, and hence we obtain eventually the diagram

$$\begin{tikzcd} M \arrow[r] \arrow[d] & \mathrm{sl}(2,\mathbb{R})^* \arrow[d] \\ M/S^1\simeq H^2 & \mathrm{sl}(2,\mathbb{R}) \arrow[l] \end{tikzcd}, \tag{3.19}$$

which gives a decomposition of the natural projection $\tau:M\to M_R$.

If a Hamiltonian H is given which is invariant under the SO(2) action, the reduced Hamiltonian H_R is determined by $H_R\circ\pi = H\circ i$. Then we have a reduced Hamiltonian system (M_R,ω_R,H_R) realized on a two-hyperboloid in $\mathbb{R}^3$. We will return to this system in the latter part of the next section.

IV. EQUIVALENCE OF THE REDUCTIONS

The reductions performed in Secs. II and III are associated with the group SU(1,1)×U(1) and SL(2,$\mathbb{R}$)×SO(2), respectively. We note here that these groups are isomorphic under the isomorphism given by

$$S = \frac{1}{\sqrt2}\begin{pmatrix}1 & 0 & 0 & 1\\ 1 & 0 & 0 & -1\\ 0 & -1 & 1 & 0\\ 0 & 1 & 1 & 0\end{pmatrix}. \tag{4.1}$$

In fact, for $g\in\mathrm{SL}(2,\mathbb{R})$ given by (3.3), we obtain

$$SgS^{-1} = \frac12\begin{pmatrix} a+d & a-d & b-c & b+c\\ a-d & a+d & b+c & b-c\\ -b+c & b+c & a+d & -a+d\\ b+c & -b+c & -a+d & a+d\end{pmatrix}, \tag{4.2}$$

which belongs to (2.21a), the elements of SU(1,1), together with

$$A = \frac12\begin{pmatrix}a+d & a-d\\ a-d & a+d\end{pmatrix},\quad B = \frac12\begin{pmatrix}-b+c & b+c\\ -b-c & b-c\end{pmatrix}, \tag{4.3}$$

and for $\rho(t)\in\mathrm{SO}(2)$ given by (3.1),

$$S\rho(t)S^{-1} = \begin{pmatrix}\cos t & 0 & \sin t & 0\\ 0 & \cos t & 0 & -\sin t\\ -\sin t & 0 & \cos t & 0\\ 0 & \sin t & 0 & \cos t\end{pmatrix}, \tag{4.4}$$

which is nothing but the matrix (2.8), the elements of U(1). We denote the right-hand side of Eq. (4.4) by $\Phi(t)$.

The isomorphism of SU(1,1)×U(1) with SL(2,$\mathbb{R}$)×SO(2) gives rise to the Lie algebra isomorphism of su(1,1) + u(1) with sl(2,$\mathbb{R}$) + so(2). We have, indeed, for $\xi\in\mathrm{sl}(2,\mathbb{R})$ given by (3.4)

$$S\xi S^{-1} = \begin{pmatrix}0 & a_2 & a_3 & a_1\\ a_2 & 0 & a_1 & a_3\\ -a_3 & a_1 & 0 & -a_2\\ a_1 & -a_3 & -a_2 & 0\end{pmatrix}, \tag{4.5}$$

which is an element of su(1,1) given by (2.22) with $c_j = a_j$, $j=1,2,3$, and $c_0=0$. As for the isomorphism u(1)≃so(2), differentiated with respect to t, Eq. (4.4) yields

$$S\dot\rho(0)S^{-1} = \begin{pmatrix}0 & G\\ -G & 0\end{pmatrix}, \tag{4.6}$$

which is a basis of u(1) given by (2.22) with $c_1=c_2=c_3=0$ and $c_0=1$.

Furthermore, the isomorphism S is symplectic. It is, in effect, easy to verify that for $p' = Sp$ with $p' = \binom{x'}{y'}$ and $p = \binom{x}{y}$ one has $\Sigma dx'_j \wedge dy'_j = \Sigma dx_j \wedge dy_j$; that is,

$$\omega(Sp,Sq) = \omega(p,q), \quad \text{for } p,q \in \mathbb{R}^4. \tag{4.7}$$

We are now in a position to show that the momentum maps associated with SU(1,1) and with SL(2,$\mathbb{R}$) are equivalent. Let $\xi \in \mathrm{sl}(2,\mathbb{R})$ be given by (3.4). Then by using (4.7), we obtain for $p' = Sp$

$$\tfrac{1}{2}\omega(\xi p, p) = \tfrac{1}{2}\omega(S\xi S^{-1}p', p'). \tag{4.8}$$

The left-hand side equals the pairing of $\xi \in \mathrm{sl}(2,\mathbb{R})$ and the momentum map associated with SL(2,$\mathbb{R}$) [see (3.10)], and the right-hand side the pairing of $S\xi S^{-1} \in \mathrm{su}(1,1)$ and the momentum map associated with SU(1,1). Thus Eq. (4.8) means the equivalence of the momentum maps. From (2.27) with $c_0 = 0$ and (3.6) with $a_j = c_j$, Eq. (4.8) implies also that

$$W_1(p) = V_1(p'), \quad W_2(p) = V_2(p'), \quad W_3(p) = V_3(p'). \tag{4.9}$$

As for W_0 and V_0, the momenta associated with U(1) and SO(2), respectively, we have

$$W_0(p) = V_0(p') = \tfrac{1}{2}H_2(p'). \tag{4.10}$$

This equation shows that the level surface $H_2 = h$ in Sec. II is diffeomorphic with the momentum manifold $W_0 = l/2$ in Sec. III, so that the latter is the three-hyperboloid H^3.

Thus we have proved that the reductions from $(\mathbb{R}^4,\omega)$ to (M_R,ω_R), performed in Secs. II and III, are equivalent; that is, in each case, the momentum manifold M is diffeomorphic to a three-hyperboloid, and the orbit manifold M_R for the S^1 action $[S^1 \simeq \mathrm{U}(1) \simeq \mathrm{SO}(2)]$ is realized as one sheet of a two-sheeted two-hyperboloid in $\mathbb{R}^3$, which is identified with an adjoint orbit of the group $\mathrm{SU}(1,1) \simeq \mathrm{SL}(2,\mathbb{R})$ in $\mathrm{su}(1,1) \simeq \mathrm{sl}(2,\mathbb{R})$. Of course, the reduced symplectic form ω_R is unique by the reduction, so that it is the same in each case.

On having made clear the equivalence of the reductions, we proceed to a detailed discussion on SO(2)-invariant Hamiltonian systems which was skipped over in Sec. III. Let K be an SO(2)-invariant Hamiltonian, which is assumed to be a finite sum of even number degree homogeneous polynomials in (x_1,x_2,y_1,y_2); $K = \Sigma K_{2k}$. Then by the linear symplectic mapping S defined by (4.1), K is transformed into a U(1)-invariant Hamiltonian H determined by $K = H \circ S$. In fact, for the Hamiltonian K such that $K \circ \rho(t) = K$, one has $H \circ \Phi(t) = H$ on account of Eq. (4.4). The Hamiltonian H is, of course, a finite sum of even number degree homogeneous polynomials; $H = \Sigma H_{2k}$. According to the result in Sec. II F, the H, U(1)-invariant, is a polynomial function of V_j, $j = 0,\dots,3$. Hence K becomes a polynomial function of W_j, $j = 0,\dots,3$, owing to Eqs. (4.9) and (4.10). Thus we come to the conclusion that polynomial Hamiltonians $K = \Sigma K_{2k}$ invariant under SO(2) are polynomials in the variables W_j, $j = 0,\dots,3$. The reduced Hamiltonian K_R on M_R, determined by $K_R \circ \pi = K \circ i$, is therefore a polynomial function in $(W_j)_R$, $j = 1,2,3$, where the $(W_j)_R$ are determined by $(W_j)_R \circ \pi = W_j \circ i$ and equal to the right-hand sides of Eq. (3.14).

Then, Hamilton's equations for the reduced system (M_R,ω_R,H_R), which are usually described in terms of (r,p_r), can be expressed as Euler's equations in a similar manner to Eq. (2.62).

We conclude this section with a remark on a dynamical group. We have shown in Sec. II G that $\mathrm{SO}_0(1,2)$ is a dynamical group for the reduced Hamiltonian system (M_R,ω_R,H_R). For the same reason as in Sec. II G, $\mathrm{SO}_0(1,2)$ is also a dynamical group for the reduced Hamiltonian system (M_R,ω_R,K_R). If $K = 2W_3$, $2K_R$ becomes equal to $2(W_3)_R$. We may call $2(W_3)_R$, $(W_3)_R$ being given in the right-hand side of (3.14), the Hamiltonian for the radial harmonic oscillator. If we choose $K = W_1 + W_3$, we obtain $K_R = \tfrac{1}{2}(p_r^2 + l^2/r^2)$, a radial free particle Hamiltonian. If $K = 2W_1$, K_R becomes a radial repulsive oscillator Hamiltonian. Therefore, $\mathrm{SO}_0(1,2)$ may be called a dynamical group for the radial harmonic oscillator, a radial free particle, or the radial repulsive oscillator.

Addendum: After this manuscript was completed, the author's attention was drawn to Ref. 8 (MR 84g #58043) in which part of this article (Euler's equation) was discussed in a different manner. Interest in this article, however, centers on the dynamical group $\mathrm{SO}_0(1,2) \cong \mathrm{SL}(2,\mathbb{R})/\mathbb{Z}_2$. Further, the reduction performed in this article will be shown in a future paper to have a quantum analog; that is, a realization of SL(2,$\mathbb{R}$) in harmonic oscillator annihilation and creation operators on $L^2(\mathbb{R}^2)$ proves to be reducible by the S^1 action to unitary irreducible representations of SL(2,$\mathbb{R}$) on $L^2(\mathbb{R}^+;r\,dr)$.

[1]M. Kummer, Commun. Math. Phys. **48**, 54 (1976); *ibid.* **58**, 85 (1978); *Classical Mechanics and Dynamical Systems*, edited by R. L. Devaney and Z. H. Nitecki (Dekker, New York, 1981), pp. 91–107.
[2]R. C. Churchill, M. Kummer, and D. L. Rod, J. Differential Eq. **49**, 359 (1983).
[3]R. Cushman and D. L. Rod, Physica D **6**, 105 (1982).
[4]R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin–Cummings, Reading, MA, 1978), 2nd ed.
[5]A. A. Kirillov, *Elements of the Theory of Representations* (Springer, New York, 1976).
[6]T. Iwai, J. Math. Phys. **23**, 1088 (1982).
[7]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1950).
[8]R. Cushman, Hadronic J. **5**, 2109 (1982).

A pair of commuting scalars for $G(2) \supset SU(2)\times SU(2)$

J. W. B. Hughes and J. Van der Jeugt[a),b)]
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 25 May 1984; accepted for publication 28 September 1984)

$G(2) \supset SU(2)\times SU(2)$ is a two-missing-labels problem, and therefore in order to give a complete and orthogonal specification of states of irreducible representations of $G(2)$ in an $SU(2)\times SU(2)$ basis, one needs to find a pair of commuting Hermitian operators which are scalar with respect to the $SU(2)\times SU(2)$ subalgebra. A theorem due to Peccia and Sharp states that there are, apart from the Lie algebra invariants, twice as many functionally independent scalars as missing labels. Here two commuting $SU(2)\times SU(2)$ scalars are obtained, both of sixth order in the $G(2)$ basis elements. They are in fact combinations of five scalars of different tensorial types, indicating that the functionally independent ones are in general insufficient to provide the lowest-order commuting scalars. An expression for the sixth-order invariant of $G(2)$ is also obtained.

I. INTRODUCTION

The problem of classifying states of irreducible representations (IR) of a Lie algebra G with respect to a subalgebra H is one which has been considered by many authors in recent years. In any physical application of a symmetry algebra G it is desirable to obtain a set of basis states which are orthogonal and which are uniquely labeled by suitably chosen parameters. This can be achieved by choosing the parameters to be the eigenvalues of a complete set of commuting Hermitian operators, and this is partially achieved by choosing the invariants of G and the subalgebra H together with the internal state labeling operators of H. If no degeneracies occur, i.e., if each IR of H occurs with multiplicity 1 in the reducible representation of H obtained when the IR under consideration of G is restricted to H, then the above set of operators will be complete and a unique sepcification of the basis states will have been achieved. Such is the case, for instance, for $SO(5) \supset SO(4)$ (see Refs. 1 and 2).

In general, however, degeneracies do occur and further labeling operators need to be found. The most convenient method of obtaining these is to find operators constructed from the enveloping algebra of G which are scalar with respect to H, i.e., which commute with all elements of H. Finally, if $G \supset H$ is an r-missing-labels problem, one tries to find a subset of r mutually commuting scalars.

A theorem due to Peccia and Sharp[3] states that the number of functionally independent scalars for the case where G and H are semisimple Lie algebras is exactly double the number of missing labels. Thus for the simplest state labeling problem considered, namely the one-missing-label problem $SU(3) \supset SO(3)$, there are two functionally independent scalars, either of which can be chosen to provide the missing label. The problem of determining their eigenvalues was first solved by Hughes[4,5] and by Judd *et al.*[6] Hughes used shift operator techniques developed by Hughes and Yadegar.[7] These techniques have been considerably refined lately and reapplied to $SU(3) \supset SO(3)$ by De Meyer *et al.*[8–10]

The simplest, and most exhaustively considered, two-missing-labels problem is that of $SU(4) \supset SU(2)\times SU(2)$. Here one needs to find a pair of independent commuting $SU(2)\times SU(2)$ scalars, and this was first accomplished by Moshinsky and Nagel.[11] Tables listing their numerical eigenvalues were set up by Quesne[12] for low-dimensional IR of SU(4), and Van der Jeugt *et al.*[13] used shift operator techniques to obtain general expressions for the eigenvalues of these operators. An independent set of commuting operators has also been constructed by Quesne[14] and Partensky and Maguin.[15] The commuting scalars for this problem were of third or fourth degree in the basis elements of SU(4).

More recently the problem of obtaining missing-label operators for $O(p) \supset O(p-2)\times O(2)$ has been considered by Bincer,[16] and Van der Jeugt[17] has discussed a procedure for obtaining a pair of commuting scalars for $G \supset [SU(2)]^n$, although his method is not immediately applicable to the case where the scalars are of greater than second degree in the elements of $G - [SU(2)]^n$.

The method employed by some of these authors for obtaining the scalars is to obtain an integrity basis for $G \supset H$, i.e., a finite number of elementary H tensors, in terms of which all others may be expressed as stretched products.[6,18] For $SU(4) \supset SU(2)\times SU(2)$, let the $SU(2)\times SU(2)$ basis elements be denoted by $S_i, T_i, i = 0, \pm 1$, and the remaining basis elements, which form a nine-dimensional irreducible $SU(2)\times SU(2)$ tensor representation, by Q_{ij}, $i,j = 0, \pm 1$. An $SU(2)\times SU(2)$ scalar of degree s,t,q, respectively, in the S_i, T_i, and $Q_{i,j}$ is denoted by $C^{(stq)}$. An integrity basis consists, apart from the three SU(4) and two $SU(2)\times SU(2)$ invariants, of a further seven scalars of type $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, $C^{(112)}$, and $C^{(113)}$, $C^{(204)}$, $C^{(024)}$. Since the problem is a two-missing-labels one, according to the theorem of Peccia and Sharp[3] there exist from this set only four functionally independent scalars, and Quesne[12] showed that they could be chosen to be the ones of type $C^{(112)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$. Finally, from these four one can find two independent pairs of commuting operators namely $\{C^{(202)}, C^{(022)}\}$, and the pair $\{C^{(111)}, C^{(202)} + C^{(022)} - C^{(112)}\}$ originally found by Moshinsky and Nagel. Note that none of these is of greater than fourth overall degree, nor of greater than second degree in the tensor components $Q_{i,j}$.

In this paper we consider the two-missing-labels problem $G(2) \supset SU(2)\times SU(2)$. We denote the $SU(2)\times SU(2)$ basis elements by j_i, k_i and the remaining eight elements,

---

[a)] Research Assistant N.F.W.O., Belgium.
[b)] Permanent address: Seminarie voor Wiskundige Natuurkunde R.U.G., Krijgslaan 281-S9, B-9000 Gent, Belgium.

 0022-2488/85/050894-07$02.50

which form an irreducible $SU(2)\times SU(2)$ tensor, by $R_{\mu\nu}$, $\mu = \pm\frac{3}{2}, \pm\frac{1}{2}$ and $\nu = \pm\frac{1}{2}$. For $G(2)$, apart from the second- and sixth-order invariants and the two $SU(2)\times SU(2)$ invariants, the scalars of degree not greater than 6 are[18] $C^{(202)}$, $C^{(112)}$, $C^{(312)}$, $C^{(204)}$, $C^{(024)}$, and $C^{(114)}$. Peccia and Sharp's theorem tells us for this case that four of these are functionally independent; these may be chosen as $C^{(202)}$, $C^{(112)}$, $C^{(312)}$, and $C^{(024)}$. However, as we shall see, in order to construct a commuting pair of operators, we shall need in addition to these four the scalar $C^{(204)}$. We do not need $C^{(114)}$, but we cannot rule out the possibility of an independent pair of commuting scalars of sixth degree which includes it. Nor can we exclude the possibility that a pair of commuting scalars of higher than sixth degree can be constructed from the four functionally independent ones, but this paper does show that if one wants the lowest-degree commuting scalars then the functionally independent scalars are not always sufficient.

In Sec. II we give the commutation relations for $G(2)$ and the method of constructing irreducible representations of $SU(2)\times SU(2)$ of higher than the first degree in the $R_{\mu\nu}$ and the construction from them of the corresponding scalar operators. In Sec. III we construct explicitly the scalars of the above types together with $C^{(006)}$, and in Sec. IV we give the formulas in terms of them for the Hermitian commuting scalars Y_1 and Y_2, together with a brief description of the computer method employed in finding them. For completeness we also give the form of the sixth-order invariant, which does not appear to have been given in the literature previously. We give Y_1, Y_2 and I_6 explicitly in terms of the basis elements of $G(2)$ in Tables I, II, and III.

No attempt is made in this paper to obtain eigenvalues of Y_1 and Y_2, as was done in the case of $SU(4) \supset SU(2)\times SU(2)$ by Van der Jeugt *et al.*,[13] but it is intended to use shift operator techniques to tackle this problem in a later paper.

II. THE LIE ALGEBRA $G(2)$

A basis for the Lie algebra $G(2)$ may be chosen to consist of the generators $(j_0, j_\pm)$, $(k_0, k_\pm)$ of the subalgebra $SU_j(2)\times SU_k(2)$ together with an eight-dimensional irreducible tensor representation of $SU_j(2)\times SU_k(2)$, which we denote by $R_{\mu\nu}$, $\mu = \pm\frac{3}{2}, \pm\frac{1}{2}$, $\nu = \pm\frac{1}{2}$. These satisfy the commutation relations

$$\begin{aligned}
&[j_0, j_\pm] = \pm j_\pm, \quad [j_+, j_-] = 2j_0,\\
&[k_0, k_\pm] = \pm k_\pm, \quad [k_+, k_-] = 2k_0,
\end{aligned} \tag{2.1}$$

$$\begin{aligned}
&[j_0, R_{\mu\nu}] = \mu R_{\mu\nu}, \quad [j_\pm, R_{\mp 1/2,\nu}] = 2R_{\pm 1/2,\nu},\\
&[j_\pm, R_{\pm 1/2,\nu}] = \sqrt{3} R_{\pm 3/2,\nu},\\
&[j_\pm, R_{\mp 3/2,\nu}] = \sqrt{3} R_{\mp 1/2,\nu},\\
&[k_0, R_{\mu\nu}] = \nu R_{\mu\nu}, \quad [k_\pm, R_{\mu,\mp 1/2}] = R_{\mu,\pm 1/2},
\end{aligned} \tag{2.2}$$

together with the mutual commutation relations of the $R_{\mu\nu}$

$$\begin{aligned}
&[R_{3/2,1/2}, R_{-1/2,-1/2}] = -(1/2\sqrt{3}) j_+,\\
&[R_{3/2,-1/2}, R_{-1/2,1/2}] = (1/2\sqrt{3}) j_+,\\
&[R_{1/2,1/2}, R_{1/2,-1/2}] = \tfrac{1}{3} j_+,\\
&[R_{-3/2,-1/2}, R_{1/2,1/2}] = -(1/2\sqrt{3}) j_-,\\
&[R_{-3/2,1/2}, R_{1/2,-1/2}] = (1/2\sqrt{3}) j_-,\\
&[R_{-1/2,1/2}, R_{-1/2,-1/2}] = -\tfrac{1}{3} j_-,\\
&[R_{3/2,1/2}, R_{-3/2,1/2}] = -\tfrac{1}{2} k_+,\\
&[R_{1/2,1/2}, R_{-1/2,1/2}] = \tfrac{1}{2} k_+,\\
&[R_{3/2,-1/2}, R_{-3/2,-1/2}] = \tfrac{1}{2} k_-,\\
&[R_{1/2,-1/2}, R_{-1/2,-1/2}] = -\tfrac{1}{2} k_-,\\
&[R_{3/2,1/2}, R_{-3/2,-1/2}] = \tfrac{1}{2}(j_0 + k_0),\\
&[R_{3/2,-1/2}, R_{-3/2,1/2}] = \tfrac{1}{2}(-j_0 + k_0),\\
&[R_{1/2,1/2}, R_{-1/2,-1/2}] = -\tfrac{1}{6}(j_0 + 3k_0),\\
&[R_{1/2,-1/2}, R_{-1/2,1/2}] = \tfrac{1}{6}(j_0 - 3k_0).
\end{aligned} \tag{2.3}$$

All commutators not given in the above equations vanish.

The Hermiticity conditions satisfied by the basis elements in order that representations exponentiate to unitary representations of the Lie group are

$$j_0^\dagger = j_0, \quad j_\pm^\dagger = j_\mp, \quad k_0^\dagger = k_0, \quad k_\pm^\dagger = k_\mp, \tag{2.4}$$

$$R_{\mu\nu}^\dagger = (-1)^{\mu+\nu} R_{-\mu-\nu}. \tag{2.5}$$

All tensor operators in this paper will in fact satisfy the Hermiticity conditions (2.5) with R replaced by the tensor under consideration and μ,ν in the corresponding ranges.

The invariants of $SU_j(2)\times SU_k(2)$ are

$$J^2 = j_+ j_- + j_0^2 - j_0, \quad K^2 = k_+ k_- + k_0^2 - k_0 \tag{2.6}$$

and the second-order invariant of $G(2)$ is

$$\begin{aligned}
I_2 = {}& R_{3/2,1/2} R_{-3/2,-1/2} - R_{3/2,-1/2} R_{-3/2,1/2}\\
&- R_{1/2,1/2} R_{-1/2,-1/2} + R_{1/2,-1/2} R_{-1/2,1/2}\\
&+ \tfrac{1}{6} J^2 + \tfrac{1}{2} K^2 - \tfrac{2}{3} j_0.
\end{aligned} \tag{2.7}$$

The expression for the sixth-order invariant I_6 contains 730 terms and is given in Table III. Its derivation is discussed in Sec. IV.

In the following section we shall obtain the tensor operators from which we construct the scalar operators needed to find the commuting scalars. The method used to obtain them is as follows: Suppose $T^{[C,D]}_{\gamma,\sigma}$ and $\overline{T}^{[\overline{C},\overline{D}]}_{\bar{\gamma},\bar{\sigma}}$ are, respectively, $(2C+1)(2D+1)$- and $(2\overline{C}+1)(2\overline{D}+1)$-dimensional irreducible tensor representations of $SU_j(2)\times SU_k(2)$. Then we construct the irreducible $(2j+1)(2k+1)$-dimensional Kronecker product tensor representation $(T\overline{T})^{[j,k]}_{\alpha,\beta}$, by means of the formula[19]

$$\begin{aligned}
(T\overline{T})^{[j,k]}_{\alpha,\beta} = {}& \sum_{\substack{\gamma,\bar{\gamma}\\ \delta,\bar{\delta}}} \sqrt{(2j+1)(2k+1)}\,(-1)^{\alpha+\beta}\\
&\times \begin{pmatrix} C & \overline{C} & j\\ \gamma & \bar{\gamma} & -\alpha \end{pmatrix} \begin{pmatrix} D & \overline{D} & k\\ \delta & \bar{\delta} & -\beta \end{pmatrix}\\
&\times T^{[C,D]}_{\gamma,\delta} \overline{T}^{[\overline{C},\overline{D}]}_{\bar{\gamma},\bar{\delta}},
\end{aligned} \tag{2.8}$$

where $\left(\begin{smallmatrix} C & \overline{C} & j\\ \gamma & \bar{\gamma} & -\alpha \end{smallmatrix}\right)$ is a 3-j symbol. Provided $T^{[C,D]}_{\gamma,\delta}$ and $\overline{T}^{[\overline{C},\overline{D}]}_{\bar{\gamma},\bar{\delta}}$ satisfy Hermiticity conditions of type (2.5), then so will the $(T\overline{T})^{[j,k]}_{\alpha,\beta}$. Given a $(2j+1)(2k+1)$-dimensional tensor $(R^r)^{[j,k]}_{\alpha,\beta}$, $\alpha = -j, \dots, j$, $\beta = -k, \dots, k$, of degree r in the $R_{\mu\nu}$, then, provided j and k are both integral, one may construct a scalar $C^{[j,k,r]}$ of degree $[j,k,r]$ in the j_i, k_i, $R_{\mu\nu}$ by adapting to the case of $SU_j(2)\times SU_k(2)$ a formula first given by Hughes and Yadegar[7] for the case of an irreducible tensor representation of a single SU(2). In that paper a formula was given for operators which shift up and down between different IR of SU(2), and, provided the tensor representation had dimension $(2j+1)$ with j integral, there exists a zero shift [i.e., an SU(2) scalar] operator which is given as a special case of the formula. For a single $SU_j(2)$ and tensor $T^{[j]}_\alpha$, the scalar operator is given by the formula

$$O_0=\gamma_0(j;l_1,m_1)T^{[j]}_0+\sum_{\alpha=1}^{j}[\gamma_\alpha(j;l_1,m_1)T^{[j]}_\alpha j^\alpha_- +(-1)^j\gamma_\alpha(j;l_1,-m_1)T^{[j]}_{-\alpha}j^\alpha_+], \tag{2.9}$$

where $l_1(l_1+1)$ and m_1 are eigenvalues of J^2 and j_0 for a state $|l_1,m_1\rangle$ of an IR of $SU_j(2)$ upon which O_0 acts and, for $\alpha=0,...,j$,

$$\gamma_\alpha(j;l_1,m_1)=(-1)^{j+3l_1-m_1}\times\left[\frac{(j!)^2(2l_1+j+1)!(l_1-m_1-\alpha)!(l_1+m_1)!}{(2j)!(2l_1-j)!(l_1+m_1+\alpha)!(l_1-m_1)!}\right]^{1/2}\times\begin{pmatrix} j & l_1 & l_1\\ \alpha & -\alpha-m_1 & m_1\end{pmatrix}. \tag{2.10}$$

It was shown in that paper that l_1 and m_1 occur only to positive integral power, and l_1 in fact occurs only in the combination $l_1(l_1+1)$, so m_1 and $l_1(l_1+1)$ can always be replaced by the operators j_0 and J^2. The above expression is therefore, contrary to appearance, not dependent on its actions on a particular state $|l_1,m_1\rangle$ of an IR of $SU_j(2)$. In fact, O_0 is of degree j in the j_i, and of degree 1 in the tensor components $T^{[j]}_\alpha$.

It is not difficult to adapt this to the case of $SU_j(2)\times SU_k(2)$. Given the tensor $(R^r)^{[j,k]}_{\alpha,\beta}$, we define

$$A^{[j,k]}_{0,\beta}=\gamma_0(j;l_1,m_1)(R^r)^{[j,k]}_{0,\beta}+\sum_{\alpha=1}^{j}[\gamma_\alpha(j;l_1,m_1)(R^r)^{[j,k]}_{\alpha,\beta}j^\alpha_- +(-1)^j\gamma_\alpha(j;l_1,-m_1)(R^r)^{[j,k]}_{-\alpha,\beta}j^\alpha_+]. \tag{2.11}$$

Then the required Hermitian scalar of degree $[j,k,r]$ in the j_i, k_i, $R_{\mu\nu}$ is given by

$$C^{[j,k,r]}=\gamma_0(k;l_2,m_2)A^{[j,k]}_{0,0}+\sum_{\beta=1}^{k}[\gamma_\beta(k;l_2,m_2)A^{[j,k]}_{0,\beta}k^\beta_- +(-1)^k\gamma_\beta(k;l_2,-m_2)A^{[j,k]}_{0,-\beta}k^\beta_+], \tag{2.12}$$

where the γ coefficients are as given in Eqs. (2.10) and $l_1(l_1+1),m_1,l_2(l_2+1),m_2$ are replaced by J^2,j_0,K^2, and k_0.

III. CONSTRUCTION OF THE SCALARS

We now use the general method given in Sec. II to determine the scalars out of which the commuting scalars Y_1 and Y_2 and the invariant I_6 will be constructed. We shall adopt the following shorthand notation:

$$Q_0=C^{(202)},\quad S_0=C^{(112)},\quad T_0=C^{(312)},\quad U_0=C^{(204)}$$
$$V_0=C^{(024)},\quad W_0=C^{(114)},\quad P_0=C^{(006)}. \tag{3.1}$$

We shall obtain the scalars by first constructing the appropriate tensor using Eq. (2.8) and then using Eqs. (2.10)–(2.12), except that in order to minimize the number of square roots and fractions they will generally be multiplied by overall scaling constants.

First, using (2.8), we obtain the tensor of type $(R^2)^{[2,0]}$ with highest components

$$q_{\pm2}=-R_{\pm3/2,\pm1/2}R_{\pm1/2,\mp1/2}+R_{\pm3/2,\mp1/2}R_{\pm1/2,\pm1/2}. \tag{3.2}$$

The other components may be obtained from this by using the commutation rules (2.1)–(2.3).

Then use of (2.10)–(2.12) gives

$$Q_0=-3q_{+2}j^2_- -3q_{-2}j^2_+ +3q_{+1}j_-(2j_0-1) -3q_{-1}j_+(2j_0+1)+\sqrt{6}q_0(J^2-3j^2_0). \tag{3.3}$$

Similarly, the tensor of type $(R^2)^{[1,1]}$ has highest components

$$s_{\pm11}=2(\sqrt{3}r_{\pm3/2,1/2}R_{\mp1/2,1/2}-R^2_{\pm1/2,1/2}), \tag{3.4}$$

and then S_0 is given by

$$\sqrt{10}S_0=s_{11}k_-j_- -\sqrt{2}s_{10}k_0j_- -s_{1\,-1}k_+j_- +s_{-1\,-1}k_+j_+ +\sqrt{2}s_{-10}k_0j_+ -s_{-11}k_-j_+ +\sqrt{2}s_{0\,-1}k_+j_0+2s_{00}k_0j_0-\sqrt{2}\,s_{01}k_-j_0. \tag{3.5}$$

Next, the tensor of type $(R^2)^{[3,1]}$ has highest components

$$t_{\pm31}=2\sqrt{5}R^2_{\pm3/2,1/2}, \tag{3.6}$$

and T_0 is then given by

$$2\sqrt{5}T_0=\sum_{i=0,\pm1}\{t_{3i}j^3_- -\sqrt{6}t_{2i}j^2_-(j_0-1)-\sqrt{3/5}t_{1i}j_-(J^2-5j^2_0+5j_0-2)+(2/\sqrt{5})t_{0i}j_0(3J^2-5j^2_0-1) +\sqrt{3/5}t_{-1i}j_+(J^2-5j^2_0-5j_0-2)-\sqrt{6}t_{-2i}j^2_+(j_0+1)-t_{-3i}j^3_+\}\bar{k}_{-i}, \tag{3.7}$$

where we have used for brevity the notation $\bar{k}_1=k_+$, $\bar{k}_0=\sqrt{2}k_0$, $\bar{k}_{-1}=-k_-$.

The above three tensor types were obtained from the Kronecker product of two $R^{[3/2,1/2]}$ tensors. In order to obtain the tensors of fourth order in the $R_{\mu\nu}$ we make use of the tensors already obtained. Firstly, to obtain the tensor of type $(R^4)^{[20]}$, we use the Kronecker product of two tensors of type $(R^2)^{[20]}$. Application of Eq. (2.8) yields the highest components

$$u_{\pm2}=2q_{\pm2}q_0+2q_0q_{\pm2}-6q^2_{\pm1}, \tag{3.8}$$

and then U_0 can be written down in analogy with Eq. (3.3),

$$U_0 = -3u_{+2}j_-^2 - 3u_{-2}j_+^2 + 3u_{+1}j_-(2j_0 - 1) - 3u_{-1}j_+(2j_0+1) + \sqrt{6}u_0(J^2 - 3j_0^2). \quad (3.9)$$

For the tensor of type $(R^4)^{[02]}$, we use the Kronecker product of two tensors of type $(R^2)^{[1,1]}$ to obtain

$$v_{\pm 2} = \sqrt{6}(s_{1\,\pm 1}s_{-1\,\pm 1} + s_{-1\,1}s_{1\,\pm 1} - s_{0\,\pm 1}^2), \quad (3.10)$$

and then V_0 is given by

$$V_0 = 3v_{+2}k_-^2 + 3v_{-2}k_+^2 - 3v_{+1}k_-(2k_0 - 1) + 3v_{-1}k_+(2k_0 + 1) - \sqrt{6}v_0(K^2 - 3k_0^2). \quad (3.11)$$

These scalars suffice for the construction of Y_1 and Y_2, but for I_6 we shall also need W_0 and P_0. First, to obtain the tensor of type $(R^4)^{[11]}$, we again take the Kronecker product of two tensors of type $(R^2)^{[11]}$, which yields the highest components

$$w_{\pm 11} = \pm s_{\pm 11}s_{00} \pm s_{00}s_{\pm 11} \mp s_{\pm 10}s_{01} \mp s_{01}s_{\pm 10}. \quad (3.12)$$

Then in analogy with (3.5) we obtain

$$W_0 = w_{11}k_-j_- - \sqrt{2}w_{10}k_0j_- - w_{1-1}k_+j_- + w_{-1-1}k_+j_+ + \sqrt{2}w_{-10}k_0j_+ - w_{-11}k_-j_+ + \sqrt{2}w_{0-1}k_+j_0 + 2w_{00}k_0j_0 - \sqrt{2}w_{01}k_-j_0. \quad (3.13)$$

Finally, in order to construct the scalar P_0 we first construct the tensor of type $(R^3)^{[1/2,\,1/2]}$ with highest component

$$\begin{aligned} p_{1/2,\,1/2} &= 3(R_{3/2,\,1/2}R_{-3/2,\,1/2}R_{1/2,\,-1/2} \\ &\quad - R_{3/2,\,-1/2}R_{-3/2,\,1/2}R_{1/2,\,1/2}) \\ &\quad - \sqrt{3}(R_{3/2,\,1/2}R_{-1/2,\,1/2}R_{-1/2,\,-1/2} \\ &\quad - R_{3/2,\,-1/2}R_{-1/2,\,1/2}R_{-1/2,\,1/2}) \\ &\quad + (R_{1/2,\,1/2}R_{1/2,\,1/2}R_{-1/2,\,1/2} \\ &\quad - R_{1/2,\,1/2}R_{1/2,\,-1/2}R_{-1/2,\,1/2}) \\ &\quad - \tfrac{1}{3}R_{1/2,\,1/2}(4j_0 - 3k_0) + R_{1/2,\,-1/2}k_+ \\ &\quad - R_{-1/2,\,1/2}j_+. \end{aligned} \quad (3.14)$$

By taking the Kronecker product of this tensor with itself we can extract the tensor type $(R^6)^{[00]}$. This has only one component which is identical to P_0. This is

$$P_0 = p_{1/2,\,1/2}p_{-1/2,\,-1/2} + p_{-1/2,\,-1/2}p_{1/2,\,1/2}$$

This completes the calculation of the scalars needed to calculate Y_1, Y_2, and I_6. The method of calculating them may seem somewhat cumbersome, however, as was mentioned earlier, the scalars are particular zero-shift cases of more general shift operators which have the same tensorial type as the scalars and which connect different IR of $SU_j(2) \times SU_k(2)$ within an IR of $G(2)$. These are the tools which it is intended to employ to calculate the eigenvalues of the scalars in a later paper and their construction also depends on the prior construction of the various tensor components listed in this section. Furthermore, computer programs have recently been developed by De Meyer and Vanden Berghe[20] for calculating the different tensorial types of scalars which are based on the method employed in this paper and which also employ subroutines for calculating 3-j symbols.

based on the method employed in this paper and which also employ subroutines for calculating 3-j symbols.

IV. THE COMMUTING SCALARS Y_1 AND Y_2

In order to search for a pair of commuting operators from the set of scalars $\{Q_0, S_0, T_0, U_0, V_0\}$, we used FORTRAN programs developed by De Meyer, Vanden Berghe, and Van der Jeugt at Gent University. In order to avoid decimals, the following code was used for the basis elements:

$$\begin{aligned} &1 \equiv 2j_0, \quad 2 \equiv j_-, \quad 3 \equiv j_+, \quad 4 \equiv 2k_0, \quad 5 \equiv k_-, \\ &6 \equiv k_+, \quad 7 \equiv 2\sqrt{3}R_{-1/2,\,-1/2}, \quad 8 \equiv 2\sqrt{3}R_{-1/2,\,1/2}, \\ &9 \equiv 2\sqrt{3}R_{1/2,\,-1/2}, \quad 10 \equiv 2\sqrt{3}R_{1/2,\,1/2}, \\ &11 \equiv 2R_{-3/2,\,-1/2}, \quad 12 \equiv 2R_{-3/2,\,1/2}, \\ &13 \equiv 2R_{3/2,\,-1/2}, \quad 14 \equiv 2R_{3/2,\,1/2}. \end{aligned} \quad (4.1)$$

With these rescaled basis elements the structure constants for $G(2)$ all become integers.

To facilitate comparison of terms in different scalars and their commutators, the terms are put in a standard order in which the basis elements with numerically higher codes are written to the left of those with numerically lower codes. For instance, when written in standard ordered numerical code, the invariant I_2 given in Eq. (2.7) becomes

$$24I_2 = 6(14,11) - 6(13,12) - 2(10,7) + 2(9,8) + 12(6,5) + 3(4,4) + 4(3,2) + (1,1) - 6(4) - 10(1). \quad (4.2)$$

The essential part of the FORTRAN program is a subroutine which replaces a nonstandard ordered polynomial form in the basis elements by the corresponding standard ordered polynomial together with all the extra lower-degree polynomial forms, themselves standard ordered, incurred by interchanging basis elements and making use of the commutation relations in terms of the structure constants.

These programs were used to calculate all possible commutators of scalars of degree not greater than 6. Since Q_0 and S_0 are fourth degree, commutators of products of these with I_2, J^2, and K^2 were also needed. We then searched for linear combinations of these commutators in which all highest- (i.e., 11th-) degree terms vanished, and then corrections were made to eliminate the lower-degree terms.

Fortunately we were able to find a vanishing linear combination of commutators which could be expressed as a commutator of linear combinations of the scalars. These two linear combinations, which are therefore the commuting scalars Y_1 and Y_2, are given in terms of the scalars of Sec. III by the formulas

$$Y_1 = V_0 - 5\sqrt{6}S_0(4K^2 - 3), \quad (4.3)$$

$$Y_2 = 6U_0 + 12\sqrt{6}T_0 + 9\sqrt{6}S_0(4J^2 - 3) - 2\sqrt{2}Q_0(6I_2 - 3K^2 - J^2 + 9). \quad (4.4)$$

$24\sqrt{6}Y_1$ and $16\sqrt{6}Y_2$, which contain, respectively, 254 and 264 standard ordered terms, are given in Tables I and II. They both contain 110 terms of degree 6, which may just be a coincidence; it is possible, however, that there is some hidden reason for this. Note that the contributions to Y_1 and Y_2

TABLE I. The terms of $24\sqrt{6}\,Y_1$ are given in standard order in terms of the numerical code of Eq. (4.2). The coefficients are written on the right of the basis elements.

14	14	12	12	5	5	-162
14	13	12	11	4	4	-108
14	12	9	7	4	4	18
14	11	9	8	6	5	-36
14	8	8	7	5	4	36
14	7	6	5	4	2	72
13	12	10	6	5	4	-54
13	12	6	5	4	1	-108
13	11	9	7	6	6	108
13	8	7	7	6	4	-36
12	10	10	10	5	5	-24
12	9	9	9	6	4	-12
11	10	9	9	6	4	-36
10	10	8	8	5	5	6
10	9	8	8	5	4	-6
10	8	6	5	5	1	-24
9	9	8	7	6	4	6
9	7	6	6	5	1	24
7	7	6	6	5	3	48
14	13	11	11	6	0	-162
14	12	5	4	1	0	-324
14	11	4	4	4	0	162
14	7	6	5	2	0	36
13	12	9	8	4	0	18
13	11	10	7	6	0	54
13	8	8	7	4	0	-12
12	10	9	9	4	0	-12
11	10	10	9	4	0	-12
10	10	7	7	4	0	1
10	9	4	4	2	0	32
10	7	4	4	1	0	-6
9	3	4	4	1	0	-6
8	7	4	4	3	0	-32
4	4	4	3	2	0	48
14	8	5	2	0	0	-324
13	11	6	1	0	0	162
11	9	6	3	0	0	252
10	7	4	4	0	0	-4
9	7	6	4	0	0	-24
6	5	3	2	0	0	32
13	11	6	0	0	0	-972
6	5	1	0	0	0	592
4	1	0	0	0	0	352
14	14	12	11	5	4	162
14	13	11	11	6	4	-162
14	12	6	5	5	1	216
14	11	9	3	4	4	18
14	3	7	7	6	5	24
14	7	4	4	4	2	18
13	12	10	7	6	5	-36
13	12	4	4	4	1	-27
13	11	6	6	5	1	-216
13	8	6	5	4	2	72
12	10	10	9	5	4	36
12	9	6	5	4	3	-72
11	10	6	5	4	3	-72
10	10	3	7	5	4	-6
10	9	3	7	6	5	-8
10	3	5	4	4	1	-6
9	9	7	7	6	6	6
9	7	6	4	4	1	6
7	7	6	4	4	3	12
14	12	10	7	5	0	54
14	11	10	3	5	0	54
14	11	4	4	1	0	54
14	7	4	4	2	0	-84
13	12	9	7	6	0	54
13	11	9	6	6	0	54
13	3	7	7	6	0	-36
12	10	5	4	3	0	-360
11	10	9	9	6	0	-36
10	10	5	4	2	0	-84
10	8	6	5	5	0	48
9	9	8	8	4	0	1
9	7	6	6	5	0	-48
7	7	6	4	3	0	-36
14	12	5	4	0	0	1944
14	7	4	2	0	0	-192
13	8	4	2	0	0	-102
10	10	5	2	0	0	72
10	7	4	1	0	0	-9
9	7	6	1	0	0	-18
4	4	4	1	0	0	-168
10	3	5	0	0	0	12
4	4	4	0	0	0	-192
4	0	0	0	0	0	720
14	14	11	11	6	5	54
14	12	10	8	5	5	108
14	12	5	4	4	1	54
14	11	9	7	6	4	54
14	8	7	7	4	4	-12
13	13	12	12	6	5	54
13	12	10	7	4	4	18
13	11	10	8	6	5	-36
13	11	6	4	4	1	-54
13	8	4	4	4	2	18
12	10	9	9	6	5	24
12	9	4	4	4	3	-18
11	10	4	4	4	3	-18
10	10	7	7	6	5	-2
10	9	8	7	4	4	4
10	7	6	5	4	1	12
9	9	6	6	5	2	-48
8	8	6	5	5	3	-48
14	14	12	11	5	0	-162
14	12	9	8	5	0	54
14	11	10	7	4	0	18
14	8	8	7	5	0	-36
13	13	12	12	4	0	-27
13	12	6	5	4	0	648
13	11	6	6	5	0	1296
13	8	6	5	2	0	96
12	9	9	9	6	0	-12
11	10	6	5	3	0	-96
10	9	8	8	5	0	6
10	8	5	4	4	0	12
9	9	8	7	6	0	6
9	7	6	4	4	0	-12
6	6	5	5	1	0	576
14	12	5	1	0	0	270
13	12	6	5	0	0	-96
13	7	6	2	0	0	-252
10	9	4	2	0	0	44
9	9	6	2	0	0	48
8	8	5	3	0	0	-72
4	4	3	2	0	0	-112
10	7	4	0	0	0	2
4	4	1	0	0	0	148
14	14	11	11	4	4	-27
14	12	10	7	5	4	-54
14	11	10	8	5	4	-54
14	11	6	5	4	1	-108
14	8	6	5	5	2	-144
13	13	12	12	4	4	-27
13	12	9	5	6	5	-36
13	11	10	8	4	4	18
13	8	8	8	5	4	12
13	7	7	7	6	6	-24
12	10	9	9	4	4	-12
11	10	10	10	5	4	12
11	9	9	9	6	6	-24
10	10	7	7	4	4	1
10	9	7	7	6	4	6
10	7	4	4	4	1	3
9	9	6	4	4	2	-12
8	8	5	4	4	3	-12
14	14	11	11	4	0	-27
14	12	9	7	4	0	18
14	11	9	8	4	0	18
14	8	7	7	4	0	-12
13	13	12	11	6	0	-162
13	12	4	4	4	0	162
13	11	6	4	4	0	324
13	8	4	4	2	0	-84
12	9	6	5	3	0	-96
11	10	4	4	3	0	84
10	9	8	7	4	0	4
10	8	5	4	1	0	36
9	9	6	4	2	0	36
9	7	6	4	1	0	-12
6	5	4	4	1	0	288
14	11	6	5	0	0	96
13	12	4	4	0	0	-276
12	10	5	3	0	0	324
10	8	5	4	0	0	-24
9	8	6	5	0	0	32
8	7	4	3	0	0	-44
14	12	5	0	0	0	-1620
9	8	4	0	0	0	2
4	3	2	0	0	0	-160
14	13	12	12	5	4	162
14	12	9	8	5	4	-54
14	11	10	7	6	5	-36
14	11	4	4	4	1	-27
14	8	5	4	4	2	-36
13	13	12	11	6	4	-162
13	12	9	8	4	4	18
13	11	10	7	6	4	54
13	8	8	7	6	5	24
13	7	6	6	5	2	144
12	10	6	5	5	3	144
11	10	10	9	6	5	24
11	9	6	6	5	3	-144
10	10	6	5	5	2	48
10	9	6	5	4	2	-48
9	9	8	8	6	5	-2
9	8	6	5	4	1	12
8	7	6	5	4	3	48
14	13	12	12	5	0	-162
14	12	6	5	5	0	-1296
14	11	9	7	6	0	54
14	8	5	4	2	0	360
13	12	10	8	5	0	54
13	12	4	4	1	0	54
13	11	6	4	1	0	108
13	7	6	4	2	0	-216
12	9	4	4	3	0	84
11	9	6	4	3	0	216
10	9	7	7	6	0	6
10	7	6	5	4	0	-24
9	8	6	5	4	0	-24
8	8	5	4	3	0	84
6	5	4	3	2	0	192
14	11	4	4	0	0	-372
13	12	4	1	0	0	81
12	9	4	3	0	0	102
10	8	5	1	0	0	-30
9	8	4	4	0	0	-4
7	7	6	3	0	0	-48
14	11	4	0	0	0	-534
9	7	6	0	0	0	-12
6	5	0	0	0	0	-288
14	13	12	11	6	5	216
14	12	9	7	6	5	-36
14	11	10	7	4	4	18
14	8	8	8	5	5	-24
14	7	7	7	6	4	-12
13	13	11	11	6	6	-162
13	12	9	7	6	4	54
13	11	9	6	6	4	54
13	8	8	7	4	4	-12
13	7	6	4	4	2	36
12	10	5	4	4	3	36
11	10	10	9	4	4	-12
11	9	6	4	4	3	-36
10	10	5	4	4	2	12
10	9	4	4	4	2	-12
9	9	8	8	4	4	1
9	8	4	4	4	1	3
8	7	4	4	4	3	12
14	13	12	11	4	0	-108
14	12	5	4	4	0	-324
14	11	6	5	4	0	648
14	7	7	7	6	0	-12
13	12	10	7	4	0	18
13	11	10	8	4	0	18
13	8	8	5	5	0	-12
12	10	10	9	5	0	-36
11	10	10	10	5	0	-12
10	10	8	7	5	0	6
10	9	6	5	2	0	-16
10	7	4	4	4	0	-6
9	8	4	4	4	0	-6
8	7	6	5	3	0	16
4	4	4	4	1	0	36
14	11	4	1	0	0	31
13	11	6	4	0	0	-648
11	10	4	3	0	0	102
10	7	6	5	0	0	32
9	8	4	1	0	0	-9
6	5	4	1	0	0	-1248
13	12	4	0	0	0	-438
6	5	4	0	0	0	-768
4	4	0	0	0	0	528

of highest, i.e., fourth, degree in the $R_{\mu\nu}$ are of tensor type $T^{[024]}$ and $T^{[204]}$, respectively, in loose analogy with results obtained by Van der Jeugt[17] for cases of Lie algebra $G \supset [\mathrm{SU}(2)]^n$ for which there were commuting scalars of degree 2 in the tensor basis elements [see Eq. (2.9) of that paper]. Also, the fact that the sixth-degree terms of fourth degree in the $R_{\mu\nu}$ in Y_1 and Y_2 are in both cases precisely 66 in number may or may not be a coincidence.

Finally, using similar (but much less time consuming) programs, we obtained the following expression for the sixth-order invariant:

$$
\begin{aligned}
I_6 = {} & 864P_0 + 324W_0 - 54\sqrt{6}V_0 - 36\sqrt{6}U_0 - 432T_0 \\
& - 72S_0(90I_2 + 27J^2 - 135K^2 + 134) \\
& + 8\sqrt{3}Q_0(18I_2 - 3J^2 - 45K^2 + 11) + 216I_2^2J^2 \\
& + 1944I_2^2K^2 + 72I_2J^4 - 432I_2J^2K^2 - 1944I_2K^4 \\
& - 10J^6 - 198K^2J^4 + 810K^4J^2 + 486K^6 \\
& + 1092I_2J^2 + 1431I_2K^2 - 1566K^4 - 1932K^2J^2 \\
& - 82J^4 + 3240K^2 + 876J^2.
\end{aligned}
\tag{4.5}
$$

TABLE II. Terms of $16\sqrt{6}\,Y_2$.

14	14	11	11	3	2	36
14	13	12	11	3	2	-72
14	13	8	7	2	2	-72
14	12	9	7	1	1	36
14	11	10	9	2	1	72
14	10	7	7	2	1	-36
14	7	4	3	2	2	-144
13	13	8	8	2	2	36
13	12	10	9	2	1	72
13	11	10	10	2	1	-72
13	10	8	7	2	1	36
13	7	6	3	2	2	-288
12	10	9	7	3	1	-36
12	8	5	3	3	1	288
11	10	4	3	3	2	144
10	10	7	7	3	2	-12
10	8	5	3	2	1	-144
9	9	6	2	1	1	72
8	7	4	3	1	1	-72
14	13	11	6	2	0	36
14	12	5	3	2	0	-144
14	11	4	3	2	0	72
14	9	8	7	2	0	-36
13	13	12	8	2	0	-36
13	12	4	3	2	0	72
13	11	8	8	3	0	72
13	9	8	8	2	0	36
12	10	5	3	1	0	288
12	7	4	3	3	0	-216
11	10	3	1	1	0	72
10	10	5	2	1	0	216
10	7	4	3	2	0	-72
9	7	4	3	1	0	-72
4	3	2	1	1	0	-288
14	8	5	2	0	0	72
13	11	6	1	0	0	-324
12	9	4	3	0	0	-108
11	8	3	3	0	0	-144
10	7	3	2	0	0	408
9	7	6	1	0	0	36
4	3	2	1	0	0	288
14	7	2	0	0	0	-936
10	7	1	0	0	0	-204
3	2	1	0	0	0	-672
14	14	11	11	1	1	-18
14	13	12	11	1	1	36
14	13	4	2	2	2	288
14	12	7	7	3	1	72
14	11	9	8	3	2	72
14	10	5	2	2	1	288
14	7	4	2	1	1	36
13	13	6	2	2	2	288
13	12	10	7	3	2	72
13	11	10	8	3	2	-72
13	10	4	2	2	1	-144
13	7	6	2	1	1	72
12	10	5	3	3	2	-288
12	7	4	3	3	1	-144
11	10	4	3	1	1	-36
10	10	7	7	1	1	6
10	8	5	1	1	1	36
9	8	4	3	2	1	72
7	7	6	3	1	1	-72
14	13	2	2	2	0	-144
14	12	5	1	1	0	288
14	11	4	1	1	0	-144
14	9	4	2	2	0	360
13	12	12	9	3	0	-36
13	12	4	1	1	0	-144
13	11	6	3	2	0	144
13	9	6	2	2	0	720
12	9	9	8	3	0	36
12	7	3	3	1	0	72
11	9	6	3	1	0	-288
10	9	8	7	1	0	-12
9	9	8	8	1	0	6
8	7	3	3	2	0	-48
14	12	5	1	0	0	324
14	7	4	2	0	0	-36
13	10	2	2	0	0	1008
12	9	3	1	0	0	936
10	10	5	2	0	0	-144
10	7	1	1	0	0	-204
8	7	3	1	0	0	24
4	1	1	1	0	0	288
13	12	1	0	0	0	540
9	8	1	0	0	0	300
1	1	1	0	0	0	72
14	14	11	9	2	2	-144
14	13	12	9	2	2	144
14	12	11	9	3	1	36
14	12	5	3	2	1	-432
14	11	9	8	1	1	-36
14	9	8	7	2	1	36
13	13	12	12	3	2	36
13	12	12	9	3	1	-36
13	12	10	7	1	1	-36
13	11	10	8	1	1	36
13	9	8	8	2	1	-36
12	12	9	9	3	3	36
12	10	5	3	1	1	72
11	11	10	10	3	3	36
11	9	6	3	3	2	288
10	10	5	2	1	1	-72
10	7	4	3	2	1	72
9	8	4	1	1	1	-18
14	14	11	11	1	0	-18
14	12	11	9	3	0	36
14	11	11	10	3	0	-36
14	11	3	2	1	0	-432
14	9	2	2	1	0	504
13	12	11	10	3	0	36
13	12	3	2	1	0	-144
13	11	6	1	1	0	-288
13	8	3	2	2	0	-288
12	9	4	3	1	0	-144
11	10	10	7	3	0	36
11	6	4	3	3	0	-216
10	9	4	2	1	0	-216
9	9	6	2	1	0	-216
8	7	3	1	1	0	24
14	11	4	1	0	0	-162
14	7	2	1	0	0	1080
13	8	4	2	0	0	-36
12	7	3	3	0	0	288
10	9	4	2	0	0	144
9	9	6	2	0	0	144
6	5	3	2	0	0	-576
3	3	2	2	0	0	288
13	8	2	0	0	0	936
6	5	1	0	0	0	576
4	1	0	0	0	0	-144
14	14	11	7	2	1	36
14	13	12	7	2	1	-36
14	12	11	7	3	3	144
14	12	5	1	1	1	-36
14	11	8	7	3	1	-72
14	9	4	2	2	1	-144
13	13	12	12	1	1	-18
13	12	12	7	3	3	-144
13	12	8	7	3	1	-72
13	11	8	8	3	1	72
13	9	6	2	2	1	-288
12	12	5	3	3	3	288
12	9	9	8	3	1	36
11	11	6	3	3	3	-288
11	9	6	3	1	1	-72
10	9	8	7	3	2	24
10	7	4	1	1	1	-18
9	7	6	3	2	1	144
14	14	11	7	2	0	-36
14	12	9	9	2	0	-72
14	11	10	9	2	0	-72
14	11	1	1	1	0	72
14	7	3	2	2	0	288
13	12	10	9	2	0	-72
13	12	1	1	1	0	-72
13	10	8	7	2	0	-36
13	8	2	1	1	0	144
12	9	3	3	2	0	-288
11	10	9	8	3	0	-36
11	8	3	3	1	0	72
10	9	3	2	2	0	48
9	8	4	3	2	0	-72
7	7	6	3	1	0	-72
14	11	3	2	0	0	792
13	12	4	1	0	0	-162
13	8	2	1	0	0	-1080
11	10	4	3	0	0	-108
10	9	2	1	0	0	-72
9	8	4	1	0	0	18
6	5	1	1	0	0	288
3	2	1	1	0	0	-48
12	9	3	0	0	0	1008
4	4	1	0	0	0	144
3	2	0	0	0	0	-1440
14	14	7	7	2	2	36
14	13	11	10	2	2	144
14	12	9	9	2	1	-72
14	11	11	10	3	1	-36
14	11	4	3	2	1	216
14	8	5	3	2	2	288
13	13	12	10	2	2	-144
13	12	11	10	3	1	36
13	12	4	3	2	1	216
13	11	6	3	2	1	432
13	8	4	3	2	2	-144
12	11	10	9	3	3	-72
12	9	4	3	3	2	144
11	10	10	7	3	1	36
11	8	4	3	3	1	-144
10	9	8	7	1	1	-12
9	9	8	8	3	2	-12
9	7	6	1	1	1	-36
14	13	12	11	1	0	36
14	12	9	7	1	0	36
14	11	9	8	1	0	-36
14	10	7	7	2	0	36
14	7	2	1	1	0	-144
13	12	10	7	1	0	-36
13	11	10	10	2	0	72
13	10	4	2	2	0	360
12	11	3	3	3	0	144
12	9	3	1	1	0	-72
11	10	4	3	1	0	-144
11	7	6	3	3	0	-432
10	9	2	1	1	0	-24
9	7	6	3	2	0	-144
6	5	1	1	1	0	-288
14	11	1	1	0	0	-180
13	12	3	2	0	0	-792
13	7	6	2	0	0	-72
11	10	3	1	0	0	-648
10	8	5	1	0	0	-36
9	8	3	2	0	0	-600
4	4	3	2	0	0	-144
1	1	1	1	0	0	-48
11	10	3	0	0	0	-720
4	3	2	0	0	0	-288
1	1	0	0	0	0	1128
14	14	5	2	2	2	-288
14	13	11	8	2	1	-36
14	12	9	7	3	2	-72
14	11	11	8	3	3	-144
14	11	4	1	1	1	18
14	8	5	2	1	1	-72
13	13	12	8	2	1	36
13	12	11	8	3	3	144
13	12	4	1	1	1	18
13	11	6	1	1	1	36
13	8	4	2	1	1	36
12	11	4	3	3	3	-288
12	9	4	3	1	1	-36
11	10	9	8	3	1	-36
11	7	6	3	3	1	-288
10	9	4	2	1	1	72
9	9	8	8	1	1	6
8	8	5	3	1	1	72
14	13	12	7	2	0	36
14	12	7	7	3	0	72
14	11	8	7	3	0	-72
14	10	5	2	2	0	-720
13	13	12	12	1	0	-18
13	12	8	7	3	0	272
13	11	10	8	1	0	36
13	10	2	2	1	0	-360
12	10	9	7	3	0	-36
12	8	5	3	3	0	432
11	10	3	3	2	0	288
10	10	7	7	1	0	6
10	8	5	3	2	0	144
8	8	5	3	1	0	72
4	4	1	1	1	0	-72
14	9	2	2	0	0	-1152
13	12	1	1	0	0	468
12	10	5	3	0	0	216
11	9	6	3	0	0	-216
10	7	4	1	0	0	18
9	8	1	1	0	0	300
4	4	1	1	0	0	72
14	11	1	0	0	0	-252
10	9	2	0	0	0	96
4	1	1	0	0	0	144
1	0	0	0	0	0	1008

This form for I_6 is, of course, not unique since any cubic polynomial in I_2 could be added to it to yield another sixth-order invariant. The authors are not aware of any previously published listing of I_6, although it was obtained in a different form by De Meyer.[21] We give $8I_6$ in Table III. At 730 terms, including 407 of sixth degree, it is shorter by about 100 terms than that of De Meyer.

V. CONCLUSION

We have obtained in this paper a pair of commuting scalars for $G(2) \supset SU(2)\times SU(2)$ of sixth degree in the basis

TABLE III. Terms of $8I_6$.

14	14	12	12	5	5	2916
14	14	12	7	5	2	-1944
14	14	11	9	2	2	432
14	14	7	7	2	2	-108
14	13	12	11	10	7	648
14	13	12	9	2	2	-432
14	13	11	11	10	8	-648
14	13	11	8	4	2	-1620
14	12	12	9	9	9	-216
14	12	11	9	3	1	-108
14	12	10	7	5	4	972
14	12	9	8	5	4	972
14	12	9	7	1	1	-108
14	12	5	4	4	1	-972
14	11	11	10	3	1	108
14	11	10	9	8	7	72
14	11	10	7	4	4	-486
14	11	9	9	6	2	-864
14	11	9	8	1	1	126
14	11	8	7	3	1	216
14	11	4	4	4	1	486
14	11	3	3	2	2	288
14	10	7	7	2	1	108
14	9	7	7	6	2	216
14	8	8	7	5	4	-648
14	8	6	5	5	2	2592
14	7	7	7	6	1	-216
14	7	4	3	2	2	432
13	13	12	12	1	1	81
13	13	12	8	8	7	216
13	13	11	8	8	8	-216
13	12	12	10	5	3	1296
13	12	11	10	4	3	-1620
13	12	10	10	5	2	-864
13	12	10	8	5	1	756
13	12	10	7	1	1	126
13	12	9	8	3	2	-72
13	12	8	7	4	3	-216
13	12	6	5	1	1	648
13	12	4	1	1	1	-54
13	11	11	10	6	3	-1944
13	11	10	9	6	2	1296
13	11	10	7	6	4	-972
13	11	8	8	4	3	648
13	11	6	3	2	1	-1296
13	10	7	7	6	2	-432
13	9	8	8	2	1	108
13	8	8	7	6	5	-432
13	8	6	5	4	2	-1296
13	8	4	2	1	1	-108
13	7	6	2	1	1	-216
12	10	10	10	5	5	432
12	10	9	9	4	4	216
12	10	6	5	5	3	-2592
12	9	9	9	6	1	216
12	9	6	5	3	1	1728
12	8	5	3	3	1	-864
11	11	6	3	3	3	864
11	10	10	9	4	1	-216
11	10	9	9	6	1	-216
11	10	6	5	3	1	-1728
11	9	9	9	6	6	432
11	9	6	3	1	1	216
10	10	10	7	7	7	-8
10	10	7	7	6	5	144
10	10	6	5	5	2	-864
10	9	8	8	5	1	108
10	9	7	7	6	4	-108
10	8	6	5	5	1	432
10	7	6	5	3	2	288
10	7	4	3	2	1	-216
9	9	9	8	8	8	8
9	9	8	8	1	1	-15
9	9	6	4	4	2	216
9	8	4	4	4	1	-54
9	8	3	3	2	2	96
9	7	6	3	2	1	-432
8	7	6	5	4	3	-864
7	7	6	3	1	1	216
6	5	3	3	2	2	-1152
4	4	3	3	2	2	-288
3	2	1	1	1	1	12
14	14	11	7	2	0	-108
14	13	11	11	6	0	4860
14	12	9	9	2	0	1944
14	12	6	5	5	0	25272
14	11	11	10	3	0	1620
14	11	9	8	4	0	-2376
14	11	6	5	1	0	3564
14	11	3	2	1	0	324
14	9	4	2	2	0	1512
14	7	7	7	6	0	648
14	7	2	1	1	0	414
13	12	12	9	3	0	-108
13	12	10	7	1	0	-1224
13	12	6	5	4	0	-12636
13	12	4	1	1	0	891
13	11	10	8	1	0	756
13	11	6	4	4	0	-6318
13	10	4	2	2	0	-1512
13	8	8	7	4	0	-216
13	8	3	2	2	0	936
12	10	10	9	5	0	1080
14	14	12	11	9	7	-648
14	14	11	11	6	8	648
14	14	11	8	7	7	-216
14	14	5	2	2	2	864
14	13	12	11	9	9	-648
14	13	12	9	7	7	-432
14	13	11	11	6	4	2916
14	13	11	9	2	1	108
14	12	12	9	5	3	-1944
14	12	11	7	3	3	-432
14	12	10	7	5	1	324
14	12	9	8	5	1	-324
14	12	8	7	5	3	1296
14	12	5	3	2	1	1296
14	11	11	9	6	3	1296
14	11	10	9	4	2	-216
14	11	10	7	4	1	216
14	11	9	8	6	5	1296
14	11	9	7	6	4	-972
14	11	7	7	6	3	-432
14	11	4	4	3	2	648
14	11	3	2	1	1	144
14	10	5	2	2	1	-864
14	9	6	5	2	2	-1728
14	8	8	7	5	1	216
14	8	5	4	4	2	648
14	7	6	5	4	2	-1296
14	7	4	2	1	1	-108
13	13	12	11	10	8	648
13	13	12	8	4	2	-324
13	13	11	8	6	2	-1944
13	12	12	9	4	3	324
13	12	11	10	3	1	-108
13	12	10	9	8	7	-72
13	12	10	7	6	5	1296
13	12	9	9	8	8	-72
13	12	9	8	1	1	-18
13	12	8	7	3	1	216
13	12	4	4	4	1	486
13	12	3	3	2	2	-288
13	11	10	10	8	7	-216
13	11	10	8	6	5	648
13	11	10	7	6	1	-324
13	11	8	8	3	1	-216
13	11	6	1	1	1	-108
13	10	6	5	2	2	1728
13	9	8	7	6	2	216
13	8	8	7	4	4	216
13	8	6	5	2	1	-1728
13	7	7	7	6	6	432
12	12	9	9	3	3	-108
12	10	10	9	5	4	-648
12	10	9	9	4	1	216
12	10	5	4	4	3	-648
12	9	9	9	4	3	-108
12	9	4	4	4	3	324
12	7	6	5	3	3	1728
11	10	10	10	5	4	-216
11	10	10	9	5	3	216
11	10	9	8	4	3	-324
11	10	4	4	4	3	324
11	9	9	8	6	3	-432
11	8	6	5	3	3	-1728
10	10	9	8	7	7	24
10	10	7	7	4	4	9
10	10	5	4	4	2	-216
10	9	8	7	6	5	-72
10	9	7	7	6	1	108
10	8	5	4	4	1	108
10	7	6	5	1	1	-360
10	7	4	1	1	1	54
9	9	8	8	6	5	144
9	9	8	7	6	4	-108
9	9	6	2	1	1	-216
9	8	4	4	3	2	-72
9	8	3	2	1	1	48
9	7	6	1	1	1	108
8	7	4	4	4	3	-216
6	6	5	5	3	2	5184
6	5	3	2	1	1	-576
4	4	3	2	1	1	-144
1	1	1	1	1	1	1
14	13	12	12	5	0	4860
14	13	11	8	2	0	972
14	12	9	8	5	0	972
14	12	5	4	4	0	6318
14	11	10	9	2	0	-1368
14	11	9	8	1	0	-1224
14	11	4	4	4	0	-3159
14	11	1	1	1	0	-387
14	9	2	2	1	0	-1512
14	7	6	5	2	0	-7992
13	13	12	12	4	0	-972
13	12	11	10	3	0	972
13	12	9	8	4	0	-216
13	12	6	5	1	0	-3564
13	12	3	2	1	0	1404
13	11	10	7	6	0	972
13	11	6	4	1	0	-1944
13	10	2	2	1	0	1080
13	8	7	7	6	0	216
13	8	2	1	1	0	-414
12	10	9	9	4	0	-216
14	14	12	11	5	4	-2916
14	14	11	11	4	4	729
14	14	11	5	5	2	1296
14	13	12	12	9	7	648
14	13	12	11	6	5	-5832
14	13	12	8	5	2	648
14	13	11	11	6	1	-972
14	13	11	7	6	2	648
14	12	11	10	9	9	432
14	12	10	9	7	7	-216
14	12	9	9	8	7	216
14	12	9	7	6	6	648
14	12	7	7	4	4	-648
14	12	5	1	1	1	108
14	11	11	8	3	3	432
14	11	10	9	2	1	-216
14	11	10	7	5	2	-72
14	11	9	8	4	4	-162
14	11	9	7	6	1	756
14	11	6	5	4	1	1944
14	11	4	4	1	1	-162
14	11	1	1	1	1	18
14	9	8	8	5	2	-432
14	9	4	4	2	2	-432
14	8	7	7	6	5	-432
14	8	5	3	2	2	-864
14	7	6	5	2	1	1728
13	13	12	12	10	7	-648
13	13	12	11	6	4	2916
13	13	12	8	2	1	-108
13	13	8	8	2	2	-108
13	12	12	9	3	1	108
13	12	11	9	6	3	648
13	12	10	9	4	2	216
13	12	10	7	4	4	-162
13	12	9	9	6	2	-432
13	12	9	7	6	4	-972
13	12	7	7	6	3	-864
13	12	4	4	3	2	-648
13	12	3	2	1	1	-144
13	11	10	10	4	2	648
13	11	10	8	4	4	-324
13	11	9	8	6	4	-972
13	11	8	7	6	3	1296
13	10	8	8	5	2	216
13	10	4	2	2	1	432
13	9	6	2	2	1	864
13	8	8	7	4	1	216
13	8	4	4	4	2	-324
13	7	6	6	5	2	-2592
12	12	5	3	3	3	-864
12	10	10	9	5	1	-216
12	10	9	8	5	3	216
12	10	5	3	3	2	-864
12	9	9	8	3	1	-108
12	9	4	4	3	1	432
12	7	4	4	3	3	432
11	10	10	10	5	1	216
11	10	10	7	4	3	108
11	10	9	8	3	1	108
11	10	4	4	3	1	-432
11	9	6	6	5	3	2592
11	8	4	4	3	3	-432
10	10	8	8	5	5	-108
10	10	7	7	4	1	54
10	10	5	2	1	1	216
10	9	8	7	4	4	-126
10	9	6	5	4	2	864
10	8	5	3	2	1	432
10	7	4	4	4	1	-54
10	7	3	3	2	2	-96
9	9	8	8	4	4	9
9	9	8	7	6	1	-108
9	8	6	5	4	1	-216
9	8	4	4	1	1	90
9	8	1	1	1	1	6
8	8	6	5	5	3	864
8	7	4	3	1	1	216
6	6	5	5	1	1	1296
6	5	1	1	1	1	-72
4	4	1	1	1	1	-18
14	14	12	11	5	0	972
14	13	12	11	4	0	2916
14	13	2	2	2	0	432
14	12	9	7	4	0	-324
14	12	5	4	1	0	5832
14	11	10	8	5	0	540
14	11	9	7	6	0	-2484
14	11	4	4	1	0	-81
14	10	7	7	2	0	252
14	8	8	7	5	0	216
14	7	4	4	2	0	-54
13	13	12	12	1	0	-324
13	12	10	9	2	0	-1656
13	12	9	8	1	0	468
13	12	4	4	4	0	-3159
13	12	1	1	1	0	387
13	11	9	8	6	0	-2916
13	11	6	3	2	0	-648
13	9	8	8	2	0	252
13	8	6	5	2	0	4536
13	7	6	4	2	0	3888
12	10	9	7	3	0	-252
14	14	12	11	5	1	972
14	14	11	11	4	1	-486
14	14	11	7	4	2	324
14	13	12	12	5	4	-2916
14	13	12	11	4	4	1458
14	13	12	7	4	2	1620
14	13	11	10	2	2	-432
14	13	8	7	2	2	216
14	12	11	10	3	3	648
14	12	10	9	5	2	1296
14	12	9	9	4	2	-648
14	12	9	7	4	4	-324
14	12	7	7	3	1	-216
14	11	11	10	10	9	-216
14	11	10	10	7	7	72
14	11	10	8	5	4	972
14	11	10	7	1	1	-18
14	11	9	8	4	1	540
14	11	8	8	5	3	-864
14	11	6	5	3	2	2592
14	11	4	4	3	1	-648
14	10	8	7	5	2	216
14	9	8	7	4	2	324
14	9	4	2	2	1	432
14	8	7	7	4	4	216
14	8	5	2	1	1	216
14	7	4	4	4	2	-324
13	13	12	12	4	4	729
13	13	12	11	6	1	972
13	13	12	7	6	2	1296
13	13	6	2	2	2	-864
13	12	12	7	3	3	-432
13	12	11	8	3	3	-432
13	12	10	9	2	1	-216
13	12	10	7	4	1	-540
13	12	9	8	4	4	-486
13	12	9	7	6	1	-756
13	12	6	5	4	1	1944
13	12	4	4	1	1	162
13	12	1	1	1	1	-18
13	11	10	10	2	1	216
13	11	10	8	3	2	216
13	11	9	8	6	1	324
13	11	6	6	5	1	3888
13	10	8	7	4	2	-324
13	10	4	4	2	1	-432
13	8	8	8	5	4	-216
13	8	7	7	6	4	648
13	8	4	4	2	1	-432
13	7	6	4	4	2	-648
12	11	10	9	3	3	216
12	10	10	7	5	3	-432
12	10	9	7	4	4	324
12	10	5	3	1	1	-216
12	9	9	7	6	3	216
12	9	4	3	3	2	-432
12	7	4	3	3	1	432
11	10	10	9	6	5	-432
11	10	10	7	3	1	-108
11	10	9	7	6	3	216
11	10	4	3	3	2	-432
11	9	6	4	4	3	648
11	9	4	3	3	1	432
10	10	8	7	5	4	108
10	10	7	7	3	2	48
10	9	9	8	8	7	-24
10	9	8	7	3	2	-96
10	9	4	4	4	2	216
10	9	5	1	1	1	-108
10	7	4	4	3	2	72
10	7	3	2	1	1	-48
9	9	8	8	4	1	-54
9	9	7	7	6	6	-108
9	8	6	5	3	2	-288
9	8	4	3	2	1	-216
9	7	6	6	5	1	-432
8	8	5	4	4	3	216
7	7	6	6	5	3	-864
6	5	4	4	3	2	2592
4	4	4	4	3	2	324
3	3	3	2	2	2	64
14	14	11	11	4	0	972
14	13	12	11	1	0	648
14	12	11	9	3	0	-2484
14	12	9	7	1	0	756
14	12	5	3	2	0	648
14	11	10	7	4	0	216
14	11	8	7	3	0	-1656
14	11	4	3	2	0	-756
14	10	5	2	2	0	432
14	8	7	7	4	0	648
14	7	4	2	1	0	-1296
13	13	12	11	6	0	972
13	12	10	8	5	0	-2484
13	12	9	7	6	0	540
13	12	4	4	1	0	-1863
13	11	10	10	2	0	1080
13	11	8	8	3	0	1080
13	11	6	1	1	0	2430
13	9	6	2	2	0	432
13	8	4	4	2	0	3078
13	7	6	2	1	0	1728
12	10	5	4	3	0	6480
14	14	12	7	7	7	216
14	14	11	11	1	1	81
14	14	11	7	2	1	-108
14	13	12	12	5	1	-972
14	13	12	11	1	1	-162
14	13	12	7	2	1	108
14	13	11	8	8	7	432
14	13	4	2	2	2	-864
14	12	11	9	5	5	-1620
14	12	10	8	5	5	-1944
14	12	9	9	2	1	216
14	12	9	7	3	3	216
14	12	6	5	5	1	-3888
14	11	11	10	4	3	-324
14	11	10	10	5	2	-432
14	11	10	8	5	1	-756
14	11	9	9	8	8	-144
14	11	9	8	3	2	-144
14	11	8	7	4	3	216
14	11	6	5	1	1	-648
14	11	4	1	1	1	-54
14	10	7	7	4	2	-108
14	9	8	7	2	1	-108
14	8	8	8	5	5	432
14	8	7	7	4	1	-216
14	7	7	7	6	4	216
14	7	4	4	2	1	432
13	13	12	12	4	1	486
13	13	12	10	2	2	432
13	13	11	11	6	6	2916
13	12	12	10	9	9	216
13	12	11	10	10	9	-432
13	12	10	10	7	7	144
13	12	10	8	5	4	972
13	12	10	7	3	2	-144
13	12	9	8	4	1	-216
13	12	8	8	5	3	-432
13	12	6	5	3	2	-2592
13	12	4	3	2	1	-648
13	11	11	10	10	10	216
13	11	10	9	8	8	216
13	11	10	8	1	1	-108
13	11	9	7	6	6	-1944
13	11	6	4	4	1	972
13	10	8	7	2	1	-108
13	9	8	8	4	2	108
13	8	8	8	5	1	-216
13	8	7	7	6	1	216
13	8	4	3	2	2	432
13	7	6	3	3	3	864
12	11	4	3	3	3	864
12	10	9	9	6	5	-432
12	10	9	7	3	1	108
12	9	9	9	6	4	216
12	9	6	5	4	3	1296
12	9	4	3	1	1	108
11	11	10	10	3	3	-108
11	10	10	7	4	4	216
11	10	9	9	6	4	648
11	10	6	5	4	3	1296
11	10	4	3	1	1	108
11	9	6	3	3	2	-864
11	7	6	3	3	1	864
10	10	8	7	5	1	-108
10	10	7	7	1	1	-15
10	9	8	8	5	4	108
10	9	8	7	1	1	30
10	9	4	2	1	1	-216
10	7	6	5	4	1	-216
10	7	4	4	1	1	-90
10	7	1	1	1	1	-6
9	9	8	8	3	2	48
9	9	6	6	5	2	864
9	8	6	5	1	1	360
9	8	4	1	1	1	54
9	7	6	4	4	1	-108
8	8	5	3	1	1	-216
7	7	6	4	4	3	-216
6	5	4	4	1	1	648
4	4	4	4	1	1	81
3	3	2	2	1	1	48
14	14	11	11	1	0	-324
14	13	12	7	2	0	-2484
14	12	10	7	5	0	-2916
14	12	7	7	3	0	1944
14	12	5	1	1	0	-2430
14	11	10	7	1	0	468
14	11	6	5	4	0	-12636
14	11	4	1	1	0	1539
14	9	8	7	2	0	-252
14	8	5	4	2	0	-6480
14	7	3	2	2	0	-936
13	13	12	8	2	0	1620
13	12	10	7	4	0	1080
13	12	8	7	3	0	-1368
13	12	4	3	2	0	108
13	11	10	8	4	0	-324
13	11	6	6	5	0	-25272
13	10	8	7	2	0	-252
13	8	8	8	5	0	648
13	8	4	2	1	0	2160
12	11	3	3	3	0	-432
12	9	9	9	6	0	-216

12	9	9	8	3	0	252
12	9	3	1	1	0	234
11	10	10	9	4	0	648
11	10	4	4	3	0	-1134
11	9	6	3	1	0	1728
10	10	8	7	5	0	252
10	9	9	8	2	0	-96
10	9	6	5	2	0	-144
10	8	6	5	5	0	-1512
10	7	6	5	4	0	756
10	7	4	1	1	0	-81
9	9	8	7	6	0	252
9	8	6	5	1	0	-2772
9	8	3	2	1	0	-300
[illegible]	7	6	3	[illegible]	0	504
8	7	4	4	3	0	684
7	7	6	3	1	0	-1080
6	5	1	1	1	0	1640
4	3	3	2	2	0	575
1	1	1	1	1	0	-30
14	12	5	1	0	0	-7452
14	11	4	1	0	0	-3672
14	8	5	2	0	0	-3312
13	12	9	8	0	0	-2592
13	12	1	1	0	0	-2916
13	8	8	7	0	0	192
12	10	5	3	0	0	-5760
11	10	4	3	0	0	-4356
10	10	5	2	0	0	-1344
10	8	5	1	0	0	3300
10	7	1	1	0	0	388
9	8	4	1	0	0	-264
8	8	5	3	0	0	2352
6	5	4	4	0	0	5184
4	4	4	1	0	0	3672
3	3	2	2	0	0	128
14	11	1	0	0	0	-14760
13	8	2	0	0	0	360
10	7	4	0	0	0	8592
8	7	3	0	0	0	896
4	3	2	0	0	0	816
13	12	0	0	0	0	-9504
4	1	0	0	0	0	27312

12	9	6	5	3	0	216
12	8	5	3	3	0	432
11	10	10	7	3	0	252
11	10	4	3	1	0	2160
11	8	4	3	3	0	1080
10	10	7	7	4	0	-252
10	9	8	8	5	0	-468
10	9	4	4	2	0	-684
10	8	5	4	4	0	-378
10	7	6	5	1	0	2772
10	7	3	2	1	0	300
9	9	6	4	2	0	-648
9	8	4	4	4	0	189
9	8	1	1	1	0	-51
9	7	6	1	1	0	-522
8	7	4	3	1	0	-864
5	6	5	5	1	0	-12960
4	4	4	4	1	0	-310
4	3	2	1	1	0	1152
14	14	11	11	0	0	3985
14	11	10	7	0	0	-3168
14	11	3	2	0	0	-2448
14	7	4	2	0	0	10404
13	12	6	5	0	0	-1944
13	11	10	8	0	0	4404
13	9	4	2	0	0	684
12	9	4	3	0	0	-1980
11	10	3	1	0	0	2748
10	9	8	7	0	0	-1216
10	7	6	5	0	0	-8280
9	9	8	8	0	0	800
9	8	3	2	0	0	1472
8	7	4	3	0	0	240
6	5	4	1	0	0	27648
4	4	3	2	0	0	3816
3	2	1	1	0	0	1720
14	7	2	0	0	0	24
12	9	3	0	0	0	624
10	7	1	0	0	0	-272
6	5	4	0	0	0	-4320
4	1	1	0	0	0	-2088
10	7	0	0	0	0	-1792
3	2	0	0	0	0	256

12	9	4	4	3	0	-1890
12	7	4	3	3	0	-1944
11	10	9	9	6	0	1080
11	10	3	3	2	0	-936
11	8	3	3	1	0	-216
10	10	7	7	1	0	-36
10	9	8	7	4	0	36
10	9	3	2	2	0	-192
10	8	5	4	1	0	-648
10	7	4	4	4	0	189
10	7	1	1	1	0	51
9	9	6	2	1	0	-216
9	8	4	4	1	0	-585
9	7	6	6	5	0	1512
8	8	5	4	3	0	-1512
8	7	3	3	2	0	192
6	5	4	4	1	0	-6480
4	4	4	3	2	0	-2160
4	1	1	1	1	0	36
14	13	12	11	0	0	-12960
14	11	9	8	0	0	-3600
14	11	1	1	0	0	2052
14	7	2	1	0	0	-4764
13	12	4	4	0	0	4536
13	11	6	4	0	0	12636
13	8	2	1	0	0	4764
12	9	3	1	0	0	-3612
11	9	6	3	0	0	-9792
10	9	4	2	0	0	192
10	7	4	4	0	0	-2016
9	9	6	2	0	0	624
9	8	1	1	0	0	-628
8	7	3	1	0	0	-144
6	5	3	2	0	0	11808
4	4	1	1	0	0	756
1	1	1	1	0	0	348
13	12	4	0	0	0	1296
11	10	3	0	0	0	1008
9	8	4	0	0	0	2328
6	5	1	0	0	0	-62784
3	2	1	0	0	0	1776
9	8	0	0	0	0	8064
1	1	0	0	0	0	-4464

12	9	4	3	1	0	-1296
12	7	3	3	1	0	-216
11	10	9	8	3	0	-252
11	10	3	1	1	0	-234
11	7	6	3	3	0	-1296
10	10	5	4	2	0	1512
10	9	8	7	1	0	72
10	9	2	1	1	0	60
10	8	5	3	2	0	-504
10	7	4	4	1	0	901
9	9	8	8	4	0	108
9	8	8	7	3	0	96
9	8	4	3	2	0	396
9	7	6	4	4	0	378
8	8	5	3	1	0	648
8	7	3	1	1	0	-60
6	5	4	3	2	0	-13824
4	4	4	1	1	0	-324
3	3	2	2	1	0	-672
14	12	9	7	0	0	7920
14	11	6	5	0	0	1944
14	9	2	2	0	0	1344
13	13	12	12	0	0	3889
13	12	4	1	0	0	756
13	11	6	1	0	0	-324
13	7	6	2	0	0	10224
12	7	3	3	0	0	1248
11	8	3	3	0	0	-240
10	9	2	1	0	0	336
10	7	4	1	0	0	-3036
9	8	6	5	0	0	8856
9	7	6	4	0	0	108
7	7	6	3	0	0	1248
6	5	1	1	0	0	4320
4	3	2	1	0	0	-1728
14	12	5	0	0	0	67176
13	12	1	0	0	0	10440
10	9	2	0	0	0	-1664
9	8	1	0	0	0	-304
4	4	4	0	0	0	1512
1	1	1	0	0	0	-904
6	5	0	0	0	0	102528
4	0	0	0	0	0	-57408

12	9	3	3	2	0	936
11	10	10	10	5	0	-216
11	10	6	5	3	0	3240
11	9	6	4	3	0	-3888
10	10	9	7	2	0	96
10	10	5	2	1	0	-216
10	9	7	7	6	0	-468
10	8	7	7	3	0	-96
10	8	5	1	1	0	522
10	7	4	3	2	0	103
9	9	8	8	1	0	-36
9	8	6	5	4	0	756
9	8	4	1	1	0	-441
9	7	6	4	1	0	216
8	7	6	5	3	0	144
7	7	6	4	3	0	648
6	5	4	1	1	0	-2592
4	4	1	1	1	0	350
3	2	1	1	1	0	-288
14	12	5	4	0	0	-37908
14	11	4	4	0	0	5100
14	8	7	7	0	0	-96
13	12	10	7	0	0	-3024
13	12	3	2	0	0	2446
13	10	2	2	0	0	-2352
12	10	9	9	0	0	-96
11	10	10	9	0	0	192
10	10	7	7	0	0	416
10	8	5	4	0	0	1404
10	7	3	2	0	0	-704
9	8	4	4	0	0	2268
9	7	6	1	0	0	-6164
6	6	5	5	0	0	10368
4	4	4	4	0	0	648
4	1	1	1	0	0	-1152
14	11	4	0	0	0	-2376
13	11	6	0	0	0	-6264
10	8	5	0	0	0	-11640
9	7	6	0	0	0	13800
4	4	1	0	0	0	-13752
14	11	0	0	0	0	30240
4	4	0	0	0	0	8352
1	0	0	0	0	0	4352

elements. These were the lowest-degree pair that could be constructed from the scalars $C^{(202)}$, $C^{(112)}$, $C^{(312)}$, $C^{(024)}$, and $C^{(204)}$, but we cannot exclude a pair which includes $C^{(114)}$. Thus for the lowest-degree commuting scalars the functionally independent scalars, i.e., the first four of the above scalars, are insufficient, although presumably a pair of higher-degree commuting scalars could be constructed from them.

It is intended to use the shift operator techniques of Hughes and Yadegar[7] to obtain the eigenvalues of Y_1 and Y_2 for IR of $G(2)$, and these shift operators will be constructed from the various tensorial types listed in Sec. III.

ACKNOWLEDGMENTS

Both authors thank G. Vanden Berghe and H. De Meyer of the Rijksuniversiteit, Gent, for valuable discussions and encouragement, and for making available to us their FORTRAN programs.

One of us (J VdJ) would like to acknowledge partial financial support in the form of a scholarship from the British Council.

[1]R. T. Sharp and S. C. Pieper, J. Math. Phys. **9**, 663 (1968).
[2]J. W. B. Hughes, J. Math. Phys. **24**, 1015 (1983).
[3]A. Peccia and R. T. Sharp, J. Math. Phys. **17**, 1313 (1976).
[4]J. W. B. Hughes, J. Phys. A **6**, 48 (1973).
[5]J. W. B. Hughes, J. Phys. A **6**, 281 (1973).
[6]B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[7]J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).
[8]H. E. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981).
[9]H. E. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2366 (1981).
[10]J. Van der Jeugt, H. E. De Meyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1025 (1983).
[11]M. Moshinsky and J. G. Nagel, Phys. Lett. **5**, 173 (1963).
[12]C. Quesne, J. Math. Phys. **17**, 1452 (1976).
[13]J. Van der Jeugt, H. E. De Meyer, and G. Vanden Berghe, Ann. Phys. **147**, 85 (1983).
[14]C. Quesne, J. Math. Phys. **18**, 1210 (1977).
[15]A. Partensky and C. Maguin, J. Math. Phys. **19**, 511 (1978).
[16]A. M. Bincer, J. Math. Phys. **24**, 1695 (1983).
[17]J. Van der Jeugt, J. Math. Phys. **25**, 1219 (1984).
[18]R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).
[19]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).
[20]H. E. De Meyer and G. Vanden Berghe (private communication).
[21]H. E. De Meyer (private communication).

Application of generating function techniques to Lie superalgebras

R. T. Sharp,[a] J. Van der Jeugt,[b] and J. W. B. Hughes
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 20 September 1984; accepted for publication 14 December 1984)

The construction of generating functions for multiplicities of irreducible representations of Lie superalgebras from generating functions for characters is examined, and applied in order to obtain polynomial tensors and branching rules. Problems arising because of the existence of typical and atypical representations are discussed in detail. The techniques are applied to osp(1,2), spl(1,2), osp(3,2), and osp(4,2).

I. INTRODUCTION

In the last few years, generating functions have emerged as a useful tool for the solution of a number of problems in the representation theory of Lie algebras and their applications in physics.[1–6] A single generating function (GF) for a Lie algebra G of rank l has the form

$$G(A_1,\ldots,A_l) = \sum_{a_1=0}^{\infty}\cdots\sum_{a_l=0}^{\infty} f_{a_1,\ldots,a_l} A_1^{a_1}\cdots A_l^{a_l}, \tag{1.1}$$

where $A_1,\ldots,A_l$ are dummy variables which carry the Dynkin labels $(a_1,\ldots,a_l)$ of an irreducible representation (irrep) of the Lie algebra as exponents. The coefficients $f_{a_1,\ldots,a_l}$ are polynomial functions with integer coefficients in some other variables, which contain useful information about the representation $(a_1,\ldots,a_l)$. For instance, $f_{a_1,\ldots,a_l}$ may depend upon new variables $A'_1,\ldots,A'_m$, which carry as exponents the Dynkin labels $(a'_1,\ldots,a'_m)$ of irreps of a subalgebra H of G, and of which the coefficient is the multiplicity of the irrep $(a'_1,\ldots,a'_m)$ in the decomposition of $(a_1,\ldots,a_l)$. Such a generating function gives the branching rule for $G\to H$. In other cases $f_{a_1,\ldots,a_l}$ can be a polynomial in a variable U of which the exponent is the degree of polynomial tensor products of a given tensor representation of G—then we are dealing with a polynomial tensor GF. An important feature of the GF's is that they can be written in a closed form as a rational expression whose numerator is a polynomial in the dummy variables with positive integer coefficients, and whose denominator consists of products of terms $(1-X)$, where X is a product of powers of the variables. Such denominator factors are formal notations for geometric series, and the GF is of the form (1.1) only when these denominator terms are expanded in power series.

For Lie algebras there is a one-to-one correspondence between the labels $(a_1,\ldots,a_l)$ of its finite-dimensional irreducible representations and the set of multiplets of l non-negative integers. Hence the generating function techniques are a natural way of dealing with problems in the representation theory of Lie algebras.

For representations of Lie superalgebras, some difficulties arise. A first problem is that the labels $(a_1;\ldots;a_l)$ of finite-dimensional irreps of superalgebras[7,8] are not all integers, and that there are certain constraints (or "consistency conditions") on a set of numbers $(a_1;\ldots;a_l)$ if they are to correspond to an irrep of finite dimension. A second problem is that Lie superalgebras have typical and atypical representations,[7,8] which have to be treated separately. Obviously, one cannot restrict oneself to typical representations since, for instance, in the decomposition of the tensor product of two typical representations, also atypical (or, more generally, indecomposable) representations can appear.[9] In this paper we show how to deal with such difficulties.

In Sec. II we examine the general construction of GF's for Lie superalgebras, starting from a character generating function. Section III contains results for osp(1,2), a Lie superalgebra which has only typical representations. In Sec. IV generating functions are given for spl(1,2), which is a Lie superalgebra of type I (See Refs. 8 and 10). Sections V and VI give results for two superalgebras of type II (see Refs. 8 and 11), namely osp(3,2) and osp(4,2). Section VII contains some concluding remarks.

II. CONSTRUCTION OF GENERATING FUNCTIONS FOR LIE SUPERALGEBRAS

The starting point for generating functions for Lie algebras[2] is a generating function $F(\eta) = F(\eta_1,\ldots,\eta_l)$ for weights with respect to a Lie algebra G of rank l. The exponents of $\eta_1,\ldots,\eta_l$ are the components of weights in the weight space for G. One assumes that the weights in F form complete irreps of G, and hence one may write

$$F(\eta) = \sum_{\lambda} \chi_\lambda(\eta) N_\lambda, \tag{2.1}$$

where χ_λ is the character of the irrep $\lambda = (\lambda_1,\ldots,\lambda_l)$, and N_λ is the multiplicity of λ in F (N_λ usually depends upon other dummy variables). The characteristic function ξ_λ is related to χ_λ by[12]

$$\chi_\lambda = \xi_\lambda/\Delta, \tag{2.2}$$

with Δ the characteristic of the scalar irrep. Hence, we have

$$\Delta(\eta)F(\eta) = \sum_{\lambda} \xi_\lambda(\eta) N_\lambda. \tag{2.3}$$

But $\xi_\lambda(\eta)$ has only one term whose weight is in the dominant Weyl sector, say $\Pi_i \eta_i^{M_i}$, where the M_i depend linearly on λ_j (with integer coefficients). Hence we multiply (2.3) by $\Pi_i \eta_i^{-M_i} \Pi_j A_j^{\lambda_j}$, sum over $\lambda_1,\ldots,\lambda_l$ from 0 to ∞, and keep only the terms which are independent of the η_i. This simple

[a] Permanent address: Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada.
[b] Research assistant N. F. W. O. (Belgium). Permanent address: Seminarie voor Wiskundige Natuurkunde, R.U.G., Krijgslaan 281-S9, B-9000 Gent, Belgium.

 0022-2488/85/050901-12$02.50

instruction transfers a GF for weights into the GF for irreducible tensors.

For Lie superalgebras the situation is more complicated for two reasons: (1) there is no one-to-one correspondence between the set of multiplets of l integers and the labels $(a_1;...;a_l)$ of the irreps; and (2) there are typical and atypical representations, for which completely different character formulas have to be used. In fact, characters for atypical representations of classical Lie superalgebras are not known in general, except for superalgebras of series A and C (see Ref. 13).

Classical Lie superalgebras[7] are divided into two classes: class 1 includes the superalgebras $\mathrm{spl}(m,n)$ and $\mathrm{osp}(2,2n)$; and class II consists of $\mathrm{osp}(m,n)$ $(m \neq 2)$ and the exceptional algebras $\mathrm{osp}(4,2;\alpha)$, $F(4)$, and $G(3)$. One can choose a Chevalley basis for the Lie superalgebra, in which the basis elements are given by $h_i (i = 1,...,l;\ l =$ rank of the superalgebra) and elements associated with positive and negative roots. Kac[7,8] showed that it is always possible to take a set of l simple roots such that only one simple root is odd. Let e_i $(i = 1,...,l)$ be the corresponding root vectors; then the Cartan matrix

$$c_{ij} = 2(e_i, e_j)/(e_j, e_j) \tag{2.4}$$

is determined by a Dynkin diagram and (e_i, e_j) is the bilinear form in weight space induced from a fixed nondegenerate bilinear form on L (see Ref. 8). Kac gives a table of Dynkin diagrams for all classical Lie superalgebras.[9] They consist of a set of nodes, connected by lines, and only one node (the colored node) corresponds to the odd simple root.

Finite-dimensional irreducible representations of Lie superalgebras are determined by their highest weight λ, and characterized by the "Dynkin labels" $(a_1;...;a_l)$ associated with the nodes of the Dynkin diagram. All numbers a_i are non-negative integers, except for the label a_s associated with the odd node. For Lie superalgebras of class I, a_s can be any complex number, and vice versa; with every complex number and set of $l - 1$ non-negative integers there corresponds one finite-dimensional irrep. In fact, class I superalgebras have a u(1) subalgebra in their Lie algebraic part, and a_s is the u(1) label associated with the highest weight. When dealing with generating functions, however, it is natural to consider only integer values of labels (because labels appear as exponents of some dummy variables in polynomial expressions). Therefore, we shall restrict ourselves to the representations for which $a_s \in \mathbb{Z}$. This restriction is not as drastic as it seems at first sight; branching rules for irreps with $a_s \notin \mathbb{Z}$ can be easily deduced from those with a_s an integer. Polynomial tensors in an irreducible representation are most interesting when the representation is the defining or the adjoint representation, and these irreps have $a_s \in \mathbb{Z}$.

For Lie superalgebras of class II, a_s cannot belong to a continuous range of values, as was the case for class I, but we still do not have the property $a_s \in \mathbb{Z}$. However, the irreps are now also characterized by the $l - 1$ labels $a_i (i \neq s)$ and a label b, which is a linear combination of some a_i's (including a_s) and which satisfies $b \in \mathbb{Z}$. Kac[7,8] gives a list of these linear combinations of the a_i's for all superalgebras of class II. So, instead of using $(a_1;...;a_l)$ as labels, we rather use $(\lambda_1;...;\lambda_l) = (a_1;...;a_{s-1};\ b;\ a_{s+1};...;a_l)$. This is a set of l non-negative integers, and they are easy to deal with in the GF technique. Unfortunately, not every set of l non-negative integers corresponds to an irrep of the Lie superalgebra: there are certain "consistency conditions" for low values of b (see Kac[7,8] for general consistency conditions, and Secs. V and VI for examples). This does not really hamper the GF technique; one can sum over all integer values of $a_1,...,a_{s-1}$, b, $a_{s+1},...,a_l$, and in the final result disregard the terms which do not correspond to irreps of the Lie superalgebra.

The most peculiar feature of Lie superalgebras is that they have so-called typical and atypical representations[7,8] [except for the superalgebras $\mathrm{osp}(1,2n)$, which have only typical irreps]. Typical representations have properties analogous to those of finite-dimensional irreps of Lie algebras, and are relatively easy to handle. Atypical representations are much harder to deal with. One cannot avoid atypical representations by restricting oneself to typical irreps; in the decomposition of the tensor product of two typical representations, atypical parts can appear. Let $L = L_0 + L_1$ be a Lie superalgebra with M odd positive roots; then the decomposition of a typical irrep of L into representations of L_0 contains in general 2^M L_0 irreps. An atypical representation does not satisfy this property; the atypical irreps in Secs. IV–VI contain 2^{M-1} L_0 irreps. In fact, Lie superalgebras also have indecomposable (i.e., reducible but not completely reducible) representations. The invariant space and factor space of such indecomposable representations consist of two "neighboring" atypical representations; the indecomposable representations have then the same shape as a typical representation. Obviously, the weights of such an indecomposable representation are the same as the weights of the direct sum of two atypical irreducible representations. In the GF technique, one deals with transforming weights into irreducible tensors. Therefore, whenever an indecomposable representation appears (e.g., in the decomposition of a tensor product), the GF technique transforms it into two atypical irreps. So, instead of finding an indecomposable representation, we find its invariant space and its factor space.

Let us finally discuss how to construct GF's for Lie superalgebras. If $L = L_0 + L_1$ is a Lie superalgebra, let Δ denote the set of roots of L, Δ_0 the even roots, Δ_1 the odd roots, and $\overline{\Delta}_1$ the odd roots β for which 2β is not an even root. A superscript $+$ implies the positive roots in each set. We define

$$\rho_0 = \frac{1}{2} \sum_{\alpha \in \Delta_0^+} \alpha, \quad \rho_1 = \frac{1}{2} \sum_{\beta \in \Delta_1^+} \beta, \quad \rho = \rho_0 - \rho_1. \tag{2.5}$$

Then an irrep with highest weight λ is typical if and only if[8]

$$(\lambda + \rho, \beta) \neq 0, \quad \forall \beta \in \overline{\Delta}_1^{\,+}. \tag{2.6}$$

In particular, we have[8]

$$(\beta, \beta) = 0 \Leftrightarrow \beta \in \overline{\Delta}_1. \tag{2.7}$$

The typicality conditions (2.6) can be translated in terms of linear conditions for the Dynkin labels $(a_1;...;a_l)$ of the representation; these conditions are summarized by Kac[8] for all classical Lie superalgebras. If λ is the highest weight of a typical representation of L, then the character for this representation is given by[8]

$$\chi_\lambda = K^{-1}\sum_{w\in W}\epsilon(w)\eta^{w(\lambda+\rho)}, \tag{2.8}$$

$$K^{-1} = \frac{\Pi_{\beta\in\Delta_1^+}(\eta^{\beta/2}+\eta^{-\beta/2})}{\Pi_{\alpha\in\Delta_0^+}(\eta^{\alpha/2}-\eta^{-\alpha/2})}, \tag{2.9}$$

where $\eta = (\eta_1,...,\eta_l)$ are dummy variables carrying the components of weights as exponents, W is the Weyl group of the even part L_0, and $\epsilon(w)$ is the sign of w. Character formulas for atypical representations are not known in general, and for the particular cases where they are known they look quite different from the expression (2.8). Therefore, difference prescriptions must be given in order to obtain GF's for typical or atypical representations. For the Lie superalgebras we deal with in Secs. III–VI, typical as well as atypical representations are known[14–20], and we find the following expression for the character of the atypical representations of L [where L = osp(1,2), spl(1,2), osp(3,2), or osp(4,2)]. Let $\beta_i\in\bar{\Delta}_1^+ = \{\beta_1,...,\beta_n\}$, λ be the highest weight of a representation, and

$$(\lambda+\rho,\beta_i) = 0. \tag{2.10}$$

Then this representation is atypical (of type i: at i), and

$$\chi_\lambda^{(\text{at i})} = \frac{\Pi_{\beta\in\Delta_1^+\setminus\bar{\Delta}_1^+}(\eta^{\beta/2}+\eta^{-\beta/2})}{\Pi_{\alpha\in\Delta_0^+}(\eta^{\alpha/2}-\eta^{-\alpha/2})}$$

$$\times\sum_{\beta_j\in\bar{\Delta}_1^+}(\eta^{\beta_1/2}+\eta^{-\beta_1/2})\cdots\widehat{(\eta^{\beta_j/2}+\eta^{-\beta_j/2})}$$

$$\times\cdots(\eta^{\beta_n/2}+\eta^{-\beta_n/2})\sum_{w\in W_{ij}}\epsilon(w)\eta^{w(\lambda+\rho+\frac{1}{2}\beta_i)}, \tag{2.11}$$

where $W_{ij} = \{w\in W\,|\,w(\beta_j) = \pm\beta_i\}$, and $\widehat{\ }$ is written above the term to be deleted.

Just as for Lie algebras, the starting point is now a generating function $F(\eta) = F(\eta_1,...,\eta_l)$ for weights, such that the weights in $F(\eta)$ form complete irreducible representations of the Lie superalgebra L. This includes of course typical (t) and atypical (at) representations

$$F(\eta) = \sum_{\lambda(\text{t})}\chi_\lambda(\eta)N_\lambda + \sum_i\sum_{\mu(\text{at i})}\chi_\mu^{(\text{at i})}(\eta)N_\mu. \tag{2.12}$$

Multiply $F(\eta)$ by $K(\eta)$

$$K(\eta) = \frac{\eta^\rho\Pi_{\alpha\in\Delta_0^+}(1-\eta^{-\alpha})}{\Pi_{\beta\in\Delta_1^+}(1+\eta^{-\beta})}. \tag{2.13}$$

In this multiplication, a denominator factor of the form $(1+\eta^{-\beta})$ is a formal notation for the power series expansion $1-\eta^{-\beta}+\eta^{-2\beta}-\cdots = \Sigma_{k=0}^\infty(-\eta^{-\beta})^k$. This transforms (2.12) into

$$K(\eta)F(\eta) = \sum_{\lambda(t)}\Big(\sum_{w\in W}\epsilon(w)\eta^{w(\lambda+\rho)}\Big)N_\lambda$$

$$+\sum_i\sum_{\mu(\text{at i})}K(\eta)\chi_\mu^{(\text{at i})}(\eta)N_\mu. \tag{2.14}$$

The coefficient of N_λ has only one term with weight in the dominant Weyl sector, say $\Pi_i\eta_i^{M_i}$, and the weight coordinates are chosen such that the M_i depend linearly on the representation labels $\lambda_1,...,\lambda_l$ with integer coefficients. Before we do the same trick as with GF's for Lie algebras, we have to investigate whether the dominant terms (i.e., terms with weights in the dominant Weyl sector) in the second part of (2.14) do not interfere with the dominant terms in the first part. The dominant terms in the second part of (2.14) come from

$$\sum_i\sum_{\mu(\text{at i})}\frac{\eta^{\mu+\rho}}{(1+\eta^{-\beta_i})}N_\mu, \tag{2.15}$$

which gives terms of the form

$$\eta^{\mu+\rho}-\eta^{\mu-\beta_i+\rho}+\eta^{\mu-2\beta_i+\rho}-\cdots. \tag{2.16}$$

But, since μ is of type (at i), $(\mu+\rho,\beta_i)=0$, and (2.7) then shows $(\mu+\rho-k\beta_i,\beta_i)=0$. Hence, all the terms in (2.16) correspond to highest weights of atypical representations. Consequently, the dominant terms in the second part of (2.14) will never interfere with dominant terms corresponding to typical representations. We therefore obtain the following prescription: multiply (2.14) by $\Pi_i\eta_i^{-M_i}A_i^{\lambda_i}$, sum over $\lambda_1,..., \lambda_l$ from 0 to ∞, and keep only the terms which are independent of the η_i. This will transform a GF $F(\eta)$ for weights into a GF $G'(A)$ for typical representations. Because we sum over all λ_i values from 0 to ∞, $G'(A)$ consists of three parts.

(1) This part is a generating function $G(A)$ where the exponents of $A_1,..., A_l$ are labels of typical representations only.

(2) This part is a function where the exponents of $A_1,..., A_l$ correspond to atypical representations only. This part, however, does not give the correct multiplicity for the atypical irreps since (2.16) shows that there is interference among the atypical representations themselves.

(3) This part is a function where the exponents of $A_1,..., A_l$ are labels of nonexistent representations (i.e., labels which do not satisfy the consistency conditions). These terms appear because we summed over all non-negative integer values of $\lambda_1,..., \lambda_l$. However, the atypicality conditions and the consistency conditions are well known in terms of the labels $(\lambda_1;...;\lambda_l)$, so it is easy to subtract parts (2) and (3) from $G'(A)$, leaving only $G(A)$. So, we have given a general prescription for finding a GF for multiplicities of typical tensors, starting from a GF for weights. This solves the problem only partially; we also need to know how to find the GF part for atypical representations. It is fairly difficult to treat this in general, but Secs. IV–VI show how to deal with this problem in practice.

III. GENERATING FUNCTIONS FOR OSP(1,2)

The Lie superalgebra osp(1,2) contains as even part the Lie algebra so(3) and as odd part the two-dimensional so(3) tensor (spinor). All representations of osp(1,2) are typical. The osp(1,2) irreps are known to be "dispin,"[14,15] i.e., when decomposed into so(3) irreps they reduce as $(a)+(a-1)$ [we use twice the angular momentum label for so(3) irreps]. The Dynkin label of the osp(1,2) irrep is then (a) $(a\in\mathbf{Z}_+)$; (0) is the trivial representation, (1) the defining representation, and (2) the adjoint representation. There are two possible gradings for a representation space: the so(3) parts (a) even and $(a-1)$ odd $[(a)_0+(a-1)_1]$, or (a) odd and $(a-1)$ even $[(a)_1+(a-1)_0]$.

The character (2.8) for a representation (a) of osp(1,2) is given by

$$\chi_{(a)} = (\eta^{a+1} - \eta^{-a-1} + \eta^{a} - \eta^{-a})/(\eta - \eta^{-1}). \quad (3.1)$$

From this we obtain the character generator

$$F(\eta; A) = \sum_{a=0}^{\infty} \chi_{(a)} A^{a} = \frac{1+A}{(1-A\eta)(1-A\eta^{-1})}. \quad (3.2)$$

Since osp(1,2)$\supset$ so(3), (3.2) is also a GF for so(3) weights [of complete so(3) irreps] and (2.3) shows how to transform so(3) weights into so(3) tensors:

$$G(L; A) = \left[(\eta - \eta^{-1})F(\eta; A) \sum_{l=0}^{\infty} \eta^{-l-1} L^{l}\right]_{\mathrm{ex}(\eta^0)}. \quad (3.3)$$

The ex(η^0) is an instruction to keep only the terms independent of η. We obtain

$$G(L; A) = (1 + A)/(1 - AL), \quad (3.4)$$

which is the branching rule GF for osp(1,2) irreps (a) into so(3) irreps (l). Let $G(L; A) = \Sigma_{a=0}^{\infty}(\Sigma_l c_{al} L^l) A^a$; then c_{al} is the multiplicity of the so(3) irrep (l) in the decomposition of (a). Clearly, (3.4) gives the dispin structure again.

If $H(A)$ is an osp(1,2) GF, then the corresponding so(3) GF $K(L)$ is found by "substituting" the osp(1,2)$\to$so(3) branching rule GF (3.4)

$$\begin{aligned} K(L) &= [H(A)G(L; A^{-1})]_{\mathrm{ex}(A^0)} \\ &= H(L) + [H(L) - H(0)]/L. \end{aligned} \quad (3.5)$$

This implies

$$H(A) = [AK(A) + K(-1)]/(A+1). \quad (3.6)$$

Hence, if an so(3) GF $K(L)$ is given, of which we know that the so(3) tensors form complete parts of osp(1,2) tensors, then (3.6) gives a prescription to transform $K(L)$ into an osp(1,2) GF $H(A)$.

Instead of going through so(3), we can also transform a GF $F(\eta)$ for osp(1,2) weights directly into a GF $G(A)$ for osp(1,2) tensors by using (2.14):

$$G(A) = \{F(\eta)[(1-\eta)/(1-A\eta)]\}_{\mathrm{ex}(\eta^0)}. \quad (3.7)$$

Let us now consider some GF's for polynomial tensors. In the tensor product of a given irreducible tensor, we maintain only the "supersymmetric" parts (i.e., the parts which are symmetric with respect to the even basis and antisymmetric with respect to the odd basis); when applied to the adjoint representation this gives the enveloping algebra. The weight generating function for polynomial tensors in the osp(1,2) irrep (1), consisting of the so(3) parts $(1)_0 + (0)_1$, is given by

$$F_{(1)}(\eta; U) = (1 + U)/(1 - U\eta)(1 - U\eta^{-1}). \quad (3.8)$$

Prescription (3.7) transforms this into the GF

$$G_{(1)}(U; A) = 1/(1 - UA). \quad (3.9)$$

In (3.8) and (3.9), U carries the degree of the polynomial tensor product in the irrep (1). If we choose the opposite grading for (1), namely $(\bar{1}) = (1)_1 + (0)_0$, then the character GF is

$$F_{(\bar{1})}(\eta; U) = (1 + U\eta)(1 + U\eta^{-1})/(1 - U), \quad (3.10)$$

and the GF for polynomial tensors is given by

$$G_{(\bar{1})}(U; A) = 1 + (UA + U^2)/(1 - U). \quad (3.11)$$

Polynomial tensors in the osp(1,2) representation $(2) = (2)_0 + (1)_1$ are generated by

$$G_{(2)}(U; A) = (1 + U^2 A)/(1 - U^2)(1 - UA^2). \quad (3.12)$$

Since $(2) = (2)_0 + (1)_1$ is the adjoint representation of osp(1,2), (3.12) is the GF for osp(1,2) tensors contained in the enveloping algebra of osp(1,2). The scalars in the enveloping algebra are then generated by $(1 - U^2)^{-1}$, which shows that there is only one independent invariant, of second degree in the generators. The opposite grading $(\bar{2}) = (2)_1 + (1)_0$ gives as polynomial tensor GF

$$G_{(\bar{2})}(U; A) = 1 + (UA^2 + U^2A^2 + U^3 + U^3A)/(1 - UA). \quad (3.13)$$

We also determined the GF's for polynomial tensors in the seven dimensional osp(1,2) representation.

$(3) = (3)_0 + (2)_1$:

$$G_{(3)}(U; A) = \frac{[1 + U^2A^3 + U^3(1 + A + A^3 + A^4) + U^4 A + U^6 A^4]}{(1 - U^4)(1 - UA^3)(1 - U^2A^2)}, \quad (3.14)$$

$(\bar{3}) = (3)_1 + (2)_0$:

$$\begin{aligned} G_{(\bar{3})}(U; A) &= (1 - U^2)^{-1} + [UA^3 + U^2(A + A^4) + U^3(A^2 + A^4) \\ &\quad + U^4(1 + A + A^3)]/(1 - U^2)(1 - UA^2). \end{aligned} \quad (3.15)$$

Finally, we consider the Clebsch–Gordan GF for osp(1,2). In (3.2), the character GF $F(\eta; A)$ is given. Since the weights of the tensor product of two irreps (a_1) and (a_2) are given by the superposition of the weights of (a_1) on those of (a_2), and (3.7) shows us how to transform weights into tensors, we find as the Clebsch–Gordan GF

$$G(A_1, A_2; A) = \left[F(\eta; A_1)F(\eta; A_2)\frac{(1-\eta)}{(1-\eta A)}\right]_{\mathrm{ex}(\eta^0)},$$

or

$$G(A_1, A_2; A) = \frac{(1 + A_1A_2A)}{(1 - A_1A_2)(1 - A_1A)(1 - A_2A)}. \quad (3.16)$$

In the expansion

$$G(A_1, A_2; A) = \sum_{a_1=0}^{\infty} \sum_{a_2=0}^{\infty} \left(\sum_a c_{a_1a_2a} A^a\right) A_1^{a_1} A_2^{a_2},$$

$c_{a_1a_2a}$ is the multiplicity of the representation (a) in the tensor product of (a_1) and (a_2).

IV. GENERATING FUNCTIONS FOR SPL(1,2)

The even part of spl(1,2) is u(1) + su(2). The odd part consists of two spinor representations with respect to su(2), with u(1) eigenvalues $+1$ and -1, respectively, and denoted by (1,1) and $(-1,1)$. The positive roots are given by $\Delta_0^+ = \{(0,2)\}$ and $\Delta_1^+ = \{(1,1), (1,-1)\}$. Irreducible representations of spl(1,2) have been studied by Scheunert *et al.*,[15] Marcu,[16] and Hughes.[17] Let $(a_1; a_2)$ be the Dynkin labels of an spl(1,2) irrep; then $a_1 \in \mathbb{C}$ and $a_2 \in \mathbb{N}$. For reasons discussed in Sec. II we restrict ourselves to representations with $a_1 \in \mathbb{Z}$, and we denote the labels by $(k; j) = (a_1; a_2)$.

When decomposed into u(1) + su(2) irreps (b, l), k is the maximum b value and j the corresponding su(2) label l. Also, (k, j) are the components of the highest weight of $(k;j)$ in the weight space determined by the positive roots.

The elements of $\bar{\Delta}_1^+$ are (1,1) and (1, − 1), and with (2.6) they give rise to the atypicality conditions

$$k = j + 2 \quad \text{(at 1)}, \tag{4.1}$$

$$k = -j \quad \text{(at 2)}, \tag{4.2}$$

respectively. The asymmetry in (4.1) and (4.2) arises because of the conventional[8] labeling of the irreps: if the irreps of spl(1,2) were labeled by $(k';j')$ with j' = maximum l value and k' the corresponding b value in the decomposition into u(1) + su(2) irreps (b,l), then the atypicality conditions would read $k' = j'$ and $k' = -j'$.

By means of (2.8) and (2.11), the characters for typical and atypical representations are determined [η_1 and η_2 carry the weights in the u(1) and su(2) direction, respectively]

$$\chi^{(t)}_{(k;j)} = \frac{(1+\eta_1\eta_2^{-1})(1+\eta_1^{-1}\eta_2^{-1})\eta_1^{k-1}(\eta_2^{j+1} - \eta_2^{-j-1})}{(1-\eta_2^{-2})} \tag{4.3}$$

$$(k \in \mathbf{Z},\ j \in \mathbf{N},\ k \neq j+2,\quad k \neq -j),$$

$$\chi^{(\text{at }1)}_{(k;j)} = \frac{(1+\eta_1^{-1}\eta_2)\eta_1^k\eta_2^j - (1+\eta_1^{-1}\eta_2^{-1})\eta_1^k\eta_2^{-j-2}}{(1-\eta_2^{-2})} \tag{4.4}$$

$$(k = j+2,\ j \in \mathbf{N}),$$

$$\chi^{(\text{at }2)}_{(k;j)} = \frac{(1+\eta_1^{-1}\eta_2^{-1})\eta_1^k\eta_2^j - (1+\eta_1^{-1}\eta_2)\eta_1^k\eta_2^{-j-2}}{(1-\eta_2^{-2})} \tag{4.5}$$

$$(k = -j,\ j \in \mathbf{N}).$$

Note that (4.5) also includes the scalar representation (0;0).

Character generating functions can now be determined from (4.3)–(4.5). The character GF for typical representations $(k;j)$ with $k \geqslant 0$ (t +) is found by taking

$$\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \chi^{(t)}_{(k;j)} K^k J^j, \tag{4.6}$$

which is

$$\frac{\eta_1^{-1}\eta_2(1+\eta_1\eta_2^{-1})(1+\eta_1^{-1}\eta_2^{-1})}{(1-\eta_1 K)(1-\eta_2 J)(1-\eta_2^{-1}J)}. \tag{4.7}$$

In (4.7), the contribution in K^0J^0 and in the atypical irreps $(j+2;j)$ must be subtracted. This results in

$$\begin{aligned} F^{(t+)}(K, J; \eta_1, \eta_2) &= \eta_1^{-1}\eta_2(1+\eta_1\eta_2^{-1})(1+\eta_1^{-1}\eta_2^{-1}) \\ &\quad \times [1/(1-\eta_1 K)(1-\eta_2 J)(1-\eta_2^{-1}J) \\ &\quad - K^2\eta_1^2/(1-\eta_1\eta_2 KJ)(1-\eta_1\eta_2^{-1}KJ) - 1]. \end{aligned} \tag{4.8}$$

The character generating function for typical representations $(k;j)$ with $k<0$ (t −) is found by summing

$$\sum_{k=-1}^{-\infty}\sum_{j=0}^{\infty} \chi^{(t)}_{(k;j)} K^k J^j,$$

and subtracting the contribution from the atypical representations $(-j;j)$. This gives

$$\begin{aligned} F^{(t-)}(K, J; \eta_1, \eta_2) &= \eta_1^{-2}\eta_2(1+\eta_1\eta_2^{-1})(1+\eta_1^{-1}\eta_2^{-1}) \\ &\quad \times [K^{-1}/(1-\eta_1^{-1}K^{-1})(1-\eta_2 J)(1-\eta_2^{-1}J) \\ &\quad - K^{-1}/(1-\eta_1^{-1}\eta_2^{-1}K^{-1}J)(1-\eta_1^{-1}\eta_2 K^{-1}J)]. \end{aligned} \tag{4.9}$$

Equations (4.4) and (4.5) are used in order to obtain character GF's for atypical representations

$$F^{(\text{at }1)}(K, J; \eta_1, \eta_2) = \frac{\eta_1^2 K^2[1+\eta_1^{-1}(\eta_2+\eta_2^{-1}) - KJ]}{(1-\eta_1\eta_2 KJ)(1-\eta_1\eta_2^{-1}KJ)}, \tag{4.10}$$

$$F^{(\text{at }2)}(K, J; \eta_1, \eta_2) = \frac{1+\eta_1^{-2}K^{-1}J}{(1-\eta_1^{-1}\eta_2 K^{-1}J)(1-\eta_1^{-1}\eta_2^{-1}K^{-1}J)}. \tag{4.11}$$

The complete character generating function is then

$$F(K, J; \eta_1, \eta_2) = F^{(t+)} + F^{(t-)} + F^{(\text{at }1)} + F^{(\text{at }2)}, \tag{4.12}$$

and when expanded in power series, the coefficient of K^kJ^j in (4.12) is the character of the spl(1,2) irrep $(k;j)$.

The GF $F(K, J; \eta_1, \eta_2)$ can be used to obtain the GF for branching rules of spl(1,2) into u(1) + su(2). Such a function is transformed to u(1) + su(2) tensors by means of

$$[F(K, J; \eta_1, \eta_2)(1-\eta_2^2)(1-\eta_2 L)^{-1}]_{\text{ex}(\eta_2^0)}. \tag{4.13}$$

This leads to

$$\begin{aligned} H(K, J; \eta_1, L) &= (\eta_1^{-1}J + \eta_1^{-1}L + \eta_1^{-2} + 1)[(1-\eta_1 K)(1-JL)]^{-1} + K^{-1}(\eta_1^{-2}J + \eta_1^{-2}L + \eta_1^{-1} + \eta_1^{-3})[(1-\eta_1^{-1}K^{-1}) \\ &\quad \times (1-JL)]^{-1} - (1+\eta_1^{-2}+\eta_1^{-1}L) - (\eta_1^2K^3J + K^2) \\ &\quad \times (1-\eta_1 KJL)^{-1} - (\eta_1^{-2}K^{-1}L + \eta_1^{-3}K^{-2}J + \eta_1^{-1}K^{-1} + \eta_1^{-3}K^{-1} \\ &\quad - \eta_1^{-2}K^{-1}J - 1)(1-\eta_1^{-1}K^{-1}JL)^{-1}. \end{aligned} \tag{4.14}$$

When appropriately expanded in power series, (4.14) is equal to

$$H(K, J; \eta_1, L) = \sum_{k=-\infty}^{\infty}\sum_{j=0}^{\infty}\Big(\sum_{n,l} c_{kj,nl}\eta_1^n L^l\Big)K^kJ^j, \tag{4.15}$$

and $c_{kj,nl}$ is the multiplicity of the u(1) + su(2) irrep $(n;l)$ in the decomposition of the spl(1,2) irrep $(k;j)$.

Finally we consider the problem of transforming an spl(1,2) weight GF $F(\eta_1, \eta_2)$, of which we may assume that the weights in $F(\eta_1, \eta_2)$ form complete irreps of spl(1,2), into a GF for spl(1,2) representations. In Sec. II we showed that

$$\left[F(\eta_1,\eta_2)\frac{(1-\eta_2^{-2})\eta_1\eta_2^{-1}}{(1+\eta_1\eta_2^{-1})(1+\eta_1^{-1}\eta_2^{-1})(1-\eta_1^{-1}K)(1-\eta_2^{-1}J)}\right]_{\mathrm{ex}(\eta^0)}=\sum_{k,j=0}^{\infty}N_{kj}K^kJ^j, \tag{4.16}$$

and

$$\left[F(\eta_1,\eta_2)\frac{(1-\eta_2^{-2})\eta_1\eta_2^{-1}}{(1+\eta_1\eta_2^{-1})(1+\eta_1^{-1}\eta_2^{-1})(1-\eta_1K^{-1})(1-\eta_2^{-1}J)}\right]_{\mathrm{ex}(\eta^0)}=\sum_{k,j=0}^{\infty}N_{kj}K^{-k}J^j \tag{4.17}$$

gives the correct multiplicity N_{kj} (where N_{kj} is a function of some other variables) for all typical representations $(k;j)$ with $k\geqslant 0$ $(k\neq j+2)$ and for all typical irreps $(k;j)$ with $k\leqslant 0$ $(k\neq -j)$, respectively. Consider now the multiplication of $F(\eta_1,\eta_2)$ by

$$(1-\eta_2^{-2})\eta_1\eta_2^{-1}/(1+\eta_1\eta_2^{-1}), \tag{4.18}$$

where $(1+\eta_1\eta_2^{-1})^{-1}=1-\eta_1\eta_2^{-1}+\eta_1^2\eta_2^{-2}-\cdots$. In (4.19) we list the terms with weight in the dominant Weyl sector (i.e., with η_2 exponent positive) after a character is multiplied by (4.18)

$$\chi^{(\mathrm{t})}_{(k;j)}:\eta_1^k\eta_2^j+\eta_1^{k-1}\eta_2^{j-1}, \tag{4.19a}$$

$$\chi^{(\mathrm{at}\,1)}_{(k;j)}:\eta_1^k\eta_2^j, \tag{4.19b}$$

$$\chi^{(\mathrm{at}\,2)}_{(k;j)}:(1+\eta_1^{-1}\eta_2^{-1})[\eta_1^{-j+1}\eta_2^{j-1}-\eta_1^{-j+2}\eta_2^{j-2}+\cdots+(-1)^j\eta_1\eta_2]+(-1)^j. \tag{4.19c}$$

Since (4.19a) and (4.19c) do not contain any terms with weight of the form $(k,j)=(j+2,j)$ it follows that

$$\left[F(\eta_1,\eta_2)\frac{(1-\eta_2^{-2})\eta_1\eta_2^{-1}}{(1+\eta_1\eta_2^{-1})(1-\eta_1^{-1}K)(1-\eta_2^{-1}J)}\right]_{\mathrm{ex}(\eta^0)} \tag{4.20}$$

gives a GF with the correct multiplicity N_{kj} for at 1 representations $(k;j)=(j+2,j)$. Of course, (4.20) also contains terms K^kJ^j corresponding to typical representations (for which it gives the wrong answer). It is easy, however, to keep only the atypical part in (4.20). Indeed,

$$G^{(\mathrm{at}\,1)}(K,J)=\left[F(\eta_1,\eta_2)\frac{(1-\eta_2^{-2})\eta_1\eta_2^{-1}A^2}{(1+\eta_1\eta_2^{-1})(1-\eta_1^{-1}A^{-1}K)(1-\eta_2^{-1}AJ)}\right]_{\mathrm{ex}(\eta^0A^0)} \tag{4.21}$$

takes out the part of (4.20) consisting purely of terms $K^{j+2}J^j$. Hence, (4.21) is the GF corresponding to at 1 representations.

We make a similar analysis for the multiplication by

$$(1-\eta_2^{-2})/(1+\eta_1^{-1}\eta_2^{-1}). \tag{4.22}$$

The terms in the dominant Weyl sector after multiplying a character by (4.22) are given by

$$\chi^{(\mathrm{t})}_{(k;j)}:\eta_1^k\eta_2^j+\eta_1^{k-1}\eta_2^{j+1}, \tag{4.23a}$$

$$\chi^{(\mathrm{at}\,1)}_{(k;j)}:(1+\eta_1\eta_2^{-1})[\eta_1^{j+1}\eta_2^{j+1}-\eta_1^j\eta_2^j+\cdots+(-1)^j\eta_1\eta_2]+(-1)^{j+1}, \tag{4.23b}$$

$$\chi^{(\mathrm{at}\,2)}_{(k;j)}:\eta_1^k\eta_2^j. \tag{4.23c}$$

Expression (4.23a) does not contain any weights of the form $(-j,j)$, and (4.23b) only gives a contribution $\eta_1^0\eta_2^0$ of the form $\eta_1^{-j}\eta_2^j$. Hence, it follows that

$$G'(K,J)=\left[F(\eta_1,\eta_2)\frac{(1-\eta_2^{-2})}{(1+\eta_1^{-1}\eta_2^{-1})(1-\eta_1A^{-1}K^{-1})(1-\eta_2^{-1}AJ)}\right]_{\mathrm{ex}(\eta^0A^0)} \tag{4.24}$$

is the GF for at 2 representations $(-j;j)$ if $j>0$. So, we have in fact

$$G^{(\mathrm{at}\,2)}(K,J)=G'(K,J)-G'(0,0). \tag{4.25}$$

Only the multiplicity of the scalar representation remains to be determined. Since $F(\eta_1,\eta_2)$ consists of characters of spl(1,2) irreps, we see from (4.3)–(4.5) that

$$[F(\eta_1,\eta_2)(1-\eta_2^{-2})]_{\mathrm{ex}(\eta^0)}=N_{00}+N_{11}. \tag{4.26}$$

But N_{11} follows from (4.16), hence N_{00} is immediately determined from (4.26). The total GF is then found as follows: we subtract the atypical parts in (4.16) and (4.17), and add the functions $G^{(\mathrm{at}\,1)}$, $G^{(\mathrm{at}\,2)}$, and N_{00}.

This procedure is now applied in order to obtain polynomial tensor GF's in a given representation $(k';j')$ of spl(1,2). Let α be the even weights, and β be the odd weights of $(k';j')$; then the weight GF for polynomial tensors in $(k';j')$ is given by

$$F_{(k';j')}(\eta_1,\eta_2;U)=\frac{\Pi_\beta(1+U\eta^\beta)}{\Pi_\alpha(1-U\eta^\alpha)}. \tag{4.27}$$

In (4.27), U carries the degree in the tensor product. When $F_{(k';j')}(\eta_1,\eta_2;U)$ is transformed into the corresponding GF $G_{(k';j')}(K,J;U)$ for spl(1,2) tensors, then

$$G_{(k';j')}(K,J;U)=\sum_{k,j}\left(\sum_u c_{kju}U^u\right)K^kJ^j, \tag{4.28}$$

and c_{kju} is the multiplicity of the irrep $(k;j)$ in the supersymmetric tensor product of u copies of $(k';j')$. Note that $(k';j')$ admits two possible gradings: one for which the highest weight state is even [denoted by $(k';j')$], and one for which the highest weight state is odd [denoted by $(\overline{k';j'})$]. We determined the GF's (4.28) for $(k';j')=(1;0)$, $(\overline{1;0})$, and $(1;1)$

$$G_{(1;0)}(K,J;U)=1+UK/(1-UJ), \tag{4.29}$$

$$G_{(\overline{1;0})}(K,J;U)=(1+U^2+U^2JK^{-1})/(1-U^2)$$

$$+ UK/(1-U^2)(1-UK) + U^3K^{-1}/(1-U^2)(1-UK^{-1}), \tag{4.30}$$

$$G_{(1;\,1)}(K,J;U) = (1-U^2)^{-1} + (U^2+U^3+U^2K^2 + U^2K^{-1}J)/(1-U)(1-U^2) + [UKJ + U^3(J^2+K^2J^2) + U^4KJ]/(1-U) \times (1-U^2)(1-UJ^2). \tag{4.31}$$

Since (1; 1) is the adjoint representation of spl(1,2), (4.31) is the GF for irreducible tensors contained in the enveloping algebra of spl(1,2). In fact, (4.31) gives the complete structure of the spl(1,2) enveloping algebra decomposed into spl(1,2).

Another application consists in obtaining the Clebsch–Gordan GF for spl(1,2). Let us consider, for instance, the tensor product of two at 1 representations $(k_1;j_1)$ and $(k_2;j_2)$. The weight GF for their Clebsch–Gordan series is given by

$$F^{(\mathrm{at}\,1)}(K_1,J_1;\eta_1,\eta_2)F^{(\mathrm{at}\,1)}(K_2,J_2;\eta_1,\eta_2), \tag{4.32}$$

where $F^{(\mathrm{at}\,1)}$ is determined in (4.10). Then we use the above-mentioned procedure to transform (4.32) into a GF for spl(1,2) tensors. This results in

$$G(K_1,J_1;K_2,J_2;K,J) = \frac{K^4K_1^2K_2^2}{(1-K^2K_1K_2J_1J_2)(1-KK_1JJ_1)(1-KK_2JJ_2)} + \frac{K^3K_1^2K_2^2J}{(1-KK_1JJ_1)(1-KK_2JJ_2)}. \tag{4.33}$$

The coefficient $c_{k_1j_1k_2j_2kj}$ of $K_1^{k_1}J_1^{j_1}K_2^{k_2}J_2^{j_2}K^kJ^j$ in the expansion is the multiplicity of the spl(1,2) irrep $(k;j)$ in the decomposition of the tensor product of the two atypical representations $(k_1;j_1)$ and $(k_2;j_2)$.

V. GENERATING FUNCTIONS FOR OSP(3,2) AND BRANCHING RULES TO SUBALGEBRAS

The even part of osp(3,2) is sp(2) + so(3)$\simeq$su(2) + su(2). The odd part consists of a tensor of type (1,2) [i.e., spinor with respect to sp(2), vector with respect to so(3)]. The positive roots are given by $\Delta_0^+ = \{(2,0),\ (0,2)\}$ and $\Delta_1^+ = \{(1,-2),\ (1,2),\ (1,0)\}$. Irreducible representations of osp(3,2) have been studied in detail.[18,20] If $(a_1;a_2)$ are the Kac–Dynkin labels of an osp(3,2) irrep, then the label b (discussed in Sec. II) is given by

$$b = a_1 - \tfrac{1}{2}a_2. \tag{5.1}$$

We use the labels $(q;p) = (b;a_2)$ for the osp(3,2) irreps. Then q and p can be any non-negative integer, and the only consistency condition is

$$q = 0 \Rightarrow p = 0, \tag{5.2}$$

implying that representations of the form $(0;p)$ $(p\neq 0)$ do not exist. The elements of $\bar{\Delta}_1^+$ are $(1,-2)$ and $(1,2)$, and with (2.6) they give rise to the atypicality conditions

$$p + 2q = 0, \tag{5.3}$$

$$p - 2q + 2 = 0, \tag{5.4}$$

respectively. The first condition (5.3) is satisfied by the trivial representation (0; 0) only; the second condition (5.4) is satisfied by representations $(q;\,2q-2)$ $(q = 1,2,3,\ldots)$. All other representations $(q;p)$ are typical. The eigenvalues of the sp(2) and so(3) diagonal operators on the highest weight of an osp(3,2) irrep $(q;p)$ are q and p, respectively.

The character (2.8) for a typical representation is given by [η_1 and η_2 carry the weights in the sp(2) and so(3) direction, respectively]

$$\chi_{(q;\,p)} = \frac{(\eta_1^{1/2}\eta_2 + \eta_1^{-1/2}\eta_2^{-1})(\eta_1^{1/2}+\eta_1^{-1/2})(\eta_1^{1/2}\eta_2^{-1} + \eta_1^{-1/2}\eta_2)}{(\eta_1-\eta_1^{-1})(\eta_2-\eta_2^{-1})} \times (\eta_1^{q-1/2}\eta_2^{p+1} - \eta_1^{q-1/2}\eta_2^{-p-1} - \eta_1^{-q+1/2}\eta_2^{p+1} + \eta_1^{-q+1/2}\eta_2^{-p-1}). \tag{5.5}$$

The character (2.11) for an atypical representation $(q;p) = (q;\,2q-2)$ with $q\geqslant 2$ is

$$\chi^{(\mathrm{at})}_{(q;\,p)} = [(\eta_1^{1/2}\eta_2^{-1} + \eta_1^{-1/2}\eta_2)(\eta_1^q\eta_2^{p+2} + \eta_1^{-q}\eta_2^{-p-2}) - (\eta_1^{1/2}\eta_2 + \eta_1^{-1/2}\eta_2^{-1})(\eta_1^{-q}\eta_2^{p+2} + \eta_1^q\eta_2^{-p-2})] \times [(\eta_1^{1/2}-\eta_1^{-1/2})(\eta_2-\eta_2^{-1})]^{-1}. \tag{5.6}$$

Equation (5.6) is not valid for the "truncated" atypical representation (1; 0). So there remains

$$\chi^{(\mathrm{at})}_{(1;\,0)} = \eta_1 + \eta_1^{-1} + \eta_2^2 + 1 + \eta_2^{-2}, \tag{5.7}$$

$$\chi_{(0;\,0)} = 1. \tag{5.8}$$

Character GF's can now be determined. From (5.5) we calculate $\Sigma_{q=0}^{\infty}\Sigma_{p=0}^{\infty}\chi_{(q;\,p)}Q^qP^p$, and subtract in this result the nonexistent part corresponding to representations $(0;p)$. We find

$$F^{(\mathrm{t})}(\eta_1,\eta_2;Q,P) = \frac{(Q+Q^2)(\eta_1+\eta_1^{-1}+\eta_2^2+\eta_2^{-2})}{(1-\eta_1Q)(1-\eta_1^{-1}Q)(1-\eta_2P)(1-\eta_2^{-1}P)}. \tag{5.9}$$

This is the character GF for typical representations. In fact, in (5.9) the contribution coming from atypical representations $(q;\,2q-2)$ should be subtracted. It is much easier, however, to leave (5.9) as it stands; we shall use the expression (5.9) to obtain other results and then only in the final results subtract the contribution coming form atypical irreps. Analogously, the summation over (5.6) gives

$$F^{(\mathrm{at})}_{(q;\,p)}(\eta_1,\eta_2;Q,P) = \{(-\eta_2^2-\eta_2^{-2}) + Q[1+(\eta_1+1+\eta_1^{-1}) \times (\eta_2^2+1+\eta_2^{-2})] + P(\eta_2+\eta_2^{-1}) - PQ(\eta_1 + 1 + \eta_1^{-1})(\eta_2+\eta_2^{-1})\}/(1-\eta_1Q) \times (1-\eta_1^{-1}Q)(1-\eta_2P)(1-\eta_2^{-1}P). \tag{5.10}$$

Again, in (5.10) only the part with $p=2q-2$ should be maintained, but it is easier to work with (5.10) as it stands and keep the atypical part of it only after the function has been transformed into another GF.

The character generating functions can now be used to find branching rule generating functions. We first consider the chain osp(3,2)$\to$sp(2) + so(3). Since η_1(resp. η_2) carry the sp(2) [resp. so(3)] weights, we determine

$$\left[F^{(t)}(\eta_1,\eta_2;Q,P)\left(\frac{1-\eta_1^2}{1-\eta_1 T}\right)\left(\frac{1-\eta_2^2}{1-\eta_2 S}\right)\right]_{\mathrm{ex}(\eta^0)}, \quad (5.11)$$

where T (resp. S) carry the sp(2) [resp. so(3)] representation labels. This gives the GF

$$[QT+QS^2+QP^2+PQ^2ST+Q^2S^2+Q^2P^2 \quad (5.12a)$$
$$+Q^2PS+Q^3][(1-QT)(1-PS)]^{-1}$$
$$-Q \quad (5.12b)$$

Obviously, $-Q$ corresponds to the atypical representation (1; 0) and is therefore deleted. Equation (5.12a) is the GF for branching rules for typical osp(3,2) irreps $(q;p)$ into sp(2) + so(3) irreps (t, s), and (5.12a) gives the correct branching rule only when $p\neq 2q-2$. If $p=2q-2$, the representation is atypical, and then (5.10) is used to obtain the branch-

$$(QT+QS^2+Q^2PST+Q^2S^2)/(1-QT)(1-PS). \quad (5.13)$$

The GF (5.13) gives the correct branching rule for all atypical representations $(q;p)=(q;2q-2)$, including (1;0). It is easy to subtract the typical part from (5.13). There remains

$$(QT+QS^2+P^2Q^2S^2T+P^2Q^2S^4)/(1-P^2QS^2T). \quad (5.14)$$

The final GF is now found by subtracting the atypical part in (5.12a), adding (5.14), and adding the result for the scalar representation. This gives

$$G(Q,P;T,S)$$
$$=1+(QT+QS^2+QP^2+PQ^2ST+Q^2S^2+Q^2P^2$$
$$+Q^2PS+Q^3)/(1-QT)(1-PS)$$
$$-(Q^2P^2T+Q^2P^2+P^2Q^2S^2$$
$$+P^4Q^3S^4)/(1-QP^2TS^2). \quad (5.15)$$

When expanded, the GF $G(Q,P;T,S)=\Sigma_{q,p}(\Sigma_{t,s}c_{qpts}\times T^tS^s)Q^qP^p$ gives the multiplicity c_{qpts} of a representation (t, s) of sp(2) + so(3) in the decomposition of any (typical and atypical) osp(3,2) irrep $(q;p)$. It follows from (5.15) that a general typical irrep $(q;p)$ $(q\geqslant 3,\ p\geqslant 2)$ decomposes in the sp(2) + so(3) irreps (q,p), $(q-1,p-2)$, $(q-1,p)$, $(q-1,p+2)$, $(q-2,p-2)$, $(q-2,p)$, $(q-2,p+2)$, and $(q-3,p)$, and that a general atypical irrep $(q;p)=(q;2q-2)$ $(q\geqslant 2)$ decomposes in (q,p), $(q-1,p)$, $(q-1,p+2)$, and $(q-2,p+2)$, a result confirmed in Refs. 18 and 20.

It is easy to calculate the sp(2) + so(3) tensors in the enveloping algebra of osp(3,2). The weight-generating function for polynomial tensors in the adjoint representation of osp(3,2) is

$$\frac{(1+V\eta_1)(1+V\eta_1\eta_2^2)(1+V\eta_1\eta_2^{-2})(1+V\eta_1^{-1})(1+V\eta_1^{-1}\eta_2^2)(1+V\eta_1^{-1}\eta_2^{-2})}{(1-U\eta_1^2)(1-U)(1-U\eta_1^{-2})(1-U\eta_2^2)(1-U)(1-U\eta_2^{-2})}, \quad (5.16)$$

where U carries the degree in the even generators, and V the degree in the odd generators. Transforming (5.16) into sp(2) + so(3) tensors (t, s) by means of an instruction as in (5.11) gives

$$G(U,V;T,S)=[(1+V^2+V^4+V^6)+(V+V^3+V^5)(S^2T+2S^2TU+S^2TU^2+TU+TU^2)$$
$$+(V^2+V^4)(S^2T^2+2S^2T^2U+S^2T^2U^2+2S^2U+2S^2U^2+S^4+S^4U+T^2U+T^2U^2+2U^2)$$
$$+V^3(S^4T+2S^4TU+S^4TU^2+S^2TU+2S^2TU^2+S^2TU^3+T^3+T^3U+TU+2TU^2$$
$$+TU^3)]/(1-U^2)^2(1-US^2)(1-UT^2). \quad (5.17)$$

From (5.17) we obtain the GF for sp(2) + so(3) invariants in the enveloping algebra of osp(3,2):

$$(1+V^2+V^4+V^6+2U^2V^2+2U^2V^4)/(1-U^2)^2. \quad (5.18)$$

The two denominator factors correspond to the sp(2) and so(3) second-order invariants, and (5.18) shows that they are the only functionally independent subalgebra invariants.

The Lie superalgebra osp(3,2) contains osp(1,2) as a maximal subalgebra. In this chain the adjoint representation $(q;p)=(2;0)$ of osp(3,2) decomposes as $(\overline{3})+(2)$ of osp(1,2), where $(\overline{3})=(3)_1+(2)_0$, and $(2)=(2)_0+(1)_1$ is the adjoint representation of osp(1,2). It is now an easy exercise to calculate the osp(1,2) scalars in the enveloping algebra of osp(3,2). Indeed, the GF $G_{(2)}(V;A)$ for polynomial tensors in (2) is given in (3.12), and the GF $G_{(\overline{3})}(U;A)$ for polynomial tensors in $(\overline{3})$ is given in (3.15). Hence, the GF for osp(1,2) invariants in the osp(3,2) enveloping algebra is

$$[G_{(2)}(V;A^{-1})G_{(\overline{3})}(U;A)]_{\mathrm{ex}(A^0)}, \quad (5.19)$$

or, explicitly,

$$\frac{1+U^4+V(U^3+U^5)+V^2(U^2+U^3+2U^4)+V^3(U+2U^3+2U^4+U^5)+V^4(U^2+U^5)}{(1-U^2)(1-V^2)(1-U^2V^2)}, \quad (5.20)$$

where V denotes the degree in the osp(1,2) generators and U the degree in the osp(1,2) tensor $(\overline{3})$. The denominator in (5.20) is interpreted as follows: V^2 corresponds to the osp(1,2) Casimir, and U^2 and U^2V^2 are the osp(3,2) second- and fourth-order Casimirs. The fact that no other denominators appear in (5.20) shows that osp(3,2)$\to$osp(1,2) is a zero missing label problem.

We can also determine the branching rules for osp(3,2) irreps into representations of its maximal osp(1,2) subalgebra. For this task, we make use of the branching rule GF

$G(Q, P; T, S)$ into sp(2) + so(3) irreps (t, s); the maximal so(3) (labeled by l) contained in sp(2) + so(3) is projected out by means of[6]

$$G(Q, P; L) = [G(Q, P; T^{-1}, S^{-1}) \times [(1-SL)(1-TL)(1-ST)]^{-1}]_{\mathrm{ex}(S^0T^0)}, \tag{5.21}$$

and then (3.6) is used to transform a GF $G(Q, P; L)$ for so(3) irreps (l) into a GF $H(Q, P; A)$ for osp(1,2) irreps (a). We apply this technique to the atypical and typical parts separately. Using (5.14), the GF for branching rules for atypical osp(3,2) irreps $(q; p) = (q; 2q-2)$ into osp(1,2) irreps (a) is given by

$$H'(Q, P; A) = \frac{P^2Q^2A}{(1-P^2QA)} + \frac{Q^2A^2 + P^2Q^2A^4}{(1-P^2QA)(1-P^2QA^3)}. \tag{5.22}$$

For typical representations, we start with the GF (5.12a). Including also the trivial representation, we find

$$H(Q, P; A) = 1 + PQ/(1-PA)(1-PQ) + (QA^2 + Q^2A^2 + PQ^2A^2)/(1-PA)(1-QA) + \frac{Q^3 + Q^3A + PQ^2A + P^2Q^2A}{(1-PA)(1-QA)(1-PQ)}. \tag{5.23}$$

When expanded in power series, the GF's (5.22) and (5.23) have the form

$$\sum_{q,p}\Big(\sum_a c_{qpa}A^a\Big)Q^qP^p, \tag{5.24}$$

and c_{qpa} is the multiplicity of the osp(1,2) irrep (a) contained in an irrep $(q; p)$ of osp(3,2). Note that (5.23) is only valid for typical representations; the "atypical terms" appearing in the expansion of (5.23) must be disregarded. Instead, for the decomposition of an atypical irrep one uses (5.22). It is easy to verify that for typical (resp. atypical) representations of osp(3,2), we have

$$c_{qpa} \leqslant 2 \quad (\text{resp. } c_{qpa} \leqslant 1). \tag{5.25}$$

Cases with $c_{qpa} = 2$ do occur, and hence there is degeneracy in the chain osp(3,2)→osp(1,2). However, the multiplicity is always less than 2, and in labeling states of an irrep, labels which assume only a finite number of values are ignored.[21] This shows again that osp(3,2)→osp(1,2) is a zero missing label problem.

Let us finally consider the problem of transforming an osp(3,2) weight GF into a GF for osp(3,2) irreps. If $F(\eta_1, \eta_2)$ is a GF for osp(3,2) weights [of complete osp(3,2) irreps], then the discussion in Sec. II shows that the prescription

$$\left[F(\eta_1, \eta_2)\frac{(1-\eta_1^{-1})(1-\eta_2^{-2})}{(1+\eta_1^{-1}\eta_2^{-2})(1+\eta_1^{-1}\eta_2^2)(1-\eta_1^{-1}Q)(1-\eta_2^{-1}P)}\right]_{\mathrm{ex}(\eta^0)} = \sum_{q,p} N_{qp}Q^qP^p \tag{5.26}$$

gives the correct multiplicity N_{qp} for typical representations $(q; p)$ but will usually give the wrong multiplicity for atypical irreps $(q; 2q-2)$ and for the scalar irrep $(0; 0)$. Consider now the multiplication of a weight function f by

$$(1-\eta_1^{-1})(1-\eta_2^{-2})/(1+\eta_2^{-1}\eta_2^2) \tag{5.27}$$

(the denominator is a formal notation for a power series expansion) and keep only the terms with weight in the dominant Weyl sector. In (5.28) these terms in the dominant Weyl sector are listed for f equal to the character (5.5) of a typical irrep, the character (5.6) of an atypical irrep, the character (5.7) of $(1; 0)$, and the character (5.8) of $(0; 0)$:

$$\chi^{(t)}_{(q;p)}: \quad \begin{array}{ll} \eta_1^q\eta_2^p + \eta_1^{q-1}\eta_2^{p-2} & (q\geqslant 2), \\ \eta_1\eta_2^p + \eta_2^{p-2} - \eta_2^p & (q=1), \end{array} \tag{5.28a}$$

$$\chi^{(at)}_{(q;p)}: \quad \eta_1^q\eta_2^p, \tag{5.28b}$$

$$\chi^{(at)}_{(1;0)}: \quad \eta_1, \tag{5.28c}$$

$$\chi_{(0;0)}: \quad 1. \tag{5.28d}$$

None of the weights in (5.28a) correspond to highest weights of atypical representations. Hence, we have shown that the GF

$$\left[F(\eta_1, \eta_2)\frac{(1-\eta_1^{-1})(1-\eta_2^{-2})}{(1+\eta_1^{-1}\eta_2^2)(1-\eta_1^{-1}Q)(1-\eta_2^{-1}P)}\right]_{\mathrm{ex}(\eta^0)} \tag{5.29}$$

gives the correct multiplicity N_{qp} for atypical representations $(q; p) = (q; 2q-2)$. Only the multiplicity N_{00} of the scalar representation remains to be determined. Since $F(\eta_1, \eta_2)$ consists of weights of complete osp(3,2) irreps, we see from (5.5)–(5.7) that

$$[F(\eta_1, \eta_2)(1-\eta_1^{-1})(1-\eta_2^{-2})]_{\mathrm{ex}(\eta^0)} = N_{00} - N_{10} + N_{20} + N_{12}. \tag{5.30}$$

But N_{10} follows from (5.29), and N_{20} and N_{12} from (5.26). Hence the multiplicity N_{00} of the scalar representation can be determined from (5.30). Then, it is easy to correct the GF (5.26) by means of (5.29) for the atypical irreps and by means of (5.30) for the scalar irrep. The resulting GF then gives the correct multiplicity for all osp(3,2) irreps.

We apply this procedure to find polynomial tensors in a given osp(3;2) irrep $(q; p)$. There are two possible gradings for a representation space: (1) highest weight state [= state with highest t value in the reduction to sp(2) + so(3)] is even $(q; p)$; and (2) highest weight state is odd $(\overline{q; p})$. The weight generating function is

$$F(\eta_1, \eta_2; U) = \frac{\Pi_\beta(1+U\eta^\beta)}{\Pi_\alpha(1-U\eta^\alpha)}, \tag{5.31}$$

where α are the even weights and β the odd weights of the given tensor representation. We determined the polynomial tensor GF's for the representations $(1; 0)$, $(\overline{1; 0})$, $(1; 1)$ and $(2; 0)$. The results are (U carries the degree of the tensor product in the given representation)

$$G_{(1;0)}(Q, P; U) = 1/(1-UQ), \tag{5.32}$$

$$G_{(\overline{1;0})}(Q, P; U) = 1/(1-U^2) + UQ/(1-U^2)(1-UP^2), \tag{5.33}$$

$$G_{(1;\,1)}(Q,P;U)=U^2/(1-U^2)+(1+U^4Q^3)/(1-U^2)(1-UPQ), \tag{5.34}$$

$$\begin{aligned}G_{(2;\,0)}(Q,P;U)=&\frac{U^2+U^4+U^6}{(1-U^2)^2}+\frac{1+U^6}{(1-U^2)(1-U^4)(1-UQ^2)}\\&+\frac{QU^5+Q^2U^3+Q^3U^4+P^2QU^2+P^2Q^2(U^2+U^5+U^6)+P^2Q^3U^4}{(1-U^2)^2(1-UP^2)(1-UQ^2)}\\&+\frac{Q(U^4+U^5+U^6)+P^2Q^2(U^4+U^5+U^6+U^8)+P^4Q^3U^5}{(1-U^2)^2(1-U^3P^4Q^2)}.\end{aligned} \tag{5.35}$$

Since (2; 0) is the adjoint representation of osp(3,2), (5.35) is the GF for irreducible tensors contained in the enveloping algebra of osp(3,2). When expanded as

$$G_{(q';\,p')}(Q,P;U)=\sum_{q,p}\Big(\sum_u c_{qpu}U^u\Big)Q^qP^p, \tag{5.36}$$

c_{qpu} is the multiplicity of the osp(3,2) irrep $(q;p)$ in the supersymmetric tensor product of u copies of the irrep $(q';p')$.

VI. GENERATING FUNCTIONS FOR OSP(4,2) AND BRANCHING RULES TO SUBALGEBRAS

We will not treat osp(4,2) as explicitly as the previous examples. Most of the results will be given only for one class of representations, namely the typical irreps.

The even part of osp(4,2) is sp(2) + so(4)$\simeq$su(2) + su(2) + su(2). The odd part consists of a tensor of type (1,1,1) with respect to the three su(2) subalgebras (i.e., of spinor–spinor–spinor type). The positive roots of osp(4,2) are given by $\Delta_0^+=\{(2,0,0),(0,2,0),(0,0,2)\}$ and $\Delta_1^+=\{(1,1,1),(1,1,-1),(1,-1,1),(1,-1,-1)\}$. Irreducible representations of osp(4,2) have been studied in Refs. 19 and 20. Let $(a_1;a_2;a_3)$ be the Kac–Dynkin labels of an osp(4,2) irrep; then the label b (introduced in Sec. II) is given by

$$b=a_1-\tfrac{1}{2}a_2-\tfrac{1}{2}a_3. \tag{6.1}$$

We use the labels $(p;q;r)=(b;a_2;a_3)$ for the osp(4,2) irreps. The labels p, q, and r can be any non-negative integer, but we have the following consistency conditions for small p:

$$p=1\Rightarrow q=r, \tag{6.2}$$

$$p=0\Rightarrow q=r=0. \tag{6.3}$$

When an irrep $(p;q;r)$ is decomposed into sp(2) + so(4) irreps (s,t,u), p is the maximum s value, and q and r the corresponding (t,u) values. Also, p, q, and r are the eigenvalues of the sp(2) + so(4) diagonal operators on the highest weight state of the irrep $(p;q;r)$.

The elements of $\bar{\Delta}_1^+$ are (1,1,1), (1,1, − 1), (1, − 1,1), and (1, − 1, − 1), and they give rise to the atypicality conditions

$$2p-q-r-4=0\quad(\text{at }1), \tag{6.4}$$

$$2p-q+r-2=0\quad(\text{at }2), \tag{6.5}$$

$$2p+q-r-2=0\quad(\text{at }3), \tag{6.6}$$

$$2p+q+r=0, \tag{6.7}$$

respectively. The fourth condition (6.7) is satisfied by the trivial representation only.

The characters for typical and atypical representations are given by (2.8) and (2.11). An explicit form of the character of a typical representation, with η_1,η_2,η_3 carrying the weights, is given by

$$\begin{aligned}\chi^{(t)}_{(p;\,q;\,r)}=&(\eta_1^{1/2}\eta_2^{1/2}\eta_3^{1/2}+\eta_1^{-1/2}\eta_2^{-1/2}\eta_3^{-1/2})(\eta_1^{1/2}\eta_2^{1/2}\eta_3^{-1/2}+\eta_1^{-1/2}\eta_2^{-1/2}\eta_3^{1/2})\\&\times(\eta_1^{1/2}\eta_2^{-1/2}\eta_3^{1/2}+\eta_1^{-1/2}\eta_2^{1/2}\eta_3^{-1/2})(\eta_1^{1/2}\eta_2^{-1/2}\eta_3^{-1/2}+\eta_1^{-1/2}\eta_2^{1/2}\eta_3^{1/2})\\&\times[(\eta_1-\eta_1^{-1})(\eta_2-\eta_2^{-1})(\eta_3-\eta_3^{-1})]^{-1}[\eta_1^{p-1}\eta_2^{q+1}\eta_3^{r+1}-\eta_1^{-p+1}\eta_2^{q+1}\eta_3^{r+1}\\&-\eta_1^{p-1}\eta_2^{-q-1}\eta_3^{r+1}-\eta_1^{p-1}\eta_2^{q+1}\eta_3^{-r-1}+\eta_1^{-p+1}\eta_2^{-q-1}\eta_3^{r+1}+\eta_1^{-p+1}\eta_2^{q+1}\eta_3^{-r-1}\\&+\eta_1^{p-1}\eta_2^{-q-1}\eta_3^{-r-1}-\eta_1^{-p+1}\eta_2^{-q-1}\eta_3^{-r-1}].\end{aligned} \tag{6.8}$$

The character generating function

$$F(P,Q,R;\eta_1,\eta_2,\eta_3)=\sum_{p,q,r=0}^{\infty}\chi_{(p;\,q;\,r)}P^pQ^qR^r \tag{6.9}$$

reads

$$\begin{aligned}&F(P,Q,R;\eta_1,\eta_2,\eta_3)\\&=[-1+P(\eta_1+\eta_1^{-1})][2+\eta_1^2+\eta_1^{-2}+\eta_2^2+\eta_2^{-2}+\eta_3^2+\eta_3^{-2}+\eta_1\eta_2\eta_3+\eta_1^{-1}\eta_2^{-1}\eta_3^{-1}\\&\quad+\eta_1\eta_2\eta_3^{-1}+\eta_1^{-1}\eta_2^{-1}\eta_3+\eta_1\eta_2^{-1}\eta_3+\eta_1^{-1}\eta_2\eta_3^{-1}+\eta_1^{-1}\eta_2\eta_3+\eta_1\eta_2^{-1}\eta_3^{-1}][(1-\eta_1P)(1-\eta_1^{-1}P)\\&\quad\times(1-\eta_2Q)(1-\eta_2^{-1}Q)(1-\eta_3R)(1-\eta_3^{-1}R)]^{-1}.\end{aligned} \tag{6.10}$$

When (6.10) is expanded in the form (6.9), it gives as coefficient of $P^pQ^qR^r$ the character of the irrep $(p;q;r)$ only when $(p;q;r)$ is a typical representation.

The character GF (6.10) is used in order to obtain the branching rule GF for osp(4,2)$\to$sp(2) + so(4)$\simeq$su(2) + su(2) + su(2). Since η_1, η_2, and η_3 carry the weights of sp(2) and su(2) + su(2), we have to determine

$$\left[F(P,Q,R;\eta_1,\eta_2,\eta_3)\frac{(1-\eta_1^2)(1-\eta_2^2)(1-\eta_3^2)}{(1-\eta_1 S)(1-\eta_2 T)(1-\eta_3 U)}\right]_{\mathrm{ex}(\eta^0)}, \tag{6.11}$$

where S carries the sp(2) label (s) and T, U carry the so(4)$\simeq$su(2) + su(2) labels (t,u). This gives rise to

$$\begin{aligned}G^{(t)}&(P,Q,R;S,T,U)\\&=P^2QRTU/(1-QT)(1-RU)+[1+(P+P^3)(QR+RT+TU+QU)\\&\quad+P^2(Q^2+T^2+R^2+U^2+PSQT+PSRU)+P^4]/(1-PS)(1-QT)(1-RU).\end{aligned} \tag{6.12}$$

When (6.12) is expanded in the form

$$\sum_{p,q,r}\Big(\sum_{s,t,u}c_{pqrstu}S^sT^tU^u\Big)P^pQ^qR^r,$$

c_{pqrstu} is the multiplicity of the sp(2) + so(4) irrep (s,t,u) in the decomposition of the osp(4,2) irrep $(p;q;r)$ if and only if $(p;q;r)$ are the labels of a typical representation of osp(4,2) (with $p\geqslant 2$). When $p=1$, we must have $q=r$, and the decomposition reads

$$(PS+PQR+PTU)/(1-QT)(1-RU), \tag{6.13}$$

and when $p=0$, then $q=r=0$, which is the trivial representation. The decomposition of atypical representations can be obtained by making use of (2.11) or of the analysis in Ref. 19. We find

$$\begin{aligned}G^{(\mathrm{at}\,1)}(P,Q,R;S,T,U)=[(1-PS)(1-QT)(1-RU)]^{-1}[1+P(TU+TR+QU)\\+P^2(T^2+U^2+PSQRTU)+P^3TU],\end{aligned} \tag{6.14}$$

$$\begin{aligned}G^{(\mathrm{at}\,2)}(P,Q,R;S,T,U)=[(1-PS)(1-QT)(1-RU)]^{-1}[1+P(QR+PSTR+TU)\\+P^2(T^2+R^2+PSRU)+P^3TR],\end{aligned} \tag{6.15}$$

$$\begin{aligned}G^{(\mathrm{at}\,3)}(P,Q,R;S,T,U)=[(1-PS)(1-QT)(1-RU)]^{-1}[1+P(QR+PSQU+UT)\\+P^2(U^2+Q^2+PSQT)+P^3QU].\end{aligned} \tag{6.16}$$

In the expansion of (6.14), only the terms $P^pQ^qR^r$ with $2p-q-r-4=0$ have the correct coefficient, and similar restrictions hold for (6.15) and (6.16).

Another maximal subalgebra of osp(4,2) is osp(1,2) + su(2). The adjoint representation of osp(4,2), (2;0;0), decomposes as $(2,0)+(0,2)+(\bar{1},2)$ [in (a,l), a is the osp(1,2) label and l the su(2) label]. It is easy to use the sp(2) + so(4) branching rule GF (6.12) as an intermediate step: first we decompose the (t,u) irreps of so(4) into su(2) irreps (l) [su(2) is the maximal subalgebra of so(4)] by means of [6]

$$G'(P,Q,R;S,L)=[G^{(t)}(P,Q,R;S,T^{-1},U^{-1})/(1-TL)(1-UL)(1-TU)]_{\mathrm{ex}(T^0U^0)}, \tag{6.17}$$

and then we use (3.6)

$$G^{(t)}(P,Q,R;A,L)=[AG'(P,Q,R;A,L)+G'(P,Q,R;-1,L)]/(A+1). \tag{6.18}$$

The final result reads

$$\begin{aligned}G^{(t)}(P,Q,R;A,L)&=\frac{1+P(QR+Q^2+R^2)+P^2(Q^2+R^2)+P^3+P^3QRA}{(1-PA)(1-QL)(1-RL)(1-QR)}\\&\quad+\frac{L(Q+R)(P+P^3A)+L^2(P+P^2)}{(1-PA)(1-QL)(1-RL)}+\frac{L^2P^2QR}{(1-QL)(1-RL)}.\end{aligned} \tag{6.19}$$

The expansion of (6.19), $\Sigma_{p,q,r}(\Sigma_{a,l}c_{pqral}A^aL^l)P^pQ^qR^r$, gives the multiplicity c_{pqral} of the osp(1,2) + su(2) irrep (a,l) in the decomposition of any osp(1,2) irrep $(p;q;r)$ if $(p;q;r)$ are the labels of a typical representation.

The osp(1,2) + su(2) scalars in the enveloping algebra of osp(4,2) can be determined as follows. We know that the adjoint representation of osp(4,2) decomposes into the irreps (2,0), (0,2), and $(\bar{1},2)$ of osp(1,2) + su(2). Polynomial tensors in (2,0) are given by (3.12)

$$G_{(2,0)}(U_1;A)=(1+U_1^2A)/(1-U_1^2)(1-U_1A^2), \tag{6.20}$$

and polynomial tensors in (0,2) by [6]

$$G_{(0,2)}(U_2;L)=[(1-U_2^2)(1-U_2L^2)]^{-1}. \tag{6.21}$$

Let α be the even weights of $(\bar{1},2)$, and β the odd weights; then the polynomial tensor GF for $(\bar{1},2)$ is given by

$$G_{(\bar{1},2)}(V;A,L)=\left[\frac{\Pi_\beta(1+V\eta^\beta)}{\Pi_\alpha(1-V\eta^\alpha)}\frac{(1-\eta_1)}{(1-\eta_1A)}\frac{(1-\eta_2^2)}{(1-\eta_2L)}\right]_{\mathrm{ex}(\eta^0)}. \tag{6.22}$$

The GF for osp(1,2) + su(2) scalars in the enveloping algebra of osp(4,2) is obtained as follows:

$$G(U_1,U_2,V)=[G_{(\bar{1},2)}(V;A,L)G_{(2,0)}(U_1;A^{-1})G_{(0,2)}(U_2;L^{-1})]_{\mathrm{ex}(A^0L^0)} \tag{6.23}$$

or, explicitly,

$$G(U_1, U_2, V) = [(1-U_1^2)(1-U_2^2)(1-V^2)]^{-1}\{1+(1-U_1V)^{-1}[VU_1U_2^2+V^2(U_1^2+U_2^2+U_1U_2) + V^3(U_1+U_1U_2+U_2^3+U_1^2U_2^2)+V^4(1+U_2^2+U_1U_2+U_1U_2^2+U_1^2U_2^2) + V^5(U_1^2+U_2+U_1U_2+U_1U_2^2)+V^6(U_2^2+U_1U_2^2)]\}. \tag{6.24}$$

In (6.24), U_1(resp. U_2) carry the degree in the osp(1,2) [resp. su(2)] generators, and V gives the degree in the osp(1,2) + su(2) tensor (1,2).

VII. CONCLUSIONS

The GF technique for Lie algebras has been extended to Lie superalgebras. Only a few modifications of the technique are required, if one deals only with typical representations of Lie superalgebras. Difficulties arise when one deals with the atypical representations, mainly because a general formula for the character of an atypical irrep for Lie superalgebras is not known. Also, the GF technique does not distinguish between indecomposable representations and direct sums of atypical representations of Lie superalgebras.

The structure of the enveloping algebras of osp(1,2), spl(1,2), and osp(3,2) has been determined in (3.12), (4.31), and (5.35). From these GF's, we obtain the following property: if a typical tensor (λ) occurs in the enveloping algebra, then it occurs as often as (λ) has states of zero weight [modulo the multiplication of (λ) by invariants]. This is a weak equivalent for superalgebras of a theorem of Kostant[3,22] for Lie algebras. For Lie superalgebras it seems to hold only for typical representations, because we have found that atypical tensors occurring in the enveloping algebra of osp(3,2) appear four times as often as they have states of zero weight.

For a Lie superalgebra $L = L_0 + L_1$, representations are usually studied in an L_0 basis (i.e., in the chain $L \to L_0$). In this paper, other Lie superalgebra–subalgebra chains, such as osp(3,2)$\to$osp(1,2) and osp(4,2)$\to$osp(1,2) + su(2), have been analyzed in detail for the first time. In this context, a classification of the maximal sub-(super-)algebras of all simple Lie superalgebras (such as was achieved for Lie algebras in the paper of Dynkin[23]) would be very interesting and useful.

ACKNOWLEDGMENTS

One of us (R.T.S.) would like to acknowledge receipt of a visiting fellowship grant from the U.K. S. E. R. C.

J. Vd. J. would like to thank the British Council for financial support in the form of a scholarship.

[1]B. R. Judd, W. Miller, J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[2]R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).
[3]M. Couture and R. T. Sharp, J. Phys. A: Math. Nucl. Gen. **13**, 1925 (1980).
[4]C. Bodine and R. Gaskell, J. Math. Phys. **23**, 2217 (1982).
[5]M. Couture, Y. Giroux, and R. T. Sharp, J. Phys. A: Math. Nucl. Gen. **17**, 715 (1984).
[6]J. Patera and R. T. Sharp, "Generating functions for characters of group representations and their applications," in *Lecture Notes in Physics*, Vol. 94 (Springer, New York, 1979).
[7]V. G. Kac, Adv. Math. **26**, 8 (1977).
[8]V. G. Kac, "Representations of classical Lie superalgebras," in *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978).
[9]M. Marcu, J. Math. Phys. **21**, 1284 (1980).
[10]J. P. Hurni and B. Morel, J. Math. Phys. **24**, 157 (1983).
[11]J. P. Hurni and B. Morel, J. Math. Phys. **23**, 2236 (1982).
[12]H. Weyl, Math. Z. **23**, 271 (1925), reprinted in H. Weyl, *Selecta* (Birkhauser, Basel, 1956).
[13]D. A. Leites, Functional Anal. Appl. **14**, 106 (1980).
[14]J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).
[15]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[16]M. Marcu, J. Math. Phys. **21**, 1277 (1980).
[17]J. W. B. Hughes, Ann. Israel Phys. Soc. **3**, 320 (1980).
[18]J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).
[19]J. Van der Jeugt, J. Math. Phys. **26**, 913 (1985).
[20]R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Nucl. Gen. 16, 473 (1983).
[21]A. Peccia and R. T. Sharp, J. Math. Phys. **17**, 1313 (1976).
[22]B. Kostant, Am. J. Math. **85**, 327 (1963).
[23]E. B. Dynkin, Am. Math. Soc. Transl. **6**, 245 (1965).

Irreducible representations of the exceptional Lie superalgebras $D(2,1;\alpha)$

Joris Van der Jeugt[a),b)]
Queen Mary College, School of Mathematical Sciences, Mile End Road, London E1 4NS, England

(Received 4 June 1984; accepted for publication 28 September 1984)

The shift operator technique is used to give a complete analysis of all finite- and infinite-dimensional irreducible representations of the exceptional Lie superalgebras $D(2,1;\alpha)$. For all cases, the star or grade star conditions for the algebra are investigated. Among the finite-dimensional representations there are no star and only a few grade star representations, but an infinite class of infinite-dimensional star representations is found. Explicit expressions are given for the "doublet" representation of $D(2,1;\alpha)$. The one missing label problem $D(2,1;\alpha) \rightarrow \mathrm{su}(2) + \mathrm{su}(2) + \mathrm{su}(2)$ is discussed in detail and solved explicitly.

I. INTRODUCTION

In the last 20 years, algebraic structures called Lie superalgebras have appeared in various contexts in physics as well as in mathematics. A classification of all simple Lie superalgebras was obtained by Kac[1] and other authors.[2,3] For a survey of the physical applications of Lie superalgebras, we refer to Corwin *et al.*[4]

Finite-dimensional representations of superalgebras have been studied in general,[5] and there have also been a number of case studies (for example, see Ref. 6 and references therein). Infinite-dimensional representations have been classified only for the orthosympletic superalgebras osp(1,2) (see Ref. 7) and osp(3,2) (see Ref. 8).

The classical simple Lie superalgebras consist of the spl(m,n) and the osp$(m,2n)$ families, the so-called strange series $P(n)$ and $Q(n)$, and the exceptional algebras $F(4)$, $G(3)$, and $D(2,1;\alpha)$ (see Refs. 1 and 3). The general linear superalgebras and the orthosymplectic families have been the subject of several papers.[6,9,10] The exceptional superalgebras $F(4)$ and $G(3)$ have been studied by DeWitt and Van Nieuwenhuizen.[11] The algebras $D(2,1;\alpha)$ are a one-parameter family of 17-dimensional nonisomorphic Lie superalgebras, which contain $D(2,1) = \mathrm{osp}(4,2)$ as a special case (when $\alpha = 1$). Among the exceptional Lie superalgebras, the $D(2,1;\alpha)$ series is certainly the most curious one, mainly because they do not have a Lie-algebraic counterpart (i.e., there are general linear, orthogonal, symplectic, and exceptional simple Lie algebras, but there does not exist a one-parameter family of nonisomorphic simple Lie algebras). It is the aim of this paper to study the $D(2,1;\alpha)$ algebras and to give a detailed analysis of their finite- and infinite-dimensional irreducible representations (irreps).

The even part of the superalgebra $D(2,1;\alpha)$ is $L_{\bar{0}} = \mathrm{su}(2) + \mathrm{su}(2) + \mathrm{su}(2)$, and the odd part consists of the tensor product of the two-dimensional tensor representations of the three su(2) components. The complex parameter α appears only in the anticommutation relations among the components of the tensor operators. The $D(2,1;\alpha)$ algebras are sometimes denoted by $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ (see Ref. 3), and we give the relation between the two notations in Sec. II.

The representations of $D(2,1;\alpha)$ are studied in the reduction scheme $D(2,1;\alpha) \rightarrow \mathrm{su}(2) + \mathrm{su}(2) + \mathrm{su}(2)$. We show that in general any $D(2,1;\alpha)$ irrep decomposes into 16 irreducible representations of the subalgebra, a result which was proven in a more general way by Kac[5] for the finite-dimensional representations. The basis states of a $D(2,1;\alpha)$ irrep are denoted by $|s,m_s;t,m_t;u,m_u;\lambda\rangle$, where s,t, and u are the three su(2) representation labels; m_s, m_t, and m_u are the three su(2) weight labels; and λ is a supplementary label, the necessity of which is explained in Sec. IV. The analysis of $D(2,1;\alpha)$ irreps is based on the construction of operators $A^{i,j,k}$ ($i,j,k = \pm\frac{1}{2}$) which shift the s and m_s, t and m_t, and u and m_u labels of a basis state by i, j, and k, respectively. Such shift operators have been studied in a general context by Hughes and Yadegar.[12]

The generalization of a Hermitian operation for a Lie algebra is a star or grade star operation for a Lie superalgebra.[13] We investigate the possible star and grade star operations for $D(2,1;\alpha)$ and find that each of the eight Hermitian operations on the even part of $D(2,1;\alpha)$ can be extended to an adjoint (i.e., star or grade star) operation for the superalgebra. Then we consider whether the corresponding irreps of $D(2,1;\alpha)$ are star (resp. grade star) representations. We always choose the representation space to be a graded Hilbert space[13] (i.e., with nondegenerate positive definite inner product $\langle\,|\,\rangle$).

For the finite-dimensional irreps, only grade star representations are allowed. We find, however, that in general the finite-dimensional irreps do not satisfy the grade star conditions. For instance, for osp(4,2) only two finite-dimensional representations are actually grade star irreps, namely the one-dimensional trivial representation (0;0;0) and the six-dimensional standard representation $(\frac{1}{2};0;0)$ [the notation $(p;q;r)$ for a $D(2,1;\alpha)$ irrep is explained in Sec. IV].

Among the infinite-dimensional representations there are several classes of star representations. Special attention is paid to a particular representation that reduces into only two subalgebra irreps, whereas a general infinite-dimensional star representation reduces into 16 subalgebra representations. For this "doublet" representation, all matrix elements of the generators of $D(2,1;\alpha)$ are given explicitly. We show that only for $\alpha = 1$, i.e., for osp(4,2), can the doublet representation be realized in terms of elements of $\mathscr{H}(\mathbb{C},\mathbb{C}^4)$. This is the space of holomorphic functions $f:\mathbb{C}\rightarrow\mathbb{C}^4$ with components f_i $(i = 1,2,3,4)$ satisfying $\int(\Sigma_{i=1}^4 |f_i(z)|^2)\ \exp(-|z|^2)$

[a)] Research assistant at National Fonds voor Wetenschappelijk Onderzoek, Belgium.
[b)] Permanent address: Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium.

$d\lambda(z) < \infty$, where λ is the Lebesgue measure on $\mathbb{C}$. The osp(4,2) generators are then realized as operators acting on $\mathscr{H}(\mathbb{C},\mathbb{C}^4)$. This result is in fact the analog of the metaplectic representations for osp(1,2) (see Ref. 7) and osp(3,2) (see Ref. 8).

II. THE LIE SUPERALGEBRAS $D(2,1;\alpha)$

Scheunert[3] denotes the $D(2,1;\alpha)$ algebras as $\Gamma(\sigma_1,\sigma_2,\sigma_3)$, where σ_1,σ_2, and σ_3 are three complex parameters. At first sight this gives the impression of a three-parameter family of superalgebras. However, there is the condition

$$\sigma_1+\sigma_2+\sigma_3=0, \tag{2.1}$$

and the property that $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ is isomorphic to $\Gamma(\sigma_1',\sigma_2',\sigma_3')$ if

$$\sigma_i'=\lambda\,\sigma_{\pi(i)} \quad (i=1,2,3), \tag{2.2}$$

where $\lambda\in\mathbb{C}-\{0\}$ and π is a permutation of $\{1,2,3\}$. This shows that the algebras $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ form essentially a one-parameter family. In this section we give the (anti-) commutation relations for the basis elements of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$, and we give the connection with the more usual notation $D(2,1;\alpha)$.

The even part of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ is su(2) + su(2) + su(2), with basis $s_i,t_i,u_i (i=0,\pm)$, respectively, and with nonvanishing commutators

$$\begin{aligned}&[s_0,s_\pm]=\pm s_\pm,\ [t_0,t_\pm]=\pm t_\pm,\\ &[u_0,u_\pm]=\pm u_\pm,\ [s_+,s_-]=2s_0,\\ &[t_+,t_-]=2t_0,\quad [u_+,u_-]=2u_0.\end{aligned} \tag{2.3}$$

The odd part consists of the tensor product of three two-dimensional su(2) tensor operators, with components $R_{i,j,k}$ $(i,j,k=\pm\frac12)$. The commutation relations between even and odd generators are given by

$$[s_0,R_{i,j,k}]=iR_{i,j,k},\quad [s_\pm,R_{\mp1/2,j,k}]=R_{\pm1/2,j,k}, \tag{2.4}$$

and analogous expressions for $[t,R]$ and $[u,R]$. Finally, the nonvanishing anticommutation relations among the odd basis elements are

$$\begin{aligned}&[R_{\pm1/2,-1/2,-1/2},R_{\pm1/2,1/2,1/2}]=\mp2\sigma_1 s_\pm,\\ &[R_{\pm1/2,-1/2,1/2},R_{\pm1/2,1/2,-1/2}]=\pm2\sigma_1 s_\pm,\\ &[R_{-1/2,\pm1/2,-1/2},R_{1/2,\pm1/2,1/2}]=\mp2\sigma_2 t_\pm,\\ &[R_{-1/2,\pm1/2,1/2},R_{1/2,\pm1/2,-1/2}]=\pm2\sigma_2 t_\pm,\\ &[R_{-1/2,-1/2,\pm1/2},R_{1/2,1/2,\pm1/2}]=\mp2\sigma_3 u_\pm,\\ &[R_{-1/2,1/2,\pm1/2},R_{1/2,-1/2,\pm1/2}]=\pm2\sigma_3 u_\pm,\\ &[R_{-1/2,\mp1/2,\mp1/2},R_{1/2,\pm1/2,\pm1/2}]\\ &\quad=2(\sigma_1 s_0\pm\sigma_2 t_0\pm\sigma_3 u_0),\\ &[R_{-1/2,\mp1/2,\pm1/2},R_{1/2,\pm1/2,\mp1/2}]\\ &\quad=2(-\sigma_1 s_0\mp\sigma_2 t_0\pm\sigma_3 u_0).\end{aligned} \tag{2.5}$$

The $D(2,1;\alpha)$ algebra is usually defined by means of its Cartan matrix A (see Refs. 1 and 14)

$$A=(a_{ij})=\begin{bmatrix}0&1&\alpha\\-1&2&0\\-1&0&2\end{bmatrix}\quad (\alpha\in\mathbb{C}-\{0,-1\}), \tag{2.6}$$

and its generators e_i,f_i,h_i $(i\in I=\{1,2,3\})$ satisfying the relations

$$\begin{aligned}&[e_i,f_j]=\delta_{ij}h_i,\quad [h_i,h_j]=0,\\ &[h_i,e_j]=a_{ij}e_j,\quad [h_i,f_j]=-a_{ij}f_j,\\ &\deg h_i=0;\ \deg e_i=\deg f_i=0,\ \text{for } i=2,3,\\ &\deg e_1=\deg f_1=1.\end{aligned} \tag{2.7}$$

The basis elements of $D(2,1;\alpha)$ are then given by

$$\begin{aligned}&s_0=(2h_1-h_2-\alpha h_3)/2(1+\alpha),\\ &s_+=i[[[e_1,e_2],e_3],e_1]/(1+\alpha),\\ &s_-=i[[[f_1,f_2],f_3],f_1]/(1+\alpha),\\ &t_0=\tfrac12 h_2,\quad t_+=e_2,\quad t_-=f_2,\\ &u_0=\tfrac12 h_3,\quad u_+=e_3,\quad u_-=f_3;\end{aligned} \tag{2.8}$$

$$\begin{aligned}&R_{1/2,1/2,1/2}=\lambda\,[[e_1,e_2],e_3],\\ &R_{-1/2,-1/2,-1/2}=i\lambda\,[[f_1,f_2],f_3]\\ &R_{1/2,1/2,-1/2}=-\lambda\,[e_1,e_2],\\ &R_{-1/2,-1/2,1/2}=-i\lambda\,[f_1,f_2],\\ &R_{1/2,-1/2,1/2}=-\lambda\,[e_1,e_3],\\ &R_{-1/2,1/2,-1/2}=-i\lambda\,[f_1,f_3],\\ &R_{1/2,-1/2,-1/2}=\lambda e_1,\\ &R_{-1/2,1/2,1/2}=i\lambda f_1.\end{aligned} \tag{2.9}$$

In the above, λ is an arbitrary factor, which produces a multiplication by $i\lambda^2/2$ for the corresponding $(\sigma_1,\sigma_2,\sigma_3)$ values of the algebra. Hence, if we take $\lambda=\sqrt2\,\exp(i\pi/4)$, then the $D(2,1;\alpha)$ basis elements (2.8) and (2.9) satisfy the $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ relations (2.3)–(2.5) with

$$(\sigma_1,\sigma_2,\sigma_3)=(1+\alpha,-1,-\alpha). \tag{2.10}$$

In the following sections we shall very often use the notation of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$; obviously, all these results are immediately transformed for $D(2,1;\alpha)$ by means of (2.10).

III. INVARIANTS AND SUBALGEBRA SCALARS

$D(2,1;\alpha)$ possesses three independent invariants I_2,I_4, and I_6. In this section we shall give expressions for I_2 and I_4, define some scalar operators with respect to $D(2,1;\alpha)_{\bar0}$, and determine some useful expressions between invariants and subalgebra scalars.

The invariants of the Lie subalgebra of $D(2,1;\alpha)$ are

$$S^2=s_+s_-+s_0^2-s_0,\quad T^2=t_+t_-+t_0^2-t_0,$$
$$U^2=u_+u_-+u_0^2-u_0. \tag{3.1}$$

Furthermore, we define

$$\begin{aligned}&(R\times R)_{i,j,k}^{[p,q,r]}\\ &\quad=\sum\langle\tfrac12 i_1\tfrac12 i_2|pi\rangle\langle\tfrac12 j_1\tfrac12 j_2|qj\rangle\\ &\qquad\times\langle\tfrac12 k_1\tfrac12 k_2|rk\rangle R_{i_1,j_1,k_1}R_{i_2,j_2,k_2},\end{aligned} \tag{3.2}$$

where $\langle\,|\,\rangle$ is an su(2) Clebsch–Gordan coefficient. The second-order invariant is then given by

$$I_2=\sqrt2(R\times R)_{0,0,0}^{[0,0,0]}-2\sigma_1S^2-2\sigma_2T^2-2\sigma_3U^2, \tag{3.3}$$

or explicitly

$$
\begin{aligned}
I_2 = {} & R_{1/2,1/2,1/2} R_{-1/2,-1/2,-1/2} \\
& + R_{1/2,-1/2,-1/2} R_{-1/2,1/2,1/2} \\
& + R_{-1/2,1/2,-1/2} R_{1/2,-1/2,1/2} \\
& + R_{-1/2,-1/2,1/2} R_{1/2,1/2,-1/2} \\
& - 2\sigma_1 S^2 - 2\sigma_2 T^2 - 2\sigma_3 U^2.
\end{aligned} \tag{3.4}
$$

We introduce the notation

$$
\begin{aligned}
& s_{\pm 1} = \pm \frac{1}{\sqrt{2}} s_\pm, \quad t_{\pm 1} = \pm \frac{1}{\sqrt{2}} t_\pm, \\
& u_{\pm 1} = \pm \frac{1}{\sqrt{2}} u_\pm,
\end{aligned} \tag{3.5}
$$

and define

$$
\begin{aligned}
C^{(110,2)} &= 2 \sum_{i,j=-1}^{1} (R \times R)_{i,j,0}^{[1,1,0]} s_{-i} t_{-j}, \\
C^{(101,2)} &= 2 \sum_{i,k=-1}^{1} (R \times R)_{i,0,k}^{[1,0,1]} s_{-i} u_{-k}, \\
C^{(011,2)} &= 2 \sum_{j,k=-1}^{1} (R \times R)_{0,j,k}^{[0,1,1]} t_{-j} u_{-k}.
\end{aligned} \tag{3.6}
$$

The operators (3.6) are subalgebra scalars, i.e., they commute with all the generators of su(2) + su(2) + su(2). They are not independent, since the following relation exists:

$$
\begin{aligned}
I_4 = {} & 4\sqrt{2}(C^{(110,2)} + C^{(101,2)} + C^{(011,2)}) \\
& + 8I_2(S^2 + T^2 + U^2) + 16(\sigma_1 S^4 + \sigma_2 T^4 + \sigma_3 U^4) \\
& - 32(\sigma_1 S^2 + \sigma_2 T^2 + \sigma_3 U^2),
\end{aligned} \tag{3.7}
$$

where S^4 stands for $(S^2)^2$, etc. Note that (3.7) gives an explicit expression for I_4. We also obtained another relation, quadratic in the operators (3.6)

$$
\begin{aligned}
& \sqrt{2} C^{(110,2)} U^2 [\sqrt{2} C^{(110,2)} + 2I_2 - 4(\sigma_1 S^2 + \sigma_2 T^2 - \sigma_3 U^2)] \\
& \quad + \sqrt{2} C^{(101,2)} T^2 [\sqrt{2} C^{(101,2)} \\
& \quad + 2I_2 - 4(\sigma_1 S^2 - \sigma_2 T^2 + \sigma_3 U^2)] \\
& \quad + \sqrt{2} C^{(011,2)} S^2 [\sqrt{2} C^{(011,2)} + 2I_2 \\
& \quad - 4(-\sigma_1 S^2 + \sigma_2 T^2 + \sigma_3 U^2)] + 4I_2 S^2 T^2 U^2 \\
& \quad \times [I_2 + 4(\sigma_1 S^2 + \sigma_2 T^2 + \sigma_3 U^2)] \\
& \quad + 16 S^2 T^2 U^2 [(\sigma_1 S^2 + \sigma_2 T^2 + \sigma_3 U^2)^2 \\
& \quad - 4(\sigma_1^2 S^2 + \sigma_2^2 T^2 + \sigma_3^2 U^2)] = 0.
\end{aligned} \tag{3.8}
$$

Note that (3.7) and (3.8) show that at most one of the operators (3.6) is independent of S^2, T^2, U^2, I_2, and I_4. On the other hand, since $D(2,1;\alpha) \supset$ su(2) + su(2) + su(2) is a one missing label problem (see Sec. IV), we expect that there is at least one subalgebra scalar, independent of the subalgebra invariants and the superalgebra invariants, whose eigenvalue can then be used to distinguish between independent states with the same subalgebra labels. The question of the missing label and a labeling operator L will be discussed in Sec. V.

IV. $D(2,1;\alpha)$ IRREPS AND SHIFT OPERATORS FOR $D(2,1;\alpha) \supset$ su(2) + su(2) + su(2)

Irreducible representations of $D(2,1;\alpha)$ reduce into the direct sum of a set of subalgebra irreps, when restricted to su(2) + su(2) + su(2). Irreps of su(2) + su(2) + su(2) can be labeled by (s,t,u) where $s(s+1)$, $t(t+1)$, and $u(u+1)$ are the eigenvalues of the subalgebra invariants (3.1) S^2, T^2, and U^2, respectively. From Kac[5] and a basis for the universal enveloping algebra of $D(2,1;\alpha)$, it is easy to find the reduction rule for $D(2,1;\alpha) \to$ su(2) + su(2) + su(2). Let p be the maximum s value in the reduction of a $D(2,1;\alpha)$ irrep, and let (q,r) be the corresponding (t,u) values; then $(p;q;r)$ is a good set of labels to specify the $D(2,1;\alpha)$ irrep, and the decomposition into su(2) + su(2) + su(2) is given by

$$
(p;q;r) \to \sum_{(s,t,u) \in \mathscr{T}} \mu_{s,t,u} (s,t,u), \tag{4.1}
$$

where

$$
\begin{aligned}
\mathscr{T} = \{ & (p,q,r), (p - \tfrac{1}{2}, q \pm \tfrac{1}{2}, r \pm \tfrac{1}{2}), \\
& (p - \tfrac{1}{2}, q \pm \tfrac{1}{2}, r \mp \tfrac{1}{2}), (p-1, q \pm 1, r), \\
& (p-1, q, r \pm 1), (p-1, q, r), (p - \tfrac{3}{2}, q \pm \tfrac{1}{2}, r \pm \tfrac{1}{2}), \\
& (p - \tfrac{3}{2}, q \pm \tfrac{1}{2}, r \mp \tfrac{1}{2}), (p-2, q, r) \}.
\end{aligned} \tag{4.2}
$$

Here, $\mu_{s,t,u}$ denotes the multiplicity of the (s,t,u) representation. For a general $D(2,1;\alpha)$ irrep, all $\mu_{s,t,u}$ are equal to 1, except $\mu_{p-1,q,r} = 2$. Hence, in the middle of this lattice of (s,t,u) nodes, there is a twofold degeneracy. This implies that the subalgebra labels are not sufficient to classify the states of a $D(2,1;\alpha)$ irrep uniquely. Therefore, an operator L, commuting with all subalgebra generators, will be introduced in Sec. V. Then, for $(s,t,u) = (p-1,q,r)$, the two different L eigenvalues λ_1 and λ_2 will solve the degeneracy problem. The states of a $(p;q;r)$ representation of $D(2,1;\alpha)$ are then well defined as the common eigenstates of $S^2, s_0, T^2, t_0, U^2, u_0$, and L, and are denoted as

$$
|s,m_s;t,m_t;u,m_u;\lambda\rangle, \tag{4.3}
$$

where m_s, m_t, and m_u are the s_0, t_0, and u_0 eigenvalues, respectively. If $(s,t,u) \neq (p-1,q,r)$, λ is very often omitted since for those cases it is unneccessary.

For completeness we give the correpondence between our labels $(p;q;r)$ and the Kac–Dynkin[5] labels $(a_1, a_2, a_3; b = (2a_1 - a_2 - \alpha a_3)/(1+\alpha))$ for a $D(2,1;\alpha)$ irrep

$$
p = b/2, \quad q = a_2/2, \quad r = a_3/2. \tag{4.4}
$$

Let $(p;q;r)$ be a $D(2,1;\alpha)$ representation with basis states (4.3). Then the operator $\hat{s}$ is defined by means of

$$
\hat{s}|s,m_s;t,m_t;u,m_u;\lambda\rangle = s|s,m_s;t,m_t;u,m_u;\lambda\rangle. \tag{4.5}
$$

An equivalent definition is, of course, $\hat{s} = [S^2 + \frac{1}{4}]^{1/2} - \frac{1}{2}$. Operators $\hat{t}$ and $\hat{u}$ are defined in an analogous way.

Next, we define shift operators $O^{i,j,k}$ $(i,j,k = \pm \frac{1}{2})$. Such operators have been studied in general by Hughes and Yadegar[12] and have been used to classify representations of osp(1,2) (see Ref. 7) and osp(3,2) (see Ref. 8). They have also been used to study representations of Lie algebras (see Refs. 15 and 16 and references therein). For one su(2) algebra with two-dimensional tensor operator $R_{\pm 1/2}$, we have the expressions[12]

$$
\begin{aligned}
O^{1/2} &= -R_{1/2}(\hat{s} + s_0 + 1) - R_{-1/2} s_+, \\
O^{-1/2} &= R_{-1/2}(\hat{s} + s_0) - R_{1/2} s_-.
\end{aligned} \tag{4.6}
$$

The actual shift operators for $D(2,1;\alpha)$

$\supset$su(2) + su(2) + su(2) are obtained from (4.6) by means of a product rule. For instance,

$$
\begin{aligned}
O^{-1/2,-1/2-1/2} &= -R_{1/2,1/2,1/2}s_-t_-u_- + R_{1/2,1/2,-1/2}s_-t_-(\hat u + u_0)\\
&\quad + R_{1/2,-1/2,1/2}s_-(\hat t + t_0)u_-\\
&\quad - R_{1/2,-1/2,-1/2}s_-(\hat t + t_0)(\hat u + u_0)\\
&\quad + R_{-1/2,1/2,1/2}(\hat s + s_0)t_-u_-\\
&\quad - R_{-1/2,1/2,-1/2}(\hat s + s_0)t_-(\hat u + u_0)\\
&\quad - R_{-1/2,-1/2,1/2}(\hat s + s_0)(\hat t + t_0)u_-\\
&\quad + R_{-1/2,-1/2,-1/2}(\hat s + s_0)(\hat t + t_0)(\hat u + u_0). \qquad (4.7)
\end{aligned}
$$

The expressions for the other shift operators $O^{i,j,k}$ follow immediately from (4.6).

The main property of shift operators is[12]

$$
O^{i,j,k}|s,m_s;t,m_t;u,m_u;\lambda\rangle \propto \sum_{\lambda'}|s+i,m_s+i;t+j,m_t+j;u+k,m_u+k;\lambda'\rangle, \qquad (4.8)
$$

which shows that they shift an eigenstate (4.3) into just one (or two, in the twofold degenerate case) eigenstate(s) with unique well-determined subalgebra labels. This is, for instance, in contrast with the operators $R_{i,j,k}$, whose action upon an eigenstate (4.3) gives a linear combination of states

$$|s',m_s+i;t',m_t+j;u',m_u+k;\lambda'\rangle,$$

with

$$s' = s \pm i, \quad t' = t \pm j, \quad u' = u \pm k.$$

It is convenient to use normalized shift[12] operators $A^{i,j,k}$, which are related to the above operators $O^{i,j,k}$ by

$$
\begin{aligned}
A^{i,j,k} = O^{i,j,k}\,&[(\hat s + s_0 + \tfrac12 + i)(\hat t + t_0 + \tfrac12 + j)\\
&\times(\hat u + u_0 + \tfrac12 + k)]^{-1/2}. \qquad (4.9)
\end{aligned}
$$

The whole analysis of $D(2,1;\alpha)$ representations is based upon relations between quadratic products of shift operators of the form $A^{i'j'k'}A^{ijk}$. Such a product is called a "scalar product" if $i'+i=j'+j=k'+k=0$; otherwise, it is a nonscalar product. Note that a scalar shift operator product is a subalgebra scalar, since it commutes with all su(2) + su(2) + su(2) basis elements. The relations between scalar and nonscalar products are summarized in the Appendix.

The states of irreducible representations of $D(2,1;\alpha)$ are connected to each other by means of shift operators. If two states of a representation are not connected to one another by a shift operator (i.e., if the matrix element of the shift operator between those two states vanishes), then the corresponding reduced matrix element of the tensor $R^{[1/2,1/2,1/2]}$ vanishes,[12] showing that the states are not connected to each other by the generators of the superalgebra. Hence, the structure of a $D(2,1;\alpha)$ irrep follows from the analysis of the shift operator matrix elements, which can be deduced from (A1)–(A6).

The fact that the squares of all shift operators $A^{i,j,k}$ vanish [relation (A1)], shows again that a $D(2,1;\alpha)$ irrep $(p;q;r)$ decomposes in general into the (s,t,u) irreps given by (4.2). This, of course, is now true for finite- as well as for infinite-dimensional representations of $D(2,1;\alpha)$. In the infinite-dimensional case, (s,t,u) [and hence also $(p;q;r)$] can be a triplet of real negative numbers.

On account of relations like (A6), it is easy to calculate the I_2 and I_4 eigenvalues for a given $D(2,1;\alpha)$ irrep $(p;q;r)$. We find

$$\langle I_2\rangle = -2[\sigma_1 p(p-1) + \sigma_2 Q^2 + \sigma_3 R^2], \qquad (4.10)$$

$$
\begin{aligned}
\langle I_4\rangle = -16\{&\sigma_1 P^4 + \sigma_2 Q^4 + \sigma_3 R^4 - 2\sigma_3 P^2 Q^2\\
&- 2\sigma_2 P^2 R^2 - 2\sigma_1 Q^2 R^2\\
&- 2(\sigma_2 Q^2 + \sigma_3 R^2)(2p+1)\\
&- 2\sigma_1 p[2Q^2 + 2R^2 + (p+1)(2p-1)]\}, \qquad (4.11)
\end{aligned}
$$

where we use the obvious shorthand notation $P^2 = p(p+1), Q^2 = q(q+1)$, and $R^2 = r(r+1)$. Then relations (A4)–(A6) can be used in order to determine the matrix elements of the shift operator products for a general $(p;q;r)$ representation. When (s,t,u) and $(s-\frac12, t+j, u+k)$ correspond to nondegenerate subalgebra irreps in the $(p;q;r)$ decomposition, i.e, when (s,t,u) and $(s-\frac12,t+j,u+k)$ belong to $\mathscr{T} - \{(p-1,q,r)\}$, the following result is obtained:

$$
\begin{aligned}
A^{1/2,-j,-k}A^{-1/2,j,k}|s,t,u\rangle &= -8jk f_{p-1}^{-1/2}(s) f_q^{j}(t) f_r^{k}(u)\\
&\quad\times[\sigma_1 p + \sigma_2\delta_j(q) + \sigma_3\delta_k(r)]\,|s,t,u\rangle, \qquad (4.12)
\end{aligned}
$$

where

$$
f_a^{1/2}(l) = \begin{cases} 2l+2, & \text{if } l \geq a,\\ 2l+1, & \text{if } l < a;\end{cases}
$$

$$
f_a^{-1/2}(l) = \begin{cases} 2l+1, & \text{if } l > a,\\ 2l, & \text{if } l \leq a;\end{cases}
$$

$$
\delta_i(l) = \begin{cases} -l, & \text{if } i = \frac12,\\ l+1, & \text{if } i = -\frac12.\end{cases}
$$

In (4.12), we have adopted the notation $|s,t,u\rangle$ for a basis state $|s,m_s;t,m_t;u,m_u;\lambda\rangle$ (4.3), since λ is unnecessary and the expression of the matrix element is independent of the m values. This notation ($|s,t,u\rangle$ or $|s,t,u;\lambda\rangle$) will also be used in the following expressions.

The actions of shift operator products, which connect a nondegenerate irrep (s,t,u) with the twofold degenerate irrep $(p-1,q,r)$, are given by

$$
\begin{aligned}
&A^{\pm1/2,\pm1/2,\pm1/2}A^{\mp1/2,\mp1/2,\mp1/2}|p-1\pm\tfrac12,q\pm\tfrac12,r\pm\tfrac12\rangle\\
&= -8\{\sigma_1 p(p-1)(2qr+q+r) + \sigma_2 q(q+1)(2pr-r+p-1) + \sigma_3 r(r+1)(2pq-q+p-1)\}\,|p-1\pm\tfrac12,q\pm\tfrac12,r\pm\tfrac12\rangle, \qquad (4.13)
\end{aligned}
$$

$$
\begin{aligned}
&A^{\pm1/2,\pm1/2,\mp1/2}A^{\mp1/2,\mp1/2,\pm1/2}|p-1\pm\tfrac12,q\pm\tfrac12,r\mp\tfrac12\rangle\\
&= -8\{\sigma_1 p(p-1)(-2qr-q-r-1) + \sigma_2 q(q+1)(-2pr+r-p) + \sigma_3 r(r+1)(2pq-q+p-1)\}
\end{aligned}
$$

$$\times|p-1\pm\tfrac12,q\pm\tfrac12,r\mp\tfrac12\rangle, \tag{4.14}$$

$$A^{\pm 1/2,\mp 1/2,\pm 1/2}A^{\mp 1/2,\pm 1/2,\mp 1/2}|p-1\pm\tfrac12,q\mp\tfrac12,r\pm\tfrac12\rangle$$
$$= -8\{\sigma_1 p(p-1)(-2qr-q-r-1)+\sigma_2 q(q+1)(2pr+p-r-1)+\sigma_3 r(r+1)(-2pq+q-p)\}$$
$$\times|p-1\pm\tfrac12,q\mp\tfrac12,r\pm\tfrac12\rangle, \tag{4.15}$$

$$A^{\pm 1/2,\mp 1/2,\mp 1/2}A^{\mp 1/2,\pm 1/2,\pm 1/2}|p-1\pm\tfrac12,q\mp\tfrac12,r\mp\tfrac12\rangle$$
$$= -8\{\sigma_1 p(p-1)(2qr+q+r)+\sigma_2 q(q+1)(-2pr+r-p)+\sigma_3 r(r+1)(-2pq+q-p)\}|p-1\pm\tfrac12,q\mp\tfrac12,r\mp\tfrac12\rangle. \tag{4.16}$$

Irreducible representations of $D(2,1;\alpha)$, or $\Gamma(\sigma_1,\sigma_2,\sigma_3)$, can now be completely analyzed by making use of (4.12)–(4.16).

V. ANALYSIS OF IRREDUCIBLE REPRESENTATIONS OF $D(2,1;\alpha)$

We first consider finite-dimensional representations of $D(2,1;\alpha)$. For such irreps, the (s,t,u) components must satisfy

$$s,t,u\in\tfrac12\mathbb{N}=\{0,\tfrac12,1,\tfrac32,\ldots\}, \tag{5.1}$$

and hence also $(p;q;r)$ belongs to this set. If $p\geq 2$, $q\geq 1$, and $r\geq 1$, the $(p;q;r)$ irrep decomposes into the (s,t,u) representations given by (4.2), unless, according to (4.12), one of the following conditions is satisfied:

$$\sigma_1 p-\sigma_2 q-\sigma_3 r=0, \tag{5.2}$$

$$\sigma_1 p+\sigma_2(q+1)-\sigma_3 r=0, \tag{5.3}$$

$$\sigma_1 p-\sigma_2 q+\sigma_3(r+1)=0, \tag{5.4}$$

$$\sigma_1 p+\sigma_2(q+1)+\sigma_3(r+1)=0. \tag{5.5}$$

In the latter case several of the matrix elements (4.12) vanish and the 16-dimensional lattice of (s,t,u) values splits up into two eight-dimensional lattices. For instance, in the case of (5.2) the $(p;q;r)$ irrep consists only of

$$\begin{aligned}\mathscr{T}_1=\{&(p,q,r),(p-\tfrac12,q\pm\tfrac12,r\mp\tfrac12),\\&(p-\tfrac12,q-\tfrac12,r-\tfrac12),\\&(p-1,q,r),(p-1,q,r-1),(p-1,q-1,r),\\&(p-\tfrac32,q-\tfrac12,r-\tfrac12)\};\end{aligned}$$

they are not connected to the other (s',t',u') values in $\mathscr{T}$ since (4.12) shows that

$$A^{1/2,-1/2,-1/2}A^{-1/2,1/2,1/2}|s,t,u\rangle=0,\quad \text{if } (s,t,u)\in\mathscr{T}_1.$$

The other eight (s',t',u') values form again an irreducible representation, which is denoted by $(p-\tfrac12;q+\tfrac12;r+\tfrac12)$ and which obviously satisfies again condition (5.2). In fact, (5.2)–(5.5) correspond exactly with the four atypicality conditions for $D(2,1;\alpha)$, also given by Kac.[5] If neither of them is satisfied, then $(p;q;r)$ is a typical representation decomposing into 16 subalgebra irreps; if one of the conditions (5.2)–(5.5) is satisfied, we have a so-called atypical representation,[5] which is in general reducible but indecomposable. Because, in this paper, only irreducible representations are considered, we find the "irreducible parts" of the atypical representations.

If $p<2$, $q<1$, or $r<1$, then we find a truncated (s,t,u) lattice: only those (s,t,u) values in (4.2) for which none of the elements is negative appear in the decomposition of $(p;q;r)$. Moreover, if $p=1$ $(q,r>0)$, or if $q=0$ $(p>1,r>0)$, or if $r=0$ $(p>1,q>0)$, then the multiplicity of $(s,t,u)=(p-1,q,r)$ is only 1. When $q=r=0$ $(p>1)$, the representation $(p-1,q,0)$ does not appear in the decomposition of $(p;0,0)$. Another example of truncated representations is given by the adjoint representation: if $(p;q;r)=(1;0;0)$ then it decomposes as $\{(1,0,0),(\tfrac12,\tfrac12,\tfrac12),(0,1,0),(0,0,1)\}$; this is the 17-dimensional adjoint representation. There are, however, two exceptions to such truncations, as can be verified immediately from (4.12).

(1) If $p=0$, all matrix elements in (4.12) must vanish, and hence we have to require $q=r=0$. This representation $(0;0;0)$ is the one-dimensional trivial representation of $D(2,1;\alpha)$.

(2) If $p=\tfrac12$, the shift operators acting on $s=p-\tfrac12$ must vanish. There remain two possibilities.

(a) $\sigma_2(2q+1)=\sigma_3(2r+1)$. (5.6)

Then the irrep $(\tfrac12;q;r)$ decomposes into $(\tfrac12,q,r)$, $(0,q+\tfrac12,r+\tfrac12)$, and $(0,q-\tfrac12,r-\tfrac12)$ (the latter does not appear if $q=0$ or $r=0$).

(b) $\sigma_2(2q+1)=-\sigma_3(2r+1)$. (5.7)

Then the irrep $(\tfrac12;q;r)$ decomposes into $(\tfrac12,q,r)$, $(0,q+\tfrac12,r-\tfrac12)$ (missing if $r=0$), and $(0,q-\tfrac12,r+\tfrac12)$ (missing if $q=0$).

Kac[5] also gave the supplementary conditions (5.6) and (5.7) when $p=\tfrac12$, but did not give the peculiar structure of the corresponding representations. For osp(4,2), condition (5.6) includes the case $q=r=0$, which gives the six-dimensional natural representation decomposing into $(\tfrac12,0,0)$ and $(0,\tfrac12,\tfrac12)$. Note that for $\text{osp}(4,2)=D(2,1;1)=\Gamma(2,-1,-1)$ the two lowest-dimensional nontrivial representations have dimensions 6 and 17 (the natural and the adjoint representation), but due to (5.6) and (5.7) other representations with dimensions between 6 and 17 might occur for $\alpha\neq 1$. For instance, if $\alpha=2$, then according to (5.6), the lowest-dimensional representation of $D(2,1;2)=\Gamma(3,-1,-2)$ is the 10-dimensional irrep $(\tfrac12;\tfrac12;0)$ decomposing into $(\tfrac12,\tfrac12,0)$ and $(0,1,\tfrac12)$. More generally, if α is a positive integer, then $D(2,1;\alpha)$ has a $(4\alpha+2)$-dimensional irrep $(\tfrac12;(\alpha-1)/2;0)$ which reduces into $(\tfrac12,(\alpha-1)/2,0)$ and $(0,\alpha/2,\tfrac12)$. If α is a negative integer $(\alpha<-1)$, then $D(2,1;\alpha)$ has the $(-4\alpha-2)$-dimensional irrep $(\tfrac12;-(\alpha+1)/2;0)$ decomposing into $(\tfrac12,-(\alpha+1)/2,0)$ and $(0,-(\alpha+2)/2,\tfrac12)$.

From (4.12)–(4.16), the matrix elements of the shift operator $A^{i,j,k}$ can be determined, up to some arbitrary multiplicative constants (which become arbitrary phase factors if a Hermiticity condition is given for the algebra and if the representation is star or grade star). For the missing label problem, we refer to expressions (5.12)–(5.16). Then we make use of relation (5.8) (see Ref. 12) and the Wigner–Eckart[17] theorem (5.9)

$$\langle s+i,t+j,u+k;\lambda' \| R^{[1/2,1/2,1/2]} \| s,t,u;\lambda\rangle = \left[\frac{(2s+1+2i)(2t+1+2j)(2u+1+2k)}{(2s+1)(2t+1)(2u+1)}\right]^{1/2} \times \langle s+i,m_s+i;t+j,m_t+j;u+k,m_u+k; \lambda' | A^{ijk} | s,m_s;t,m_t;u,m_u;\lambda\rangle, \tag{5.8}$$

$$R_{i,j,k}|s,m_s;t,m_t;u,m_u;\lambda\rangle = \sum(-1)^{s'-m_s'+t'-m_t'+u'-m_u'} \times \begin{pmatrix} s' & \frac{1}{2} & s \\ -m_s' & i & m_s \end{pmatrix} \begin{pmatrix} t' & \frac{1}{2} & t \\ -m_t' & j & m_t \end{pmatrix} \times \begin{pmatrix} u' & \frac{1}{2} & u \\ -m_u' & k & m_u \end{pmatrix} \times \langle s',t',u';\lambda' \| R^{[1/2,1/2,1/2]} \| s,t,u;\lambda\rangle s',t',u';\lambda'\rangle. \tag{5.9}$$

This gives us the proper action of all tensor components $R_{i,j,k}$ upon the basis states. Together with the well-known actions of the subalgebra generators upon the states (4.3), this determines the explicit forms of the representatives of all the basis elements of the superalgebra for all irreducible representations.

The infinite-dimensional representations of $D(2,1;\alpha)$ decompose into infinite-dimensional irreps of the subalgebra su(2) + su(2) + su(2). Several cases are possible, depending on which su(2) part is infinite-dimensional. For the star or grade star representations considered in Sec. VI, the subalgebra satisfies Hermiticity conditions corresponding to those for an su(2) or su(1,1) algebra. The unitary su(2) irreps are finite dimensional, and labeled by $l \in \frac{1}{2}\mathbb{N}$, whereas the unitary irreps of su(1,1) are infinite dimensional, and labeled by $l \in \mathbb{R}$ (or, more generally, also by the complex number $l = -\frac{1}{2} + i\rho, \rho \in \mathbb{R}$) (see Ref. 18). If the even part of $D(2,1;\alpha)$ is, in an obvious notation, su(1,1) + su(2) + su(2), then the label p may be a real (negative) number, but q and r still belong to $\frac{1}{2}\mathbb{N}$. Because of the symmetric contents of the three su(2) parts in $\Gamma(\sigma_1,\sigma_2,\sigma_3)$, we shall discuss in detail only the cases su(1,1) + su(2) + su(2), su(2) + su(1,1) + su(1,1), and su(1,1) + su(1,1) + su(1,1). For all the cases, the general $(p;q;r)$ irreps still decompose into the (s,t,u) representations given by (4.2), except when one of the atypicality conditions (5.2)–(5.5) is satisfied, in which case the 16-dimensional lattice of (s,t,u) values splits up into two eight-dimensional parts. The only difference with the finite-dimensional case is that now no such truncations of the (s,t,u) lattice appear when one of the labels corresponds to su(1,1) (because such a label can take on all real values). We discuss this for the following possibilities.

(1) su(1,1) + su(2) + su(2). Now $s \in \mathbb{R}$, but $t,u \in \frac{1}{2}\mathbb{N}$, and this also applies for p,q, and r. The (s,t,u) of $\mathscr{T}$ (4.2) for which t and/or u would become negative are deleted. This gives the structure of all $D(2,1;\alpha)$ irreps in an su(1,1) + su(2) + su(2) basis.

Interesting cases occur when "truncations" arise (because of small q or r) simultaneously with one of the atypicality conditions (5.2)–(5.5). For instance, when $q=r=0$ and p satisfies $\sigma_1 p + \sigma_3 = 0$ (and $\sigma_1 p + \sigma_2 \neq 0$), then it is easy to see that the irrep $(p;0;0)$ decomposes only into $(p,0,0) + (p-\frac{1}{2},\frac{1}{2},\frac{1}{2}) + (p-1,0,0)$.

When $(q,r) = (\frac{1}{2},0)$, and $\sigma_1 p - \sigma_2 q = 0$ [or, in the $D(2,1;\alpha)$ notation: $p = -1/2(1+\alpha)$], then the irrep $(-1/2(1+\alpha);\frac{1}{2};0)$ decomposes into the doublet of subalgebra irreps $(-1/2(1+\alpha),\frac{1}{2},0) + (-(\alpha+2)/2(1+\alpha),0,\frac{1}{2})$. Because of their simple structure, these doublet representations will be considered in detail in Section VII. A similar doublet representation is $(-\alpha/2(1+\alpha);0;\frac{1}{2})$, decomposing into $(-\alpha/2(1+\alpha),0,\frac{1}{2}) + (-(2\alpha+1)/2(1+\alpha),\frac{1}{2},0)$.

(2) su(2) + su(1,1) + su(1,1). Now $s \in \frac{1}{2}\mathbb{N}$, $t,u \in \mathbb{R}$, and only those (s,t,u) values for which s is negative are deleted from $\mathscr{T}$. Similarly to the previous case, there are special representations because of the simultaneous appearance of atypicality conditions and "truncations." For instance, if $p=1$ and $q + \alpha r + (1+\alpha) = 0$, then the irrep $(1;q;r)$ decomposes into $(1,q,r),(\frac{1}{2},q\pm\frac{1}{2},r-\frac{1}{2})$, $(\frac{1}{2},q-\frac{1}{2},r+\frac{1}{2})$, $(0,q,r)$, $(0,q-1,r)$, and $(0,q,r-1)$. When $p=\frac{1}{2}$, there are again the two possibilities (5.6) or (5.7) with their corresponding reduction rule.

(3) su(1,1) + su(1,1) + su(1,1). In this case, all labels can be real and negative, and consequently no truncations arise. All $(p;q;r)$ irreps $(p,q,r \in \mathbb{R})$ reduce as prescribed in (4.2), except when one of the conditions (5.2)–(5.5) is satisfied, in which case the representations split into two parts, both consisting of the sum of eight subalgebra irreps.

The construction (5.8) and (5.9) remains valid in the infinite-dimensional case, if one takes the "analytic continuations" (see Ref. 18, pp. 195–206) of the expressions for the Wigner $3j$ symbols. However, one has to exclude certain possibilities because of the appearance of denominators like $(2s+1)$, etc. This problem has been discussed in Ref. 8. A detailed study for $D(2,1;\alpha)$ finally showed that we have to exclude the infinite-dimensional irreps $(p;q;r)$ for which the $\mathrm{su}(1,1)_s$ label p is $0,\frac{1}{2}$, or 1, or for which the $\mathrm{su}(1,1)_t$ label q or $\mathrm{su}(1,1)_u$ label r is $0, -\frac{1}{2}$, or -1.

Finally, we consider the missing label problem for general $(p;q;r)$ irreps of $D(2,1;\alpha)$. As we mentioned before, $(p-1,q,r)$ is the only subalgebra irrep in the reduction of $(p;q;r)$ which appears with multiplicity 2. We define the operator

$$L = A^{1/2,1/2,1/2} A^{-1/2,-1/2,-1/2}. \tag{5.10}$$

Obviously, L is an su(2) + su(2) + su(2) scalar operator, and hence a good candidate for the labeling operator. In terms of the scalars (3.6), L is given by

$$L = -\sqrt{2}C^{(110,2)}\hat{u} - \sqrt{2}C^{(101,2)}\hat{t} - \sqrt{2}C^{(011,2)}\hat{s} - 2I_2\hat{s}\hat{t}\hat{u} - 4\hat{s}\hat{t}\hat{u}[\sigma_1(\hat{s}+1)(\hat{s}+2) + \sigma_1(\hat{t}+1)(\hat{t}+2) + \sigma_3(\hat{u}+1)(\hat{u}+2)]. \tag{5.11}$$

Furthermore, we define states $|p-1,q,r;i\rangle$ $(i=1,2)$ by means of

$$A^{-1/2,-1/2,-1/2}|p-\tfrac{1}{2},q+\tfrac{1}{2},r+\tfrac{1}{2}\rangle = |p-1,q,r;1\rangle, \tag{5.12}$$

$$A^{1/2,1/2,1/2}|p-\tfrac{3}{2},q-\tfrac{1}{2},r-\tfrac{1}{2}\rangle = |p-1,q,r;2\rangle. \tag{5.13}$$

Then (A1) implies

$$L\,|p-1,q,r;1\rangle = 0, \tag{5.14}$$

and (4.13) produces

$$L\,|p-1,q,r;2\rangle = \lambda_2|p-1,q,r;2\rangle, \tag{5.15}$$

where

$$\lambda_2 = -8\{\sigma_1 p(p-1)(2qr+q+r) + \sigma_2 q(q+1)(2pr-r+p-1) + \sigma_3 r(r+1)(2pq+p-q-1)\}. \tag{5.16}$$

Suppose we have a $(p;q;r)$ irrep for which (5.16) is nonzero. Then the above expressions show that we have constructed the two independent eigenstates of L, $|p-1,q,r;\lambda_i\rangle = |p-1,q,r;i\rangle$ $(i=1,2)$, with eigenvalues $\lambda_1=0$ and λ_2 given by (5.16). This solves the missing label problem.

When, for some special cases, (5.16) turns out to be zero, it is easy to see that another choice of L (for instance, $L = A^{1/2,1/2,-1/2}A^{-1/2,-1/2,1/2}$) and of $|p-1,q,r;i\rangle$ gives rise to a similar result with λ_2 one of the expressions in (4.14)–(4.16). If the expression in (4.13) is zero, at least one of the expressions in (4.14)–(4.16) is nonzero, and then the corresponding operator L is a good labeling operator.

VI. STAR AND GRADE STAR REPRESENTATIONS

The equivalents of Hermitian operations for Lie algebras are star and grade star operations for Lie superalgebras.[13] For a Lie superalgebra $L = L_{\bar 0} + L_{\bar 1}$, the operation † (resp.‡) which maps L_α into L_α $(\alpha = \bar 0, \bar 1)$, is a star (resp. grade star), operation if

$$(aA+bB)^\dagger = a^*A^\dagger + b^*B^\dagger,$$
$$\text{resp. } (aA+bB)^\ddagger = a^*A^\ddagger + b^*B^\ddagger,$$
$$[A,B]^\dagger = [B^\dagger,A^\dagger], \quad \text{resp. } [A,B]^\ddagger = (-1)^{\alpha\beta}[B^\ddagger,A^\ddagger],$$
$$(A^\dagger)^\dagger = A, \quad \text{resp. } (A^\ddagger)^\ddagger = (-1)^\alpha A, \tag{6.1}$$

for all elements (resp. for all homogeneous elements) A and B of L and for all complex numbers a,b. The notation * denotes the complex conjugate, and α(resp. β) are the degrees of A and B.

Definition (6.1) implies that the restriction of a star or grade star operation to the even part $L_{\bar 0}$ is a Hermitian operation of the Lie algebra $L_{\bar 0}$. Therefore, we consider all possible Hermitian operations on the even part, and investigate whether it is possible to extend them to a star or grade star operation for the Lie superalgebra. The even part of $D(2,1;\alpha)$ is $\mathrm{su}(2)_s + \mathrm{su}(2)_t + \mathrm{su}(2)_u$, and we shall consider only those Hermitian operations which map the elements of each su(2) subalgebra into the same su(2) [i.e., $\mathrm{su}(2)_s \to \mathrm{su}(2)_s$, etc.]. Since there are two distinct operations for an su(2) algebra, denoted by

$$\begin{aligned} &\mathrm{su}(2): && s_0^\dagger = s_0, \quad s_\pm^\dagger = s_\mp\,, \\ &\mathrm{su}(1,1): && s_0^\dagger = s_0, \quad s_\pm^\dagger = -s_\mp\,, \end{aligned} \tag{6.2}$$

we have eight different Hermitian operations for $\mathrm{su}(2)+\mathrm{su}(2)+\mathrm{su}(2)$. We find the following result for $D(2,1;\alpha) = \Gamma(\sigma_1,\sigma_2,\sigma_3)$: If and only if all σ_i are real (or if and only if $\alpha \in \mathbb{R}$), each of the eight Hermitian operations for $D(2,1;\alpha)_{\bar 0}$ can be extended in two possible ways (i.e., $\epsilon = 1$ or $\epsilon = -1$) to a star (S) or grade star (GS) operation for $D(2,1;\alpha)$. In (6.3) we summarize for which cases there is a GS or S operation, and we give their explicit form:

(1) su(2) + su(2) + su(2):GS

$$s_0^\ddagger = s_0,\ s_\pm^\ddagger = s_\mp;\quad t_0^\ddagger = t_0,\ t_\pm^\ddagger = t_\mp;$$
$$u_0^\ddagger = u_0,\ u_\pm^\ddagger = u_\mp;\quad R^\ddagger_{i,j,k} = \epsilon(8ijk)R_{-i,-j,-k};$$

(2) su(1,1) + su(2) + su(2):S

$$s_0^\dagger = s_0,\ s_\pm^\dagger = -s_\mp;\quad t_0^\dagger = t_0,\ t_\pm^\dagger = -t_\mp;$$
$$u_0^\dagger = u_0,\ u_\pm^\dagger = -u_\mp;\quad R^\dagger_{i,j,k} = \epsilon(4jk)R_{-i,-j,-k};$$

(3),(4) su(2) + su(1,1) + su(2) and su(2) + su(2) + su(1,1):S, and the explicit operations are the analog of (2);

(5) su(2) + su(1,1) + su(1,1):GS

$$s_0^\ddagger = s_0,\ s_\pm^\ddagger = s_\mp;\quad t_0^\ddagger = t_0,\ t_\pm^\ddagger = -t_\mp;$$
$$u_0^\ddagger = u_0,\ u_\pm^\ddagger = -u_\mp;\quad R^\ddagger_{i,j,k} = \epsilon(2i)R_{-i,-j,-k};$$

(6),(7) su(1,1) + su((2) + su(1,1) and su(1,1) + su(1,1) + su(2):GS, and the explicit expressions are the analog of (5);

(8) su(1,1) + su(1,1) + su(1,1): S

$$s_0^\dagger = s_0,\quad s_\pm^\dagger = -s_\mp;\quad t_0^\dagger = t_0,\quad t_\pm^\dagger = -t_\mp;$$
$$u_0^\dagger = u_0,\quad u_\pm^\dagger = -u_\mp;\quad R^\dagger_{i,j,k} = \epsilon R_{-i,-j,-k}$$

where $\epsilon = \pm 1$, and $i,j,k \in \{-\tfrac{1}{2},\tfrac{1}{2}\}$. (6.3)

Let ρ be a representation of the Lie superalgebra L into a graded representation space $V = V_{\bar 0} + V_{\bar 1}$, with positive definite nondegenerate Hermitian form $\langle\,|\,\rangle$ satisfying

$$\langle V_{\bar 0}|V_{\bar 1}\rangle = \{0\}, \tag{6.4}$$

such that V is a graded Hilbert space.[13] Let pl(V) be the graded vector space[1] of linear mappings of V into itself. For every homogeneous element A of pl(V), the adjoint operator $A^\dagger$ (resp. grade adjoint operator $A^\ddagger$) is defined by[13]

$$\langle A^\dagger x|y\rangle = \langle x|Ay\rangle, \quad \forall x,y \in V,$$
$$\text{resp.} \langle A^\ddagger x|y\rangle = (-1)^{\alpha\xi}\langle x|Ay\rangle,$$
$$\forall x \in V_\xi, \quad \forall y \in V, \tag{6.5}$$

where $\alpha = $ degree (A). Then, the representation $\rho: L \to \mathrm{pl}(V)$ is a star (resp. grade star) representation if for all $A \in L_\alpha (\alpha = \bar 0, \bar 1)$

$$\rho(A^\dagger) = (\rho(A))^\dagger, \quad \text{resp.} \rho(A^\ddagger) = (\rho(A))^\ddagger. \tag{6.6}$$

We now investigate which of the representations, considered in Sec. V, are star or grade star representations. In all cases, the representation space V is spanned by the basis vectors (4.3), which are the mutual eigenvectors of the set of commuting operators $W = \{S^2, s_0, T^2, t_0, U^2, u_0, L\}$. It is easy to verify that for the eight possibilities given in (6.3), all operators in W are self-adjoint, i.e., they always satisfy $A^\dagger = A$ or $A^\ddagger = A$ (for $A \in W$). This restricts the choice of an inner product on V, and it is natural to define

$$\langle s',m_s';t',m_t';u',m_u';\lambda'|s,m_s;t,m_t;u,m_u;\lambda\rangle = \delta_{ss'}\delta_{m_sm_s'}\delta_{tt'}\delta_{m_tm_t'}\delta_{uu'}\delta_{m_um_u'}\delta_{\lambda\lambda'}. \tag{6.7}$$

In our discussion of star and grade star representations, we shall consider only the cases (1), (2), (5), and (8) of (6.3), since the remaining possibilities are quite analogous due to the symmetry of the three su(2) parts in $\Gamma(\sigma_1,\sigma_2,\sigma_3)$. Furthermore, because of (2.10) and (2.2), we can restrict ourselves to the $D(2,1;\alpha)$ algebras with $\alpha>0$ [in fact $D(2,1;\alpha)\simeq D(2,1;-1-\alpha)\simeq D(2,1;-1-1/\alpha)$].

A. Finite-dimensional grade star representations

Only the first choice in (6.3) corresponds to finite-dimensional representations of $D(2,1;\alpha)$. We investigate which of the finite-dimensional representations are in fact grade star representations. For this, we make use of the adjoint operations of the shift operators. Making use of Ref. 12, we obtain in case (1) of (6.3) ($i,j,k=\pm\frac{1}{2}$)

$$\begin{aligned}(A^{i,j,k})^{\ddagger}(2\hat{s}+1)(2\hat{t}+1)(2\hat{u}+1) &= -\epsilon(8ijk)A^{-i,-j,-k}(2\hat{s}-2i+1)(2\hat{t}-2j+1)\\ &\quad\times(2\hat{u}-2k+1).\end{aligned} \tag{6.8}$$

This implies, since the $A^{i,j,k}$ are odd operators, that

$$\begin{aligned}\langle s,t,u;\lambda|A^{i,j,k}A^{-i,-j,-k}|s,t,u;\lambda\rangle &= \sum_{\lambda'}(-\epsilon)(8ijk)(-1)^{\sigma(s,t,u)}\\ &\quad\times\frac{(2s-2i+1)(2t-2j+1)(2u-2k+1)}{(2s+1)(2t+1)(2u+1)}\\ &\quad\times|\langle s-i,t-j,u-k;\lambda'|A^{-i,-j,-k}|s,t,u;\lambda\rangle|^2,\end{aligned} \tag{6.9}$$

where $\sigma(s,t,u)$ is the degree of the state $|s,t,u;\lambda\rangle$. Relation (6.9) provides us with the positivity and negativity conditions for the matrix elements of the shift operator products. We discuss the case $\epsilon=1$ and $\sigma(p,q,r)=\bar{0}$—the remaining possibilities give rise to analogous results. First we consider the general case in which the $D(2,1;\alpha)$ irrep reduces into 16 subalgebra representations whose (s,t,u) values are given by (4.2). From (4.12) and (6.9) we obtain

$$\begin{aligned}|\langle p-\tfrac{1}{2},q+\tfrac{1}{2},r-\tfrac{1}{2}|A^{-1/2,1/2,-1/2}|p,q,r\rangle|^2 &= [(2p+1)^2/p](2q+1)(2r+1)\\ &\quad\times[\sigma_1p-\sigma_2q+\sigma_3(r+1)],\end{aligned} \tag{6.10}$$

$$\begin{aligned}|\langle p-1,q+1,r|A^{-1/2,1/2,-1/2}|p-\tfrac{1}{2},q+\tfrac{1}{2},r+\tfrac{1}{2}\rangle|^2 &= -2\frac{(2p)^2}{(2p-1)}(2q+2)\frac{(2r+2)^2}{(2r+1)}\\ &\quad\times[\sigma_1p-\sigma_2q+\sigma_3(r+1)].\end{aligned} \tag{6.11}$$

Obviously, the positivity conditions in (6.10) and (6.11) cannot be satisfied simultaneously, and hence the representation is not a grade star representation. In fact, a detailed analysis of all possibilities, making use of similar arguments to (6.10) and (6.11), shows that the only grade star representations are (1) the trivial representation (0;0;0); and (2) for $\alpha\in\mathbb{N}$, the irreps $(\frac{1}{2};(\alpha-1)/2;0)$, with even states $|\frac{1}{2},\pm\frac{1}{2};(\alpha-1)/2,m_t;0,0\rangle$ and odd states $|0,0;\alpha/2,m_t;\frac{1}{2},\pm\frac{1}{2}\rangle$. The conclusion is the same as for osp(3,2) (see Ref. 8): only a few finite-dimensional irreps of $D(2,1;\alpha)$ are grade star representations.

B. Infinite-dimensional star representations in the case su(1,1) + su(2) + su(2)

We prefer to work with the $O^{i,j,k}$ operators instead of the $A^{i,j,k}$, because of the complicated internal structure of infinite-dimensional su(2) irreps. The star conditions [6.3(2)] imply

$$\begin{aligned}(O^{i,j,k})^{\dagger}(2\hat{s}+1)(2\hat{t}+1)(2\hat{u}+1) &= -\epsilon(4jk)O^{-i,-j,-k}\\ &\quad\times(2\hat{s}-2i+1)(2\hat{t}-2j+1)(2\hat{u}-2k+1),\end{aligned} \tag{6.12}$$

from which we obtain

$$\begin{aligned}&(2s+1)(2t+1)(2u+1)\\ &\quad\times\langle s,t,u;\lambda|O^{i,j,k}O^{-i,-j-k}|s,t,u;\lambda\rangle\\ &= \sum_{\lambda'}(-\epsilon)(4jk)(2s-2i+1)(2t-2j+1)\\ &\quad\times(2u-2k+1)|\langle s-i,t-j,u-k;\lambda'|\\ &\quad\times O^{-i,-j,-k}|s,t,u;\lambda\rangle|^2.\end{aligned} \tag{6.13}$$

We consider a general $(p;q;r)$ irrep, and obtain from (6.13) and (4.12) (in the case $\epsilon=1$)

$$\begin{aligned}&|\langle p-\tfrac{1}{2},m_p-\tfrac{1}{2};q+\tfrac{1}{2},m_q+\tfrac{1}{2};r+\tfrac{1}{2},m_r+\tfrac{1}{2}|O^{-1/2,1/2,1/2}|p,m_p;q,m_q;r,m_r\rangle|^2\\ &\quad=2[(2p+1)^2/(2p)](2q+1)(2r+1)(\sigma_1p-\sigma_2q-\sigma_3r)(p+m_p)(q+m_q+1)(r+m_r+1),\end{aligned} \tag{6.14}$$

$$\begin{aligned}&|\langle p-\tfrac{1}{2},m_p-\tfrac{1}{2};q+\tfrac{1}{2},m_q+\tfrac{1}{2};r-\tfrac{1}{2},m_r-\tfrac{1}{2}|O^{-1/2,1/2,-1/2}|p,m_p;q,m_q;r,m_r\rangle|^2\\ &\quad=2[(2p+1)^2/(2p)](2q+1)(2r+1)[\sigma_1p-\sigma_2q+\sigma_3(r+1)](p+m_p)(q+m_q+1)(r+m_r),\end{aligned} \tag{6.15}$$

$$\begin{aligned}&|\langle p-\tfrac{1}{2},m_p-\tfrac{1}{2};q-\tfrac{1}{2},m_q-\tfrac{1}{2};r+\tfrac{1}{2},m_r+\tfrac{1}{2}|O^{-1/2,-1/2,1/2}|p,m_p;q,m_q;r,m_r\rangle|^2\\ &\quad=2[(2p+1)^2/(2p)](2q+1)(2r+1)[\sigma_1p+\sigma_2(q+1)-\sigma_3r](p+m_p)(q+m_q)(r+m_r+1),\end{aligned} \tag{6.16}$$

$$\begin{aligned}&|\langle p-\tfrac{1}{2},m_p-\tfrac{1}{2};q-\tfrac{1}{2},m_q-\tfrac{1}{2};r-\tfrac{1}{2},m_r-\tfrac{1}{2}|O^{-1/2,-1/2,-1/2}|p,m_q;q,m_q;r,m_r\rangle|^2\\ &\quad=2[(2p+1)^2/(2p)](2q+1)(2r+1)[\sigma_1p+\sigma_2(q+1)+\sigma_3(r+1)](p+m_p)(q+m_q)(r+m_r).\end{aligned} \tag{6.17}$$

Let us first suppose that $\sigma_1p-\sigma_2q-\sigma_3r\geq0$, or in the $D(2,1;\alpha)$ notation: $(1+\alpha)p+q+\alpha r\geq0$. Then (6.14) implies that $p(p+m_p)\geq0$. Making use of this last inequality in (6.15)–(6.17) gives, together with the assumption, the four conditions

$$\begin{aligned}&(1+\alpha)p+q+\alpha r\geq0,\\ &(1+\alpha)p+q-\alpha(r+1)\geq0,\\ &(1+\alpha)p-(q+1)+\alpha r\geq0,\\ &(1+\alpha)p-(q+1)-\alpha(r+1)\geq0.\end{aligned} \tag{6.18}$$

Since $q,r \in \frac{1}{2}\mathbb{N}$ and $\alpha > 0$, this means that p must be a positive real number, so that $p(p+m_p) \geq 0$ implies that the m_p values are bounded from below. This shows that we have a positive discrete su(1,1) representation D^+, labeled by p, and with minimum m_p value $\underline{m}_p = p + 1$ [for a brief review of infinite-dimensional unitary representations of su(1,1), see Refs. 8 or 18]. All the other shift operator matrix elements are analyzed in a similar way, and we find the following result.

The representation $(p;q;r)$, with $p > 0$ and $q,r \in \frac{1}{2}\mathbb{N}$, and such that the four conditions (6.18) are satisfied, is an infinite-dimensional star representation of $D(2,1;\alpha)$ for which the basis states are given by $|s,m_s;\ t,m_t;u,m_u;\lambda\rangle$ with $(s,t,u) \in \mathscr{T}(4.2)$, $m_s = s+1, s+2,..., m_t = -t, -t+1,..., +t$, and $m_u = -u, -u+1,..., +u$. In other words, the representation decomposes into su(1,1) + su(2) + su(2) irreps which consist of the direct product of a positive discrete D^+ with two finite-dimensional su(2) representations. There is a second general solution, consisting of representations $(p;q;r)$ with $p < 0$, $q,r \in \frac{1}{2}\mathbb{N}$, satisfying

$$(1+\alpha)p + q + \alpha r \leq 0,$$
$$(1+\alpha)p + q - \alpha(r+1) \leq 0,$$
$$(1+\alpha)p - (q+1) + \alpha r \leq 0, \tag{6.19}$$
$$(1+\alpha)p - (q+1) - \alpha(r+1) \leq 0.$$

Such irreps are again star representations of $D(2,1;\alpha)$, and the (s,t,u) representations in which they decompose consist of the direct product of a D^+ (with minimum m value $\underline{m}_s = -s$) with two finite-dimensional su(2) irreps labeled by t and u. Also the truncated or the atypical representations are star representations, if they satisfy (6.18) or (6.19). In particular we consider the doublet representations $(-1/2(1+\alpha);\frac{1}{2};0)$. Because in this case the two corresponding s values are between -1 and 0, we find two solutions for the star conditions: (1) states $|-1/2(1+\alpha), m_s;\frac{1}{2}, \pm\frac{1}{2};0,0\rangle$ with $\underline{m}_s = 1/2(1+\alpha)$, connected to $|-(\alpha+2)/2(1+\alpha), m_s'; 0,0;\frac{1}{2}, \pm\frac{1}{2}\rangle$ with $\underline{m}_s' = (\alpha+2)/2(1+\alpha)$; (2) states $|-1/2(1+\alpha), m_s;\ \frac{1}{2}, \pm\frac{1}{2};0,0\rangle$ with $\underline{m}_s = (2\alpha+1)/2(1+\alpha)$, connected to $|-(\alpha+2)/2(1+\alpha), m_s'; 0,0;\frac{1}{2}, \pm\frac{1}{2}\rangle$ with $\underline{m}_s' = \alpha/2(1+\alpha)$. The explicit form for the representatives of the $D(2,1;\alpha)$ elements in the doublet representations will be given in Sec. VII.

For $\epsilon = -1$, we find similar results: all the representations are again star representations, but the su(1,1) part now consists of a negative discrete series D^- instead of D^+.

C. Grade star conditions in the case su(2) + su(1,1) + su(1,1)

The grade star operation [6.3(5)] leads to the following property for the shift operators:

$$(O^{i,j,k})^\ddagger(2\hat{s}+1)(2\hat{t}+1)(2\hat{u}+1) = -\epsilon(2i)O^{-i,-j,-k}(2\hat{s}-2i+1)(2\hat{t}-2j+1)(2\hat{u}-2k+1), \tag{6.20}$$

from which we deduce

$$(2s+1)(2t+1)(2u+1)\langle s,t,u;\lambda|O^{i,j,k}O^{-i,-j,-k}|s,t,u;\lambda\rangle$$
$$= \sum_{\lambda'}(-\epsilon)(2i)(-1)^{\sigma(s,t,u)}(2s-2i+1)(2t-2j+1)(2u-2k+1)|\langle s-i,t-j,u-k;\lambda'|O^{-i,-j,-k}|s,t,u;\lambda\rangle|^2, \tag{6.21}$$

where $\sigma(s,t,u)$ is the degree of the state $|s,t,u;\lambda\rangle$.

Let us consider a general irrep $(p;q;r)$, with $p \in \frac{1}{2}\mathbb{N}$ and $q,r \in \mathbb{R}^+$. From (4.12) and (6.21) we determine the following expressions:

$$|\langle p-\tfrac{1}{2},m_p-\tfrac{1}{2};q+\tfrac{1}{2},m_q+\tfrac{1}{2};r+\tfrac{1}{2},m_r+\tfrac{1}{2}|O^{-1/2,1/2,1/2}|p,m_p;q,m_q;r,m_r\rangle|^2$$
$$= 2(-1)^{\sigma(p,q,r)}\epsilon[(2p+1)^2/(2p)](2q+1)(2r+1)(\sigma_1 p - \sigma_2 q - \sigma_3 r)(p+m_p)(q+m_q+1)(r+m_r+1), \tag{6.22}$$

$$|\langle p-\tfrac{1}{2},m_p-\tfrac{1}{2};q+\tfrac{1}{2},m_q+\tfrac{1}{2};r-\tfrac{1}{2},m_r-\tfrac{1}{2}|O^{-1/2,1/2,-1/2}|p,m_p;q,m_q;r,m_r\rangle|^2$$
$$= -2(-1)^{\sigma(p,q,r)}\epsilon[(2p+1)^2/(2p)](2q+1)(2r+1)[\sigma_1 p - \sigma_2 q + \sigma_3(r+1)](p+m_p)(q+m_q+1)(r+m_r), \tag{6.23}$$

$$|\langle p-1,m_p-1;q+1,m_q+1;r,m_r|O^{-1/2,1/2,-1/2}|p-\tfrac{1}{2},m_p-\tfrac{1}{2};q+\tfrac{1}{2},m_q+\tfrac{1}{2};r+\tfrac{1}{2},m_r+\tfrac{1}{2}\rangle|^2$$
$$= -2(-1)^{\sigma(p-1/2,q+1/2,r+1/2)}\epsilon[(2p)^2/(2p-1)](2q+1)[(2r+2)^2/(2r+1)][\sigma_1 p - \sigma_2 q + \sigma_3(r+1)]$$
$$\times(p+m_p-1)(q+m_q+2)(r+m_r+1). \tag{6.24}$$

We first investigate the case $\epsilon = +1$ and $\sigma(p,q,r) = \bar{0}$. The latter condition implies $\sigma(p-\frac{1}{2},q+\frac{1}{2},r+\frac{1}{2}) = \bar{1}$. Suppose that $\sigma_1 p - \sigma_2 q - \sigma_3 r \geq 0$. Then (6.22) leads to

$$(2q+1)(2r+1)(q+m_q+1)(r+m_r+1) \geq 0. \tag{6.25}$$

Making use of this result in (6.23) gives

$$\sigma_1 p - \sigma_2 q + \sigma_3(r+1) \leq 0. \tag{6.26}$$

But under these conditions the right-hand side of (6.24) turns out to be negative, which leads to a contradiction. Hence, the irrep under consideration is not a grade star representation. We have the same conclusion for $\sigma_1 p - \sigma_2 q - \sigma_3 r \leq 0$, or for the other possible choices for ϵ and $\sigma(p,q,r)$. A detailed analysis showed also that the truncated and the atypical representations violate the positivity conditions. As a consequence, none of the infinite-dimensional $D(2,1;\alpha)$ irreps in an su(2) + su(1,1) + su(1,1) basis are grade star representations.

D. Star conditions in the case su(1,1) + su(1,1) + su(1,1)

The star conditions [6.3(8)] imply

$$(O^{i,j,k})^\dagger(2\hat{s}+1)(2\hat{t}+1)(2\hat{u}+1)$$
$$= -\epsilon O^{-i,-j,-k}(2\hat{s}-2i+1)(2\hat{t}-2j+1)$$

$$\times(2\hat{u}-2k+1), \tag{6.27}$$

$$\begin{aligned}(2s+1)(2t+1)(2u+1)\\ \times\langle s,t,u;\lambda\,|O^{i,j,k}O^{-i,-j,-k}|s,t,u;\lambda\rangle\\ =\sum_{\lambda'}(-\epsilon)(2s-2i+1)(2t-2j+1)(2u-2k+1)\\ \times|\langle s-i,t-j,u-k;\lambda'|O^{-i,-j,-k}|s,t,u;\lambda\rangle|^2.\end{aligned} \tag{6.28}$$

Just as for the previous cases, we make use of this and (4.12) in order to obtain the positivity conditions for the shift operator matrix elements. These conditions give rise to the following relations for p, q, and r:

$$\sigma_1 p-\sigma_2 q-\sigma_3 r\geq 0, \tag{6.29a}$$

$$\sigma_1 p-\sigma_2 q+\sigma_3(r+1)\leq 0, \tag{6.29b}$$

$$\sigma_1 p+\sigma_2(q+1)-\sigma_3 r\leq 0, \tag{6.29c}$$

$$\sigma_1 p+\sigma_2(q+1)+\sigma_3(r+1)\geq 0. \tag{6.29d}$$

But (6.29a) and (6.29d) imply $\sigma_1(2p-1)\geq 0$, whereas (6.29b) and (6.29c) imply $\sigma_1(2p-1)\leq 0$. This shows that there is no admissable solution for p,q, and r which satisfies (6.29). Hence, the infinite-dimensional $D(2,1;\alpha)$ irreps in an su(1,1) + su(1,1) + su(1,1) basis are not star representations.

VII. THE DOUBLET REPRESENTATION FOR $D(2,1;\alpha)$

The doublet representation, considered in Sec. VI B, is a star representation of $D(2,1;\alpha)$. Because of its simple structure and the connection with the metaplectic representation,[7,8] we study this infinite-dimensional representation in detail.

We denote the states of the doublet representation $(-1/2\,(1+\alpha);\frac{1}{2};0)$ by

$$\varphi^{\pm}_{\alpha,n}=\left|\frac{-1}{2(1+\alpha)},\frac{2n\alpha+2n+1}{2(1+\alpha)};\frac{1}{2},\pm\frac{1}{2};0,0\right\rangle \quad \text{(even)}, \tag{7.1}$$

$$\psi^{\pm}_{\alpha,n}=\left|\frac{-(\alpha+2)}{2(1+\alpha)},\frac{2n\alpha+2n+\alpha+2}{2(1+\alpha)};0,0;\frac{1}{2},\pm\frac{1}{2}\right\rangle \quad \text{(odd)}, \tag{7.2}$$

where $n=0,1,2,\ldots,\infty$. The actions of the generators upon the basis states are obtained by means of the method explained in Sec. V. This gives us the following explicit form of the doublet representation for $D(2,1;\alpha)$ [in (7.4), only the nonvanishing actions are given]:

$$\begin{aligned}s_0\,\varphi^{\pm}_{\alpha,n}&=\frac{2n\alpha+2n+1}{2(1+\alpha)}\varphi^{\pm}_{\alpha,n},\\ s_0\,\psi^{\pm}_{\alpha,n}&=\frac{2n\alpha+2n+\alpha+2}{2(1+\alpha)}\psi^{\pm}_{\alpha,n},\\ s_+\,\varphi^{\pm}_{\alpha,n}&=\left[\frac{(n+1)(\alpha n+n+1)}{(1+\alpha)}\right]^{1/2}\varphi^{\pm}_{\alpha,n+1},\\ s_-\,\varphi^{\pm}_{\alpha,n}&=-\left[\frac{n(\alpha n+n-\alpha)}{(1+\alpha)}\right]^{1/2}\varphi^{\pm}_{\alpha,n-1},\\ s_+\,\psi^{\pm}_{\alpha,n}&=\left[\frac{(n+1)(\alpha n+n+\alpha+2)}{(1+\alpha)}\right]^{1/2}\psi^{\pm}_{\alpha,n+1},\\ s_-\,\psi^{\pm}_{\alpha,n}&=-\left[\frac{n(\alpha n+n+1)}{(1+\alpha)}\right]^{1/2}\psi^{\pm}_{\alpha,n-1},\end{aligned} \tag{7.3}$$

$$\begin{aligned}&t_0\,\varphi^{\pm}_{\alpha,n}=\pm\tfrac{1}{2}\varphi^{\pm}_{\alpha,n},\quad t_\pm\,\varphi^{\mp}_{\alpha,n}=\varphi^{\pm}_{\alpha,n},\\ &u_0\,\psi^{\pm}_{\alpha,n}=\pm\tfrac{1}{2}\psi^{\pm}_{\alpha,n},\quad u_\pm\,\psi^{\mp}_{\alpha,n}=\psi^{\pm}_{\alpha,n},\\ &t_\mu\,\psi^{\pm}_{\alpha,n}=u_\mu\,\varphi^{\pm}_{\alpha,n}=0\quad(\mu=0,\pm);\\ &R_{1/2,\pm 1/2,k}\,\varphi^{\mp}_{\alpha,n}=\pm[2(\alpha n+n+1)]^{1/2}\psi^{(k)}_{\alpha,n},\\ &R_{-1/2,\pm 1/2,k}\,\varphi^{\mp}_{\alpha,n}=\mp[2n(\alpha+1)]^{1/2}\psi^{(k)}_{\alpha,n-1},\\ &R_{1/2,j,\pm 1/2}\,\psi^{\mp}_{\alpha,n}=\pm[2(n+1)(\alpha+1)]^{1/2}\varphi^{(j)}_{\alpha,n+1},\\ &R_{-1/2,j,\pm 1/2}\,\psi^{\mp}_{\alpha,n}=\mp[2(n\alpha+n+1)]^{1/2}\varphi^{(j)}_{\alpha,n},\end{aligned} \tag{7.4}$$

where the indices (j) or (k) denote the sign of j and k, respectively. It is easy to verify the (anti-) commutation relations (2.3)–(2.5) and the star condition [6.3(2)].

The form of the doublet representation suggests that this irrep may be realized in terms of functions of one complex variable with four components, i.e., functions f: $\mathbb{C}\to\mathbb{C}^4$ with components f_i $(i=1,2,3,4)$. We denote f by $[f_1 f_2 f_3 f_4]^t$, where t is the transpose. We found, however, that such a realization is possible only for $\alpha=1$, that is, for the case $D(2,1;1)\equiv\mathrm{osp}(4,2)$. The normalized basis states (7.1) and (7.2) are then realized as follows:

$$\begin{aligned}\varphi^{+}_{1,n}&=e^{i\pi n/2}[(2n)!]^{-1/2}z^{2n}\,[1\,0\,0\,0]^t,\\ \varphi^{-}_{1,n}&=e^{i\pi n/2}[(2n)!]^{-1/2}z^{2n}[0\,1\,0\,0]^t,\\ \psi^{+}_{1,n}&=e^{i\pi(2n+1)/4}[(2n+1)!]^{-1/2}\\ &\quad\times z^{2n+1}[0\,0\,1\,0]^t,\\ \psi^{-}_{1,n}&=e^{i\pi(2n+1)/4}[(2n+1)!]^{-1/2}\\ &\quad\times z^{2n+1}[0\,0\,0\,1]^t.\end{aligned} \tag{7.5}$$

These expressions are elements of $\mathscr{H}(\mathbb{C},\mathbb{C}^4)$, the space of holomorphic functions f:$\mathbb{C}\to\mathbb{C}^4$, which satisfy

$$\int\left(\sum_{i=1}^4|f_i(z)|^2\right)\exp(-|z|^2)d\lambda\,(z)<\infty, \tag{7.6}$$

where λ is the Lebesgue measure on $\mathbb{C}$. The inner product (6.7) for the representation space can be expressed as

$$\begin{aligned}\langle f|g\rangle&=\frac{1}{\pi}\int f(z)^\dagger g(z)\exp(-|z|^2)d\lambda\,(z)\\ &=\frac{1}{\pi}\int\left(\sum_{i=1}^4 f_i(z)^* g_i(z)\right)\\ &\quad\times\exp(-|z|^2)d\lambda\,(z),\end{aligned} \tag{7.7}$$

for $f,g\in\mathscr{H}(\mathbb{C},\mathbb{C}^4)$.

The basis elements of $D(2,1;1)$ are then operators acting in the space $\mathscr{H}(\mathbb{C},\mathbb{C}^4)$. We write their 4×4-matrix expressions in terms of 2×2-block matrices, making use of the Pauli[17] matrices $\sigma_\pm=\frac{1}{2}(\sigma_1\pm i\sigma_2),\sigma_3$ and of the shorthand notation

$$e_{11}=\begin{bmatrix}1&0\\0&0\end{bmatrix},\quad e_{22}=\begin{bmatrix}0&0\\0&1\end{bmatrix}.$$

We find

$$s_0=\frac{1}{2}z\frac{d}{dz}+\frac{1}{4},\quad s_+=\frac{i}{2}z^2,\quad s_-=\frac{i}{2}\frac{d^2}{dz^2},$$

$$t_0=\frac{1}{2}\begin{bmatrix}\sigma_3&0\\0&0\end{bmatrix},\quad t_\pm=\begin{bmatrix}\sigma_\pm&0\\0&0\end{bmatrix},$$

$$u_0=\frac{1}{2}\begin{bmatrix}0 & 0\\ 0 & \sigma_3\end{bmatrix},\quad u_\pm=\begin{bmatrix}0 & 0\\ 0 & \sigma_\pm\end{bmatrix},$$

$$
\begin{aligned}
R_{1/2,1/2,1/2}&=\sqrt{2}\,e^{i\pi/4}z\begin{bmatrix}0 & \sigma_+\\ \sigma_+ & 0\end{bmatrix},\\
R_{1/2,1/2,-1/2}&=\sqrt{2}\,e^{i\pi/4}z\begin{bmatrix}0 & -e_{11}\\ e_{22} & 0\end{bmatrix},\\
R_{1/2,-1/2,1/2}&=-\sqrt{2}\,e^{i\pi/4}z\begin{bmatrix}0 & -e_{22}\\ e_{11} & 0\end{bmatrix},\\
R_{1/2,-1/2,-1/2}&=-\sqrt{2}\,e^{i\pi/4}z\begin{bmatrix}0 & \sigma_-\\ \sigma_- & 0\end{bmatrix},\\
R_{-1/2,-1/2,-1/2}&=-\sqrt{2}\,e^{3i\pi/4}\frac{d}{dz}\begin{bmatrix}0 & \sigma_-\\ \sigma_- & 0\end{bmatrix},\\
R_{-1/2,-1/2,1/2}&=\sqrt{2}\,e^{3i\pi/4}\frac{d}{dz}\begin{bmatrix}0 & e_{22}\\ -e_{11} & 0\end{bmatrix},\\
R_{-1/2,1/2,-1/2}&=-\sqrt{2}\,e^{3i\pi/4}\frac{d}{dz}\begin{bmatrix}0 & e_{11}\\ -e_{22} & 0\end{bmatrix},\\
R_{-1/2,1/2,1/2}&=\sqrt{2}\,e^{3i\pi/4}\frac{d}{dz}\begin{bmatrix}0 & \sigma_+\\ \sigma+ & 0\end{bmatrix}.
\end{aligned}
\tag{7.8}
$$

Similar realizations have been found for osp(1,2) (see Ref. 7) and osp(3,2) (see Ref. 8). For osp(1,2), the so-called metaplectic representation could be realized in terms of elements of $\mathscr{H}(\mathbb{C},\mathbb{C})$, whereas for osp(3,2), its "metaplectic" representation was given in terms of $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$. It is remarkable that only for $\alpha=1$, i.e., for $D(2,1;1)=\mathrm{osp}(4,2)$, can the doublet representation be realized in this way. The sp(2;R) labels for the doublet representation of osp(4,2) in the reduction to su(1,1) + su(2) + su(2)$\simeq$sp(2;R) + so(4) are $-\frac{1}{4}$ and $-\frac{3}{4}$ [see (7.1) and (7.2)], just as was the case for osp(1,2) and osp(3,2).

VIII. CONCLUSIONS

Finite- and infinite-dimensional irreducible representations of the exceptional Lie superalgebras $D(2,1;\alpha)$ have been classified. The star and grade star conditions for the algebra and for the representations have been investigated in detail. For the finite-dimensional case, only grade star representations are possible. It appears that only a few finite-dimensional irreps turn out to be grade star representations, a result similar to that for $B(1,1)$ (see Refs. 8 and 19) and $B(0,2)$ (see Ref. 19). Among the infinite-dimensional irreps we find a large class of star representations. They are in fact $D(2,1;\alpha)$ irreps in an sp(2;R) + so(4) basis. One of them, the doublet representation, has been constructed explicitly in Sec. VII.

For a finite-dimensional typical representation, it is easy to verify from (4.2) that the dimension is given by

$$\dim(p;q;r)=16(2p-1)(2q+1)(2r+1). \tag{8.1}$$

This expression was also given by Kac,[5] and shows that the dimensions of the typical representations are the same for all $D(2,1;\alpha)$. Since we have given the reduction rule also for all atypical and truncated representations, it is easy to determine their dimensions. In particular, we have shown that the dimensions of some irreps of $D(2,1;\alpha)$ are dependent on α. For instance, the two lowest-dimensional nontrivial irreps of $D(2,1;1)$ have dimensions 6 and 17, whereas for $D(2,1;2)$ they have dimensions 10 and 17. This property shows more than anything else that two $D(2,1;\alpha)$ algebras with different α values are not equivalent.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. J. W. B. Hughes (Mathematics Department, Queen Mary College, London) for helpful discussions and for reading the manuscript.

The British Council is acknowledged for a maintenance grant. This research project was partially supported by Science and Engineering Research Council Grant No. GR/C/33147.

APPENDIX: SHIFT OPERATOR PRODUCTS

A shift operator product $A^{i'j'k'}A^{ijk}$ is called of type $(i'+i, j'+j, k'+k)$. We find the simplest relations when $|i'+i|=|j'+j|=|k'+k|=1$

$$A^{ijk}A^{ijk}=0. \tag{A1}$$

Because of the symmetry of the three su(2) parts in $\Gamma(\sigma_1,\sigma_2,\sigma_3)$, and since there exist several transformation rules for shift operator product relations,[12,15,16] it is enough to give the relations of type (1,1,0), (1,0,0), and (0,0,0):

(1) (1,1,0)

$$A^{i,j,-1/2}A^{i,j,1/2}\hat{u}+A^{i,j,1/2}A^{i,j,-1/2}(\hat{u}+1)=0; \tag{A2}$$

(2) (1,0,0)

$$A^{i,1/2,1/2}A^{i,-1/2,-1/2}(\hat{t}+\hat{u}+1)+A^{i,1/2,-1/2}A^{i,-1/2,1/2}(2\hat{t}+1)\hat{u}+A^{i,-1/2,1/2}A^{i,1/2,-1/2}\hat{t}(2\hat{u}+1)=0; \tag{A3}$$

(3) (0,0,0)

$$
\begin{aligned}
&A^{1/2,1/2,1/2}A^{-1/2,-1/2,-1/2}+A^{-1/2,-1/2,-1/2}A^{1/2,1/2,1/2}\\
&\quad-\tfrac{1}{8}I_4+2I_2\,[S^2+T^2+U^2-\hat{s}\hat{t}-\hat{s}\hat{u}-\hat{t}\hat{u}-\hat{s}-\hat{t}-\hat{u}-1]\\
&\quad+2\{\sigma_1S^4+\sigma_2T^4+\sigma_3U^4+2\sigma_1S^2(3\hat{t}\hat{u}-\hat{s}\hat{t}-\hat{s}\hat{u}-\hat{s}+\hat{t}+\hat{u})\\
&\quad+2\sigma_2T^2(3\hat{s}\hat{u}-\hat{s}\hat{t}-\hat{t}\hat{u}-\hat{t}+\hat{s}+\hat{u})+2\sigma_3U^2(3\hat{s}\hat{t}-\hat{t}\hat{u}-\hat{s}\hat{u}-\hat{u}+\hat{s}+\hat{t})\}=0;
\end{aligned}
\tag{A4}
$$

$$
\begin{aligned}
&A^{1/2,-1/2,1/2}A^{-1/2,1/2,-1/2}+A^{-1/2,1/2,-1/2}A^{1/2,-1/2,1/2}-\tfrac{1}{8}I_4+2I_2\,[S^2+T^2+U^2+\hat{s}\hat{t}+\hat{t}\hat{u}-\hat{s}\hat{u}+\hat{t}\,]\\
&\quad+2\{\sigma_1S^4+\sigma_2T^4+\sigma_3U^4+2\sigma_1S^2(-3\hat{t}\hat{u}+\hat{s}\hat{t}-\hat{s}\hat{u}-2\hat{u}-\hat{t}-1)\\
&\quad+2\sigma_2T^2(3\hat{s}\hat{u}+\hat{s}\hat{t}+\hat{t}\hat{u}+2\hat{s}+2\hat{u}+\hat{t}+1)+2\sigma_3U^2(-3\hat{s}\hat{t}+\hat{t}\hat{u}-\hat{s}\hat{u}-2\hat{s}-\hat{t}-1)\}=0;
\end{aligned}
\tag{A5}
$$

$$\begin{aligned}
&A^{-1/2,1/2,-1/2}A^{1/2,-1/2,1/2}(2\hat{u}+1)(\hat{s}-\hat{t})-A^{-1/2,-1/2,-1/2}A^{1/2,1/2,1/2}(2\hat{s}+1)(\hat{u}+\hat{t}+1)+A^{-1/2,-1/2,1/2}A^{1/2,1/2,-1/2}\\
&\quad\times(2\hat{t}+1)(\hat{s}-\hat{u})-(\hat{s}+1)(2\hat{t}+1)(2\hat{u}+1)\{-\tfrac{1}{8}I_4+2I_2(S^2+T^2+U^2+\hat{u}\hat{t}-\hat{s}\hat{t}-\hat{s}\hat{u}-2\hat{s}-1)\\
&\quad+2\,[\sigma_1S^4+\sigma_2T^4+\sigma_3U^4+2\sigma_1\hat{s}(\hat{s}-1)(\hat{u}\hat{t}-\hat{s}\hat{u}-\hat{s}\hat{t}-2\hat{s}-1)-2\sigma_1S^2\\
&\quad+2\sigma_2T^2(\hat{t}\hat{u}-\hat{s}\hat{u}-\hat{s}\hat{t})+2\sigma_3U^2(\hat{t}\hat{u}-\hat{s}\hat{u}-\hat{s}\hat{t})]\}=0.
\end{aligned}\tag{A6}$$

[1]V. G. Kac, Adv. Math. **26**, 8 (1977).
[2]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626 (1976).
[3]M. Scheunert, *Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979).
[4]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[5]V. G. Kac, *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), pp. 597–626.
[6]R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Gen. **16**, 473 (1983).
[7]J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).
[8]J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).
[9]A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1810 (1981).
[10]J. P. Hurni and B. Morel, J. Math. Phys. **24**, 157 (1983).
[11]B. S. DeWitt and P. Van Nieuwenhuizen, J. Math. Phys. **23**, 1953 (1982).
[12]J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).
[13]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[14]V. G. Kac, Commun. Math. Phys. **53**, 31 (1977).
[15]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, Ann. Phys. **147**, 85 (1983).
[16]G. Vanden Berghe, H. De Meyer, and J. Van der Jeugt, Lett. Nuovo Cimento **33**, 120 (1982).
[17]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1957).
[18]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974), pp. 182–188.
[19]Q. Han and H. Sun, Commun. Theor. Phys. **2** (3), 1137 (1983).

On U(1)-gauge fields and their geometrical interpretation

J. Beckers and V. Hussin[a)]

Université de Liège, Physique Théorique et Mathématique, Institut de Physique au Sart Tilman, Bâtiment B.5, B-4000 Liège 1, Belgium

(Received 17 July 1984; accepted for publication 28 September 1984)

The Harnad–Shnider–Vinet study of symmetry properties on gauge fields in terms of invariant connections on principal fiber bundles is reviewed in the simple U(1)-gauge theory. It is extended to the case of invariant electromagnetic fields admitting nontrivial extensions of their symmetry groups. Some specific examples are discussed.

I. INTRODUCTION

During the last decade, differential geometry[1] became a very useful language in theoretical physics and more particularly in gauge theories. Classical electromagnetism and spinor electrodynamics [a theory subtended by the gauge group U(1)] have been formulated in terms of one- and two-forms, fiber bundles, associated connections, etc. A nice report on these tools has been given by Drechsler and Mayer.[2]

More recently, symmetries in arbitrary gauge theories have been developed using basic notions of differential geometry by Bergmann–Flaherty[3] and Forgacs–Manton.[4] Some interesting contributions have also been obtained, on the one hand, on symmetries and conservation laws in gauge theories by Jackiw–Manton[5] and, on the other hand, on invariance conditions for gauge fields under group actions by Harnad, Shnider, and Vinet.[6,7] All these studies mix space-time and gauge symmetries or transformations. They are developed in principle for arbitrary gauge groups and have to give, in particular, all the well-known results when the gauge group U(1) is under consideration, i.e., when we are dealing with classical interactions through electromagnetic fields. So, a simple question arises and is the starting point of the contents of this paper: Can we recover all the well-known results of the U(1)-gauge theory from the above geometrical approaches dealing with arbitrary gauge groups? If, happily, the answer is essentially positive, we want to emphasize some nontrivial points dealing with invariant electromagnetic fields[8] or invariant potentials,[9] through compensating gauges,[10] group extensions,[11] and associated factor sets[12] (or exponents).

In the U(1)-gauge theory, the electromagnetic potentials are the gauge fields and their symmetries are determined through the coupling of space-time and gauge transformations. For example, Janner–Janssen[10] have studied compensating gauge transformations in connection with an extension by $\mathbf{R}$ of the symmetry group associated with a given electromagnetic field. From gauge theories based on arbitrary gauge groups as discussed by Forgacs–Manton,[4] there are specific assumptions on the compensating gauges which can only be satisfied in the U(1)-case if the symmetry group of the electromagnetic field admits no nontrivial extensions by $\mathbf{R}$. This is not the more general context as we shall see in the following.

Moreover, let us point out the Harnad–Shnider–Vinet approach[7] dealing with the geometrical interpretation of symmetry properties on gauge fields in terms of invariant connections on principal fiber bundles. Here also, the interpretation is realized within the Forgacs–Manton hypotheses so that, in the U(1) context, it is not complete when the symmetry groups admit nontrivial extensions by $\mathbf{R}$.

Let us finally notice that analogous difficulties have already been pointed out and partially solved by Henneaux[13] and Duval–Horvathy.[14] Here we plan to give a more general answer based on the Harnad–Shnider–Vinet developments.[7]

In Sec. II, we just recall some elements issued by the works of Janner–Janssen[10] and Forgacs–Manton[4], restricted to the U(1)-gauge theory in order to pick out symmetry properties on potentials associated with a given invariant electromagnetic field. Section III is devoted to the Harnad–Shnider–Vinet approach,[7] using fiber bundle techniques with U(1) as the structure group but with the extension by $\mathbf{R}$ of the symmetry group (of an electromagnetic field) as the group of automorphisms on the bundles. In this way, we classify the fiber bundles and the invariant connection forms leading to the interpretation of invariance conditions on potentials and fields. Section IV contains some examples and comments in order to apply our developments. The procedure is summarized and applied to two specific examples, the first one for constant and uniform electromagnetic fields (IV A) and the second one for arbitrary electromagnetic fields (IV B). Some comments are added in connection with physical approaches leading to complete sets of constants of motion associated with charged particles moving in external electromagnetic fields.[15] Such a study and all the results lead to a deeper insight into the geometrical aspects of U(1)-gauge field symmetries.

II. INVARIANT ELECTROMAGNETIC FIELDS AND POTENTIALS

Let $f{:}G \times M \to M$ be a differentiable left action of a connected Lie group G on a smooth manifold M. This action induces a homomorphism $\xi \to X^{\xi}$ from the Lie algebra $\mathbf{G}$ of G into that of vector fields on M. Since G is connected, the G invariance is equivalent to infinitesimal invariance.

Firstly, let us consider a G-invariant closed two-form F on M with values in $\mathbf{R}$, i.e.,

$$L_{X^{\xi}}F = 0, \quad \forall \xi \in \mathbf{G}, \tag{2.1}$$

where L denotes the Lie derivative. In addition, if U is an arbitrary contractible open subset of M, we have $F|_U = dA$ for some one-form A on U. From Eq. (2.1), we deduce that $dL_{X^{\xi}}A = 0$ and then we may choose a linear mapping $W{:}\mathbf{G} \to C_{\infty}(U,\mathbf{R})$ such that

[a)] Chercheur Institut Interuniversitaire des Sciences Nucléaires.

 0022-2488/85/050925-04$02.50

$$L_{X^\xi}A = dW_\xi, \quad \forall \xi \in \mathbb{G}. \tag{2.2}$$

As easily seen, the differential of

$$c(\xi,\xi') = L_{X^\xi}W_{\xi'} - L_{X^{\xi'}}W_\xi - W_{[\xi,\xi']} \tag{2.3}$$

vanishes on U so that c is in fact a skew-symmetric mapping from $\mathbb{G}\times\mathbb{G}$ into $\mathbb{R}$. It is actually a two-cocycle of $\mathbb{G}$ for the trivial representation on $\mathbb{R}$ (see Ref. 16). Indeed, we have

$$c([\xi,\xi'],\xi'') + c([\xi',\xi''],\xi) + c([\xi'',\xi],\xi') = 0. \tag{2.4}$$

Of course, the above A and W are not unique and other choices would modify c, at most, by adding a coboundary. Thus, what we really get in this way is a cohomology class $c(F) = [c] \in H^2(\mathbb{G},\mathbb{R},0)$ of the second cohomology space of the trivial representation of $\mathbb{G}$ on $\mathbb{R}$, which, clearly, does not depend on U and hence is globally attached to F.

If $c(F) = 0$, then c is the coboundary of some linear mapping $c':\mathbb{G}\to\mathbb{R}$ and $W' = W + c'$ satisfies the equation

$$L_{X^\xi}W'_{\xi'} - L_{X^{\xi'}}W'_\xi - W'_{[\xi,\xi']} = 0. \tag{2.5}$$

This equation has been completely solved[17] when M is an orbit G/G_0 of G. In this case, the de Rahm cohomology class of F is a characteristic class of the principal G_0 bundle $G(M,G_0)$.

For later use, we need another form of Eqs. (2.2) and (2.3). Set $\psi:\mathbb{G}\to C_\infty(U,\mathbb{R})$ defined by

$$\psi_\xi = W_\xi - i(X^\xi)A, \quad \forall \xi \in \mathbb{G}, \tag{2.6}$$

where i denotes the interior product. Then, as easily seen, Eqs. (2.2) and (2.3) read, respectively,

$$i(X^\xi)F = d\psi_\xi \tag{2.7}$$

and

$$F(X^\xi,X^{\xi'}) = \psi_{[\xi,\xi']} - c(\xi,\xi'). \tag{2.8}$$

Secondly, we turn to apply the previous discussion to invariant electromagnetic potentials and fields. The manifold M is now some open region of the Minkowski space-time $\mathbb{R}^4$ and F represents an electromagnetic field in M. We denote by G the group of geometrical symmetries of F. In this context, A represents an electromagnetic potential associated with F, and the mapping W, as a function from $\mathbb{G}\times U$ into $\mathbb{R}$, is the usual compensating gauge transformation.[10] Then, the G invariance produces some class $c(F)$ and both $c(F)$ and W are ruled by Eq. (2.3).

However, if we apply the Forgacs–Manton considerations[4] to the case of the U(1)-gauge theory, we recover Eq. (2.5) rather than Eq. (2.3). In fact, their considerations are valid under some specific assumptions on the compensating gauge transformations which imply the vanishing of our class $c(F)$. In addition, Eq. (2.5) is a necessary condition for the existence[7] of an interpretation of the symmetry properties for the field F and its potentials in terms of invariant connection one-forms on a U(1)-principal bundle.

Our main goal in this paper is to furnish a method allowing us to work out the general case for which $c(F)\neq 0$. Roughly speaking, it consists in killing $c(F)$ by enlarging the symmetry group G of F, thus reducing the problem to the previous one. In fact, $H^2(\mathbb{G},\mathbb{R},0)$ classifies the central extensions of $\mathbb{G}$ by $\mathbb{R}$ up to equivalence,[11,16] and the new symmetry group is essentially an extension of G associated with the nontrivial extension of $\mathbb{G}$ corresponding to $c(F)$. This kind of approach has been initiated in some particular cases (see, for example, the works of Henneaux[13] and Duval–Horvathy[14]).

III. PRINCIPAL BUNDLES AND INVARIANT POTENTIALS

As known,[2] the field F and the potentials A are presented in terms of a connection one-form ω on a principal bundle P over M with structure group U(1), the gauge group of electrodynamics. One has $F = i\Omega$, where Ω is the curvature form viewed as a two-form on M [note that U(1) is abelian with the algebra $i\mathbb{R}$]; if $\pi:U\to P$ is a section, then $A = i\pi^*\omega$ is a potential for F over U.

For a G-invariant F, the problem is to lift on P the action of G on M in such a way that ω becomes a G-invariant form, reflecting thus the symmetries of F at the level of its potentials. This problem has been solved[7] when $c(F) = 0$, and then when G admits no nontrivial extension by R. U(1) bundles over M admitting a lift of the action of G on M have been classified by Harnad, Shnider, and Vinet.[7] In the transitive case, i.e., when M is a homogeneous space G/G_0, these bundles are the U(1) bundles $G^\lambda(M,\mathrm{U}(1))$ associated with the principal G_0 bundle $G(M,G_0)$ and with the homomorphisms $\lambda:G_0\to\mathrm{U}(1)$. On the other hand, the G-invariant connection one-forms on the bundle $G^\lambda(M,\mathrm{U}(1))$ are classified using the well-known Wang theorem.[1]

We now turn to the study of a G-invariant F, for which $c(F)\neq 0$. As quoted at the end of Sec. II, there is no direct interpretation of F in terms of a G-invariant connection on some U(1) bundle over M.

Let $\bar{\mathbb{G}}$ denote the central extension of $\mathbb{G}$ by $\mathbb{R}$ corresponding to $c(F)$. As a set, $\bar{\mathbb{G}}$ is the product $\mathbb{G}\times\mathbb{R}$ while its Lie algebra structure is given by

$$[\bar\xi,\bar\xi'] = ([\xi,\xi'],c(\xi,\xi')). \tag{3.1}$$

[From now on, we will denote by $\bar\xi = (\xi,\eta)$ a typical element of $\bar{\mathbb{G}}$.] If we denote by $\bar G$ the simply connected Lie group with the Lie algebra $\bar{\mathbb{G}}$, there is a unique homomorphism $\phi:\bar G\to G$ whose differential at e coincides with the projection $\bar{\mathbb{G}}\to\mathbb{G}$ onto the first factor. Consequently, $\bar G$ operates on M through the action of G by

$$\bar f_{\bar g} = f_{\phi(\bar g)}. \tag{3.2}$$

In particular, we have $X^{\bar\xi} = X^\xi$, $\forall\bar\xi = (\xi,\eta)\in\bar{\mathbb{G}}$. This definition (3.2) thus implies that F is $\bar G$ invariant.

Now, if $\bar W_{\bar\xi} = W_\xi + \eta$, then $\bar W$ is a compensating gauge for A, i.e.,

$$L_{X^{\bar\xi}}A = d\bar W_{\bar\xi}, \tag{3.3}$$

and it satisfies Eq. (2.5). It follows that the class $\bar c(F)$ vanishes and hence that we may apply the results of Harnad–Shnider–Vinet[7] and Wang,[1] provided we replace G by its extension $\bar G$ associated with $c(F)$.

Let us finally notice that, in such a context, the Wang theorem becomes the following. There is a one-to-one correspondence between the set of $\bar G$-invariant connections on $\bar G^{\bar\lambda}$, where $\bar\lambda:\bar G_0\to\mathrm{U}(1)$ is a homomorphism, and the set of linear mappings $\Lambda:\bar{\mathbb{G}}\to i\mathbb{R}$ such that

$$\Lambda|_{\bar{\mathbb{G}}_0} = \bar\lambda_*, \quad \Lambda|_{[\bar{\mathbb{G}}_0,\bar{\mathbb{G}}]} = 0. \tag{3.4}$$

Using the definition of A in terms of ω, we also have

$$A(\bar{\xi}) = A((\xi,\eta)) = i(W(\xi,x_0) - A(X^{\xi}_{x_0}) + \eta) \quad (3.5a)$$

or

$$A(\bar{\xi}) = i(\bar{W}_{\bar{\xi}}(x_0) - A(X^{\bar{\xi}}_{x_0})). \quad (3.5b)$$

Equations (3.5) may then be rewritten in the form

$$A((\xi,\eta)) = i(\Psi(\xi,x_0) + \eta), \quad (3.6)$$

showing clearly the connection between A and Ψ defined by Eq. (2.6). Let us end this section by noticing that, using the bracket (3.1), Eq. (2.8) becomes

$$A([\bar{\xi},\bar{\xi}']) = -iF(X^{\xi}_{x_0},X^{\xi'}_{x_0}). \quad (3.7)$$

IV. EXAMPLES AND COMMENTS

Through some specific examples, let us illustrate the geometrical interpretation of the preceding symmetry properties. Let us at once mention the example studied by Lecomte[17] of an induction produced by a magnetic monopole. The fiber bundle interpretation of the associated symmetries is realized with the quantization condition of the magnetic charge. Such an example does work with a symmetry group admitting only trivial extensions by $\mathbb{R}$, and the Harnad–Shnider–Vinet method can then be directly applied.

Here we are more particulary interested in examples where the symmetry groups of electromagnetic fields admit nontrivial extensions by $\mathbb{R}$. Then, let us distinguish a specific example in the case of constant and uniform fields[8] (Sec. IV A) and another one in the case of arbitrary fields (Sec. IV B), i.e., a magnetic induction produced by an unlimited line wire carrying steady currents.

Let us summarize the procedure leading to the expected interpretation in both cases.

For a given field F and an associated potential A, we determine their symmetry properties by using the elements of Sec. II. These results permit us to characterize the algebra $\bar{\mathbb{G}}_F$, i.e., an extension by $\mathbb{R}$ of the algebra $\mathbb{G}_F$ associated with the symmetries of the field F.

To the algebra $\bar{\mathbb{G}}_F$, there corresponds a simply connected Lie group denoted $\bar{G}_F$ and we search for the stabilizer $\bar{G}_0$ in $\bar{G}_F$ of a given point in M and for the corresponding algebra $\bar{\mathbb{G}}_0$.

We then determine the homomorphisms $\bar{\lambda}_*:\bar{\mathbb{G}}_0\to i\mathbb{R}$ and the applications $A:\bar{\mathbb{G}}_F\to i\mathbb{R}$ satisfying the relations (3.4), (3.5), and (3.7).

Finally, we get the homomorphisms $\bar{\lambda}:\bar{G}_0\to \mathrm{U}(1)$ leading to the construction of the bundles $\bar{G}^{\bar{\lambda}}(M,\mathrm{U}(1))$.

A. Constant and uniform electromagnetic fields

We are working on $M\equiv\mathbb{R}^4$ with the Minkowski metric. Let us recall that, from the Bacry–Combe–Richard results,[8] there essentially exist two nonequivalent constant and uniform fields, the parallel ($F_\parallel$) and orthogonal ($F_\perp$) fields. Here let us only discuss the first one. The study of the $F_\perp$ field can be realized in an analogous way and we leave this study as an exercise for the interested reader.

The $F_\parallel$ field is chosen as

$$F_\parallel \equiv (\mathbf{E},\mathbf{B}), \quad \mathbf{E} = (0,0,E), \quad \mathbf{B} = (0,0,B), \quad (4.1)$$

which corresponds to the two-form

$$F_\parallel = E\, dz\wedge dt + B\, dx\wedge dy. \quad (4.2)$$

The symmetry group $G_\parallel$ of this field[8] is a subgroup of the Poincaré group whose elements are associated with spatial rotations around the z axis, pure Lorentz transformations along the z axis, and space-time translations. The corresponding six-dimensional Lie algebra is

$$\mathbb{G}_\parallel = \{J^3,K^3,P^\mu(\mu=0,1,2,3)\}, \quad (4.3)$$

with the only nonzero commutators

$$[J^3,P^1] = -P^2, \quad [J^3,P^2] = P^1,$$
$$[K^3,P^0] = P^3, \quad [K^3,P^3] = P^0. \quad (4.4)$$

Now, if we choose the gauge-symmetrical potential $A_\parallel$ associated with $F_\parallel$ and corresponding to the one-form

$$A_\parallel = \tfrac{1}{2}E(z\,dt - t\,dz) + \tfrac{1}{2}B(x\,dy - y\,dx), \quad (4.5)$$

the compensating gauge transformation W can easily be obtained by using Eq. (2.2) and the following realization for the generators of $\mathbb{G}_\parallel$

$$J^3 = xD_y - yD_x, \quad K^3 = tD_z + zD_t,$$
$$P^0 = D_t, \quad P^1 = -D_x, \quad P^2 = -D_y, \quad P^3 = -D_z. \quad (4.6)$$

In fact, we have, up to an additive constant,

$$W(J^3,x) = W(K^3,x) = 0, \quad W(P^0,x) = -\tfrac{1}{2}Ez,$$
$$W(P^1,x) = -\tfrac{1}{2}By, \quad W(P^2,x) = \tfrac{1}{2}Bx, \quad W(P^3,x) = -\tfrac{1}{2}Et. \quad (4.7)$$

So that, from Eq. (2.3), the two-cocyle c takes the only nonzero values

$$c(P^0,P^3) = -E, \quad c(P^1,P^2) = -B. \quad (4.8)$$

Then, the extension $\bar{\mathbb{G}}_\parallel$ is generated by the new generators $\mathscr{J}^3$, κ^3, π^μ, and I associated with those of $\bar{\mathbb{G}}_\parallel$ and R, respectively. The only nonzero commutation relations characterizing $\bar{\mathbb{G}}_\parallel$ are

$$[\mathscr{J}^3,\pi^1] = -\pi^2, \quad [\mathscr{J}^3,\pi^2] = \pi^1,$$
$$[\kappa^3,\pi^0] = \pi^3, \quad [\kappa^3,\pi^3] = \pi^0, \quad (4.9)$$
$$[\pi^0,\pi^3] = \alpha EI, \quad [\pi^1,\pi^2] = \alpha BI,$$

where α is an arbitrary real parameter.

With regard to the second step, if x_0 is the origin of coordinates in M, the stabilizer of x_0 ($\bar{G}_{0,\parallel}$) in $\bar{G}_\parallel$ is the direct product of $G_{0,\parallel}$ by $\mathbb{R}$, where $G_{0,\parallel}$ is the homogeneous Lorentz subgroup of $G_\parallel$. The algebra $\bar{\mathbb{G}}_{0,\parallel}$ is then abelian and generated by $\mathscr{J}^3$, κ^3, and I.

In the third step, we search for the homomorphisms $\bar{\lambda}_*:\bar{\mathbb{G}}_{0,\parallel}\to i\mathbb{R}$, defined by

$$\bar{\lambda}_*(\mathscr{J}^3) = i\nu, \quad \bar{\lambda}_*(\kappa^3) = i\rho, \quad \bar{\lambda}_*(I) = i\sigma, \quad (4.10)$$

with ν, ρ, σ, $\in$, $\mathbb{R}$. Now the applications $A:\bar{\mathbb{G}}_\parallel\to i\mathbb{R}$ can be determined. Indeed, Eq. (3.4) implies that

$$A(\mathscr{J}^3) = i\nu, \quad A(\kappa^3) = i\rho, \quad A(I) = i\sigma, \quad (4.11)$$

and

$$A(\pi^\mu) = 0, \quad \forall\mu. \quad (4.12)$$

Moreover, having to satisfy Eqs. (3.5) and (3.7), it is easy to show that the parameters ν and ρ in Eqs. (4.10) and (4.11) vanish and, in the commutators (4.9), we have the condition

$$\alpha = 1/\sigma. \quad (4.13)$$

Finally, the homomorphism $\bar{\lambda}:\bar{G}_{0,\parallel}\rightarrow U(1)$ is defined by

$$\bar{\lambda}\,(e^{\kappa I+\nu\kappa^3-\theta\mathscr{J}^3})=e^{i\kappa\sigma},\quad \kappa\in\mathbb{R}. \tag{4.14}$$

Consequently, the symmetry properties of the $F_\parallel$ field and of the associated potentials are interpreted in terms of those of $\bar{G}_\parallel$-invariant connection one-forms $\bar{\omega}$ on the principal fiber bundle $\bar{G}^{\bar{\lambda}}(M,U(1))$. The extension $\bar{\mathbb{G}}_\parallel$ is characterized by the commutators (4.9) with the α value (4.13). Such connection one-forms are obtained from the application $\Lambda:\bar{\mathbb{G}}_\parallel\rightarrow i\mathbb{R}$, defined by

$$\Lambda\,(\mathscr{J}^3)=\Lambda\,(\kappa^3)=\Lambda\,(\pi^\mu)=0$$

and

$$\Lambda\,(I)=i\sigma,\quad \sigma\in\mathbb{R}. \tag{4.15}$$

Let us notice that, in fact, the bundle $\bar{G}^{\bar{\lambda}}(M,U(1))$ is trivial.

B. Nonconstant fields

Let us consider the example of a magnetic field $\mathbf{B}$ produced by an unlimited line wire carrying a steady current strength j. This field can be written

$$\mathbf{B}=(j\mu_0/2\pi r^2)(-y,x,0), \tag{4.16}$$

with $r=(x^2+y^2)^{1/2}$ when the z axis of the reference frame coincides with the line wire. The field $\mathbf{B}$ is then defined for the points of $\mathbb{R}^3$ not located on the z axis so that we are working on $M=\mathbb{R}^3\setminus\{(0,0,z):z\in\mathbb{R}\}$. Such a field corresponds to the two-form

$$F=(j\mu_0/2\pi r^2)(x\,dz\wedge dx-y\,dy\wedge dz). \tag{4.17}$$

In cylindrical coordinates (r,ϕ,z) it takes the form

$$F=(j\mu_0/2\pi r)\,dz\wedge dr. \tag{4.18}$$

It is then easy to show that this closed form is invariant under a group G whose elements are associated with spatial rotations around the z axis, translations along the z axis, and dilatations in the xy plane. Indeed, if the generators of $\mathbb{G}$ are realized by

$$J^3=xD_y-yD_x=D_\phi,\quad P^3=-D_z,$$
$$D=xD_x+yD_y=rD_r, \tag{4.19}$$

we have

$$L_XF=0,\quad \forall X\in\{J^3,P^3,D\}. \tag{4.20}$$

Now with a potential one-form

$$A=(j\mu_0/2\pi)\ln r\,dz, \tag{4.21}$$

we get the associated compensating gauge transformation

$$W(J^3,x)=W(P^3,x)=0,\quad W(D,x)=-(j\mu_0/2\pi)z. \tag{4.22}$$

This leads through Eq. (2.3) to the cocycle c

$$c(J^3,P^3)=c(J^3,D)=0,\quad c(P^3,D)=(j\mu_0/2\pi), \tag{4.23}$$

and we get the algebra $\bar{\mathbb{G}}=\{\mathscr{J}^3,\pi^3,\mathscr{D},I\}$. It is characterized by the only nonzero commutator

$$[\pi^3,\mathscr{D}]=\alpha(j\mu_0/2\pi)I, \tag{4.24}$$

where α is an arbitrary real parameter.

In this context, the stabilizer G_0 in G of some point in M is necessarily reduced to the unity of G. The group $\bar{G}_0$ is then identified to $\mathbf{R}$ and the algebra $\bar{\mathbb{G}}_0$ is generated by I.

Now the homomorphism $\bar{\lambda}_*:\bar{\mathbb{G}}_0\rightarrow i\mathbb{R}$ is defined by

$$\bar{\lambda}_*(I)=i\sigma,\quad \sigma\in\mathbb{R}, \tag{4.25}$$

and the application $\Lambda:\bar{\mathbb{G}}\rightarrow i\mathbb{R}$ satisfying Eqs. (3.4), (3.5), and (3.7) is defined, for example, by

$$\Lambda\,(\mathscr{J}^3)=\Lambda\,(\pi^3)=\Lambda\,(\mathscr{D})=0$$

and

$$\Lambda\,(I)=i\sigma, \tag{4.26}$$

if the point $x_0\in M$ is chosen as $x_0=(1,0,0)$.

In conclusion, the homomorphism $\bar{\lambda}:\bar{G}_0\rightarrow U(1)$ is defined by

$$\bar{\lambda}\,(e^{\kappa I})=e^{i\kappa\sigma},\quad \kappa\in\mathbb{R}, \tag{4.27}$$

and the interpretation in terms of $\bar{G}$-invariant connection one-forms on $\bar{G}^{\bar{\lambda}}(M,U(1))$ is again obtained. The extension $\bar{\mathbb{G}}$ is characterized by the only nonzero commutator

$$[\pi^3,\mathscr{D}]=(j\mu_0/2\pi\sigma)I. \tag{4.28}$$

Let us insist on the fact that the homomorphism $\bar{\lambda}$ can be extended to a smooth function $\delta:\bar{G}\rightarrow U(1)$ defined by

$$\delta(e^{\kappa I+\theta\mathscr{J}^3+a\pi^3+\rho\mathscr{D}})=e^{\kappa\sigma}, \tag{4.29}$$

such that

$$\delta(\bar{g}\bar{g}_0)=\delta(\bar{g})\bar{\lambda}\,(\bar{g}_0),\quad \forall\bar{g}\in\bar{G},\quad \bar{g}_0\in\bar{G}_0.$$

Then from Corollary 1 in the Harnad–Shnider–Vinet approach,[7] the bundle $\bar{G}^{\bar{\lambda}}$ is also trivial over M.

Let us end this section by noticing that these geometrical developments have an immediate physical meaning in connection with the determination of complete sets of constants of motion associated with the description of charged particles moving in external electromagnetic fields.[15] In fact, if we limit ourselves to the example of Sec. (IV A), it has been shown that the generators of the extended algebra $\bar{\mathbb{G}}_\parallel$ associated with the electromagnetic field $F_\parallel$ give rise to the constants of motion deduced from the Hamiltonian or Lagrangian formalisms.[15] These physical results and their present geometrical interpretation lead to a deep understanding of different aspects of the U(1)-gauge fields and more generally of the U(1)-gauge theory.

ACKNOWLEDGMENT

We want to express our hearty thanks to P. Lecomte for very helpful discussions and for his kind interest.

[1] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1.

[2] W. Drechsler and M. E. Mayer, "Fiber Bundle Techniques in Gauge Theories," in *Lecture Notes in Physics*, Vol. 67 (Springer–Verlag, New York, 1977).

[3] P. G. Bergmann and E. J. Flaherty Jr., J. Math. Phys. **19**, 212 (1978).

[4] P. Forgacs and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).

[5] R. Jackiw and N. S. Manton, Ann. Phys. **127**, 257 (1980).

[6] J. Harnad and L. Vinet, Phys. Lett. B **76**, 589 (1978); J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931 (1979).

[7] J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980).

[8] H. Bacry, Ph. Combe, and J. L. Richard, Nuovo Cimento A **67**, 267 (1970).

[9] J. Beckers and V. Hussin, J. Math. Phys. **24**, 1286 (1983); **24**, 1295 (1983).

[10] A. Janner and T. Janssen, Physica **53**, 1 (1971).

[11] V. Bargmann, Ann. Math. **59**, 1 (1954); M. Hammermesh, *Group Theory* (Addison–Wesley, Reading, MA, 1962), Sec. 12.7.

[12] J. M. Levy-Leblond, *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. 2, p. 221.

[13] M. Henneaux, J. Math. Phys. **23**, 830 (1982).

[14] C. Duval and P. Horvathy, Ann. Phys. **142**, 10 (1982).

[15] J. Beckers and V. Hussin, Phys. Rev. D **29**, 2814 (1984).

[16] L. Michel *Applications of Mathematics to Problems in Theoretical Physics* (Gordon and Breach, New York, 1967).

[17] P. Lecomte, C. R. Acad. Sci. Paris Ser. A **291**, 135 (1980).

Linearization of nonlinear differential equations by means of Cauchy's integral

Helmut Kröger
Département de Physique, Université Laval, Québec, P.Q., Canada G1K 7P4

(Received 2 August 1984; accepted for publication 7 December 1984)

It is suggested to transform a class of nonlinear differential equations with a holomorphic type of nonlinearity into linear integrodifferential equations. The method is presented in detail for first-order ordinary differential equations. The transformed equation is studied and is found to have a unique solution with an analytical representation. In a numerical test calculation rapid convergence of an approximate solution of the linearized equation towards the reference solution is found. The method can be applied to higher-order ordinary and partial differential equations. The transformation can be generalized also to operator-valued differential equations.

I. INTRODUCTION

In recent years there has been growing interest in nonlinear phenomena in physics. These are mostly described by nonlinear partial differential equations (PDE's). We want to mention as an example Liouville's equation, which has significant applications in electrostatics,[1] hydrodynamics,[2–4] and cosmology.[5] Recently it has also been studied in particle physics in connection with monopole theories.[6] Another example is the sine–Gordon equation, which plays a role in differential geometry,[7] nonlinear optics,[8] plasma physics,[9] superconductivity,[10] and particle physics.[11] Finally let us mention the dynamical equations in quantum chromodynamics (QCD).

The importance of these equations stresses the need for solution methods. There are two principal methods to solve nonlinear PDE's in 1 + 1 (space + time) dimensions, namely the Bäcklund transformation and the inverse scattering method.[12] The latter has been developed only in one space dimension, while the Bäcklund transformation is known for certain PDE's in higher dimensions.[13] If one considers nonlinear PDE's in field theory, one deals with operator-distribution-valued equations. The difficulties to solve these equations are partly due to the appearance of functions of operators.

A review of methods to solve nonlinear differential equations (DE's) and study stability and bifurcation is given in Ref. 14. Here we want to propose a new approach to treat nonlinear DE's. We suggest to transform a nonlinear (quasilinear) DE into a linear integrodifferential equation, solve this equation, and reconstruct the solution of the original equation. When dealing with operator-value nonlinear PDE's it is considered as an advantage if one deals with a linear equation instead. Because our approach ultimately aims to treat those equations, it has been conceived to apply in principle to operator-valued DE's. In addition, for c-number-valued DE's, a transformation to a linear equation may be useful.

The present paper is considered as a first step towards this goal. We present the method in some detail for first-order nonlinear DE's. We show that the solution of the transformed linear equation is unique. It is straightforward to reconstruct the solution of the original equation (Sec. II). An analytical expression for the solution of the linearized equation is given. Properties of the kernel of the linearized equation are studied (Sec. III). Linear functional equations for the solution of the linearized equation are given in Sec. IV. For one sample case the method has been tested numerically, and rapid convergence to the reference solution was found (Sec. V). In Sec. VI, the method is generalized to higher-order nonlinear ordinary differential equations (ODE's). Section VII deals with nonlinear PDE's. For a Laplace equation with a holomorphic nonlinearity (Liouville equation, sine–Gordon equation) three kinds of assumptions are discussed which lead to a linearized equation.

II. FIRST-ORDER ORDINARY DIFFERENTIAL EQUATION

Assume $G_{\rm int}$, $G_{\rm med}$, $G_{\rm out}$, G denote domains in the complex plane, as shown in Fig. 1, with $G = (\bar{G}_{\rm int} \cup \bar{G}_{\rm med} \cup \bar{G}_{\rm out})^0$. Let S denote an oriented closed Jordan curve in $G_{\rm out}$. For the sake of technical simplicity we choose S to be a circle around the origin. The origin is assumed to be interior to $G_{\rm int}$.

Let $H(z)$ be a holomorphic function in G. Consider the first-order differential equation

$$\frac{d}{dx} f(x) = H(f(x)), \tag{2.1}$$

where x runs over a finite real interval $D = [x_1, x_2]$ and f satisfies the initial condition

$$f(x_1) = f_1. \tag{2.2}$$

We are interested in a unique, well-behaved solution of Eqs. (2.1) and (2.2).

It is known that there is a disk D_1: $|z - x_1| < r$ in the complex plane, where Eq. (2.1) has a unique holomorphic

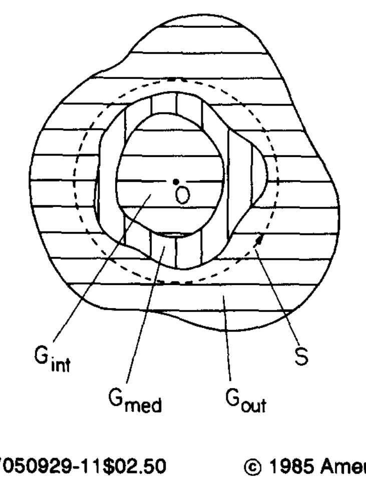


FIG. 1. Schematic plot of the domains $G_{\rm int}$, $G_{\rm med}$, $G_{\rm out}$ and the contour curve S, as defined in Sec. II.

 0022-2488/85/050929-11$02.50

solution, obeying the initial condition Eq. (2.2). A proof is given, e.g., in Ref. 15 (Theorems 2.2.1 and 2.3.1).

Moreover, we assume

$$f(D_1)\subset G_{\text{int}}. \tag{2.3}$$

Thus, we can consider the extension of Eq. (2.1) into the complex domain D_1, which we denote by

$$\frac{d}{dz}f(z)=H(f(z)), \tag{2.4}$$

with the initial value condition, where $z_1=x_1$,

$$f(z_1)=f_1. \tag{2.5}$$

We define for $\xi\in G_{\text{out}}$, $\eta\in G_{\text{int}}$ (see Fig. 1)

$$\Omega(\xi,\eta)=1/(\xi-\eta). \tag{2.6}$$

Hence, in particular, for $\eta=f(z)$, $z\in D_1$, we obtain

$$\frac{d}{dz}\Omega(\xi,f(z))=\frac{H(f(z))}{(\xi-f(z))^2}. \tag{2.7}$$

For each $\epsilon>0$ let $S_\epsilon=\{\xi(1+\epsilon)|\xi\in S\}$ denote a "radially inflated" curve and G_ϵ the corresponding domain bounded by the curve S_ϵ. For each $\xi\in S$, the function $1/(\xi(1+\epsilon)-\eta)^2$ is holomorphic as a function of η in $G_{\epsilon/2}$. Cauchy's theorem can be applied and yields then for $\xi\in S$, $\epsilon>0$, $\eta\in G_{\epsilon/2}$

$$\frac{H(\eta)}{(\xi(1+\epsilon)-\eta)^2}=\frac{1}{2\pi i}\int_S d\xi'\,\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\,\frac{1}{\xi'-\eta}. \tag{2.8}$$

Due to Eq. (2.3) and G_{int} having a positive distance from S by assumption, we can choose in particular $\eta=f(z)$ and perform the ϵ limit on the left-hand side (lhs) of Eq. (2.8). Thus, one obtains, using Eq. (2.7), for each $\xi\in S$, $z\in D_1$

$$\frac{d}{dz}\Omega(\xi,f(z))=\lim_{\epsilon\to+0}\frac{1}{2\pi i}\int_S d\xi'\,\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\times\Omega(\xi',f(z)). \tag{2.9}$$

The initial condition Eq. (2.5) now reads for each $\xi\in S$

$$\Omega(\xi,f(z_1))=\Omega(\xi,f_1)=1/(\xi-f_1). \tag{2.10}$$

We claim that Eqs. (2.9) and (2.10) give a transformation of the original nonlinear differential equation with respect to the variable z into a linear but singular integrodifferential equation with respect to the variables ξ, z. To make this more explicit, we define D_ω to the class of functions $\omega(\xi,z)$, with $\xi\in G_{\text{out}}$, $z\in D_1$, which are holomorphic in $G_{\text{out}}\times D_1$, which satisfy for $\xi\in S$, $z\in D_1$

$$\frac{d}{dz}\omega(\xi,z)=\lim_{\epsilon\to+0}\frac{1}{2\pi i}\int_S d\xi'\,\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\,\omega(\xi',z) \tag{2.11}$$

and possess the initial value

$$\omega(\xi,z_1)=1/(\xi-f_1), \tag{2.12}$$

for all $\xi\in S$. We claim the following.

Proposition 1: Defining for $\xi\in G_{\text{out}}$, $z\in D_1$

$$\omega'(\xi,z)=\Omega(\xi,f(z)) \tag{2.13}$$

and requiring for $\xi\in S$ the initial condition

$$\omega'(\xi,z_1)=1/(\xi-f_1), \tag{2.14}$$

we obtain an element $\omega'\in D_\omega$, i.e., D_ω is not empty.

Proof: Here, $\Omega(\xi,f(z))$ is holomorphic in $G_{\text{out}}\times D_1$ because $f(z)$ is holomorphic in D_1, $\Omega(\xi,\eta)$ is holomorphic in $G_{\text{out}}\times G_{\text{int}}$, and $f(D_1)\subset G_{\text{int}}$ by assumption. Equation (2.9) shows that ω', given by Eq. (2.13), is an explicit solution of Eq. (2.11).

With Eq. (2.11) we have obtained the desired equation, which obviously is a linear, singular integrodifferential equation. The following questions arise.

(a) Are there any other elements in D_ω apart from that given by Proposition 1?

(b) If the answer to (a) is no, how can one recover from the solution $\omega(\xi,z)$ of Eqs. (2.11) and (2.12) the solution of our original Eqs. (2.4) and (2.5)?

(c) What is the advantage of this transformation?

Part (a) is answered by Proposition 2.

Proposition 2: The element ω' given by Proposition 1 is the only element of D_ω.

The proof is given in Appendix A.

Concerning part (b), one immediately obtains for $z\in D_1$

$$\begin{aligned}f(z)&=\frac{1}{2\pi i}\int_S d\xi'\,\frac{\xi'}{\xi'-f(z)}\\&=\frac{1}{2\pi i}\int_S d\xi'\,\xi'\,\Omega(\xi',f(z))\\&=\frac{1}{2\pi i}\int_S d\xi'\,\xi'\,\omega(\xi',z).\end{aligned} \tag{2.15}$$

To answer (c), we consider it as an advantage that linear operator methods can be employed to obtain solutions. This will be useful in particular for applications to nonlinear partial differential equations. We want to mention here that Eq. (2.11) formally looks like the Schrödinger equation

$$-\frac{\hbar}{i}\frac{\partial}{\partial t}\psi=H\psi, \tag{2.16}$$

where H denotes a time-independent Hamilton operator.

We introduce the operator A defined on $\mathscr{A}(G_{\text{out}})$, the space of holomorphic functions in G_{out}, for all $\xi\in S$, by

$$(Ag)(\xi)=\lim_{\epsilon\to+0}\frac{1}{2\pi i}\int_S d\xi'\,\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\,g(\xi'). \tag{2.17}$$

By construction $\omega(\xi,z)$ is with respect to the ξ dependence an element of $\mathscr{A}(G_{\text{out}})$ for each $z\in D_1$.

Thus Eq. (2.11) can be written as

$$\frac{d}{dz}\omega(\circ,z)=A\omega(\circ,z). \tag{2.18}$$

III. SERIES EXPANSION AND ANALYTICAL EXPRESSION OF ω

The solution of the Schrödinger equation (2.16) is

$$\psi(t)=e^{(-i/\hbar)H(t-t_1)}\,\psi(t_1), \tag{3.1}$$

if H is time independent. Because the operator A, defined by Eq. (2.17), affects only the ξ dependence, one might expect from Eq. (2.18) by analogy to the Schrödinger equation that ω, the solution of Eqs. (2.11) and (2.12), could be written in a closed form

$$\omega(\circ,z)=e^{A(z-z_1)}\,\omega_1(\circ,z), \tag{3.2}$$

where

$$\omega_1(\xi,z)=\omega(\xi,z_1)=1/(\xi-f_1). \tag{3.3}$$

To show that this is actually true in some domain and to state its precise meaning we have Proposition 3.

Proposition 3: The operator A defined by Eq. (2.17) for $g\in\mathscr{A}(G_{\rm out})$ and restricted to $(Ag)(\xi)$ with $\xi\in S$ can be extended to $\hat{A}$ such that

$$(\hat{A}g)(\xi)=(Ag)(\xi), \tag{3.4}$$

for all $\xi\in S$. Here, $(\hat{A}g)(\xi)$ is defined for all $\xi\in G_{\rm out}$ and $\hat{A}g\in\mathscr{A}(G_{\rm out})$. This extension is unique.

Moreover, the domain of $\hat{A}$ can be extended to $\mathscr{A}(G_{\rm out}\times D_1)$ (which we denote by the same symbol $\hat{A}$ and one has

$$\hat{A}\,\mathscr{A}(G_{\rm out}\times D_1)\subset\mathscr{A}(G_{\rm out}\times D_1), \tag{3.5}$$

$$\left[\hat{A},\frac{d}{dz}\right]=0. \tag{3.6}$$

The solution of Eqs. (2.11) and (2.12) can be written for $\xi\in G_{\rm out}$, $z\in D_1$

$$\omega(\xi,z)=\sum_{\nu=0}^{\infty}\frac{1}{\nu!}(\hat{A}^{\nu}\omega_1)(\xi,z)(z-z_1)^{\nu}, \tag{3.7}$$

where ω_1 is given by Eq. (3.3). The expansion converges uniformly in ξ,z in any double disk $D_a\times D_b$ in $G_{\rm out}\times D_1$. Expression (3.2) is a formal way of writing Eq. (3.7). The proof is given in Appendix B.

IV. FUNCTIONAL EQUATIONS

In this section we seek information on ω, the solution of Eqs. (2.11) and (2.12). Here also the influence of ω outside the disk of convergence will play a role. We are going to construct functionals and calculate the mapping of ω under those functionals. One expects, however, that the set of these functionals is not large enough to render complete information on ω. We will introduce a Hilbert space $\mathscr{H}$ and transform Eq. (2.11) to

$$(1-B)\omega=g, \tag{4.1}$$

where $\omega,g\in\mathscr{H}$, $B:\mathscr{H}\to\mathscr{H}$, such that B^+ exists. For every $h\in\mathscr{H}$ we will define a functional F_h via

$$\langle F_h|=\langle(1-B^+)h|, \tag{4.2}$$

which yields

$$F_h(\omega)=\langle h|g\rangle. \tag{4.3}$$

In order to apply it to Eq. (2.11), we define a parametrization of the curve S via

$$\xi=R\exp(i\phi),\quad \phi\in[-\pi,+\pi] \tag{4.4}$$

and

$$\omega(\phi,x)=1/[R\exp(i\phi)-f(x)], \tag{4.5}$$

$$\omega_1(\phi,x)=\omega(\phi,x_1). \tag{4.6}$$

Thus Eq. (2.11) can be written as an integral equation

$$\omega(\phi,x)=\omega_1(\phi,x)+\lim_{\epsilon\to+0}\frac{1}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x-x')\times\int_{-\pi}^{\pi}d\phi'\,\frac{\exp(i\phi')H(R\exp(i\phi'))}{(\exp(i\phi)(1+\epsilon)-\exp(i\phi'))^2}\,\omega(\phi',x'). \tag{4.7}$$

We consider the following Hilbert space:

$$\mathscr{H}=L_2[-\pi,\pi]\times[x_1,x_2]=\left\{h\;\Big|\int_{-\pi}^{\pi}d\phi\int_{x_1}^{x_2}dx|h(\phi,x)|^2\text{ exists}\right\}, \tag{4.8}$$

with its usual scalar product denoted by $\langle\cdot,\cdot\rangle$ and we introduce

$$\mathscr{H}_0=\{h\,|\,h\in\mathscr{H},\ h(\phi,x)\text{ is }2\pi\text{ periodic and infinitely many times differentiable with respect to }\phi\}. \tag{4.9}$$

Here, $\mathscr{H}_0$ is a non-normalizable space, but it is dense in $\mathscr{H}$. We define for each $\epsilon>0$ the operator B_ϵ on $\mathscr{H}$ by

$$(B_\epsilon h)(\phi,x)=\frac{1}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x-x')\times\int_{\pi}^{\pi}d\phi'\,\frac{\exp(i\phi')H(R\exp(i\phi'))}{(\exp(i\phi)(1+\epsilon)-\exp(i\phi'))^2}\,h(\phi',x') \tag{4.10}$$

and B on $\mathscr{H}_0$ by

$$(Bh_0)(\phi,x)=\lim_{\epsilon\to+0}(B_\epsilon h_0)(\phi,x). \tag{4.11}$$

Then we have Proposition 4.

Proposition 4: For each $\epsilon>0$, B_ϵ and B both map $\mathscr{H}_0$ into $\mathscr{H}_0$. The Hilbert adjoint B_ϵ^+ is defined on $\mathscr{H}$ and maps $\mathscr{H}_0$ into $\mathscr{H}_0$. There is a "Hilbert adjoint" B^+ of B defined on $\mathscr{H}_0$ mapping into $\mathscr{H}_0$.

The proof is given in Appendix C.

Now Eq. (4.7) can be written, using Eqs. (4.6) and (4.11),

$$\omega=\omega_1+B\omega, \tag{4.12}$$

which is of the form as Eq. (4.1). Obviously ω and ω_1 are elements of $\mathscr{H}_0$. We can define functionals and give the mapping of the solution ω under these functionals. Corresponding to each $h\in\mathscr{H}_0$ we define F_h by Eq. (4.2), which yields, when applied to the solution,

$$\langle F_h|\omega\rangle=\langle h|\omega_1\rangle, \tag{4.13}$$

which is of the form of Eq. (4.3).

V. NUMERICAL SOLUTION OF THE LINEARIZED EQUATION

In this section a numerical application of the linearization method shall be described.

Let us consider the following example of Eq. (2.1):

$$\frac{d}{dx}f(x)=f^2(x),\quad x\in\mathbb{R}, \tag{5.1}$$

with the initial value condition

$$f(0)=-i. \tag{5.2}$$

The unique solution is given by

$$f(x)=1/(i-x). \tag{5.3}$$

For all real-valued arguments one finds the range of f to be bounded;

$$|f(x)|\leqslant 1. \tag{5.4}$$

In order to apply the linearization method it is suitable to choose a contour path S to be a circle around the origin with radius $R=2$. We want to solve Eq. (2.11) approximately. Basically we are going to seek an approximate solution in the form of Eq. (3.2), but with a continuous ξ dependence replaced by a discrete ξ dependence, thus A being replaced by a finite-dimensional matrix. The details of the construction are given in the following. For each function h, holomorphic in G, the Cauchy integral

$$h(z)=\frac{1}{2\pi i}\int_S d\xi\,\frac{h(\xi)}{\xi-z} \tag{5.5}$$

can be approximated by discretization of the path S in the following way:

$$h(z)\cong\frac{1}{2\pi i}\sum_{j=1}^{N}\Delta\xi_j\,\frac{h(\xi_j)}{\xi_j-z}, \tag{5.6}$$

where

$$\xi_j\in S,\quad \Delta\xi_j=\xi_{j+1}-\xi_j,\quad \xi_{N+1}=\xi_1,\quad j=1,\ldots,N. \tag{5.7}$$

We define

$$\omega_j(x)=\frac{\Delta\xi_j}{2\pi i}\,\omega(\xi_j,x) = \frac{1}{2\pi i}\,\frac{\Delta\xi_j}{\xi_j-f(x)},\quad j=1,\ldots,N. \tag{5.8}$$

The following property of the ω_j turns out to be useful. Each product $\omega_i(x)\omega_j(x)$ can be expressed, at least approximately, by a linear combination of the $\omega_k(x)$ with coefficients γ independent of x, i.e.,

$$\omega_i(x)\omega_j(x)\cong\sum_{k=1}^{N}\gamma_{ijk}\omega_k(x),\quad i,j=1,\ldots,N. \tag{5.9}$$

To establish this relation, first consider the following case.

(a) $i\neq j$, i.e., $\xi_i\neq\xi_j$:

$$\begin{aligned}\omega_i(x)\omega_j(x) &=\frac{1}{2\pi i}\,\frac{\Delta\xi_i}{\xi_i-f(x)}\,\frac{1}{2\pi i}\,\frac{\Delta\xi_j}{\xi_j-f(x)}\\ &=\frac{1}{2\pi i(\xi_j-\xi_i)}\left[\frac{1}{2\pi i}\,\frac{\Delta\xi_i\Delta\xi_j}{\xi_i-f(x)}-\frac{1}{2\pi i}\,\frac{\Delta\xi_i\Delta\xi_j}{\xi_j-f(x)}\right]\\ &=\alpha_{ij}\omega_i(x)+\alpha_{ji}\omega_j(x),\end{aligned} \tag{5.10}$$

where

$$\alpha_{ij}=\frac{1}{2\pi i}\,\frac{\Delta\xi_j}{\xi_j-\xi_i}. \tag{5.11}$$

In this case Eq. (5.9) holds exactly. Now consider the next case.

(b) $i=j$.

The application of Eqs. (5.5) and (5.6) to the unit function $h(z)=1$ and putting $z=f(x)$ gives

$$\begin{aligned}1&=\frac{1}{2\pi i}\int_S d\xi\,\frac{1}{\xi-f(x)}\\ &\cong\frac{1}{2\pi i}\sum_{j=1}^{N}\Delta\,\xi_j\,\frac{1}{\xi_j-f(x)}=\sum_{j=1}^{N}\omega_j(x),\end{aligned} \tag{5.12}$$

and multiplying by ω_i and using Eq. (5.10) for $i\neq j$ yields

$$\omega_i^2(x)\cong\omega_i(x)-\sum_{j=1}^{N}(1-\delta_{ij})(\alpha_{ij}\omega_i(x)+\alpha_{ji}\omega_j(x)). \tag{5.13}$$

Thus the γ coefficients of Eq. (5.9) are defined by Eqs. (5.11) and (5.13).

Now we construct an approximate set of linear first-order differential equations

$$\begin{aligned}\frac{d}{dx}\,\omega_i(x)&=\frac{1}{2\pi i}\,\frac{\Delta\xi_i}{(\xi_i-f(x))^2}\,\frac{d}{dx}f(x)\\ &=\frac{2\pi i}{\Delta\xi_i}\,\omega_i^2(x)f^2(x).\end{aligned} \tag{5.14}$$

Equation (5.6) yields for $h(z)=z^2$, with $z=f(x)$,

$$f^2(x)\cong\frac{1}{2\pi i}\sum_{j=i}^{N}\Delta\xi_j\,\frac{\xi_j^2}{\xi_j-f(x)}=\sum_{j=1}^{N}\xi_j^2\,\omega_j(x). \tag{5.15}$$

From Eqs. (5.14), (5.15), and (5.9) one implies

$$\begin{aligned}\frac{d}{dx}\,\omega_i(x)&\cong\frac{2\pi i}{\Delta\xi_i}\sum_{j=1}^{N}\gamma_{iij}\omega_j(x)\sum_{k=1}^{N}\xi_k^2\,\omega_k(x)\\ &\cong\frac{2\pi i}{\Delta\xi_i}\sum_{j,k,l=1}^{N}\gamma_{iij}\,\xi_k^2\,\gamma_{jkl}\omega_l(x)\\ &=\sum_{j=1}^{N}A_{ij}\omega_j(x),\end{aligned} \tag{5.16}$$

where we have introduced the coefficient matrix A, hence Eq. (5.16) reads in matrix notation

$$\frac{d}{dx}\,\boldsymbol{\omega}(x)\cong A\boldsymbol{\omega}(x). \tag{5.17}$$

Equation (5.17) has the general solution

$$\boldsymbol{\omega}(x)\cong e^{Ax}\boldsymbol{\omega}^0, \tag{5.18}$$

with $\boldsymbol{\omega}^0$ being a constant vector which can be determined from the initial value condition Eq. (5.2)

$$\begin{aligned}\omega_j^0=\omega_j(x)\,\big|_{x=0}&=\frac{1}{2\pi i}\,\frac{\Delta\xi_j}{\xi_j-f(x)}\bigg|_{x=0}\\ &=\frac{1}{2\pi i}\,\frac{\Delta\xi_j}{\xi_j+i}.\end{aligned} \tag{5.19}$$

Having found an approximate solution from Eqs. (5.18) and (5.19) it is easy to recover an approximate expression for $f(x)$ again by using Eq. (5.6) with $h(z)=z$, $z=f(x)$

$$\begin{aligned}f(x)&\cong\frac{1}{2\pi i}\sum_{j=1}^{N}\Delta\xi_j\,\frac{\xi_j}{\xi_j-f(x)}=\sum_{j=1}^{N}\xi_j\omega_j(x)\\ &=\boldsymbol{\xi}^*\circ\boldsymbol{\omega}(x).\end{aligned} \tag{5.20}$$

For numerical calculations the impractical exponential function of a matrix can be avoided by using the eigenrepresentation of A

$$A\mathbf{v}^{(k)}=\lambda^{(k)}\mathbf{v}^{(k)},\quad k=1,\ldots,N. \tag{5.21}$$

Then $\boldsymbol{\omega}^0$ can be expanded in the eigenvector basis

$$\boldsymbol{\omega}^0=\sum_{k=1}^{N}\sigma^{(k)}\mathbf{v}^{(k)}. \tag{5.22}$$

The expansion coefficients σ can be determined by comparison of Eq. (5.22) with (5.19), i.e.,

$$\sum_{k=1}^{N}\sigma^{(k)}v_j^{(k)}=\frac{1}{2\pi i}\,\frac{\Delta\xi_j}{\xi_j+i},\quad j=1,\ldots,N, \tag{5.23}$$

which forms an inhomogeneous system of linear equations for $\sigma^{(k)}$. Now Eq. (5.18) reads, using Eqs. (5.21) and (5.22)

TABLE I. Comparison of a reference solution f given by Eq. (5.3) with the solution f^{Cauchy} obtained via Eq. (5.2) from ω, the solution of the linearized equation (5.16), using $N = 15$ nodes for path discretization. Here, f^{L_2} is the best $L_2[-1, +1]$ approximation of f by the functions $\exp(\lambda^{(k)}x)$, $k = 1,\ldots,15$, where $\lambda^{(k)}$ are the eigenvalues of the matrix A given by Eq. (5.16). The f^{cont} is an analytic continuation of the Taylor series.

x	$f(x)$		$f^{\text{Cauchy}}(x)$		$f^{L_2}(x)$		$f^{\text{cont}}(x)$	
	Re	Im	Re	Im	Re	Im	Re	Im
0.0	0.000 000	− 1.000 00	− 0.210 387-9	− 0.999 999	− 0.981 782-7	− 1.000 00	0.000 000	− 1.000 00
0.1	− 0.990 099-1	− 0.990 099	− 0.990 098-1	− 0.990 098	− 0.990.096-1	− 0.990 099	− 0.990 099-1	− 0.990 099
0.2	− 0.192 308	− 0.961 538	− 0.192 308	− 0.961 538	− 0.192 308	− 0.961 538	− 0.192 308	− 0.961 538
0.3	− 0.275 229	− 0.917 431	− 0.275 229	− 0.917 430	− 0.275 230	− 0.017 431	− 0.275 229	− 0.917 431
0.4	− 0.344 828	− 0.862 069	− 0.344 827	− 0.862 068	− 0.344 827	− 0.862 069	− 0.344 828	− 0.862 069
0.5	− 0.400 000	− 0.800 000	− 0.400 001	− 0.799 998	− 0.400 000	− 0.800 000	− 0.400 000	− 0.800 000
0.6	− 0.441 176	− 0.735 294	− 0.441 221	− 0.735 267	− 0.441 177	− 0.735 294	− 0.441 176	− 0.735 294
0.7	− 0.469 799	− 0.671 141	− 0.470 563	− 0.670 605	− 0.469 798	− 0.671 141	− 0.469 799	− 0.671 141
0.8	− 0.487 805	− 0.609 756	− 0.496 592	− 0.602 726	− 0.487 805	− 0.609 756	− 0.487 805	− 0.609 756

$$\omega(x) \cong \sum_{k=1}^{N} \exp(\lambda^{(k)}x)\sigma^{(k)}\mathbf{v}^{(k)}. \tag{5.24}$$

Numerical results are shown in Table I. The exact solution $f(x)$ is compared with the Cauchy linearized approximation $f^{\text{Cauchy}}(x)$, which has been obtained from Eqs. (5.20) and (5.24), where $N = 15$ nodes have been used to discretize the contour S. Also shown is $f^{L_2}(x)$, the best approximation of $f(x)$ in the $L_2[-1, +1]$ norm by functions from the subspace $\{\exp(\lambda^{(k)}x) | k = 1,\ldots,N\}$. The table moreover displays $f^{\text{cont}}(x)$, obtained by analytic continuation of the Taylor series using the expansion points $x_1 = 0$ and 200 Taylor coefficients, $x_2 = 0.6$ and 30 coefficients, $x_3 = 1.2$ and six coefficients, and $x_4 = 1.8$ and one coefficient. The results show that the values of f^{Cauchy} have decreasing accuracy with increasing distance from the initial value $x = 0$. The radius of the disk of convergence of the exact solution is 1. Then f^{Cauchy} seems to break down at this radius. This default can be cured and the accuracy can be greatly improved with little additional effort as follows.

If we look for an approximate solution $f^{\text{Cauchy}}(x)$ in a finite real interval I which includes the initial value point $x = 0$ (but may exceed the disk of convergence) we split I into subintervals $I^{(\nu)} = [x^{(\nu)}, x^{(\nu+1)}]$, $\nu = 1,\ldots,M$. For simplicity let $x^{(1)} = 0$. Then one constructs piecewise approximations in $I^{(\nu)}$ starting with $I^{(1)}$. The idea is to use the general solution Eq. (5.18) in all intervals $I^{(\nu)}$, but with different vectors $\omega^{0(\nu)}$. In $I^{(1)}$ one keeps the former solution with

$$\omega^{0(1)} = \omega^0, \tag{5.25}$$

with ω^0 given by Eq. (5.19). From Eqs. (5.18), (5.25), and (5.20) one obtains approximately $f(x^{(2)})$

$$f(x^{(2)}) \cong \xi^* \circ e^{A(x^{(2)} - x^{(1)})}\omega^{0(1)}. \tag{5.26}$$

Then one determines a new initial vector $\omega^{0(2)}$ for $I^{(2)}$ in analogy to Eq. (5.19) by requiring

$$\omega_j^{0(2)} = \frac{1}{2\pi i}\frac{\Delta\xi_j}{\xi_j - f(x^{(2)})}, \quad j = 1,\ldots,N, \tag{5.27}$$

where $f(x^{(2)})$ is taken from Eq. (5.26). Then one proceeds to $I^{(3)}$ and obtains

$$f(x^{(3)}) \cong \xi^* \circ e^{A(x^{(3)} - x^{(2)})}\omega^{0(2)}, \tag{5.28}$$

and so on. The results are shown in Table II for the real part of the function. The imaginary part shows a similar behavior. The best approximation f^{L_2} is determined in the interval $I = [0,6]$. For f^{Cauchy}, $M = 16$ subintervals have been chosen. Also, N is the number of contour discretization nodes. Then f^{Cauchy} keeps six digits of accuracy over the interval [0,6] while the analytical continuation suddenly breaks down.

VI. ORDINARY DIFFERENTIAL EQUATIONS OF nTH ORDER

In this section it will be indicated that in analogy to Sec. II, in addition, a higher-order ordinary nonlinear differential

Table II. The same as Table I, but only for the real part of the functions. Here, N is the number of nodes for the path discretization, f^{Cauchy} is improved by recursive calculation of $M = 16$ initial values in the interval [0,6], and f^{L_2} is the best L_2 approximation in [0,6].

x	$f(x)$	$f^{\text{Cauchy}}(x)$			$f^{L_2}(x)$			$f^{\text{cont}}(x)$
		$N = 10$	$N = 15$	$N = 20$	$N = 10$	$N = 15$	$N = 20$	
0.0	0.000 000	− 0.237 541-9	+ 0.305 174-4	+ 0.210 387-9	+ 0.170 386-1	− 0.223 744-2	+ 0.590 338-2	0.000 000
0.6	− 0.441 176	− 0.441 233	− 0.441 204	− 0.441 176	− 0.436 595	− 0.440 740	− 0.439 842	− 0.441 176
1.2	− 0.491 803	− 0.492 013	− 0.491 812	− 0.491 803	− 0.494 473	− 0.492 181	− 0.490 816	− 0.491 803
1.8	− 0.424 528	− 0.424 667	− 0.424 530	− 0.424 528	− 0.424 756	− 0.424 239	− 0.425 060	− 0.424 329
2.4	− 0.355 030	− 0.355 117	− 0.355 029	− 0.355 029	− 0.352 833	− 0.355 316	− 0.355 665	− 0.359 356
3.0	− 0.300 000	− 0.300 058	− 0.299 999	− 0.300 000	− 0.301 213	− 0.299 700	− 0.298 886	div
3.6	− 0.257 880	− 0.257 920	− 0.257 879	− 0.257 880	− 0.258 957	− 0.258 180	− 0.258 476	
4.2	− 0.225 322	− 0.225 351	− 0.225 321	− 0.225 322	− 0.223 694	− 0.225 019	− 0.226 088	
4.8	− 0.199 667	− 0.199 689	− 0.199 667	− 0.199 667	− 0.200 093	− 0.199 992	− 0.198 760	
5.4	− 0.179 045	− 0.179 062	− 0.179 045	− 0.179 045	− 0.179 578	− 0.178 832	− 0.177 698	

equation can be written as a linear integrodifferential equation. We consider

$$\partial_x^n f(x) = H(f(x), \partial_x f(x), \ldots, \partial_x^{n-1} f(x)), \tag{6.1}$$

with the initial conditions

$$f(x_1) = a_1, \quad \partial_x f(x_1) = a_2, \ldots, \quad \partial_x^{n-1} f(x_1) = a_n. \tag{6.2}$$

We assume that $\partial_x^k f$, $k = 0,..,n-1$ map the domain D into $G_{\rm int}$.

Let $H(z_1,\ldots,z_n)$ be holomorphic for $(z_1,\ldots,z_n) \in G^n$. Let us define

$$y_1 = f, \quad y_2 = \partial_x f, \ldots, \quad y_n = \partial_x^{n-1} f \tag{6.3}$$

and

$$H_1(z_1,\ldots,z_n) = z_2, \quad H_2(z_1,\ldots,z_n) = z_3, \ldots,$$
$$H_{n-1}(z_1,\ldots,z_n) = z_n, \quad H_n(z_1,\ldots,z_n) = H(z_1,\ldots,z_n). \tag{6.4}$$

All H_i are holomorphic for $(z_1,\ldots,z_n) \in G^n$.

Thus Eq. (6.1) reads

$$\partial_x y_i = H_i(y_1,\ldots,y_n), \quad i = 1,\ldots,n, \tag{6.5}$$

with the initial conditions

$$y_i(x_1) = a_i. \tag{6.6}$$

We define for $\xi = (\xi_1,\ldots,\xi_n) \in G^n_{\rm out}$

$$\omega(\xi,x) = \prod_{i=1}^{n} \frac{1}{\xi_i - y_i(x)}, \tag{6.7}$$

with y_i being given by Eq. (6.3).

Then one obtains, using Cauchy's theorem,

$$\partial_x \omega(\xi,x) = \frac{1}{(2\pi i)^n} \lim_{\epsilon \to +0} \int_{S^n} d\xi' \prod_{i=1}^{n} \frac{1}{(\xi_i(1+\epsilon) - \xi'_i)} \times \sum_{j=1}^{n} \frac{H_j(\xi')}{(\xi_j(1+\epsilon) - \xi'_j)} \omega(\xi',x), \tag{6.8}$$

with

$$\int_{S^n} d\xi = \int_S d\xi_1 \int_S d\xi_2 \times \cdots \times \int_S d\xi_n, \tag{6.9}$$

in analogy to Eq. (2.11). The initial condition reads

$$\omega(\xi,x_1) = \prod_{i=1}^{n} \frac{1}{\xi_i - a_i}, \tag{6.10}$$

in analogy to Eq. (2.12).

VII. PARTIAL DIFFERENTIAL EQUATIONS

In this section it shall be demonstrated how the transformation used above can be applied to certain nonlinear partial differential equations. Let us consider the following type of PDE in three dimensions, i.e., a nonlinear Laplace equation

$$\Delta f = H(f). \tag{7.1}$$

Examples are the sine–Gordon equation and the Liouville equation.[13] Again we assume that H is a holomorphic function in G (Fig. 1) and $f(\mathbf{x}) \in G_{\rm int}$ for every $\mathbf{x}$ in a given domain D. If one tries to transform Eq. (7.1) into a linear equation in the same way it was done for an ODE, one is faced with a problem. The principle, which allowed us to transform Eq. (2.1) into a linear equation, can be stated in the following way. For each polynomial $P^{(n)}$ there is a polynomial $Q^{(m)}$ such that in some domain

$$\frac{d}{dx} P^{(n)}(f(x)) \cong Q^{(m)}(f(x)). \tag{7.2}$$

That is, the class of polynomials of $f(x)$ [the solution of Eq. (2.1)] is at least approximately closed under application of the operator d/dx [the class of holomorphic functions of $f(x)$ is exactly closed]. However, this does not hold for the PDE given by Eq. (7.1). If we define for $\xi \in G_{\rm out}$

$$\omega(\xi,\mathbf{x}) = 1/[\xi - f(\mathbf{x})], \tag{7.3}$$

we obtain from Eq. (7.1)

$$\Delta\omega(\xi,\mathbf{x}) = 2(\nabla f)^2/(\xi - f)^3 + \Delta f/(\xi - f)^2. \tag{7.4}$$

In general the term $(\nabla f)^2$ cannot be expressed as a polynomial (holomorphic function) of $f(\mathbf{x})$. In the following, three cases will be discussed which yield linear equations.

(a) We assume there is an auxiliary function H_1, which obeys

$$(\nabla f)^2 = H_1(\alpha,\beta,\gamma,f), \tag{7.5}$$

where α,β,γ are some generalized coordinates, being functions of $\mathbf{x}$, and $H_1(\xi_1,\ldots,\xi_4)$ is holomorphic for $(\xi_1,\ldots,\xi_4)$ in G^4. We define for $\xi \in G_{\rm out}$

$$\omega(\xi,\mathbf{x}) = 1/[\xi - f(\mathbf{x})]. \tag{7.6}$$

Then we obtain from Eqs. (7.4), (7.5), and (7.1)

$$\Delta\omega(\xi,\mathbf{x}) = \frac{2H_1(\alpha,\beta,\gamma,f)}{(\xi - f)^3} + \frac{H(f)}{(\xi - f)^2} = \frac{1}{2\pi i} \lim_{\epsilon \to +0} \int_S d\xi' \left[\frac{2H_1(\alpha,\beta,\gamma,\xi')}{(\xi(1+\epsilon) - \xi')^3} + \frac{H(\xi')}{(\xi(1+\epsilon) - \xi')^2} \right] \omega(\xi',\mathbf{x}), \tag{7.7}$$

which is a linear singular integropartial differential equation.

Let us give just two examples for H_1. If we consider the one-soliton solution of the sine–Gordon equation in $3+0$ dimensions, as given in Ref. 13, one obtains

$$H_1(\alpha,\beta,\gamma,f) = 2(1 - \cos(f)). \tag{7.8}$$

If we consider the first nontrivial solution of the Liouville equation in $3+0$ dimensions, which is given in Ref. 13, one obtains

$$H_1(\alpha,\beta,\gamma,f) = 2\exp(f). \tag{7.9}$$

(b) Let us consider a slight generalization of Eq. (7.1), but only in two dimensions,

$$(\partial_x^2 + \partial_y^2) f(x,y) = H(f, f_x, f_y). \tag{7.10}$$

Let H be holomorphic in G^3 and let $f(\mathbf{x})$, $f_x(\mathbf{x})$, $f_y(\mathbf{x})$ be elements for $G_{\rm int}$ for every $\mathbf{x}$ in the domain D.

Now we assume there are H_i, $i = 1,2,3$, holomorphic in G^3, which obey

$$\partial_x^2 f = H_1(f, f_x, f_y), \tag{7.11}$$

$$\partial_x, \partial_y f = H_2(f, f_x, f_y), \tag{7.12}$$

$$\partial_y^2 f = H_3(f, f_x, f_y). \tag{7.13}$$

The following example shows that this is no empty definition:

$$f(x,y)=1(x+y^2),\quad (\partial_x^2+\partial_y^2)f=-2f^2+2f^3-2ff_y,$$
$$H_1(z_1,z_2,z_3)=2z_1^3,\quad H_2(z_1,z_2,z_3)=-2z_1z_3,$$
$$H_3(z_1,z_2,z_3)=-2z_1^2-2z_1z_3. \tag{7.14}$$

We denote

$$f_1=f,\quad f_2=f_x,\quad f_3=f_y, \tag{7.15}$$
$$A_1(\mathbf{f})=f_2,\quad A_2(\mathbf{f})=H_1(\mathbf{f}),\quad A_3(\mathbf{f})=H_2(\mathbf{f}), \tag{7.16}$$
$$B_1(\mathbf{f})=f_3,\quad B_2(\mathbf{f})=H_2(\mathbf{f}),\quad B_3(\mathbf{f})=H_3(\mathbf{f}). \tag{7.17}$$

Putting

$$\omega(\boldsymbol{\xi},\mathbf{x})=\prod_{i=1}^{3}\frac{1}{\xi_i-f_i(\mathbf{x})}, \tag{7.18}$$

one obtains the linear first-order PDE's

$$\partial_x\omega(\boldsymbol{\xi},\mathbf{x})=\lim_{\epsilon\to+0}\frac{1}{(2\pi i)^3}\int_{S^3}d\boldsymbol{\xi}'\prod_{k=1}^{3}\frac{1}{(\xi_k(1+\epsilon)-\xi'_k)}\times\sum_{j=1}^{3}\frac{A_j(\boldsymbol{\xi}')}{(\xi_j(1+\epsilon)-\xi'_j)}\,\omega(\boldsymbol{\xi}',\mathbf{x}), \tag{7.19}$$

$$\partial_y\omega(\boldsymbol{\xi},\mathbf{x})=\lim_{\epsilon\to+0}\frac{1}{(2\pi i)^3}\int_{S^3}d\boldsymbol{\xi}'\prod_{k=1}^{3}\frac{1}{(\xi_k(1+\epsilon)-\xi'_k)}\times\sum_{j=1}^{3}\frac{B_j(\boldsymbol{\xi}')}{(\xi_j(1+\epsilon)-\xi'_j)}\,\omega(\boldsymbol{\xi}',\mathbf{x}). \tag{7.20}$$

If, in addition to Eq. (7.10), boundary conditions are given, linear equations can be set up to determine the functions H_i, $i=1,2,3$. Let us consider a rectangular domain $D=[x_1,x_2]\times[y_1,y_2]$ and assume the boundary conditions

$$f=a_1,\quad f_x=a_2,\quad f_y=a_3 \tag{7.21}$$

hold on the left lower border $R=[x_1,x_2]\times[y_1]\cup[x_1]\times[y_1,y_2]$, where a_i are some given functions. Integration of Eq. (7.11) yields

$$\partial_x f(x,y_1)-\partial_x f(x_1,y_1)=\frac{1}{(2\pi i)^3}\int_{S^3}d\boldsymbol{\xi}'H_1(\boldsymbol{\xi}')\int_{x_1}^{x}dx'\times\frac{1}{(\xi_1-f(x,y_1))(\xi_2-f_x(x,y_1))(\xi_3-f_y(x,y_1))} \tag{7.22}$$

or

$$F_1(x)=\frac{1}{(2\pi i)^3}\int_{S^3}d\boldsymbol{\xi}'H_1(\boldsymbol{\xi}')R_1(\boldsymbol{\xi}',x),\quad x\in[x_1,x_2], \tag{7.23}$$

where F_1,R_1 are known from the boundary conditions. Similar equations are obtained by integrating Eq. (7.13) over y and Eq. (7.12) over x or over y. Finally, the following holds:

$$H=H_1+H_3. \tag{7.24}$$

Equation (7.23) and similar equations and Eq. (7.24) set up linear equations for the determination of the functions H_i.

(c) Finally, we want to set up linear functional equations for the solution of Eq. (7.1).

We assume that the domain is real, i.e., $D\subset\mathbb{R}^3$ and $f(\mathbf{x})$, $\nabla f(\mathbf{x})$ are given on δD, the boundary of D. Let $\phi(\mathbf{x})$ be a function which is twice differentiable in D. Then Eq. (7.4) yields

$$\int_D d\mathbf{x}\,\Phi\Delta\omega=\int_D d\mathbf{x}\,\Phi\left[\frac{2(\nabla f)^2}{(\xi-f)^3}+\frac{H(f)}{(\xi-f)^2}\right]. \tag{7.25}$$

By using Green's theorem, one obtains

$$\int_D d\mathbf{x}\,\Phi\Delta\omega=\int_{\delta D}d\sigma[\Phi\nabla\omega-\omega\nabla\Phi]+\int_D d\mathbf{x}(\Delta\Phi)\omega$$
$$=g_1(\xi)+\int_D d\mathbf{x}(\Delta\Phi)\omega \tag{7.26}$$

and

$$\int_D d\mathbf{x}\,\Phi\frac{(\nabla f)^2}{(\xi-f)^3}=\frac{1}{4}\left[-\int_{\delta D}d\sigma\left[\frac{\Phi\nabla f}{(\xi-f)^2}+\frac{\nabla\Phi}{\xi-f}\right]-\int_D d\mathbf{x}\frac{\Delta\Phi}{\xi-f}\right]$$
$$=g_2(\xi)-\frac{1}{4}\int_D d\mathbf{x}(\Delta\Phi)\omega. \tag{7.27}$$

The functions g_1,g_2 are known from the boundary conditions. Equations (7.25)–(7.27) yield

$$\frac{3}{2}\int_D d\mathbf{x}(\Delta\Phi(\mathbf{x}))\omega(\xi,\mathbf{x})=-g_1(\xi)+2g_2(\xi)+\lim_{\epsilon\to+0}\frac{1}{2\pi i}\int d\xi'\times\int_D d\mathbf{x}\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\,\omega(\xi',\mathbf{x}), \tag{7.28}$$

for all $\xi\in S$ and Φ subject to the above-stated conditions. But otherwise Φ is arbitrary; it can be chosen, e.g., from an orthogonal function system. The functions g_1,g_2 depend on the boundary conditions on f and on Φ. Equation (7.28) constitutes a linear functional equation for ω.

VIII. CONCLUSIONS

We have pointed out the importance of nonlinear differential equations in many branches of physics and the need for practical solution methods. Here we have suggested a new approach to solve nonlinear differential equations, when the nonlinearity involved is of a holomorphic type. It consists of transforming the nonlinear differential equation into a linear integrodifferential or functional equation. The solution of the latter can be transformed easily into the solution of the original equation. The transformed linear equation has been studied. Its basic features are the following: It is an equation with a singular integral kernel. Its solution is unique and it can be given a closed analytical form. Linear functional equations in a Hilbert space have been given, which involve properties of the solution outside the convergence radius of its Taylor series. The kernel of the linear integrodifferential equation can be studied in a Hilbert space and it can be decomposed in a compact plus a bounded plus an unbounded operator in a simple form. In addition, a theorem on weak convergence of approximate solutions in a Hilbert space can be given. (The two latter results will be published elsewhere.) The method has been tested numerically for one example and was found to give results accurate up to the sixth digit if 20 nodes for the discretization of the contour integral have been used, which we consider as excellent convergence. All this has been done for first-order ordinary differential equations.

The method can be applied also to higher-order ODE's. Application of the method to partial differential equations is not so straightforward. Nevertheless it has been shown how linear integrodifferential equations or functional equations can be obtained using knowledge of an additional constraint equation or from boundary conditions. The type of PDE considered here is a nonlinear Laplace equation, examples of which are the sine–Gordon or the Liouville equation.

Finally we want to mention that the method can be applied also to operator-valued differential equations, if the operators are bounded. The basic tool of transformation to a linear equation, which is the Cauchy integral, can be generalized in the Dunford calculus[16] to operator-valued functions. We hope this might lead to useful applications in particle physics.

ACKNOWLEDGMENTS

It is a pleasure to thank M. Fortin and G. A. Philippin for fruitful discussions. The author is indebted to G. Leibbrandt and E. Meister for reading the manuscript and making valuable suggestions.

This work has been supported by the National Sciences and Engineering Research Council of Canada.

APPENDIX A: PROOF OF PROPOSITION 2

Here we want to give a proof of Proposition 2.

Corresponding to each $g \in \mathscr{A}(G_{\rm out})$ let us define ψ^g and Ξ^g via

$$\psi^g(\xi',\xi) = [\,g(\xi') - g(\xi)\,]/(\xi' - \xi), \tag{A1}$$

$$\Xi^g(\xi',\xi) = [\,\psi^g(\xi',\xi) - \psi^g(\xi,\xi)\,]/(\xi' - \xi), \tag{A2}$$

respectively, for all $\xi,\xi' \in G_{\rm out}$. Then both ψ^g and Ξ^g are holomorphic in $G^2_{\rm out}$. If $g \in \mathscr{A}(G_{\rm out}\times D_1)$, ψ^g and Ξ^g are defined by

$$\psi^g(\xi',\xi,z) = [\,g(\xi',z) - g(\xi,z)\,]/(\xi' - \xi), \tag{A3}$$

$$\Xi^g(\xi',\xi,z) = [\psi^g(\xi',\xi,z) - \psi^g(\xi,\xi,z)]/(\xi' - \xi), \tag{A4}$$

for all $\xi',\xi \in G_{\rm out}$, $z \in D_1$. Then both ψ^g, Ξ^g are holomorphic in $G^2_{\rm out}\times D_1$.

We claim that A defined on $\mathscr{A}(G_{\rm out})$ by Eq. (2.17) has the following representation:

$$(Ag)(\xi) = \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^g(\xi',\xi), \quad \xi \in S. \tag{A5}$$

This can be seen as follows. For $\xi \in S$ one has

$$(Ag)(\xi) = \lim_{\epsilon\to+0}\frac{1}{2\pi i}\left[\int_S d\xi'\,\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\times(g(\xi') - g(\xi)) + g(\xi)\int_S d\xi'\,\frac{H(\xi')}{(\xi(1+\epsilon)-\xi')^2}\right]. \tag{A6}$$

The second term on the right-hand side of Eq. (A6) vanishes because the integrand is holomorphic in the interior of S. Thus, we obtain

$$\begin{aligned}(Ag)(\xi) &= \lim_{\epsilon\to+0}\frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\frac{(\xi'-\xi)}{(\xi(1+\epsilon)-\xi')^2}\psi^g(\xi',\xi)\\ &= \lim_{\epsilon\to+0}\frac{1}{2\pi i}\left[\int_S d\xi'\, H(\xi')\frac{(\xi'-\xi)}{(\xi(1+\epsilon)-\xi')^2}\times(\psi^g(\xi',\xi) - \psi^g(\xi,\xi)) + \psi^g(\xi,\xi)\int_S d\xi'\times H(\xi')\frac{(\xi'-\xi)}{(\xi(1+\epsilon)-\xi')^2}\right].\end{aligned} \tag{A7}$$

Again, the second term vanishes on the rhs of Eq. (A7). Thus one has

$$\begin{aligned}(Ag)(\xi) &= \lim_{\epsilon\to+0}\frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\frac{(\xi'-\xi)^2}{(\xi(1+\epsilon)-\xi')^2}\Xi^g(\xi',\xi)\\ &= \lim_{\epsilon\to+0}\frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^g(\xi',\xi)\times\left[1 - \frac{2\xi\epsilon}{\xi(1+\epsilon)-\xi'} + \frac{\xi^2\epsilon^2}{(\xi(1+\epsilon)-\xi')^2}\right].\end{aligned} \tag{A8}$$

In the limit $\epsilon\to+0$ the last two terms in the bracket give no contribution, thus Eq. (A5) is obtained.

Now, we assume contrary to the claim of Proposition 2 that there is another $\omega'' \in D_\omega$. We put

$$\Delta(\xi,z) = \omega'(\xi,z) - \omega''(\xi,z). \tag{A9}$$

Because ω', ω'' obey Eq. (2.12) one has for $\xi \in S$

$$\Delta(\xi,z_1) = 0. \tag{A10}$$

Hence Eqs. (A3) and (A10) give for $\xi',\xi \in S$

$$\psi^\Delta(\xi',\xi,z_1) = 0. \tag{A11}$$

Similarly, Eqs. (A4) and (A11) imply for $\xi',\xi \in S$

$$\Xi^\Delta(\xi',\xi,z_1) = 0. \tag{A12}$$

Equations (2.18) and (A5) can be applied to yield for $\xi \in S$, $z \in D_1$

$$\begin{aligned}\frac{d}{dz}\Delta(\xi,z) &= \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')(\Xi^{\omega'}(\xi',\xi,z) - \Xi^{\omega''}(\xi',\xi,z))\\ &= \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^\Delta(\xi',\xi,z).\end{aligned} \tag{A13}$$

Hence Eqs. (A12) and (A13) yield for $\xi \in S$

$$\frac{d}{dz}\Delta(\xi,z)\big|_{z=z_1} = 0. \tag{A14}$$

By construction, Δ given by Eq. (A9) is a holomorphic function in $G_{\rm out}\times D_1$, i.e., $\Delta \in \mathscr{A}(G_{\rm out}\times D_1)$. If we denote $\Delta^{(n)}$ for $(\xi,z) \in G_{\rm out}\times D_1$, $n = 0,1,2,\ldots$, by

$$\Delta^{(n)}(\xi,z) = \left(\frac{d}{dz}\right)^n \Delta(\xi,z), \tag{A15}$$

then $\Delta^{(n)} \in \mathscr{A}(G_{\rm out}\times D_1)$ holds for all $n = 0,1,2,\ldots$. In addition, $\psi^{\Delta^{(n)}}, \Xi^{\Delta^{(n)}} \in \mathscr{A}(G^2_{\rm out}\times D_1)$ for all $n = 0,1,2,\ldots$ [see definitions (A3) and (A4) and the following remark].

Next we claim for $\xi \in S$, $z \in D_1$, $n = 0,1,2,\ldots,$

$$\Delta^{(n+1)}(\xi,z) = \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^{\Delta^{(n)}}(\xi',\xi,z). \tag{A16}$$

This is valid for $n = 0$ by Eq. (A13).

One immediately calculates

$$\frac{d}{dz}\psi^{\Delta^{(n)}}(\xi',\xi,z) = \psi^{\Delta^{(n+1)}}(\xi',\xi,z), \tag{A17}$$

$$\frac{d}{dz}\Xi^{\Delta^{(n)}}(\xi',\xi,z) = \Xi^{\Delta^{(n+1)}}(\xi',\xi,z), \tag{A18}$$

for $(\xi',\xi,z) \in G^2_{\rm out}\times D_1$, $n = 0,1,2,\ldots$.

Now we assume that Eq. (A16) is valid for $n = n_0$. Then we have for $\xi \in S$, $z \in D_1$,

$$\begin{aligned}\Delta^{(n_0+2)}(\xi,z) &= \frac{d}{dz}\Delta^{(n_0+1)}(\xi,z) \\ &= \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\frac{d}{dz}\Xi^{\Delta^{(n_0)}}(\xi',\xi,z) \\ &= \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^{\Delta^{(n_0+1)}}(\xi',\xi,z), \end{aligned} \tag{A19}$$

where Eq. (A18) has been used. That proves Eq. (A16) for all $n = 0,1,2,\ldots$.

Next, we claim for $\xi \in S$, $n = 0,1,2,\ldots$,

$$\Delta^{(n)}(\xi,z)|_{z=z_1} = 0. \tag{A20}$$

This is valid for $n = 0,1$ by Eqs. (A10) and (A14), respectively.

So let us assume Eq. (A20) is valid for $n = n_0$. That implies for $\xi', \xi \in S$

$$\psi^{\Delta^{(n_0)}}(\xi',\xi,z)|_{z=z_1} = 0, \tag{A21}$$

which in turn implies for $\xi', \xi \in S$

$$\Xi^{\Delta^{(n_0)}}(\xi',\xi,z)|_{z=z_1} = 0. \tag{A22}$$

Equations (A16) and (A22) yield for $\xi \in S$

$$\Delta^{(n_0+1)}(\xi,z)|_{z=z_1} = 0, \tag{A23}$$

which proves Eq. (A20) for all $n = 0,1,2,\ldots$.

Since $\Delta \in \mathscr{A}(G_{\text{out}} \times D_1)$, it can be expanded in a Taylor series with respect to the variable z, around $z = z_1$, for each $\xi \in S$

$$\Delta(\xi,z) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!}\Delta^{(\nu)}(\xi,z)|_{z=z_1}(z-z_1)^\nu. \tag{A24}$$

This series converges in the largest disk around z_1, where $\Delta(\xi,z)$ is holomorphic with respect to z. But $\Delta(\xi,z)$ is holomorphic in D_1 by the definition of D_ω.

Hence Eq. (A24) is valid for $\xi \in S$, $z \in D_1$. But Eq. (A20) implies then for $\xi \in S$, $z \in D_1$,

$$\Delta(\xi,z) = 0. \tag{A25}$$

Since $\Delta \in \mathscr{A}(G_{\text{out}} \times D_1)$, the theorem of identity of two holomorphic functions can be applied, which shows that Eq. (A25) holds for all $(\xi,z) \in G_{\text{out}} \times D_1$, which finishes the proof of Proposition 2.

APPENDIX B: PROOF OF PROPOSITION 3

Here we want to prove Proposition 3. Let us define for $g \in \mathscr{A}(G_{\text{out}})$, $\xi \in G_{\text{out}}$

$$(\hat{A}g)(\xi) = \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^g(\xi',\xi), \tag{B1}$$

where Ξ^g is given by Eq. (A2). Because $\Xi^g \in \mathscr{A}(G^2_{\text{out}})$ [see the remark after Eq. (A2)], the rhs of Eq. (B1) is a holomorphic function with respect to $\xi \in G_{\text{out}}$, i.e., $\hat{A}g \in \mathscr{A}(G_{\text{out}})$. Comparing Eq. (A5) with (B1) shows that $(Ag)(\xi)$ and $(\hat{A}g)(\xi)$ agree for $\xi \in S$. The identity theorem for holomorphic functions can be applied and yields the uniqueness of the extension $\hat{A}$ of A.

Now let us consider $g \in \mathscr{A}(G_{\text{out}} \times D_1)$, Hence $\Xi^g \in \mathscr{A}(G^2_{\text{out}} \times D_1)$ and we define analogously to Eq. (B1)

$$(\hat{A}g)(\xi,z) = \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^g(\xi',\xi,z). \tag{B2}$$

Obviously, the rhs is holomorphic with respect to $(\xi,z) \in G_{\text{out}} \times D_1$, i.e., $\hat{A}g \in (G_{\text{out}} \times D_1)$, which proves Eq. (3.5).

One has in analogy to Eq. (A18) for $\xi', \xi \in G_{\text{out}}$, $z \in D_1$

$$\frac{d}{dz}\Xi^g(\xi',\xi,z) = \Xi^{(d/dz)g}(\xi',\xi,z). \tag{B3}$$

Hence, Eqs. (B2) and (B3) imply for $\xi \in G_{\text{out}}$, $z \in D_1$

$$\begin{aligned}\left(\frac{d}{dz}\hat{A}g\right)(\xi,z) &= \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\frac{d}{dz}\Xi^g(\xi',\xi,z) \\ &= \frac{1}{2\pi i}\int_S d\xi'\, H(\xi')\Xi^{(d/dz)g}(\xi',\xi,z) \\ &= \left(\hat{A}\frac{d}{dz}g\right)(\xi,z), \end{aligned} \tag{B4}$$

which establishes Eq. (3.6).

Equation (2.18) reads more explicitly for $\xi \in S$, $z \in D_1$,

$$\frac{d}{dz}\omega(\xi,z) = (A\omega)(\xi,z). \tag{B5}$$

Using the extension $\hat{A}$ one obtains for $\xi \in G_{\text{out}}$, $z \in D_1$,

$$\frac{d}{dz}\omega(\xi,z) = (\hat{A}\omega)(\xi,z). \tag{B6}$$

Equations (B6) and (3.6) imply for $\xi \in G_{\text{out}}$, $z \in D_1$, $n = 0,1,2,\ldots$

$$\left(\frac{d}{dz}\right)^n \omega(\xi,z) = (\hat{A}^n\omega)(\xi,z). \tag{B7}$$

Because $\omega \in \mathscr{A}(G_{\text{out}} \times D_1)$ it can be expanded in a Taylor series with respect to z at $z = z_1$ for each $\xi \in G_{\text{out}}$

$$\omega(\xi,z) = \sum_{\nu=0}^{\infty}\frac{1}{\nu!}\left(\frac{d}{dz}\right)^\nu \omega(\xi,z)|_{z=z_1}(z-z_1)^\nu, \tag{B8}$$

which converges uniformly for ξ,z in any double disk $D_a \times D_b$ interior to $G_{\text{out}} \times D_1$ (see Ref. 17). Equations (B7) and (3.3) imply

$$\begin{aligned}\left(\frac{d}{dz}\right)^n \omega(\xi,z)|_{z=z_1} &= (\hat{A}^n\omega)(\xi,z)|_{z=z_1} \\ &= (\hat{A}^n\omega_1)(\xi,z)|_{z=z_1} = (\hat{A}^n\omega_1)(\xi,z). \end{aligned} \tag{B9}$$

The last equality holds because ω_1 is independent of z and application of $\hat{A}$ on ω_1 does not create any z dependence. Finally, Eqs. (B6) and (B9) yield Eq. (3.7).

APPENDIX C: PROOF OF PROPOSITION 4

Here we give the proof of Proposition 4. Let $h \in \mathscr{H}_0$. For each $\varepsilon > 0$ the integral kernel corresponding to B_ϵ given by Eq. (4.10) is a square-integrable function, hence $B_\epsilon h \in \mathscr{H}$. Moreover $(B_\epsilon h)(\phi,x)$ is 2π periodic with respect to ϕ. The same holds for $\partial^n_\phi(B_\epsilon h)(\phi,x)$, $n = 0,1,2,\ldots$, hence $B_\epsilon h \in \mathscr{H}_0$. Introducing in analogy to Eqs. (A3) and (A4)

$$\psi^h(\phi',\phi,x) = \frac{h(\phi',x) - h(\phi,x)}{\exp(i\phi') - \exp(i\phi)}, \tag{C1}$$

$$\Xi^h(\phi',\phi,x)=\frac{\psi^h(\phi',\phi,x)-\psi^h(\phi,\phi,x)}{\exp(i\phi')-\exp(i\phi)}, \tag{C2}$$

one obtains analogously to Eq. (A8)

$$(B_\epsilon h)(\phi,x)=\frac{1}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x-x')\times\int_{-\pi}^{\pi}d\phi'\exp(i\phi')H(R\exp(i\phi'))\times\Big[1-\frac{2\epsilon\exp(i\phi)}{\exp(i\phi)(1+\epsilon)-\exp(i\phi')}+\frac{\epsilon^2\exp(2i\phi)}{(\exp(i\phi)(1+\epsilon)-\exp(i\phi'))^2}\Big]\times\Xi^h(\phi',\phi,x') \tag{C3}$$

and in the ϵ limit in analogy to Eq. (A5)

$$(Bh)(\phi,x)=\frac{1}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x-x')\int_{-\pi}^{\pi}d\phi'\exp(i\phi')\times H(R\exp(i\phi'))\,\Xi^h(\phi',\phi,x'). \tag{C4}$$

Since Ξ^h is 2π periodic and infinitely many times differentiable with respect to ϕ',ϕ, so is $(Bh)(\phi,x)$ 2π periodic and infinitely many times differentiable with respect to ϕ. That implies also for $(Bh)(\phi,x)$ to be square integrable with respect to (ϕ,x), hence $Bh\in\mathscr{H}_0$.

Next we consider the Hilbert adjoint of B_ϵ. One can easily verify that $\widehat{B}_\epsilon$, defined on $\mathscr{H}$ by

$$(\widehat{B}_\epsilon g(\phi,x)=\frac{1}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x'-x)\times\int_{-\pi}^{\pi}d\phi'\,\frac{\exp(-i\phi)(H(R\exp(i\phi)))^*}{(\exp(-i\phi')(1+\epsilon)-\exp(-i\phi))^2}\times g(\phi',x'), \tag{C5}$$

fulfills the requirement of a Hilbert adjoint of B_ϵ, i.e.,

$$\langle g|B_\epsilon h\rangle=\langle\widehat{B}_\epsilon g|h\rangle, \tag{C6}$$

for all $g,h\in\mathscr{H}$, and thus

$$B_\epsilon^+=\widehat{B}_\epsilon. \tag{C7}$$

In analogy to B_ϵ one verifies that B_ϵ^+ maps $\mathscr{H}_0$ into $\mathscr{H}_0$.

Next we give the definition and study properties of a "Hilbert adjoint" of B. Let $\widehat{B}$ be defined on $\mathscr{H}_0$ by

$$(\widehat{B}h(\phi,x)=\lim_{\epsilon\to+0}(B_\epsilon^+h)(\phi,x). \tag{C8}$$

One can write Eq. (C5) in the form

$$(B_\epsilon^+h)(\phi,x)=\frac{\exp(i\phi)(H(R\exp(i\phi)))^*}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x'-x)\times\int_{-\pi}^{\pi}d\phi'\frac{\exp(2i\phi')}{(\exp(i\phi)(1+\epsilon)-\exp(i\phi'))^2}h(\phi',x'). \tag{C9}$$

In analogy to Eqs. (A8) and (C3) one has

$$(B_\epsilon^+h)(\phi,x)=\frac{\exp(i\phi)(H(R\exp(i\phi)))^*}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x'-x)\times\int_{-\pi}^{\pi}d\phi'\exp(2i\phi')\Big[1-\frac{2\epsilon\exp(i\phi)}{\exp(i\phi)(1+\epsilon)-\exp(i\phi')}+\frac{\epsilon^2\exp(2i\phi)}{(\exp(i\phi)(1+\epsilon)-\exp(i\phi'))^2}\Big]\Xi^h(\phi',\phi,x'), \tag{C10}$$

which yields in the ϵ limit in analogy to Eqs. (A5) and (C4)

$$(\widehat{B}h)(\phi,x)=\frac{\exp(i\phi)(H(R\exp(i\phi)))^*}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x'-x)\times\int_{-\pi}^{\pi}d\phi'\exp(2i\phi')\,\Xi^h(\phi',\phi,x'). \tag{C11}$$

Similarly as above one concludes that $\widehat{B}$ maps $\mathscr{H}_0$ into $\mathscr{H}_0$. Next we claim that for every $h\in\mathscr{H}$,

$$B_\epsilon h-Bh\ \overset{\|\ \|}{\underset{\epsilon\to0}{\longrightarrow}}\ 0, \tag{C12}$$

$$B_\epsilon^+h-\widehat{B}h\ \overset{\|\ \|}{\underset{\epsilon\to0}{\longrightarrow}}\ 0. \tag{C13}$$

We noted already that $(B_\epsilon h)(\phi,x)$ and $(B_\epsilon^+h)(\phi,x)$ converge pointwise to $(Bh)(\phi,x)$ and $(\widehat{B}h)(\phi,x)$, respectively. It turns out that they even converge in the sup norm which guarantees the convergence in the L_2 norm. Let us verify it for $B_\epsilon h$. According to Eqs. (C3) and (C4) we have to estimate

$$(B_\epsilon h-Bh)(\phi,x)=\frac{1}{2\pi R}\int_{x_1}^{x_2}dx'\,\theta(x-x')\int_{-\pi}^{\pi}d\phi'\exp(i\phi')\times H(R\exp(i\phi'))\Big[\frac{-2\epsilon\exp(i\phi)}{\exp(i\phi)(1+\epsilon)-\exp(i\phi')}+\frac{\epsilon^2\exp(2i\phi)}{(\exp(i\phi)(1+\epsilon)-\exp(i\phi'))^2}\Big]\Xi^h(\phi',\phi,x'). \tag{C14}$$

Similar estimates have been performed in Ref. 18.

We define

$$H_s=\sup_{\xi\in S}|H(\xi)|, \tag{C15}$$

$$\Xi_s^h=\sup_{\substack{x\in[x_1,x_2]\\ \phi,\phi'\in[-\pi,+\pi]}}|\Xi^h(\phi,\phi',x)|. \tag{C16}$$

Due to the 2π periodicity with respect to ϕ of the functions in Eq. (C14), one can change in Eq. (C14) the boundaries of ϕ' integration $[-\pi,+\pi]$ to $[\phi-\pi,\phi+\pi]$. Then we estimate

$$|(B_\epsilon h-Bh)(\phi,x)|\leqslant\frac{1}{2\pi R}H_s(x_2-x_1)\,\Xi_s^h\times\int_{\phi-\pi}^{\phi+\pi}d\phi'\,\Big|\frac{2\epsilon}{1+\epsilon-\exp(i(\phi'-\phi))}\Big|+\Big|\frac{\epsilon}{1+\epsilon-\exp(i(\phi'-\phi))}\Big|^2. \tag{C17}$$

The substitution $\phi'-\phi\to\phi''$ shows that the integral is independent of ϕ. Thus we can estimate

$$|(B_\epsilon h-Bh)(\phi,x)|\leqslant M\int_{-\pi}^{\pi}d\phi''\,2\Big|\frac{\epsilon}{1+\epsilon-\exp(i\phi'')}\Big|+\Big|\frac{\epsilon}{1+\epsilon-\exp(i\phi'')}\Big|^2. \tag{C18}$$

We split the integration interval

$$[-\pi,+\pi]=[-\pi,-\sqrt{\epsilon}]\cup[-\sqrt{\epsilon},+\sqrt{\epsilon}]\cup[+\sqrt{\epsilon},\pi]. \tag{C19}$$

In the first and third interval, the integrand can be estimated, using

$$\left|\frac{\epsilon}{1+\epsilon-\exp(i\phi'')}\right| \leqslant \left|\frac{\epsilon}{1+\epsilon-\exp(\pm i\sqrt{\epsilon})}\right| \underset{\epsilon\to 0}{\sim} \sqrt{\epsilon}. \quad \text{(C20)}$$

In the second interval, the integrand can be estimated, using

$$|\epsilon/[1+\epsilon-\exp(i\phi'')]| \leqslant 1. \quad \text{(C21)}$$

Thus the integral in Eq. (C18) tends to 0 with $\epsilon \to 0$, because either the integrand tends to 0 (first and third interval) or the integrand is bounded but the length of the interval tends to 0 (second interval). Thus we obtain

$$\sup_{\substack{\phi\in[-\pi,+\pi]\\ x\in[x_1,x_2]}} |(B_\epsilon h - Bh)(\phi,x)| \underset{\epsilon\to 0}{\longrightarrow} 0, \quad \text{(C22)}$$

which guarantees Eq. (C12). Analogously one obtains Eq. (C13). Now we have for $g, h \in \mathscr{H}_0$

$$\langle g|Bh\rangle = \langle g|(B-B_\epsilon)h\rangle + \langle B_\epsilon^+ g|h\rangle$$
$$= \langle g|(B-B_\epsilon)h\rangle + \langle (B_\epsilon^+ - \widehat{B})g|h\rangle + \langle \widehat{B}g|h\rangle. \quad \text{(C23)}$$

Due to Eqs. (C12) and (C13) the first two terms on the rhs vanish with $\epsilon \to 0$, hence

$$\langle g|Bh\rangle = \langle \widehat{B}g|h\rangle. \quad \text{(C24)}$$

We define

$$B^+ = \widehat{B} \quad \text{(C25)}$$

and call it the "Hilbert adjoint" of B.

[1]O. W. Richardson *The Emission of Electricity from Hot Bodies* (Longmans, Green, New York, 1921), pp. 50–54; M. V. Laue, Jahrb. Radioakt. Elektron. **15**, 205 (1918).

[2]H. Bateman, *Partial Differential Equations of Mathematical Physics* (Cambridge U.P., Cambridge, 1964), pp. 166–168.

[3]W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1965), pp. 180–183.

[4]S. Richardson, Mathematika **27**, 321 (1980).

[5]G. W. Walker, Proc. R. Soc. London Ser. A **91**, 410 (1915).

[6]D. Olive, "Classical Solution in Gauge Theories–Spherically Symmetric Monopoles–Lax Pairs and Toda Lattices," Imperial College Preprint ICTP/80/81, Imperial College, London.

[7]L. P. Eisenhart, *Differential Geometry of Curves and Surfaces* (Dover, New York, 1960).

[8]A. J. DeMaria, D. A. Stetser, and W. H. Glenn, Jr., Science **156**, 1557 (1967); G. J. Lamb, Jr., Phys. Lett. A **25**, 181 (1967); Rev. Mod. Phys. **43**, 99 (1971).

[9]H. Washimi and T. Taniuti, Phys. Rev. Lett. **17**, 996 (1966); F. D. Tappert, Phys. Fluids **15**, 2446 (1972).

[10]B. D. Josephson, Phys. Lett. **1**, 251 (1962); P. W. Anderson and J. M. Rowell, Phys. Rev. Lett. **10**, 230 (1963).

[11]R. Dashen, B. Hasslacher, and R. Neveu, Phys. Rev. D **10**, 4114, 4130, 4138 (1974); D **11**, 3424 (1975); R. Rajaraman, Phys. Rep. C **21**, 227 (1975). See also "Extended Systems in Field Theory," edited by J. L. Gervais and A. Neveu, Phys. Rep. C **23**, 236 (1976).

[12]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1462 (1973); **31**, 125 (1973).

[13]G. Leibbrandt, R. Morf, and S. Wang, J. Math. Phys. **21**, 1613 (1980); G. Leibbrandt, S. Wang, and N. Zamani, J. Math. Phys. **25**, 1566 (1982).

[14]*Proceedings of the International Conference on "Nonlinear Differential Equations: Invariance, Stability and Bifurcation,"* Trento, Italy, 1980, edited by P. de Mottoni and L. Salvadori (Academic, New York, 1981).

[15]E. Hille, *Ordinary Differential Equations in the Complex Domain* (Wiley, New York, 1976).

[16]W. Rudin, *Functional Analysis* (McGraw–Hill, New York, 1973).

[17]L. Bers, *Introduction to Several Complex Variables* (Courant Institute of Mathematical Sciences, New York University, New York, 1964), Chap. 1, Corollary 5; H. Behnke and P. Thullen, *Theorie der Funktionen mehrerer komplexer Veränderlichen, Ergebnisse der Mathematik and ihrer Grenzgebiete* (Springer, Berlin, 1970), Vol. 51, Chap. 3, paragraph 4.

[18]H. Kröger and C. C. Fenske, Rep. Math. Phys. **17**, 459 (1980).

Compact expression for Löwdin's alpha function

A. A. Antone
Department of Applied Mathematical Studies, University of Leeds, Leeds, England[a)]

(Received 17 July 1984; accepted for publication 21 December 1984)

A simple analytical expression for Löwdin's alpha function is derived. This expression is expected to be more convenient with the expansion of a general function about a displaced center than any other available in the literature.

I. INTRODUCTION

In many nuclear and electronic physics problems we are faced with "the problem of shifted origins," where we need to expand a function $F(\mathbf{R}) = f(R)Y_L^m(\mathbf{R})$ in terms of the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ in the form

$$f(R)Y_L^m(\mathbf{R}) = \sum_{l_1 l_2 m_1} \langle l_1,m_1;l_2,m-m_1|l_1 l_2 L m\rangle \times \alpha(l_1,l_2,L,r_1,r_2,f) Y_{l_1}^{m_1}(\Omega_1) Y_{l_2}^{m-m_1}(\Omega_2), \tag{1}$$

where

$$\mathbf{R} = \mathbf{r}_1 + \mathbf{r}_2 \tag{2}$$

and the solid harmonic $Y_L^m(\mathbf{R})$ is defined in terms of the spherical harmonic $Y_L^m(\Omega_R)$ as

$$Y_L^m(\mathbf{R}) = R^L Y_L^m(\Omega_R) = R^L Y_L^m(\theta_R,\phi_R). \tag{3}$$

Such an expansion would enable us to evaluate physical quantities like electronic energy, electric dipole transition probabilities, molecular dipole moments, and various other quantities which are usually expressed in terms of integrals of the translation operator known as the multicenter integrals.

In the special case when $\mathbf{r}_2$ lies on the z axis and is denoted in the spherical polar coordinates (r_2,θ_2,ϕ_2) by

$$(r_2,0,\phi_2) = \mathbf{a}, \tag{4}$$

Eq. (1) takes the form

$$f(R)Y_L^m(\mathbf{R}) = f(R)Y_L^m(\mathbf{r}_1 + \mathbf{a}) = \sum_{l_1 l_2} \langle l_1,m;l_2,0|l_1 l_2 L m\rangle \left\{\frac{2l_2+1}{4\pi}\right\}^{1/2} \times \alpha(l_1,l_2,L,r_1,a,f) Y_{l_1}^m(\Omega_1), \tag{5}$$

where we have used

$$Y_j^m(0,\phi) = [(2j+1)/4\pi]^{1/2}\delta_{m,0}. \tag{6}$$

Because of the history of the problem many attempts have been made in the past to derive an analytical expression for Löwdin's[1] alpha function $\alpha(l_1,l_2,L,r_1,a,f)$ appearing in Eq. (5). Recently Suzuki[2] has attempted to derive some recurrence relations that would make numerical computations easier. However, his study was based on two analytical expressions derived previously by Sharma[3] and Silverstone *et al.*,[4] respectively. Sharma's expression[5] involves the summation over four indices while that of Silverstone *et al.*[6] contains only three. Neither of these expressions visibly reduces to the well-known special case when $f(R) = 1$. The solid harmonic $Y_L^m(\mathbf{R}) = Y_L^m(\mathbf{r}_1+\mathbf{r}_2)$ is expressed by Eq. (12.41) in Talman[7] in terms of the solid harmonics $Y_{l_1}^{m_1}(\mathbf{r}_1)$ and $Y_{L-l_1}^{m-m_1}(\mathbf{r}_2)$ as

$$Y_L^m(\mathbf{r}_1+\mathbf{r}_2) = \sum_{l_1 m_1} \left\{\frac{4\pi(2L+1)!}{(2l_1+1)!(2L-2l_1+1)!}\right\}^{1/2} \times \langle l_1,m_1;L-l_1,m-m_1|l_1(L-l_1)Lm\rangle \times Y_{l_1}^{m_1}(\mathbf{r}_1) Y_{L-l_1}^{m-m_1}(\mathbf{r}_2), \tag{7}$$

which, when $\mathbf{r}_2 = \mathbf{a}$, reduces to

$$Y_L^m(\mathbf{r}_1+\mathbf{a}) = \sum_{l_1} a^{L-l_1}\left\{\frac{(2L+1)!}{(2l_1+1)!(2L-2l_1)!}\right\}^{1/2} \times \langle l_1,m;L-l_1,0|l_1(L-l_1)Lm\rangle Y_{l_1}^m(\mathbf{r}_1). \tag{8}$$

In this paper we will derive an analytical expression for the general function $\alpha(l_1,l_2,L,r_1,r_2,f)$ appearing in Eq. (1) and then deduce the special case when $\mathbf{r}_2 = \mathbf{a}$ and find the simple expression for Löwdin's alpha function in the form (we will prove later that l_2 can only take the value $l_2 = L - l_1$)

$$\alpha(l_1,L,r_1,a,f) = \left[a^{L-l_1} r_1^{l_1} \Gamma(l_1+\tfrac{3}{2})/l_1!\right] \times \left\{\frac{4(2L+1)!}{(2l_1+1)!(2L-2l_1+1)!}\right\}^{1/2} \times \int_0^\pi f(\lambda)\sin^{2l_1+1}\theta\, d\theta, \tag{9}$$

where $\lambda^2 = r_1^2 + a^2 + 2ar_1\cos\theta$ and the integral may be written as

$$\int_0^\pi f(\lambda)\sin^{2l_1+1}\theta\, d\theta = \frac{1}{(2ar_1)^{2l_1+1}} \int_{|a-r_1|}^{a+r_1} 2\lambda f(\lambda)d\lambda \times \{4a^2r_1^2 - (\lambda^2 - a^2 - r_1^2)^2\}^{l_1}. \tag{10}$$

It is worth noting at this stage that when $f(R) = 1$, Eq. (9) gives

$$\alpha(l_1,L,r_1,a,1) = \left[a^{L-l_1} r_1^{l_1} \Gamma(l_1+\tfrac{3}{2})/l_1!\right] \times \left\{\frac{4(2L+1)!}{(2l_1+)!(2L-2l_1+1)!}\right\}^{1/2} \left\{\frac{\sqrt{\pi}\, l_1!}{\Gamma(l_1+\tfrac{3}{2})}\right\} = a^{L-l_1} r_1^{l_1} \left\{\frac{4\pi(2L+1)!}{(2l+1)!(2L-2l_1+1)!}\right\}^{1/2}. \tag{11}$$

[a)] Current address: Department of Mathematics, Mu'tah University, P.O. Box 5076, Amman, Jordan.

 0022-2488/85/050940-03$02.50

Substituting Eq. (11) in Eq. (5) we reproduce Talman's relation given by Eq. (8). Similarly, in the general case our expansion given by Eq. (1) reduces to, on using Eq. (32), Talman's relation given by Eq. (7) when $f(R)=1$.

II. CALCULATION OF $\alpha(l_1,l_2,L,r_1,r_2,f)$

Using the orthonormality of the spherical harmonics we invert Eq. (1) as

$$\langle l_1,l_1;l_2,L-l_1|l_1l_2LL\rangle\alpha(l_1,l_2,L,r_1,r_2,f)$$
$$=\int f(R)Y_L^L(\mathbf{R})\overset{*}{Y}{}_{l_1}^{l_1}(\Omega_1)\overset{*}{Y}{}_{l_2}^{L-l_1}(\Omega_2)d\Omega_1\,d\Omega_2, \tag{12}$$

where we have chosen $m=L$ and $m_1=l_1$ and use the usual notation $d\Omega=\sin\theta\,d\theta\,d\phi$. In the spherical polar coordinates (R,θ,ϕ) Eq. (2) gives

$$R^2=r_1^2+r_2^2+2r_1r_2\times\{\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\phi_1-\phi_2)\} \tag{13}$$

and

$$R\sin\theta\,e^{i\phi}=r_1\sin\theta_1\,e^{i\phi_1}+r_2\sin\theta_2\,e^{i\phi_2}. \tag{14}$$

In order to evaluate the integral appearing in Eq. (12) we need Eq. (2.5.5) of Ref. 8 in the form

$$Y_j^m(\theta,\phi)=\frac{(-1)^m}{2^jj!}\left\{\frac{(2j+1)(j-m)!}{4\pi(j+m)!}\right\}^{1/2}\sin^m\theta\,e^{im\phi}\times\left(\frac{d}{d\cos\theta}\right)^{j+m}(\cos^2\theta-1)^j$$
$$=\frac{(-1)^m}{2^j}\left\{\frac{(2j+1)(j-m)!}{4\pi(j+m)!}\right\}^{1/2}\sin^m\theta\,e^{im\phi}\times\sum_x\frac{(-1)^x(2j-2x)!\cos^{j-m-2x}\theta}{x!(j-x)!(j-m-2x)!}$$
$$=(-2)^m\left\{\frac{(2j+1)(j-m)!}{4\pi(j+m)!}\right\}^{1/2}\sin^m\theta\,e^{im\phi}\times\sum_x\frac{(-1)^x\Gamma(j-x+\frac{1}{2})\cos^{j-m-2x}\theta}{x!\Gamma\{\frac{1}{2}(j-m-2x+1)\}\Gamma\{\frac{1}{2}(j-m-2x+2)\}} \tag{15}$$

where we have used the duplication formula

$$\sqrt{\pi}\,\Gamma(2z)=2^{2z-1}\Gamma(z)\Gamma(z+\tfrac{1}{2}). \tag{16}$$

When $m=j$, Eq. (15) gives

$$Y_L^L(\mathbf{R})=\frac{(-1)^L}{2^LL!}\left\{\frac{(2L+1)!}{4\pi}\right\}^{1/2}(R\sin\theta\,e^{im\phi})^L$$
$$=\frac{(-1)^L}{2^LL!}\left\{\frac{(2L+1)!}{4\pi}\right\}^{1/2}\sum_k\frac{L!}{k!(L-k)!}\times(r_1\sin\theta_1\,e^{i\phi_1})^{L-k}(r_2\sin\theta_2\,e^{i\phi_2})^k, \tag{17}$$

where we have used Eq. (14).

We also need Eq. (3.6.13) of Ref. 8 as

$$\langle l_1,l_1;l_2,L-l_1|l_1l_2LL]=\left\{\frac{(2L-1)!(2l_1)!}{(L+l_1-l_2)!(L+l_1+l_2+1)!}\right\}^{1.2}. \tag{18}$$

Using Eqs. (15), (17), and (18) we write Eq. (12) as

$$\alpha(l_1,l_2,L,r_1,r_2,f)=\sum_{kx}\frac{A(k,x)}{4\pi^2}\int_0^\pi d\theta_1\sin^{L+l_1-k+1}\theta_1\times\int_0^\pi d\theta_2\sin^{L+k-l_1+1}\cos^{l_1+l_2-L-2x}\times\int_0^{2\pi}d\phi_1\,e^{(ik+l_1-L)\phi_2}\times\int_0^{2\pi}f(R)\,e^{i(L-l_1-k)\phi_1}\,d\phi_1, \tag{19}$$

where

$$A(k,x)=\frac{r_1^{l-k}r_2^k(-1)^x\Gamma(l_2-x+\frac{1}{2})}{2^{2l_1+1}k!(L-k)!l_1!x!\Gamma\{\frac{1}{2}(l_1+l_2-L-2x+1)\}\Gamma\{\frac{1}{2}(l_1+l_2-L-2x+2)\}}\times\left\{\frac{\pi(2l_1+1)(2l_2+1)(l_1+l_2-L)!(L+l_1-l_2)!)L+l_1+l_2+1)!}{(L+l_2-l_1)!}\right\}^{1.2}. \tag{20}$$

We now put $\psi=\phi_1-\phi_2$ and notice Eq. (13) showing that R takes the same value

$$R_0^2=r_1^2+r_2^2+2r_1r_2\{\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos\phi_2\},$$

when ψ takes either of the values $\psi=-\phi_2$ or $\psi=2\pi-\phi_2$.

We therefore write

$$\int_0^{2\pi}\int_0^{2\pi}f(R)e^{i(L-l_1-k)(\phi_1-\phi_2)}\,d\phi_1\,d\phi_2=\int_0^{2\pi}d\phi_2\int_{-\phi_2}^{2\pi-\phi_2}f(R)e^{i(L-l_1-k)\psi}\,d\psi$$
$$=\int_0^{2\pi}d\phi_2\int_{R_0}^{R_0}G(R)\,dR=4\pi^2f(\rho)\delta_{\psi,0}, \tag{21}$$

where

$$\rho^2=r_1^2+r_2^2+2r_1r_2\cos(\theta_1-\theta_2). \tag{22}$$

Thus Eq. (19) takes the form

$$\alpha(l_1,l_2,L,r_1,r_2,f)=\sum_{kx}A(k,x)\int_0^\pi d\theta_1\sin^{L+l_1-k+1}\theta_1\times\int_0^\pi f(\rho)\sin^{L+k-l_1+1}\theta_2\cos^{l_1+l_2-L-2x}\theta_2\,d\theta_2. \tag{23}$$

This general result could also be represented by a double integral over a plane area in polar coordinates as

$$\alpha(l_1,l_2,L,r_1,r_2,f)=\sum_{kx}A(k,x)\int_0^\pi\int_c^bG(\rho,\theta_2)\times\sin^{L+k-l_1+1}\theta_2\cos^{l_1+l_2-L-2x}\theta_2\,\rho\,d\rho\,d\theta_2, \tag{24}$$

where

$$G(\rho,\theta_2)=\frac{f(\rho)}{r_1r_2\sin\theta_2}\sin^{2l_1+1}(\beta+\theta_2), \tag{25}$$

$$\beta=\cos^{-1}\{(\rho^2-r_1^2-r_2^2)/2r_1r_2\}, \tag{26}$$

$$c^2=r_1^2+r_2^2-2r_1r_2\cos\theta^2, \tag{27}$$

and

$$b^2=r_1^2+r_2^2+2r_1r_2\cos\theta_2. \tag{28}$$

We will now consider two special cases. The first is to show that Eq. (1) reduces to Talman's equation (7) when $f(R)=1$. The second is to show that the general result of Eq. (24) reduces to Eq. (9) when $\mathbf{r}_2=\mathbf{a}$. Both proofs start from Eq. (19).

When $f(R)=1$, Eq. (19) shows that the angular integrals vanish unless $k=L-l_1$ and the integer (l_1+l_2-L-2x) is even. Under these conditions the angular integrals give

$$\frac{4\pi^2 l_1!(L-l_1)!\sqrt{\pi}\,\Gamma\{\frac{1}{2}(l_1+l_2-L-2x+1)\}}{\Gamma(l_1+\frac{3}{2})\Gamma\{\frac{1}{2}(L+l_2-l_1-2x+3)\}} \tag{29}$$

and the summation over x could now be evaluated using Eq. (A1.2) of Ref. 8 as

$$\sum_x\frac{(-1)^x\Gamma(l_2-x+\frac{1}{2})}{x!\{\frac{1}{2}(l_1+l_2-L)-x\}!\Gamma\{\frac{1}{2}(L+l_2-l_12x+3)\}}=\frac{(-1)^{(1/2)(l_1+l_2-L)}\Gamma\{\frac{1}{2}(L+l_2-l_1+1)\}\{\frac{1}{2}(L-l_1-l_2)\}!}{\{\frac{1}{2}(l_1+l_2-L)\}!\Gamma\{\frac{1}{2}(L+l_2-l_1+3)\}}=\frac{2}{2L-2l_1+1}\delta_{L,l_1+l_2}, \tag{30}$$

where we have use the relation

$$\Gamma(n-z)/\Gamma(-z)=(-1)^n\Gamma(z-n+1)/\Gamma(z+1). \tag{31}$$

Substituting Eqs. (29) and (30) in Eq. (19) we get

$$\alpha(l_1,L-l_1,r_1,r_2,1)=r_1^{l_1}r_2^{L-l_1}\left\{\frac{4\pi(2L+1)!}{(2l_1+1)!(2L-2l_1+1)}\right\}^{1/2} \tag{32}$$

which when substituted in Eq. (1) reproduces Talman's equation (7).

For the second special case when $\mathbf{r}_2=\mathbf{a}$, that is $\theta_2=0$, we have

$$R=r_1^2+a^2+2ar_1\cos\theta_1. \tag{33}$$

Thus Eq. (19) takes the form

$$\alpha(l_1,l_2,L,r_1,a,f)=\sum_{kx}\frac{A(k,x)}{4\pi^2}\int_0^\pi f(R)\sin^{L+l_1-k+1}\theta_1\,d\theta_1\times\int_0^\pi\sin^{L+k-l_1+1}\theta_2\cos^{l_1+l_2-L-2x}\theta_2\,d\theta_2\times\int_0^{2\pi}\int_0^{2\pi}e^{i(L-l_1-k)(\phi_1-\phi_2)}d\phi_1\,d\phi_2. \tag{34}$$

Equation (34) again shows that the angular integrals vanish unless $k=L-l_1$ and the integer (l_1+l_2-L-2x) is even. Thus we have

$$\int_0^\pi\sin^{L+k-l_1+1}\theta_2\cos^{l_1+l_2-L-2x}\theta_2\,d\theta_2\times\int_0^{2\pi}\int_0^{2\pi}e^{i(L-l_1-k)(\phi_1-\phi_2)}d\phi_1\,d\phi_2$$
$$=4\pi^2\int_0^\pi\sin^{2L-2l_1+1}\theta_2\cos^{l_1+l_2-L-2x}\theta_2\,d\theta_2$$
$$=4\pi^2\frac{(L-l_1)!\Gamma\{\frac{1}{2}(l_1+l_2-L-2x+1)\}}{\Gamma\{\frac{1}{2}(L+l_2-l_1-2x+3)\}}. \tag{35}$$

Substituting Eqs. (20) and (35) in Eq. (34) we get

$$\alpha(l_1,l_2,L,r_1,a,f)=\left\{\frac{\pi(2l_1+1)(2l_2+1)(l_1+l_2-L)!(L+l_1-l_2)!)(L+l_1+l_2+1)!}{(L+l_2-l_1)!}\right\}^{1/2}\times\frac{r_1^{l_1}a^{L-l_1}}{2^{2l_1+1}l_1!l_1!}\int_0^\pi f(R)\sin^{2l_1+1}\theta_1\,d\theta_1\sum_x\frac{(-1)^x\Gamma(l_2-x+\frac{1}{2})}{x!\{\frac{1}{2}(l_1+l_2-L)-x\}!\Gamma\{\frac{1}{2}(L+l_2-l_1+3)\}}. \tag{36}$$

The above summation over x is the same as that evaluated by Eq. (30) and when substituted Eq. (36) immediately reproduces Eq. (9) on noting Eq. (16).

[1]P. O. Löwdin, Adv. Phys. **5**, 96 (1956).
[2]N. Suzuki, J. Math. Phys. **25**, 1133 (1984).
[3]R. R. Sharma, Phys. Rev. A **13**, 517 (1976).
[4]H. J. Silverstone and R. K. Moats, Phys. Rev. A **16**, 1731 (1977).
[5]See Eq. (14) in Ref. 3.
[6]See Eq. (2.2) in Ref. 2.
[7]J. D.Talman, *Special Functions* (Benjamin, New York, 1968).
[8]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960).

The Hamiltonian structure of a complex version of the Burgers hierarchy

M. Bruschi and O. Ragnisco
Dipartimento di Fisica, Università "La Sapienza," 00185 Rome, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Italy

(Received 29 March 1984; accepted for publication 14 December 1984)

In this paper we construct a class of integrable Hamiltonian nonlinear evolution equations generated by a purely differential recursion operator. It turns out that this hierarchy is a complex version of the Burgers hierarchy and can be linearized through a generalization of the Cole–Hopf transformation.

I. INTRODUCTION

In recent years, a number of remarkable results have been obtained in the field of integrable Hamiltonian systems with infinitely many degrees of freedom.[1] The essential property shared by all such systems is the existence of a twofold Hamiltonian structure, generated by a Poisson and a recursion (Nijenhuis) operator, well coupled to one another.[2] However, to our knowledge, all the integrable Hamiltonian evolution equations so far introduced have been shown to be associated with integrodifferential recursion operators.[3] Thus, we wondered whether there exists a hierarchy of nonlinear Hamiltonian evolution equations generated by a purely differential recursion operator. In Sec. II we exhibit a class of nonlinear evolution equations (NEE's) which is nothing but a complex version of the Burgers hierarchy, generated by a first-order differential recursion operator. In Sec. III we prove that this recursion operator is a Nijenhuis operator and that the whole class has a twofold Hamiltonian structure. In Sec. IV we exhibit a field transformation which, as in the Burgers case, allows one to linearize any equation in the hierarchy.

II. THE COMPLEX BURGERS HIERARCHY

Let us introduce the following linear differential operator:

$$\mathbf{L}\varphi := i\varphi_x + u_x\varphi, \tag{2.1}$$

where u and φ are complex functions of the real variable x (say, space), possibly parametrically dependent on a further real variable t (say, time). Moreover we assume that u_x and φ vanish rapidly enough as $|x|\to\infty$.

Then, the complex Burgers hierarchy is given by

$$u_t = h(\mathbf{L})\mathbf{S}(iu_x \exp(-i(u-\bar{u}))), \tag{2.2}$$

where $h(z)$ is an arbitrary entire function and the operator $\mathbf{S}$ is defined as

$$\mathbf{S}\alpha := i\exp(i(u-\bar{u}))\cdot\alpha. \tag{2.3}$$

Concerning formulas (2.2) and (2.3) we point out that, although the exponential factor $\exp(i(u-\bar{u}))$ cancels for any equation of the hierarchy, it plays an essential role in the definition of the operator $\mathbf{S}$, as will be shown in Sec. III.

The first equation in the hierarchy is obtained by setting $h(\mathbf{L}) = 1$ and reads

$$u_t = -u_x. \tag{2.4}$$

The next equation, corresponding to $h(\mathbf{L}) = \mathbf{L}$, is given by

$$u_t = -iu_{xx} - (u_x)^2. \tag{2.5}$$

In terms of the new fields $q = u_x$, it becomes

$$q_t = -iq_{xx} - 2qq_x. \tag{2.6}$$

Equation (2.6) is easily recognized to be a complex version of the well-known Burgers equation. On the other hand, it might be worthwhile to write down Eq. (2.5) in terms of the real and imaginary parts of $u = v + iw$:

$$v_t = w_{xx} - (v_x)^2 + (w_x)^2, \tag{2.7a}$$

$$w_t = -v_{xx} - 2v_xw_x. \tag{2.7b}$$

The next simplest equation, corresponding to $h(\mathbf{L}) = \mathbf{L}^2$, reads

$$u_t = u_{xxx} - 3iu_xu_{xx} - (u_x)^3. \tag{2.8}$$

In terms of q, it becomes

$$q_t = (q_{xx} - 3iqq_x - q^3)_x. \tag{2.9}$$

Equation (2.9) is clearly a complex version of the second equation in the Burgers hierarchy.[4,5] The corresponding system for v and w reads

$$v_t = v_{xxx} + 3(v_xw_x)_x + 3v_x(w_x)^2 - (v_x)^3, \tag{2.10a}$$

$$w_t = w_{xxx} - 3((v_x)^2 - (w_x)^2)_x/2 + (w_x)^3 - 3(v_x)^2w_x. \tag{2.10b}$$

III. THE HAMILTONIAN STRUCTURE

In this section, we will derive the Hamiltonian structure associated with the hierarchy (2.2). To this aim, we have at our disposal two different approaches: the first one,[6] mainly associated with the names of Gel'fand–Dikij, Lebedev–Manin, Kupershmidt–Wilson, and Adler, is of fairly algebraic nature, and relies on the properties of certain infinite-dimensional Lie algebras (i.e., the algebra of pseudodifferential operators of negative degree); the second one, essentially geometrical in nature, investigates directly the integrability structure as tensor fields defined on some infinite-dimensional differentiable manifolds.[3,7] We will follow the latter approach, relying extensively upon Refs. 3 and 7, both for the theoretical background and for the notations.

Accordingly, we will regard u as a point in the manifold $\mathscr{M}$ (henceforth denoted as "configuration space") given by the affine hyperplane of the Fréchet space $\mathbb{F} := C^\infty(\mathbb{R},\mathbb{C})$, formed by C^∞ complex-valued functions defined on the whole real axis and obeying preassigned asymptotic conditions. Correspondingly, the tangent space T_u, whose elements will be denoted by the last letters in the Greek alphabet, is the space of C^∞ complex-valued functions of the real variable x, obeying homogeneous asymptotic conditions.

 0022-2488/85/050943-03$02.50

The cotangent space T_u^*, whose elements will be denoted by the first letters in the Greek alphabet, is the same space as T_u, and can be put in duality with it through the nondegenerate bilinear form:

$$\langle \alpha,\varphi\rangle = \int_{-\infty}^{+\infty} dx(\bar{\alpha}\varphi + \alpha\bar{\varphi}) = 2\,\mathrm{Re}\int_{-\infty}^{+\infty} dx\,\bar{\alpha}\varphi. \qquad (3.1)$$

By direct calculations, one can easily prove the following propositions.

Proposition 1: The recursion operator **L**, which maps the tangent space into itself, is a *Nijenhuis operator*, i.e., it satisfies the following "zero-torsion"[7] condition:

$$\mathbf{L}'(\varphi;\mathbf{L}\psi) - \mathbf{L}'(\psi;\mathbf{L}\varphi) = \mathbf{L}[\mathbf{L}'(\varphi;\psi) - \mathbf{L}'(\psi;\varphi)], \qquad (3.2)$$

where by $\mathbf{L}'(\varphi;\psi)$ we denote the Gateaux derivative of the operator **L** evaluated at the point φ in the ψ direction.

Proposition 2: The operator **S** is a *Poisson operator*, i.e., it maps the cotangent space into the tangent space and fulfills the following two conditions:

$$\langle \alpha,\mathbf{S}\beta\rangle = -\langle \beta,\mathbf{S}\alpha\rangle \quad \text{(skew symmetricity)}, \qquad (3.3a)$$

$$\langle \alpha,\mathbf{S}'(\beta;\mathbf{S}\gamma)\rangle + \text{cyclic permutation} = 0. \qquad (3.3b)$$

Proposition 3: The Nijenhuis operator **L** and the Poisson operator **S** are well coupled, i.e., they fulfill the following conditions:

$$\mathbf{L}\cdot\mathbf{S} = \mathbf{S}\cdot\mathbf{L}^* \ [\mathbf{L}^* \text{ is the adjoint of } \mathbf{L} \text{ wrt } (3.1)], \qquad (3.4a)$$

$$\langle \alpha,\mathbf{L}'(\mathbf{S}\beta;\varphi) - \mathbf{L}'(\varphi;\mathbf{S}\beta)\rangle = \langle \beta,\mathbf{S}'(\alpha;\mathbf{L}\varphi) - \mathbf{L}'(\varphi;\mathbf{S}\alpha) - \mathbf{L}\mathbf{S}'(\alpha;\varphi)\rangle. \qquad (3.4b)$$

Thus, they endow $\mathscr{M}$ with a Poisson–Nijenhuis structure, or, equivalently, with a twofold Hamiltonian structure, defined by the Poisson operators **S** and $\mathbf{M} := \mathbf{L}\cdot\mathbf{S}$.

Proposition 4: The operator $Q_1(u) = iu_x \exp(i(\bar{u} - u))$ is a potential operator, i.e., it maps the configuration space into the tangent space and satisfies the condition

$$\langle Q_1'\cdot\varphi,\psi\rangle = \langle Q_1'\cdot\psi,\varphi\rangle. \qquad (3.5)$$

Moreover, it is *well-coupled* with **L**, i.e., it satisfies the condition

$$\langle \mathbf{L}^*Q_1'\cdot\varphi,\psi\rangle - \langle \mathbf{L}^*Q_1'\cdot\psi,\varphi\rangle = \langle Q_1,\mathbf{L}'(\varphi;\psi) - \mathbf{L}'(\psi;\varphi)\rangle. \qquad (3.6)$$

From Propositions 1 and 4 it follows then that all the operators

$$Q_n(u) := (\mathbf{L}^*)^{n-1}Q_1(u) \qquad (3.7)$$

are again potential operators.

Let us now, for simplicity, restrict considerations to the case $h(z) = z^n$. It is easily seen that any equation in the hierarchy (2.2) can be cast in the form

$$u_t = \mathbf{L}^{n-1}\mathbf{S}Q_1(u) = \mathbf{S}(\mathbf{L}^*)^{n-1}Q_1(u) = \mathbf{S}Q_n(u). \qquad (3.8)$$

Propositions 1–4 then entail that the following hold.

(i) The twofold Hamiltonian flows $H_n(u) := \mathbf{S}Q_n(u) = \mathbf{M}Q_{n-1}(u)$ are commuting flows, i.e., we have

$$[H_j,H_k] := H_j'\cdot H_k - H_k'\cdot H_j = 0. \qquad (3.9)$$

(ii) For any such flow, there exist an infinite set of integrals of motion $I_n[u]$, in involution with respect to the Poisson bracket,

$$\{F,G\} := \langle \nabla_u F,\mathbf{S}\nabla_u G\rangle, \qquad (3.10)$$

which, for instance, can be evaluated through the formula

$$I_n[u] = \int_0^1 d\lambda\,\langle Q_n(\lambda u),u\rangle. \qquad (3.11)$$

We report here the first integrals of motion:

$$I_1[u] = \int_{-\infty}^{+\infty} dx \exp[i(\bar{u} - u)](u_x + \bar{u}_x), \qquad (3.12a)$$

$$I_2[u] = \int_{-\infty}^{+\infty} dx \exp[i(\bar{u} - u)](\bar{u}_x u_x), \qquad (3.12b)$$

$$I_3[u] = \int_{-\infty}^{+\infty} dx \exp[i(\bar{u} - u)] \times [\bar{u}_x u_x(\bar{u}_x + u_x) - i(u_x\bar{u}_{xx} - \bar{u}_x u_{xx})]. \qquad (3.12c)$$

IV. SOLUTION THROUGH A LINEARIZING TRANSFORMATION

Let us now perform a change of local chart on our manifold, by means of the point transformation

$$r(x,t) = \exp(-iu(x,t)). \qquad (4.1)$$

We will show in this section that, through the transformation (4.1), which can be considered as a complex version of the well-known Cole–Hopf transformation, the whole class (2.2) is linearized. To this aim, we take advantage of the transformation properties implied by (4.1) on tangent and cotangent vectors and on field-dependent operators. We recall here just the results, referring to (2) for the details. For a given change of local chart in our manifold, given by

$$r = r(u), \qquad (4.2)$$

we have

$$\varphi_r = r'\cdot\varphi_u, \qquad (4.3a)$$

$$\alpha_u = (r')^*\cdot\alpha_r. \qquad (4.3b)$$

Accordingly, the potential operators obey the transformation law

$$Q(u) = (r')^*\cdot Q(r), \qquad (4.4a)$$

while for the tensor operators **L** and **S** we have

$$\mathbf{L}_r\cdot r' = r'\cdot\mathbf{L}_u, \qquad (4.4b)$$

$$\mathbf{S}_r = r'\cdot\mathbf{S}_u\cdot(r')^*. \qquad (4.4c)$$

Hence, for the transformation (4.1) we have the simple relations

$$\varphi_r = -ir\varphi_u, \qquad (4.5a)$$

$$\alpha_u = i\bar{r}\alpha_r, \qquad (4.5b)$$

$$\mathbf{L}_r = r\cdot\mathbf{L}\cdot r^{-1}, \qquad (4.5c)$$

$$\mathbf{S}_r = r\cdot\mathbf{S}_u\cdot\bar{r}. \qquad (4.5d)$$

Then, we can conclude that the hierachy of NEE's,

$$u_t = \mathbf{L}_u^{n-1}u_x, \qquad (4.6)$$

gives rise, in terms of the new field r, to the hierarchy of linear evolution equations:

$$r_t = (i\,\partial_x)^{n-1}r_x, \qquad (4.7)$$

whose Cauchy problem can be trivially solved by the Fourier transform.

For the class (4.7), the Nijenhuis operator is just "$i\,\partial_x$," and the Poisson operator is nothing but the imaginary unit "i." The integrals of motion $I_n[u]$, which, due to formulas

(3.11), (4.3a), (4.4a), are invariant with respect to any point transformation, are now given by

$$I_n[r] = \int_{-\infty}^{+\infty} dx\, \bar{r}(i\,\partial_x)^n r. \tag{4.8}$$

V. CONCLUSIONS

To end this paper, we would like to remark that, although the starting motivation for the present research was technical in nature, we obtained new completely integrable NEE's which, due to their mathematical simplicity, could be relevant for applications.

ACKNOWLEDGMENTS

This research has been partially supported by the Italian Ministry of Public Education.

[1]See, for instance, the basic works of V. E. Zakharov and L. D. Faddeev, Funct. Anal. Appl. **5**, 280 (1971); C. S. Gardner, J. Math. Phys. **12**, 1548 (1971); and V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 63 (1972).

[2]F. Magri, J. Math. Phys. **19**, 1167 (1978); A. Fokas and B. Fuchssteiner, Physica D **4**, 47 (1981).

[3]A table of the recursion operators for the more relevant integrable NEE's can be found in F. Magri, Lect. Notes Phys. **120**, 233 (1980).

[4]D. V. Choodnovsky and G. V. Choodnovsky, Nuovo Cimento B **40**, 339 (1977).

[5]D. Levi, O. Ragnisco, and M. Bruschi, Nuovo Cimento B **74** 33 (1983).

[6]I. M. Gel'Fand and L. A. Dikii, Usp. Math. Nauk. **30**, 67 (1975); D. R. Lebedev and Yu. A. Manin, "Gel'Fand–Dikii Hamiltonian operator and coadjoint representation of Volterra group" (to appear); Yu. A. Manin, J. Sov. Math. **11**, 1 (1979); B. A. Kupershmidt and G. Wilson, Invent. Math. **63**, 403 (1981); M. Adler, Invent. Math. **50**, 219 (1979).

[7]F. Magri and C. Morosi, "A geometrical characterization of integrable Hamiltonian systems through the theory of Poisson–Nijenhuis manifolds," Preprint Università di Milano (to appear).

Field equations and the tetrad connection

John R. Urani
Physics Department, University of Missouri—Kansas City, Kansas City, Missouri 64110
Frank J. Kutchko
Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland 20707

(Received 17 October 1983; accepted for publication 4 January 1985)

A fundamental result of Geroch is that a space-time admits a spinor structure if and only if it is parallelizable. A nonsymmetric, metric-compatible curvature-free connection is associated with a global orthonormal tetrad field on such a parallelizable space-time. This connection is used to examine reported inconsistencies for $S > \frac{1}{2}$ spinor field equations on general space-times. It is shown that the assumed Levi–Civita transport of Clifford units causes the inconsistencies at the Klein–Gordon stage. The relation of the torsion tensor of the parallelization connection to the space-time topology is indicated and the Lorentz covariance of the modified Klein–Gordon equations is demonstrated. A particularly simple plane-wave solution form for free-field equations is shown to result for locally flat space-times for which the torsion tensor is necessarily zero.

I. INTRODUCTION

A number of treatments of classical field theories on curved space-times exist in the mathematics and physics literature.[1–4] In particular, Weinberg[5] gives a very thorough, physical discussion of tensor and spinor wave equations on arbitrary space-times. Many other papers have appeared which presented detailed calculations of solutions to field equations on various background geometries.[6–8]

Few, if any, of these papers cite the severe limitations placed on the existence of global spinor solutions implied by the fundamental theorem of Geroch[9] that a space-time (M, g) admits a spinor structure if and only if it is parallelizable. On a space-time, this implies that the manifold admits a global C^k orthonormal (o-n) tetrad field ($k > 2$ assumed).

Recently, using global tetrad fields definable on such a parallelizable space-time, a covariant Dirac theory was reported[10] which generated natural torsion terms associated with the metric-compatible, torsional connection[11,12] $\widetilde{\nabla}$, defined by a tetrad field as a smooth section of the space-time's principal bundle of oriented o-n frames $O^+(M)$. The torsion appeared naturally, not in the dynamical Dirac equation, but rather in the associated Klein–Gordon equation for the Dirac amplitude $\psi(x)$, where a spin-torsion coupling term of the classic ECSK[13,14] form appeared.

The present paper examines the implications of the use of an o-n tetrad field to represent the Dirac algebra (or the Infeld–van der Waerden symbols in general treatments). In particular, reports of inconsistencies in Klein–Gordon equations for higher spin fields on curved space-times[3,4,6] are examined in light of the assumed Levi-Civita transport properties for the Dirac gammas and Infeld–van der Waerden symbols.

Alternate forms for Klein–Gordon (KG) equations are obtained for the Dirac and Rarita–Schwinger cases. These equations contain tetrad-torsion terms and are consistent without curvature restrictions on the space-time. The Lorentz covariance of the modified KG equations is exhibited in Sec. III.

Noting that the vanishing of $\widetilde{\nabla}$ torsion is possible only on locally flat space-times[10] for which $\nabla = \widetilde{\nabla}$; a general construction method for global solutions to free field equations is exhibited for such cases. These plane-wave-like solutions are shown to be a natural generalization of plane-wave functions on flat Lorentzian R^4.

We shall deal herein only with parallelizable space-times since they admit a spinor structure. Our signature will be $\eta = (- \ - \ - \ +)$ throughout.

II. THE DIRAC ALGEBRA AND THE TETRAD FIELD

Spinor theories are restricted to parallelizable space-times due to the fundamental result of Geroch[9] that a space-time admits a spinor structure if and only if it admits a global, o-n C^0 tetrad field. A global tetrad field $K_a(x)$ ($a = 1, 2, 3, 4$) is a smooth section of $O^+(M)$, the trivial principal bundle of oriented o-n frames for the space-time (M, g). An o-n tetrad field induces a metric-compatible connection[11,12] $\widetilde{\nabla}$ with curvature $\widetilde{R} = 0$ and torsion $\widetilde{\tau}$. Full $SO^+(3, 1)$ gauge freedom remains to smoothly vary the local tetrad. Each tetrad field invariantly defines a connection $\widetilde{\nabla}$ as an additional manifold structure. The vital topic of Lorentz covariance of spinor field equations is discussed in Sec. III.

In both classical and quantum gravity theories, "dynamical" tetrads and torsions are often discussed.[13] For example, Einstein's equations are often written in a tetrad basis and the tetrad is considered to be a dynamical variable.[13] Here, we introduce o-n tetrad fields on a predetermined space-time (M, g). Neither $\widetilde{\nabla}$ nor $\widetilde{\tau}$ affects the dynamics of fermions since *neither* enters the first-order dynamical equations of Dirac, Rarita–Schwinger, etc. The "field equations" for the tetrad fields are $\widetilde{\nabla}_\alpha K^{\mu}_{a} = 0$ since they parallel transport via their induced connection $\widetilde{\nabla}$. But, the $\widetilde{\nabla}$ connection is curvature-free (integrable). Thus, the tetrad field equations are not causal and hence the tetrad is not dynamical. The tetrad field is simply an assignment of a global set of Lorentz frames which, for example, might conveniently be taken to be the rest frames of a global observer congruence. An o-n tetrad field generally cannot satisfy causal (Levi-Civita) field equations, e.g., $\nabla_\alpha K^{\mu}_{a}(x) = 0$, $\nabla_\mu K^{\mu}_{a} = 0$, or $g^{\mu\nu}\nabla_\mu \nabla_\nu K^{\alpha}_{a} = 0$.

But, notable as exceptions are those extremely special cases where the Levi-Civita curvature is zero, which implies

 0022-2488/85/050946-05$02.50

that ∇ is integrable (in fact $\nabla = \tilde{\nabla}$ for such locally flat cases). This distinction between ∇ and $\tilde{\nabla}$ was the substance of the famous "distant parallelism" dialogue between Einstein and Cartan.[15]

The appearance of the torsion tensor $\tilde{\tau}$ in the *second*-order Klein–Gordon equation of spinor theories is crucial because it allows an examination of the typical assumptions used to construct $S > \frac{1}{2}$ spinor equations on curves space-times. The usual equations are well known to be torsion-free (by construction) and inconsistent[3,4] at the Klein–Gordon stage except on highly restricted space-times. This topic is examined later in this section and in Sec. III.

It might be thought that the appearance of a torsion tensor $\tilde{\tau}$ would be rare for general space-times. Just the opposite is the case. The vanishing of the torsion $\tilde{\tau}$ has strict topological criteria,[10] namely closed tetrad forms. This requires $M = R^4$ (perhaps with some points deleted) or a necessarily nontrivial $H^1(M)$ cohomology group. Closed tetrad forms then imply local flatness for (M, g) and vanishing torsion.

Choosing a global tetrad field to represent a spinor structure induces the $\tilde{\nabla}$ connection and its $\tilde{\tau}$ torsion as manifold structures. In Ref. 10, it was shown that the $\tilde{\tau}$ torsion appears in the Klein–Gordon equation of covariant Dirac theory through the usual local-coordinate representation $\gamma^\mu(x)$ of the Dirac units. The torsion $\tilde{\tau}$ couples naturally to spin in the second-order equations. The dynamical Dirac, Rarita–Schwinger, and other first-order spinor equations are unaltered. They contain no torsion terms, which was expected from the flat-R^4 case.

The classical geometrical tetrad field $K_a(x)$ is used to expand the metric as $g(x) = \eta_{ab} K^{-1a}(x) \otimes K^{-1b}(x)$, where $K^{-1a}(x) = \omega^a(x)$ denotes the global tetrad one-forms. Thus the metric satisfies the transport laws $\nabla g = 0$ *and* $\tilde{\nabla} g = 0$ and the tetrad one-forms also self-transport via $\tilde{\nabla}$ according to $\tilde{\nabla}_\alpha K_\mu^{-1a}(x) = 0$.

The explicit $\tilde{\Gamma}$ connection coefficients in local coordinates[10] are

$$\tilde{\Gamma}^\mu_{\nu\alpha} = K^\mu_a(x)\partial_\nu K_\alpha^{-1a}(x).$$

The torsion $\tilde{\tau} = \tilde{\Gamma} - \tilde{\Gamma}^T$, invariantly expressed as $\tilde{\tau} = K_a \otimes dK^{-1a}$, vanishes only on locally flat space-times[10] for which each dual field may be taken to be closed $[dK^{-1a}(x) = 0]$. This is directly seen using the directional derivative relation

$$\tilde{\nabla}_{K_a} K_b - \tilde{\nabla}_{K_b} K_a = 0 = \tilde{\tau}(K_a, K_b) + [K_a, K_b],$$

which gives globally vanishing torsion if and only if the tetrad vector fields commute globally. In those cases, the tetrad fields are expressible on a local-coordinate chart as simple coordinate gradients. The dual one-forms would then all be closed globally and local flatness follows.

Expressing the torsion in tetrad components identifies the torsion as the tetrad Lie structure functions. Using the Lie bracket $[K_a(x), K_b(x)] = f^c_{ab}(x) K_c(x)$ and an inner product with $K^{-1d}(x)$ gives that $\tilde{\tau}^d_{ab}(x) = f^d_{ba}(x)$. The global vanishing of torsion $\tilde{\tau}$ requires the global vanishing of the tetrad Lie structure functions.

The connection $\tilde{\nabla}$ is related to the unique symmetric Levi–Civita connection ∇ by the local coordinate relation[10]

$$\Gamma^\mu_{\nu\alpha} = \tilde{\Gamma}^\mu_{(\nu\alpha)} + g^{\mu\lambda} g_{\theta(\nu} \tilde{\tau}^\theta_{\alpha)\lambda}. \tag{2.1}$$

Equation (2.1) is merely the necessary ECSK relation[13,14] for any metric-compatible, torsional connection, namely $\tilde{\Gamma} = \Gamma + \tilde{\Delta}$, where the contorsion tensor $\tilde{\Delta}$ is a (1, 2) tensor made up of the torsion plus a symmetrized torsion part.

The connections ∇ and $\tilde{\nabla}$ are equal only when the torsion[16] is zero in Eq. (2.1). This requires the vanishing of Levi-Civita curvature (since then $R = \tilde{R} = 0$) and again (M, g) must be locally flat.

Equation (2.1) may be contracted to yield $\Gamma^\mu_{\alpha\mu} = \tilde{\Gamma}^\mu_{\alpha\mu}$ along with

$$g^{\mu\nu}\Gamma^\alpha_{\mu\nu} = g^{\mu\nu}\tilde{\Gamma}^\alpha_{\mu\nu} + g^{\alpha\mu}\tilde{\tau}^\nu_{\nu\mu}. \tag{2.2}$$

Several relations ($\Gamma^\mu_{\mu\alpha} = \tilde{\Gamma}^\mu_{\mu\alpha}$ and $\tilde{\tau}^\mu_{\mu\alpha} = 0$) in Ref. 10 were separately incorrect. The correct combined relation is

$$\tilde{\Gamma}^\mu_{\mu\alpha} = \Gamma^\mu_{\mu\alpha} + \tilde{\tau}^\mu_{\mu\alpha},$$

with $\tilde{\tau}^\mu_{\mu\alpha} \neq 0$ in general.

Again, the connection $\tilde{\nabla}$ is metric compatible[11,12] since $\tilde{\nabla} g = 0$. The $\tilde{\Gamma}$ definition is an integrability condition for $\tilde{\nabla}$. Thus $\tilde{\nabla}$ is curvature-free ($\tilde{R} = 0$) and hence is not a causal connection.

But, surprisingly, the connection $\tilde{\nabla}$ enters physics through spinor theories because Clifford units are needed. For example, in Dirac theory, to form the generally invariant vector field operator $\gamma^\mu(x)\partial_\mu$, the constant Dirac units γ^a are represented in a local coordinate chart via the expression $\gamma^\mu(x) = K^\mu_a(x)\gamma^a$. The Dirac operator then takes the elegant form

$$\gamma^\mu(x)\partial_\mu = \gamma^a K^\mu_a(x)\partial_\mu = \gamma^a K_a(x),$$

in the global set of o-n bases defined by a global tetrad field.

The $\gamma^\mu(x)$ satisfy the transport law $\tilde{\nabla}_\alpha \gamma^\mu(x) = 0$ since the tetrad fields all satisfy $\tilde{\nabla} K_a = 0$. In local coordinates this reads

$$\tilde{\nabla}_\alpha \gamma^\mu(x) = \tilde{\nabla}_\alpha K^\mu_a \gamma^a = (\partial_\alpha K^\mu_a + \tilde{\Gamma}^\mu_{\alpha\lambda} K^\lambda_a)\gamma^a = 0. \tag{2.3}$$

It is clear why the Dirac units should exhibit this behavior. The Dirac γ^a elements (and the entire algebra generated from them) may be assigned to each point $x \in M$ since the tangent bundle of the parallelizable manifold is trivially $M \times R^4$. All 16 algebra basis elements are $SO^+(3, 1)$ invariant. They do not depend on the local tetrad basis $K_a(x)$ chosen at any $x \in M$. However, the γ^a, represented in local coordinate charts as $\gamma^\mu(x) = K^\mu_a(x)\gamma^a$, express this constancy of γ^a, $a = 1, 2, 3, 4$ over M via the $\tilde{\nabla}$ transport law of Eq. (2.3).

Similarly, the Infeld–van der Waerden symbols σ^a and σ_a are intrinsically defined relative to o-n frames. They are assignable, as are the Dirac units, as constants over M. However, when these symbols are represented using local coordinate charts, they are expressed as $\sigma_\mu(x) = \sigma_a K_\mu^{-1a}(x)$ and $\sigma^\mu(x) = \sigma^a K^\mu_a(x)$. Thus, as in the Dirac unit case, the constancy of these symbols over M is expressed via the transport laws $\tilde{\nabla}_\alpha \sigma^\mu(x) = 0$ and $\tilde{\nabla}_\alpha \sigma_\mu(x) = 0$. In any o-n tetrad basis, the basic first-order differential operators are $\sigma^a K_a(x) = \sigma^a K^\mu_a(x)\partial_\mu$ for spin-$\frac{1}{2}$ or $\sigma^a K^\mu_a(x)\nabla_\mu$ for higher-rank spinors. Since the tetrad field operators do not in general commute, the second-order Klein–Gordon equations for arbitrary-rank spinor fields will contain torsion terms arising from the tetrad commutators. Thus, both Dirac-type and Penrose-type spinor equations will contain torsion terms at the Klein–Gordon stage.

It has been reported[3,4] that $S > \frac{1}{2}$ spinor field equations cannot be consistent except on restricted space-times such as Einstein[3,6] spaces or those with constant sectional (Riemann) curvature.[4] It is notable that these results occur when two standard assumptions are made. The first assumption is that all coordinate derivatives ∂_μ should be replaced by Levi-Civita covariant derivatives ∇_μ, which is correct for spin-$\frac{3}{2}$, $\frac{5}{2}$, ... spinor fields.[17]

The second assumption is that the Clifford units or Infeld–van der Waerden symbols[18] always parallel transport via the Levi-Civita ∇ connection.[3,4,6] It has been shown[10] that (a) the integrable tetrad transport law $\widetilde{\nabla}_\alpha \gamma^\mu(x) = 0$ is appropriate for the constant Dirac units, etc., and (b) $\nabla = \widetilde{\nabla}$ iff the space-time is parallelizable *and* locally flat.

If the Dirac units (or Infeld–van der Waerden symbols) are assumed to Levi-Civita transport globally, thereby arguing away the torsion terms in the Klein–Gordon equation, it is equivalent to assuming the integrability of the Levi–Civita connection. For such cases the torsion is zero but the space-times are locally flat and hardly very interesting. Actually, the restriction claims[3,4,6] are not without a modicum of truth since curvature-free, locally flat space-times are trivial examples of Einstein spaces of constant (zero) sectional curvature.

Hence, we find that the usual inconsistency claims[3,4] are the result of assuming the Levi-Civita transport of the Clifford units, etc., for spinor field equations. Many authors[11–13] have discussed the connection associated with o-n tetrad fields. However, the $\widetilde{\nabla}$ transport of the various algebraic units and the importance of that transport for spinor field equations had not previously been reported.

III. DIRAC AND RARITA–SCHWINGER EQUATIONS

The Dirac equation, expressed relative to a given global o-n tetrad field $K_a(x)$, $a = 1, 2, 3, 4$ is

$$(\gamma^a K_a(x) + im)\psi(x) = H(x).$$

Here γ^a are the constant Dirac matrices in any representation, $\psi(x)$ is the (general scalar/Lorentz spinor) amplitude, and $H(x)$ is a source term. In a local coordinate chart, $K_a(x) = K^\mu_a(x)\partial_\mu$ and $\gamma^a K_a(x) = \gamma^\mu(x)\partial_\mu$ in analogy with the flat-R^4 case. It should be noted that this is the usual curved-space-time Dirac equation.[17] The connection $\widetilde{\nabla}$ does not enter the first-order dynamical Dirac equation.

But, by operating with $(\gamma^a K_a(x) - im)$ and defining $(\gamma^a K_a - im)H = S$, the associated Klein–Gordon equation[10]

$$(g^{\mu\nu}\nabla_\mu \nabla_\nu + m^2)\psi + g^{\mu\nu}\tilde{\tau}^\lambda_{\mu\lambda}\,\partial_\nu \psi - (i/2)\sigma^{\mu\nu}(x)\tilde{\tau}^\alpha_{\mu\nu}\,\partial_\alpha \psi = S$$

is obtained, where $\sigma^{\mu\nu}(x) = (-i/2)\,[\gamma^\mu(x), \gamma^\mu(x)]$. The usual Klein–Gordon equation is reproduced along with two torsion terms, one a trace term and another term coupling spin to the torsion. In locally flat cases, both torsion terms vanish as in the flat R^4 case. The Lorentz covariance of the KG equation is treated at the end of this section.

The Dirac amplitude $\psi(x)$ is a cross section of the vector bundle[19] ${}_G C^4 \times O^+(M)$ where $G = D^{(1/2,0)} \otimes D^{(0,1/2)}$. The spin-$\frac{3}{2}$ Rarita–Schwinger amplitude $\psi_\alpha(x)$ (with spinor index suppressed is a cross section of the bundle ${}_G C^4 \times T^*(M) \times O^+(M)$ which is subject to the subsidiary conditions $\gamma^\mu(x)\psi_\mu(x) = 0$ and $g^{\mu\nu}(x)\nabla_\mu \psi_\nu(x) = 0$ in local coordinates.

The first-order Rarita–Schwinger equation is then

$$(\gamma^\mu(x)\nabla_\mu + im)\psi_\alpha = H_\alpha, \tag{3.1}$$

where H_α is a suitable source term.

Generating the associated Klein–Gordon equation by operating on Eq. (3.1) with $(\gamma^\nu(x)\nabla_\nu - im) = D^*$ and defining $D^*H = S$ we obtain

$$(g^{\mu\nu}\nabla_\mu \nabla_\nu + m^2)\psi_\alpha + i\sigma^{\mu\nu}\nabla_\mu \nabla_\nu \psi_\alpha + (g^{\nu\beta} + i\sigma^{\nu\beta}(x))$$

$$\times(\Gamma^\mu_{\nu\beta} - \widetilde{\Gamma}^\mu_{\nu\beta})\nabla_\mu \psi_\alpha = S_\alpha.$$

Using Eq. (2.2) and the definition of $\tilde{\tau}$, the Klein–Gordon equation becomes

$$(g^{\mu\nu}\nabla_\mu \nabla_\nu + m^2)\psi_\alpha - (i/2)\sigma^{\mu\nu}(\tilde{\tau}^\beta_{\mu\nu}\nabla_\beta \psi_\alpha)$$

$$+ (i/4)\sigma^{\mu\nu}R^{\ \lambda}_{\alpha\mu\nu}\psi_\lambda + g^{\mu\theta}\tilde{\tau}^\lambda_{\theta\lambda}\nabla_\mu \psi_\alpha = S_\alpha.$$

Again, we see the spin-torsion coupling and the torsion-trace terms (as for spin-$\frac{1}{2}$) along with the expected coupling of spin and the curvature tensor R (of ∇). The torsion terms vanish only in locally flat cases as in the spin-$\frac{1}{2}$ case. We see that for the Dirac case, the Rarita–Schwinger case, and for any spinor wave equation of the first order, the tetrad connection modifies the standard (inconsistent) Klein–Gordon equation by the addition of torsion terms. The standard first-order equations for fermion dynamics are unaltered in all cases.

A question of great importance is the Lorentz covariance of the results obtained in this section. The first impression is a pessimistic one. Because no natural parallelization exists in general, each o-n tetrad field choice would seem to give a new Klein-Gordon equation with a new torsion tensor for any given spin. This seems obvious because a change of o-n tetrad field via a pointwise Lorentz transformation $K'_b(x) = \Lambda_b^{-1a}(x)K_a(x)$ takes $\tilde{\tau} = K_a \otimes dK^{-1a}$ to a new torsion tensor $\tilde{\tau}' = K_a \otimes K^{-1a} + \Lambda_b^{-1a}K_a \otimes d\Lambda^b_c \wedge K^{-1c}$. However, the torsion terms in the above KG equations were generated naturally by the transport laws for the Dirac gammas. We must start with the Dirac operator and *generate* a new KG equation. Due to their Lorentz invariance, the Dirac units satisfy $L\Lambda^\mu_\nu \gamma^\nu L^{-1} = \gamma^\mu$, where L is an element of $D^{(1/2,0)} \otimes D^{(0,1/2)}$. The Dirac operator $\gamma^\mu(x)\partial_\mu$ relative to another tetrad choice is simply $K\gamma^\mu(x)L^{-1}\partial_\mu$ because the tensor parts are contracted. The Dirac operator is purely spinor in nature. Thus, when the KG equation is formed for the Dirac or Rarita–Schwinger cases, only derivatives of the L matrices will appear. These terms are absorbed in the standard way[5] into the covariant derivative by defining the "spin connection." The torsion term then transforms as a spinor under a change of tetrad field. This unexpected result is actually just the built-in spinor covariance, but tensor invariance, of the Dirac operator, which does its own bookkeeping, always resulting in purely spinor Lorentz covariant equations. Similar arguments[20] give the same result for general spinors based on the Infeld–van der Waerden symbols which satisfy the Lorentz invariance relation $L\Lambda_a^{-1b}\sigma^b L^{*T} = \sigma_a$, where L and L^* are elements of the $D^{(1/2,0)}$ and $D^{(0,1/2)}$ representations of SL(2, C). The Lorentz spinor and general scalar nature of the operators $\sigma^a K^\mu_a \nabla_\mu$ makes the KG torsion terms Lorentz covariant as purely spinor quantities.

IV. FREE-FIELD SOLUTIONS AND TORSION

It was shown[10] that the torsion associated with a global o-n tetrad $K_a(x)$, $a = 1, 2, 3, 4$ on a parallelizable space-time is $\tilde{\tau} = K_a \otimes dK^{-1a}$ since $\tilde{\tau}^\mu_{\nu a} = K^\mu_a(\partial_\nu K_a^{-1a} - \partial_a K_\nu^{-1a})$ in local coordinates. It is clear that in order to have vanishing torsion and $\nabla = \tilde{\nabla}$, the linear independence of the $K_a(x)$ fields implies that all four tetrad one-forms must be closed, namely $dK^{-1a} = 0$. Were all those one-forms exact, we would have[10] flat Lorentzian R^4 (perhaps with some points deleted). Consequently, for vanishing torsion on a general manifold, we must have *at least* one closed, inexact tetrad one-form. The space-time manifold must then have a nontrivial cohomology group $H^1(M) \neq 0$.

Torsion $\tilde{\tau}$ and curvature $\tilde{R}$ both zero implies that (M, g) is locally flat, which is consistent with the tetrad one-forms being locally exact (i.e., closed). Many space-times are parallelizable[9] but not locally flat. The Friedman–Robertson–Walker (FRW) space-times with topologies $R^1 \times R^2$ (open) and $R^1 \times S^3$ (closed) are all simply connected $[\pi_1(M) = 0]$ which implies $H^1(M) = 0$ in these cases. Similarly, Schwarzschild (exterior) space-time has topology $R^1 \times (R^3 - \{0\})$ so $\pi_1(M) = 0$ and $H^1(M) = 0$. Torsion $\tilde{\tau}$ is then necessarily nonzero in these cases.

In the usual (r, θ, ϕ, ct) coordinates, the torsion for the exterior Schwarzschild case has the few nonzero components $\tilde{\tau}^4_{14} = -\tilde{\tau}^4_{41} = m^*/(r^2 - 2m^* r)$, $\tilde{\tau}^2_{12} = -\tilde{\tau}^2_{21} = [1 - (r/(r - 2m^*))^{1/2}]/r$, and $\tilde{\tau}^3_{13} = -\tilde{\tau}^3_{31} = \tilde{\tau}^2_{12}$ with $m^* = Gm/c^2$.

In locally flat cases, very simple solutions exist for free-field Klein–Gordon, Dirac, Maxwell, etc., equations. The vanishing of the tetrad torsion $\tilde{\tau}$ gives $\nabla = \tilde{\nabla}$. Consequently, tetrad vector fields are also ∇ geodesic vector fields.

One may then cover M with an atlas of charts such that $M = \bigcup_j (V_j)$ where each V_j is the open domain of coordinates y^a_j, $a = 1, 2, 3, 4$ for each chart. The tetrad one-forms being closed implies that the coordinate functions on each chart may be chosen such that $dy^a_j = K^{-1a}$ on each chart. In local coordinates, this is expressible as

$$dy^a_j = \partial_\mu y^a_j \, dx^\mu = K_\mu^{-1a} \, dx^\mu .$$

In the atlas $\bigcup_j \{y^a_j, V_j\}$ itself, $\partial_a = \partial/y^a_j$ on each chart and thus $K_\mu^{-1a} = \delta^a_\mu$. In the locally flat case, the tetrad forms are harmonic since $d(K^{-1a}) = 0$ (closed) and $*d*K^{-1a} = 0$ (coclosed) since coclosed implies $\nabla_\mu K^\mu_a(x) = 0$. Recall that linear combinations $H(x) = k^a K_a(x)$, with constant coefficients k^a, satisfy $\nabla_\mu H(x) = 0$ in locally flat cases.[16] These vector fields have a global casual nature since $g(H, H) = \eta(k, k)$ is constant over M.

Plane-wave *functions* are then of the form $P(x) = \exp(ik_a y^a_j(x))$ on each chart domain V_j. See Ref. 21 for plane-wave space-times.

Note that these functions are maps from $R^4 \times M$ into C^1 rather than the topological exponential map. Clearly, $dP(x) = ik_a \, dy^a_j(x) P(x)$ on each chart V_j. In local coordinates,

$$dP(x) = ik_a \, \partial_\mu y^a_j(x) dx^\mu \, P(x) = i\hat{H}(x) P(x),$$

where $\hat{H}(x) = k_a K^{-1a}(x)$ is dual to $H(x) = k^a K_a(x)$. Clearly $d(dP(x)) = 0$ implies $d\hat{H}(x) = 0$.

The Hodge dual three-form $*dP(x)$ is $*dP(x) = iP(x)(*\hat{H}(x))$ and then $d*dP(x) = idP(x) \Lambda *\hat{H}(x) + iP(x) d*\hat{H}(x) = idP(x) \Lambda *\hat{H}(x)$ since $d*\hat{H}(x) = 0$ (equivalently $\nabla_\mu K^\mu_a = 0 = \tilde{\nabla}_\mu K^\mu_a$).

In local coordinates, we may write the above results in a form very much like special relativity. We have

$$\nabla_\mu P(x) = \partial_\mu P(x) = ik_a K_\mu^{-1a}(x) P(x),$$

since $P(x)$ is a scalar, and

$$\nabla_\nu \nabla_\mu P(x) = ik_a K_\mu^{-1a}(x)_{;\nu} P(x) - k_a k_b K_\mu^{-1a}(x) K_\nu^{-1b}(x) P(x).$$

But, $K_{\mu;\nu}^{-1a} = 0$ if $\nabla = \tilde{\nabla}$; thus we have

$$g^{\nu\mu} \nabla_\nu \nabla_\mu P(x) = -\eta(k, k) P(x).$$

We may then solve various free-field equations utilizing these $P(x)$ functions.

The scalar Klein–Gordon equation is trivially solved by $f(x) = qP(x)$, where $q \in C^1$ and $\eta(k, k) = m^2$ for the particle in question. Clearly, $H(x) = k^a K_a(x)$ is a null vector field if $m^2 = 0$.

The free Dirac equation $(\gamma^a K_a(x) + im)\psi(x) = 0$ is solved by $\psi(x) = \chi(k) P(x)$ with $\chi(x)$ the usual Dirac polarization vector in C^4 satisfying $(\gamma^a k_a + m)\chi(k) = 0$ from the Dirac equation above. The Klein–Gordon equation gives $\eta(k, k) = m^2$, as expected.

Maxwell free-field equations are similarly solved with solutions for the potential one-form $A(x)$ given by $A(x) = \hat{A}(k, x) P(x)$, where $\hat{A}(k, x) = \hat{A}_b(k) K^{-1b}(x)$ is a closed one-form since $\hat{A}_b(k)$ is a constant Lorentz four-vector parametrized by k_a. The field two-form is $F = dA = (d\hat{A})P - \hat{A} \Lambda dP = -\hat{A} \Lambda dP$. Trivially, $ddF = 0$, which gives two of Maxwell's equations.

The remaining free-field Maxwell equations result from $*d*F = 0$. Written out in local coordinates, these equations are $\nabla_\mu F^{\mu\nu} = 0$. The contravariant field tensor $F^{\mu\nu}$ is explicitly

$$F^{\mu\nu} = -i\hat{A}^b k^a (K^\mu_b K^\nu_a - K^\nu_b K^\mu_a) P(x).$$

We have (with $\nabla = \tilde{\nabla}$) that $\nabla_\nu K^\mu_a = 0$, $\nabla_\mu K^\mu_b = 0$, etc. This yields

$$\nabla_\mu F^{\mu\nu} = -i\hat{A}^b k^a (K^\mu_b K^\nu_a - K^\nu_b K^\mu_a) ik_c K_\mu^{-1c} P = (\eta(\hat{A}, k) k^a K^\nu_a - \eta(k, k) \hat{A}^b K^\nu_b) P.$$

A free-field solution is obtained if we require the flat-R^4 conditions $\eta(A, k) = \eta(k, k) = 0$. The solution obtained is then based, as expected on the global, null wave-number field $H(x) = k_a K^a(x)$.

Analogous solutions to the free Rarita–Schwinger equation are also easily constructed by the methods described above.

V. CONCLUSIONS

The connection $\tilde{\nabla}$ associated with a global o-n tetrad field has been recognized as metric compatible[11,12] for some time. Its appearance in the physical setting of Dirac theory has occured only recently.[10] This work was based strongly on the Geroch parallelizability criterion.

The trivial nature of the tangent, cotangent, linear frame, and o-n frame bundles of parallelizable space-times is very useful. Constant nonzero cross sections can be constructed for the tangent bundle, for example. The set of constant Dirac γ^a units may be assigned globally for this reason. Because the γ and σ elements (which are used in all spinor theories) are globally constant and are thus independent of the Levi–Civita geodesic spray, the γ and σ elements must transport using $\tilde{\nabla}$ rather than the ∇ of Levi-Civita. Consequently, the inconsistencies in higher spin Klein–Gordon equations[3,4,6] result due to the commonly assumed Levi-Civita transport of the Dirac and Infeld–van der Waerden symbols.

Using the purely spinor nature of the differential operators in spinor theories, it was also possible to show the rather subtle Lorentz covariance of the modified Klein–Gordon equations containing torsion.

Finally, a particularly simple plane-wave solution form was found for free-field equations on locally-flat space-times for which tetrad torsion necessarily vanishes and the two covariant derivatives coincide.

ACKNOWLEDGMENTS

J. R. U. would like to thank Ron Carlson for torsion computations and John Beem and Brian DeFacio for very valuable discussions.

[1]F. G. Friedlander, *The Wave Equation on a Curved Spacetime* (Cambridge U. P., Cambridge, 1975).

[2]G. W. Gibbons, in *General Relativity—An Einstein Centenary Survey*, edited by S. W. Hawking and I. Israel (Cambridge U. P., Cambridge, 1979), Chap. 13.

[3]H. Buchdahl, Nuovo Cimento **10**, 96 (1958); **25**, 486 (1962). Also, Classical Quantum Gravit. **1**, 189 (1984); J. Phys. A: Math. Nucl. Gen. **15**, 1 (1982); **15**, 1057 (1982).

[4]A. Fordy, Gen. Relativ. Gravit. **98**, 387 (1978).

[5]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), Chap. 12.

[6]J. Dowker, Phys. Rev. D **28**, 3013 (1983). Here as in many other papers dealing with the reported field equation inconsistencies, the proper Lorentz generators $\sigma^{\mu\nu}(x)$ should be used in a local-coordinate basis rather than the constant tetrad-basis generators σ^{ab} (labeled $\sigma^{\mu\nu}$). Also the local coordinate gradient $\nabla_\mu = \partial_\mu$ for the spin-$\frac{1}{2}$ case.

[7]S. Chandrasekhar, Ref. 2, Chap. 7.

[8]J. Centrella and R. A. Matzner, Phys. Rev. D **25**, 930 (1982).

[9]R. Geroch, J. Math. Phys. **9**, 1739 (1968); J. Math. Phys. **11**, 343 (1970).

[10]J. R. Urani and G. P. Barker, J. Math. Phys. **24**, 2407 (1983).

[11]C. T. J. Dodson, *Categories, Bundles and Spacetimes Topology* (Shiva, Orpington, England 1980), p. 149.

[12]R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds* (Dover, New York, 1980), p. 247.

[13]S. Deser and C. J. Isham, Phys. Rev. D **14**, 2505 (1976). See also, S. B. Edgar, Gen. Relativ. Gravit. **12**, 347 (1980).

[14]F. Heyl, P. von der Heyde, and G. Kerlick, Rev. Mod. Phys. **48**, 393 (1976).

[15]A. Pais, *Subtle is the Lord*... (Oxford, New York, 1982).

[16]W. Poor, *Differential Geometric Structures* (McGraw–Hill, New York, 1981), p. 99.

[17]The Dirac field is a general scalar and a Lorentz spinor, thus $\nabla_\mu \psi = \partial_\mu \psi$. A change of tetrad requires $D_\mu = \partial_\mu + \Gamma_\mu$ (Ref. 5).

[18]C. Misner, K. Thorne, and J. Wheelar, *Gravitation* (Freeman, San Francisco, 1973).

[19]S. Avis and C. Isham, *Recent Developments in Gravitation. Cargese 1978*, edited by M. Levy, and S. Deser (Plenum, New York, 1979).

[20]J. Urani, Phys. Rev. D (submitted).

[21]S. Hawking and G. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, 1973).

Numerical integration in many dimensions. I

Charles Schwartz
Department of Physics, University of California, Berkeley, California 94720

(Received 13 June 1984; accepted for publication 21 September 1984)

If a d-dimensional integral involves an integrand of the functional form $F(f_1(x_1)+f_2(x_2)+\cdots)$, then one can introduce an integral transform (Fourier or Laplace or variants on those) which allows all the integrals over the coordinates x_i to factor. Thus a d-dimensional integral is reduced to a one-dimensional integral over the transform variable. This is shown to be a very powerful and practical numerical approach to a number of problems of interest. Among the examples studied is the computation of the volume of phase space for an arbitrary collection of relativistic particles. One important aspect of the approach involves numerical integration along various contours in the complex plane.

I. INTRODUCTION

If an accurate numerical evaluation of a one-dimensional integral requires n points on a lattice, then, according to conventional wisdom, one will require n^d lattice points to evaluate a similar d-dimensional integral. This number n^d grows so rapidly as d increases that such a direct approach becomes prohibitive. Thus there has been great interest in Monte Carlo and related methods that appear to be independent of the number of dimensions. This paper reports an attempt to turn against this tide and to find some analytically based schemes for multidimensional integration that have high accuracy and systematic improvement with considerably less than n^d operations.

In Sec. II, I consider integrals that involve a function (or a few functions) of the form $F(f_1(x_1)+f_2(x_2)+\cdots+f_d(x_d))$. An integral transform is used to reduce the d-dimensional integral to a one-dimensional (or a few-dimensional) integral over the transform variable(s). An interesting aspect of this method is that one often ends up integrating numerically over some contour in the complex plane; and some examples show that this can be done quite nicely. This method is applied to computation of the relativistically invariant phase space volume for any number of particles with arbitrary masses and some total energy specified in Sec. III.

In Sec. IV, I consider integrands F whose argument is a product, rather than a sum, of functions of the different variables. Here a Mellin transform does the trick; and some further examples are given.

The philosophy guiding this work is not that one should expect a universal rule good for all types of functions. Rather, the aim is to develop a variety of techniques, each one powerful for certain classes of functions. Then, either through analysis or by trial and error, one can seek the procedure most efficient for any given problem. While the particular type of function studied in this paper may seem very special, it appears to be the most commonly encountered in studies of multidimensional integrals and is familiar in many physics problems.

In the following paper[1] two very different new techniques for multidimensional integration are presented.

II. THE TRANSFORM METHOD

Consider integrals of the form

$$I=\left(\prod_{i=1}^{d}\int g_i(x_i)dx_i\right)F\left(\sum_{i=1}^{d}f_i(x_i)\right). \tag{1}$$

Assume that we can find a suitable integral transform for the function F:

$$F(s)=\int d\sigma\,\widehat{F}(\sigma)e^{su(\sigma)}; \tag{2}$$

then the original integral becomes

$$I=\int d\sigma\,\widehat{F}(\sigma)\prod_{i=1}^{d}w_i(\sigma), \tag{3}$$

where

$$w_i(\sigma)=\int g_i(x)dx\,e^{f_i(x)u(\sigma)}. \tag{4}$$

Thus we have replaced a d-dimensional integral by $(d+1)$ one-dimensional integrals. This implies a great economy: from n^d to n^2d operations.

The choice of the integral representation (2) will depend on the nature of the function F and the range of the variables. Some examples:

$$\theta(s)s^{p-1}=\int_{C-i\infty}^{C+i\infty}\frac{d\sigma}{2\pi i}\frac{\Gamma(p)}{\sigma^p}e^{s\sigma},\quad p>0; \tag{5}$$

$$s^{-p}=\int_0^\infty d\sigma\frac{\sigma^{p-1}}{\Gamma(p)}e^{-s\sigma},\quad p>0,\quad s>0. \tag{6}$$

These can be used as in the following d-dimensional integrals:

$$\left(\prod_{i=1}^{d}\int_0^\infty dx_i\,x_i^{\alpha_i-1}e^{-\beta_i x_i}\right)\left(\sum_{i=1}^{d}x_i+\gamma\right)^{p-1} =\int_{C-i\infty}^{C+i\infty}\frac{d\sigma}{2\pi i}\frac{\rho(p)}{\sigma^p}e^{\gamma\sigma}\prod_{i=1}^{d}\frac{\Gamma(\alpha_i)}{(\beta_i-\sigma)^{\alpha_i}},\quad 0<C<\beta_i \tag{7}$$

and

$$\left(\prod_{i=1}^{d}\int_0^1 dx_i\right)\left(\sum_{i=1}^{d}a_ix_i+a_0\right)^{-p} =\int_0^\infty d\sigma\frac{\sigma^{p-1}}{\Gamma(p)}e^{-\sigma a_0}\prod_{i=1}^{d}\left(\frac{1-e^{-a_i\sigma}}{a_i\sigma}\right). \tag{8}$$

After one has chosen an integral transform the final task is to select a good contour and to have a reliable scheme for numerical integration. It is well known that one can make this last task terribly hard by choosing an unfortunate contour for integration—one where the function is very large and rapidly oscillating so that numerical accuracy is rapidly lost. However, (as is perhaps less well known) con-

 0022-2488/85/050951-04$02.50

tour integration in the complex plane can also be a very easy and well-behaved problem.

For a test I chose the representation of the gamma function given in (5):

$$\frac{1}{\Gamma(p)} = \int_{C-i\infty}^{C+i\infty} \frac{d\sigma}{2\pi i}\frac{e^{\sigma}}{\sigma^{p}}. \tag{9}$$

If you look at the integrand along the positive real axis, you see that it has a minimum at $\sigma = p$. Furthermore, if σ can go to infinity in the left half-plane, then the exponential will decay rapidly. So I chose the contour

$$\sigma = p + 1 - \cosh x + i \sinh x, \quad -\infty < x < \infty. \tag{10}$$

Finally, for the infinite integration over x, I use the simple rule

$$\int_{-\infty}^{\infty} f(x)dx \cong h \sum_{n=-\infty}^{\infty} f(x_0 + nh), \tag{11}$$

which will generally have an error decreasing exponentially fast with decreasing h, for analytic functions f (see Ref. 2). Results of this computation are given in Table I; they look quite satisfactory. I redid the computation with the alternative contour

$$\sigma = p + i \sinh x \tag{12}$$

and found that the results were about the same for large values of p but for smaller p the integration required more points to be taken for the same accuracy. (For $p = 2$ it did not work at all.) This carries the interesting lesson that some problems may get *easier* as the number of dimensions gets larger: Note the increasing number of powers of σ in the denominators of (7) and (8) as d increases; and it is this large negative exponent that helps make the final integral converge rapidly in a small domain.

TABLE I. Numerical integration for the gamma function $\Gamma(p)$. Results from the equation $1/\Gamma(p) = \int_{C-i\infty}^{C+i\infty} (d\sigma/2\pi i)(e^{\sigma}/\sigma^{p})$, with integration along the contour $\sigma = p + 1 - \cosh x + i \sinh x$ are given. The trapezoidal rule was used for integration in x, terminating when the added terms were less than one part in 10^7. The machine was accurate to six decimal figures of arithmetic. The interval h was started at 1.0 and then successively halved. The numbers in parentheses give the number of integration points used at each value of h. This could have been reduced by half by using the symmetry in x. The dot under each number indicates the place after which it ceases to be accurate.

$p = 2$		$p = 16$		$\times 10^{-13}$
0.993 299	(9)	0.111 804	(9)	
1.000 461	(15)	0.128 573	(17)	
1.000 002	(29)	0.130 805	(30)	
		0.130 767	(57)	
$p = 4$		$p = 32$		$\times 10^{-34}$
6.572 09	(9)	1.851 694	(9)	
6.006 62	(15)	0.918 359	(17)	
6.000 00	(27)	0.820 985	(33)	
		0.822 279	(63)	
$p = 8$		$p = 64$		$\times 10^{-88}$
4605.07	(9)	0.145 196	(11)	
4985.79	(15)	0.208 250	(19)	
5039.94	(28)	0.198 967	(37)	
5040.00	(53)	0.198 261	(71)	

Sometimes this integral transform technique allows one to express a complicated-looking multidimensional integral in closed form. For example, in Eq. (7), if p should be an integer or if the numbers α_i are integers, then one can write the answer in terms of residues at the poles at $\sigma = 0$ or $\sigma = \beta_i$ (for $\gamma = 0$.) A large number of multidimensional integrals of this general type have been evaluated by Fichtenholz[3] using more laborious techniques. I prefer to stress the practicality of numerical integration as illustrated here rather than struggling for "closed form" answers.

As an illustration of this last remark consider the integral

$$\int_{-\infty}^{\infty} dx\, x^2 \prod_{i=1}^{N} \frac{\sin p_i x}{p_i x}, \tag{13}$$

which was derived by Cerulus and Hagedorn[4] by an integral transform from some other multidimensional integral. Those authors showed how to evaluate this algebraically in 2^N operations by means of residues. (Sometimes this involves much cancellation between nearly equal terms. The entire integral vanishes if any one p exceeds the sum of the others.) I tried integrating (13) directly, using the rule (11), and found that it worked excellently, except for very small N.

III. PHASE SPACE INTEGRAL

I have applied this method to an interesting and difficult problem which has long concerned high-energy physicists: calculating the volume of phase space for N particles with total energy E. The relativistically invariant integral is expressed in momentum variables as

$$R_N = \left(\prod_{i=1}^{N} \int \frac{d^3p_i}{2E_{p_i}}\right) \delta^3\left(\sum_i \mathbf{p}_i\right) \delta\left(\sum_i E_{p_i} - E\right), \tag{14}$$

where

$$E_{p_i} = +(\mathbf{p}_i^2 + m_i^2)^{1/2}. \tag{15}$$

I start by introducing the Fourier integral representation of the Dirac delta functions,

$$\delta^3(\mathbf{p})\delta(E) = \int \frac{d^3x\, dt}{(2\pi)^4} e^{i\mathbf{p}\cdot\mathbf{x}} e^{-iEt}, \tag{16}$$

and then we have the separate integrals over each momentum variable which result in the modified Bessel function of order 1:

$$\int \frac{d^3p}{2E_p} e^{i\mathbf{p}\cdot\mathbf{x}} e^{-iE_p t} = 2\pi \frac{m}{(\mathbf{x}^2 - t^2)^{1/2}} K_1(m\sqrt{\mathbf{x}^2 - t^2}). \tag{17}$$

From (17) we see that the integration variable t may be taken into the lower half of the complex plane; this means that the square root expression $(\mathbf{x}^2 - t^2)^{1/2}$ always has a positive real part. Now we introduce a new integration variable σ as follows. Writing for shorthand

$$\Pi(\sigma) = \prod_{i=1}^{N} \frac{2\pi m_i}{\sigma} K_1(m_i \sigma), \tag{18}$$

the integral (14) is equal to

$$R_N = \int \frac{d^3x\, dt}{(2\pi)^4} \int_{C-i\infty}^{C+i\infty} d\sigma\, \Pi(\sigma) \frac{i\sigma/\pi}{\sigma^2 - \mathbf{x}^2 + t^2} e^{+iEt}, \tag{19}$$

where C is a small positive constant so that the contour runs to the right of the singularity at $\sigma = 0$ and to the left of the poles at $\sigma = (\mathbf{x}^2 - t^2)^{1/2}$. Now the integrals over $\mathbf{x}$ and t can be carried out, giving us another Bessel function; and the final result is

$$R_N = \frac{1}{4\pi^2 i}\int_{C-i\infty}^{C+i\infty} d\sigma \frac{\sigma^2}{E} I_1(\sigma E)\Pi(\sigma). \tag{20}$$

Thus the $3N$-[or $(3N$-$4)$-] dimensional integral (14) is transformed into a single integral (20). A few analytic remarks can be made before proceeding to discuss the numerical evaluation of this integral.

The contour of integration may be moved about since now the only singularity of the integrand occurs at the origin. If we move far to the right, the Bessel functions can be replaced by their asymptotic forms and we have the simple exponential behavior

$$\exp(E - M): \quad M = \sum_{i=1}^{N} m_i. \tag{21}$$

Thus if E is less than M, the integral is seen to vanish, which is physically correct. If the masses of the particles are all zero, we have

$$\frac{m_i}{\sigma} K_1(m_i\sigma) \to \frac{1}{\sigma^2}, \tag{22}$$

and the integral becomes

$$R_N(0) = \frac{1}{4\pi i^2}\int_{C-i\infty}^{C+i\infty} d\sigma \frac{I_1(\sigma E)}{E}\frac{(2\pi)^N}{\sigma^{2N-2}}, \tag{23}$$

which can be evaluated in terms of the pole at the origin, yielding the well-known result

$$R_N(0) = (\pi/2)^{N-1}E^{2N-4}/(N-1)!(N-2)!. \tag{24}$$

The nonrelativistic limit is gotten by taking the masses of the particles large and using the asymptotic formulas for the Bessel functions:

$$(m_i/\sigma)K_1(m_i\sigma) \to (m_i\pi/2\sigma^3)^{1/2}e^{-m_i\sigma}, \tag{25}$$

$$I_1(\sigma E) \to (1/2\pi E\sigma)^{1/2}e^{\sigma E}, \tag{26}$$

$$R_N \to \left(\prod_{i=1}^{N} m_i^{1/2}\right)\frac{1}{E^{3/2}}\pi^{(3/2)(N-1)}2^{(1/2)(N-3)}$$
$$\times \frac{1}{2\pi i}\int_{C-i\infty}^{C+i\infty} d\sigma \frac{e^{(E-m)\sigma}}{\sigma^{(3/2)(N-1)}}, \tag{27}$$

and the final integral is given by (5). Hybrid closed forms, where some of the particles are considered ultrarelativistic according to (22) and the rest are treated nonrelativistically according to (25) and (26), can also be obtained. As far as I am aware, such closed form results for (14) are new.[5]

For numerical evaluation of (20) we first choose the contour, following the earlier experience with the similar, but much simpler integral (9). Along the real axis, the integrand in (20) grows large at both small σ and large σ; so we shall choose the contour through the point $\sigma = C$ where the integrand has its minimum. Upper and lower bounds for C can be well estimated by using the approximation (26) for the function I_1 and either of the approximations (22) or (25) for the functions K_1:

$$C = [(2 - \tfrac{1}{2}\zeta)N - \tfrac{3}{2}]/(E - M), \tag{28}$$

where

$$0 < \zeta < 1.$$

In my program I let the machine choose the minimum point after three evaluations of the integrand: at the two extremes of (28) and at their midpoint.

Next is the question of how to evaluate the Bessel functions occurring in the integrand (20). I used the polynomial approximations for I_1 and K_1 given by Abramowitz and Stegun.[6] These have an advertised accuracy of about one part in 10^7 or better for *real* arguments; one could worry about their accuracy for the complex arguments needed in our integral. However, I was able to convince myself that this procedure was adequate for the present uses. Finally, there is the task of the actual numerical integration. I used the simple rule (11) after a change of variables

$$\sigma = C(1 + ixe^{x^2}), \tag{29}$$

which helps the integrand to decrease rapidly. Working to an accuracy of one part in 10^4 for the final answer I found that as few as 15 points in the numerical integration were required for large values of N (20 or more); about 40 points were needed at $N = 6$ and about 160 at $N = 3$. The program did not work for $N = 2$. I have ideas about how to change the contour so that this could be improved but it hardly seemed worthwhile. Small-N results can be calculated directly from (14) much more simply. The hard problem is for large N and here my program worked beautifully.

A few checks on the program are available. The zero-mass result (24) is one. The case of all but one particle having mass zero is another[7]; and the case of three particles of equal mass m is given by the integral [gotten directly from (14)]

$$2(1-3\gamma)^2(1+\gamma)^2\int_0^1 dx$$
$$\times \left[x(1-x)\frac{(1+3\gamma)(1-\gamma)-(1-3\gamma)(1+\gamma)x}{4\gamma^2+(1-3\gamma)(1+\gamma)x}\right]^{1/2} \tag{30}$$

where

$$\gamma = m/E,$$

and this is normalized to unity at $\gamma = 0$.

A production run for ten values of N (from $N = 3$ to $N = 30$) and nine values of M/E (from 0 to 0.5) took about 10 s of computer time and cost just over one dollar. This was for the case of all masses equal so that each Bessel function K was evaluated only once at each integration point. In general the time required will be proportional to the number of different masses; but this should still be far far less than the time for any other known method at large N.

A summary of the calculated results (for N equal mass particles) is

$$\rho_N = R_N(m)/R_N(0) \cong \rho_2 e^{-\lambda(N-2)},$$
$$\rho_2 = (1 - (M/E)^2)^{1/2}, \quad M = Nm, \tag{31}$$

and the value of λ is given by

$$\lambda \cong (M/E)^2 \ln(E/M)^2, \tag{32}$$

for small values of M/E (up to 0.1) and increases to about twice this value at $M/E = 0.5$.

IV. ANOTHER TRANSFORM USED

Multidimensional integrals that have the particular form (1) might be thought of as being terribly special. Yet, if one looks at the leading textbook on numerical integration[8] this is the most common form of examples shown. Out of a total of 26 numerical examples given by Davis and Rabinowitz for integrals in more than two dimensions, 16 are of the form (1); and all but one of the remaining examples are of the alternate form:

$$J=\left(\prod_{i=1}^{d}\int g_i(x_i)dx_i\right)F\left(\prod_{i=1}^{d}f_i(x_i)\right). \tag{33}$$

The change from a sum to a product in the argument of the function F leads us to use a Mellin transform in order to factorize the dependence on the separate coordinates x_i. The transform is

$$F(t)=\int_{C-i\infty}^{C+i\infty}\frac{d\sigma}{2\pi i}\,t^{-\sigma}\hat{F}(\sigma), \tag{34}$$

and its inverse is

$$\hat{F}(\sigma)=\int_0^\infty dt\,F(t)t^{\sigma-1}. \tag{35}$$

This leads to the single integral for J, analogous to (3),

$$J=\int\frac{d\sigma}{2\pi i}\hat{F}(\sigma)\prod_{i=1}^{d}w_i(\sigma), \tag{36}$$

where

$$w_i(\sigma)=\int dx\,g_i(x)[f_i(x)]^{-\sigma}. \tag{37}$$

One example of reduction of a d-dimensional integral of this form is

$$\left(\prod_{i=1}^{d}\int_0^1 dx_i\,x_i^q\right)F\left(\prod_{i=1}^{d}x_i\right)=\int_0^1 dt\,t^qF(t)\frac{[\ln 1/t\,]^{d-1}}{(d-1)!}. \tag{38}$$

A second example is

$$\prod_{i=1}^{d}\left(\int_0^\infty dx_i\,e^{-x_i}\right)\exp\left(-b\prod_{i=1}^{d}x_i\right)=\int\frac{d\sigma}{2\pi i}\Gamma(\sigma)\Gamma^d(1-\sigma)b^{-\sigma}, \tag{39}$$

which involves gamma functions, entering as the Mellin transform of the exponential function. The contour of integration here is parallel to the imaginary axis, passing between the poles at $\sigma=0$ and $\sigma=1$. Numerical evaluation of (39) was carried out very successfully, following the general advice given in Sec. II. Gamma functions for complex argument are readily computed by starting with the general asymptotic expansion for moderately large argument. For $d=10$ and various values of b $(2,10,100,2i,10i,100i)$ I was able to obtain six-figure accuracy with under 100 integration points.

Other examples, both analytical and numerical, were studied but need not be recorded here.

V. FURTHER REMARKS

The general technique described here may be of particular practical use in some statistical problems. If one has several independent random variables x_i, distributed according to probability functions $g_i(x_i)$, then an integral of the form (1) with $F=\delta(R-\Sigma_{i=1}^{d}x_i)$ gives the probability distribution for the sum of the variables to have the value R. Some previous work on such problems, it appears, could benefit from the present technique.[9]

In conclusion I should note some possible extensions of the method of integral transforms described above. If the multidimensional integral has not just one function F of the form shown in (1) but a few of them in product, then one could carry out an integral transform (2) for each of them. The resulting product integral in the transform variables might still be more tractable by direct integration than was the original integral.

If an analytic expression for the transform $\hat{F}$ is not available, one might evaluate this also by numerical integration of the inverse transform of F. Similarly, if the integrals w_i of (4) do not give nice closed form answers, numerical integration may be used on these one-dimensional integrals. Thus, for the Bessel functions of complex argument needed in the phase space problem, one could get them directly by numerical integration from the integral representations for these Bessel functions. (I have tried this and it works well.)

Finally, if the function F is of the form $F(f_{12}(x_1,x_2)+f_{34}(x_3,x_4)+\cdots)$, there is an obvious generalization of the method that might be useful.

ACKNOWLEDGMENTS

This work has been assisted by discussions with W. Chinowsky, D. Brillinger, A. Chorin, J. D. Jackson, and G. Lynch.

[1]C. Schwartz, J. Math. Phys. **26**, 955 (1985).

[2]C. Schwartz, J. Comput. Phys. **4**, 19 (June 1969).

[3]G. M. Fichtenholz, *Differential-und Integralrechnung*, translated from the Russian (Deutscher Verlag, Berlin, 1964), Vol. III, pp. 383–407. Almost all of the multidimensional integrals listed by Gradshteyn and Ryzhik in their "Tables" come from this source.

[4]F. Cerulus and R. Hagedorn, Nuovo Cimento Suppl. **9** (2), 646 (1958).

[5]There were some earlier attempts to use integral transform techniques for the phase space problem but they apparently did not succeed very far. See M. Kretzschmar, Ann. Rev. Nucl. Sci. **11**, 1 (1961).

[6]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

[7]M. M. Block and J. D. Jackson, Z. Phys. C **3**, 255 (1980).

[8]P. J.Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1975), Chap. 5.

[9]J.Meisner and F. Demirmen, J. R. Statist. Soc. A **144**, 1 (1981); L. A. Baxter, Comm. Statist. B **10**, 281 (1981).

Numerical integration in many dimensions. II

Charles Schwartz
Department of Physics, University of California, Berkeley, California 94720

(Received 13 June 1984; accepted for publication 21 September 1984)

Two new techniques are presented that appear to be useful in obtaining accurate numerical values for the numerical integration of fairly smooth functions in many dimensions. Both methods start with the idea of a mesh containing n points laid out in each of the d dimensions, then seek strategies that use far less than all n^d points in some systematically improved sequence of approximations.

I. EXTRAPOLATION METHOD

Suppose we have some prescription for the numerical integration of a function $f(x)$ of one variable:

$$\sum_{j=1}^{n} w_j f(x_j) = \int_a^b f(x)dx + E(n). \tag{1}$$

A high-accuracy prescription (quadrature rule) is the set of points x_j and weights w_j such that the error $E(n)$ is a small and rapidly decreasing function of n, the number of mesh points used.

Now suppose we want to integrate a function $F(x_1,x_2,...,x_d) = F(\mathbf{x})$ over the d-dimensional cube. The direct product technique would be to use the rule (1) d times:

$$\sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_d=1}^{n_d} w_{j_1} w_{j_2} \cdots w_{j_d} F(x_{j_1}, x_{j_2},...,x_{j_d}) \equiv S(n_1,n_2,...,n_d) = S(\mathbf{n}). \tag{2}$$

This computation will require a large amount of effort, since the total number of evaluations involved is

$$N = \prod_{i=1}^{d} n_i. \tag{3}$$

To see the form of the error, apply the relation (1) d times to $F(\mathbf{x})$:

$$S(\mathbf{n}) = \iint \cdots \int d^d x\, F(\mathbf{x}) + [E_1(n_1) + E_2(n_2) + \cdots + E_d(n_d)] + \text{higher-order terms}, \tag{4}$$

where the higher-order terms would be of the form of products of two or more "small" terms. This is the main result: If the errors are indeed small in each separate dimension, the leading (first-order) error term for the multidimensional computation is *additive* in contribution from each dimension.

Upon this observation we build a simple technique for eliminating the first-order errors. First, compute S for a given set of numbers n_i; then, one at a time, increase the number of mesh points used in a *single* dimension while keeping all the others fixed, and compute

$$D_i \equiv S(n_1, n_2,...,n_i,...,n_d) - S(n_1, n_2,...,n_i',...,n_d), \quad i = 1, d. \tag{5}$$

Then, from (4), we have

$$D_i \cong E_i(n_i) - E_i(n_i'); \tag{6}$$

and, if n_i' is substantially larger than n_i, we may take

$$D_i \approx E_i(n_i), \tag{7}$$

because each $E(n)$ is assumed to decrease very rapidly as n increases. Thus we computationally determine the first-order error terms and we subtract these terms out from the original computation to get the improved approximation for the integral I of $F(\mathbf{x})$:

$$I \approx S(\mathbf{n}) - \sum_{i=1}^{d} D_i. \tag{8}$$

The saving in computer time by this technique may be considerable: If N_0 is the number of evaluations needed to compute the original $S(\mathbf{n})$, and if we take each $n_i' = 2n_i$, then the additional computing effort for the result (8) is $2dN_0$; this may be compared to $2^d N_0$ which is the amount of effort needed if one doubled all the n_i at once.

This result is an extension of the basic idea in Richardson extrapolation, except that we do not assert a known form for the error function $E(n)$ but only rely upon it being rapidly decreasing.

For numerical examples I took two six-dimensional integrals of complicated form from the book by Davis and Rabinowitz[1]:

$$F_1 = x_1x_2x_3x_4x_5x_6[\log(x_1x_2x_3/x_4x_5x_6)]^2, \quad \text{integrated over the cube } (0,1)^6, \tag{9a}$$

$$F_2 = \tfrac{1}{64}\cos(3x_1x_2x_3x_4x_5(1-x_6) + \tfrac{1}{2}) \quad \text{integrated over the cube } (-1,1)^6. \tag{9b}$$

The points x_j and weights w_j used were those tabulated for Gauss–Legendre numerical quadrature.

Computed results are displayed in Table I. The column headed "Mesh" gives the set of numbers n_i used for the original $S(\mathbf{n})$ (2^6, 3^6, etc.) in each block, followed by the incremental sets ($n^{d-1}n'$) used. The column headed "Number" counts the number of function evaluations needed at each stage of the computation. (In the actual work these numbers were much reduced because of the permutation symmetry of the integrands, but that is not a general feature of the present method.) The columns headed "Error" give the fractional

TABLE I. Numerical results for the six-dimensional integrals (9a) and (9b) using Gauss–Legendre quadrature rules plus the extrapolation technique (8).

Mesh	Error $-F_1$	Number	Error $-F_2$
2^6	0.16	64	0.002 9
$2^5 4$	0.028	768	0.002 1
$2^5 6$	0.0085	1 152	
$2^5 8$	0.0040	1 536	
$2^5 10$	0.0024	1 920	
3^6	0.060	729	0.000 27
$3^5 6$	0.0076	8 736	0.000 064
$3^5 8$	0.0031	11 664	
$3^5 10$	0.0015	14 580	
4^6	0.028	4 096	0.000 014
$4^5 6$	0.0077	36 864	0.000 000 81
$4^5 8$	0.0030	49 152	0.000 000 83
$4^5 10$	0.0014	61 440	
5^6	0.014	15 625	0.000 000 56
$5^5 8$		25 000	0.000 000 006 9
6^6	0.0076[a]	46 656	0.000 000 01[a]
8^6	0.0031[a]	262 144	

[a] From R. Cranley and T. N. L. Patterson, Numer. Math. **16**, 70 (1970).

error in the numerical value of the integral (for the functions F_1 and F_2) computed.

Looking first at the results for the function F_1, we see that overall the error is not very small and decreases rather slowly: for example, look only at the sequence n^6. This is doubtless due to the logarithmic singularity in the integrand, something which the chosen quadrature rule is ill prepared to accommodate. Yet, given that overall difficulty, the present scheme is seen to be very successful at getting higher accuracy with fewer number of mesh points used: compare the accuracy at $2^5 10$ (1,920 + 64 mesh points) with that at 8^6 (262,144 mesh points.) There is a cost saving here of two orders of magnitude for the same result.

When we turn to the results for F_2 things are different. The overall accuracy is better and the convergence more rapid. This may be attributed to the analytic character of the function F_2. The improvements gained by the present extrapolation technique start out as nil (in the topmost block) but then increase rapidly, reaching almost two orders of accuracy improvement (in the fourth block) at a cost of less than twice the starting number of mesh points.

I do not have a general theory to predict when this technique will work well or how best to implement it strategically. It does appear to be quite promising, however, as a technique which one can readily experiment with, using systematic increases in the numbers n to show whether the convergence seems to be good or poor.

II. FACTORIZATION METHOD

Suppose the function $F(\mathbf{x})$ were given as a product of factors, each involving only a single coordinate,

$$f_1(x_1) f_2(x_2) \cdots f_d(x_d); \tag{10}$$

then the d-dimensional integral of F would be simply the product of d one-dimensional integrals, each one of which could be evaluated by some numerical quadrature rule such as (1). The total cost would be proportional to nd rather than the much larger number n^d.

Suppose that $F(\mathbf{x})$ may be well approximated by a factorized form (10) but the individual functions $f_i(x_i)$ are not known. Then one may construct these functions as follows. Choose some reference point (node) $\mathbf{y} = (y_1, y_2, \ldots, y_d)$, such that $F(\mathbf{y})$ is nonzero. Now tabulate the values of F walking out from this node along each one of the coordinate axes:

$$f_i(x_j) = F(y_1, y_2, \ldots, y_{i-1}, x_j, y_{i+1}, \ldots, y_d)/F(\mathbf{y}), \quad j = 1, n, \tag{11}$$

where we have chosen a normalization for the factor functions f_i such that they are equal to 1 at the node, and the points x_j would be chosen to fit the quadrature rule (1) being used. We have thus constructed the aproximation

$$F(\mathbf{x}) \approx G(\mathbf{x}) = F(\mathbf{y}) \prod_{i=1}^{d} f_i(x_i), \tag{12}$$

and the integration follows easily.

Now, to develop a generally useful method, we need to invent a sequence of approximations, like (12), such that we may approach closer and closer to the given function F. From the discussion above it is clear that we have the freedom of choice of the node point $\mathbf{y}$ from which the construction (11) follows.

A first strategy is to take a sequence of nodes $\mathbf{y}_k$, $k = 1, 2, 3, \ldots$, and then construct a sequence of product functions $G_k(\mathbf{x})$, defined by (11) and (12), where G_1 is built from the original function F, G_2 is built from the residual function $F - G_1$, G_3 is built from $F - G_1 - G_2$, etc. This procedure was tried on the two six-dimensional integrals (9a) and (9b); the results were very poor. Probably what is happening is this: At the k th stage one is fitting exactly at the point $\mathbf{y}_k$ and on the lines passing through this point but at the same time one is messing up the fit achieved at the previous node points and their lines. Thus the error can just bounce around from one region to another without being reduced.

A second strategy involved constructing a set of approximations $G_k(\mathbf{x})$, each constructed to fit the original function $F(\mathbf{x})$ at the point $\mathbf{y}_k$, independent of the others

$$G_k(\mathbf{x}) = F(\mathbf{y}_k) \prod_{i=1}^{d} f_i^k(x_i), \quad k = 1, 2, \ldots . \tag{13}$$

Then take a linear combination of these G_k to minimize the expression

$$\sum_k \Big[F(\mathbf{y}_k) - \sum_l C_l G_l(\mathbf{y}_k) \Big]^2 . \tag{14}$$

This was also tried on the same two functions (9a) and (9b) for five points; and the results were even worse than with the first strategy.

A third strategy involved a more complicated "cluster decomposition":

$$\begin{aligned} F(\mathbf{x}) = {} & F(\mathbf{y}) \prod_i f_i(x_i) + \sum_{i_1 < i_2} H^{(2)}_{i_1, i_2}(x_{i_1}, x_{i_2}) \\ & \times \prod_{i \neq i_1, i_2} h_i^{(2)}(x_i) + \sum_{i_1 < i_2 < i_3} H^{(3)}_{i_1, i_2, i_3}(x_{i_1}, x_{i_2}, x_{i_3}) \\ & \times \prod_{i \neq i_1, i_2, i_3} h_i^{(3)}(x_i) + \cdots . \end{aligned} \tag{15}$$

Here only a single node point y is used; the functions $H^{(2)}$, $H^{(3)}$, etc., span larger-dimensional subspaces and are defined to vanish when any of their arguments are on the lines passing through y. This method was tried, through third order, on the same two functions (9a) and (9b) and the results were unsatisfactory once again.

A fourth strategy works the other way: rather than building up correlations between the coordinates from the uncorrelated product (10), we start by taking the full d-dimensional space and decomposing it into a product of two subspaces

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2), \tag{16}$$

where d_1 (the dimension of $\mathbf{x}_1$) and d_2 (the dimension of $\mathbf{x}_2$) add up to d. The original function F is represented by

$$F(\mathbf{x}) = F(\mathbf{x}_1, \mathbf{x}_2) = \sum_k G_1^k(\mathbf{x}_1) G_2^k(\mathbf{x}_2). \tag{17}$$

This arrangement has a special property, which was first noticed to be true in the first strategy above only for the case $d = 2$. There is a freedom of redefinition of the functions G which leaves F unchanged:

$$G_2^k \to G_2^k + AG_2^{k'},\ G_1^{k'} \to G_1^{k'} - AG_1^k, \tag{18}$$

for any number A. With this, one can choose a series of node points

$$\mathbf{y}^k = (\mathbf{y}_1^k, \mathbf{y}_2^k),$$

and require

$$G_1^{k'}(\mathbf{y}_1^k) = G_2^{k'}(\mathbf{y}_2^k) = 0, \quad \text{for all } k' > k. \tag{19}$$

This means that we can carry out the sequential fitting described as the "first strategy" to evaluate the functions G^k [Eq. (17)]. The new advantage, from (19), is the fact that fitting at the k th node $\mathbf{y}^k$ will *not* disturb the previous fittings obtained at other nodes. The price paid for this advantage is that each G function must be evaluated at a large number of points. Still, the total number of evaluations, $n^{d_1} + n^{d_2}$, for each point $\mathbf{y}^k$ can be significantly less than the full number of mesh points $n^{d_1 + d_2}$. Some experiments were carried out using this method. The function (9a) yielded very good results after three node points; the function (9b) gave only fair results with up to six node points. A chief advantage of this method appears to be that the results tend to converge relatively smoothly; while the previous strategies would often give results that jumped around irregulary.

Obviously, the approach of this fourth strategy could be carried further: each subspace $\mathbf{x}_1$ and $\mathbf{x}_2$ could be subdivided into smaller subspaces with consequent savings in the number of evaluations needed.

It is not clear to me when these various strategies will work well and when they will fail. What are the characteristics of the function F which suggest that one or another technique will be most successful? What is the best way to choose a sequence of node points $\mathbf{y}^k$? Perhaps some later analysis or accumulation of experience may shed light on these questions. For the present I believe it is useful to have a variety of strategies which one may simply try out when an expensive multidimensional integral confronts one.

III. SUMMARY

Two new methods have been presented for trying to deal with multidimensional integrals in systematic manners that allow one to judge the accuracy in terms of experimental observations of how the computer outputs converge. The first method is based upon a simple analysis of the error terms when high-accuracy numerical quadrature rules are used. The second method has a geometric conception, with the function being fitted along sets of lines passing through selected node points in the multidimensional space. Several strategies within this second method have been described, with a success rate (at least for the rather difficult test problems studied here) that calls for considerable further work before one would be tempted to market this second method. The numerical success of the first method, on the other hand, is quite encouraging; and the first method is, furthermore, simpler to understand and to implement.

[1] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1975). Chapter 5 deals with multidimensional integrals.

On the intrinsic behavior of the internal variable in the Finslerian gravitational field

Satoshi Ikeda
Department of Mechanical Engineering, Faculty of Science and Technology, Science University of Tokyo, Noda, Chiba 278, Japan[a] *and Graduate Course of Mathematics, II-Faculty of Science, Science University of Tokyo, Shinjuku, Tokyo 162, Japan*

(Received 8 June 1984; accepted for publication 21 December 1984)

A structural extension of the gravitational field is attempted in reference to the theory of Finsler spaces: The vector y is attached to each point x as the internal variable and the intrinsic behavior of y is reflected in the whole spatial structure.

I. INTRODUCTION

It goes without saying that the gravitational field in Einstein's sense[1] has a (four-dimensional) Riemannian structure, which is wholly dominated by the Riemannian metric $\gamma_{\lambda\kappa}(x)$ $(\kappa,\lambda = 1,2,3,4)$. On the other hand, as is well known, several kinds of structural extensions of Einstein's gravitational field[1] have been investigated, such as Brans–Dicke theory,[2] Einstein–Cartan theory,[3] Weyl–Dirac theory with torsion,[4] etc. In these theories, such "non"-Riemannian quantities as (conformal) scalar, torsion, etc. have been introduced, besides $\gamma_{\lambda\kappa}(x)$, at the stage of metric or connection. However, these "non"-Riemannian fields may be regarded as "local" in the sense of Yukawa's nonlocal field theory,[5] because only the point $x(=x^\kappa;\ \kappa = 1,2,3,4)$ is adopted as the independent variable. Therefore, if some new independent variable is attached to each point, then a new "non"-Riemannian and "nonlocal" gravitational field can be realized.[6] This explains our standpoint that if we penetrate into this more microscopically than in Einstein's sense, then we may conjecture that the resulting micrograviational field in our sense does not necessarily remain Riemannian, but comes to have a "non"-Riemannian structure due to some microdegrees of freedom. So, along this line, we shall choose, in this paper, the vector y as such independent variable and consider a Finslerian generalization of the gravitational field. In particular, the intrinsic behavior of the internal variable (y) will be geometrically investigated in detail. By the way, the Riemannian structure may be regarded as "point spacelike," "macroscopic," and "local," while the Finsler structure may be regarded as "line-element spacelike," "microscopic," and "nonlocal."

II. ON THE FINSLERIAN STRUCTURE

The Finslerian nonlocalization can be realized by annexing the vector y to each point x as the internal variable.[6] The Finslerian structure itself is, of course, influenced by the intrinsic behavior of the internal variable y, so that it is necessary to treat equally those two fields existing around x: One is the external (x) field spanned by points $\{x\}$ and the other is the internal (y) field spanned by vectors $\{y\}$. The former is nothing else than the Einstein's gravitational field with the (four-dimensional) Riemannian structure, while the latter may be compared to the so-called internal space associated with each point, which has, in general, a (four-dimensional) Riemannian structure, although the internal space is premised to be flat in most physical problems. [The case of the (y) field being flat will also be considered below. See Sec. III.] Therefore, for our purpose in this paper, we must first consider a "unification" between the (x) and (y) fields and then construct a Finslerian metric $g_{\lambda\kappa}(x,y)$ by unifying the Riemannian metric $\gamma_{\lambda\kappa}(x)$ of the (x) field and the Riemannian metric $h_{ij}(y)$ of the (y) field. [If the (y) field is flat, then $h_{ij}(y)$ reduces to the Minkowskian metric.] It should be remarked here that in order to distinguish the physical function explicitly, the Greek indices $\kappa,\lambda,\ldots(=1,2,3,4)$ are used for the external quantities such as $y^\kappa, g_{\lambda\kappa}$, etc., while the Latin indices $i,j,\ldots(=1,2,3,4)$ are used for the internal quantities such as y^i, h_{ij}, etc.

So, we shall consider our unification as follows: Within the framework of the theory of Finsler spaces,[7] the tangent space at the point x $(=\text{fixed})$ is a (four-dimensional) Riemannian space spanned by tangent vectors such as $\{y\}$ and is governed by its Riemannian metric such as $h_{ij}(y)$, where the system (i) is chosen properly. And it is known[7] that a Finsler metric such as $h_{\lambda\kappa}(x,y)$ can be made locally not to depend on x [i.e., $h_{\lambda\kappa}(x,y) \mapsto h_{ij}(y)$] under suitable conditions, where the system (i) must be chosen properly. Therefore, in our case, although the (y) field is not necessarily regarded as the tangent space, the system (i) of y^i and $h_{ij}(y)$ may be likened to the above-mentioned system (i). Of course, the system (κ) is a general one. Then, the internal quantities y^i and h_{ij} are brought to the external quantities y^κ and $h_{\lambda\kappa}$ through the following mapping relations:

$$y^\kappa = e^\kappa_i(x)y^i,$$
$$h_{\lambda\kappa}(x,y) = e^i_\lambda(x)e^j_\kappa(x)h_{ij}(y), \tag{2.1}$$

where the quantity e denotes physically the mapping operator and resembles geometrically the coordinate transformation matrix. [If the (y) field is flat, then e becomes a function of (x,y), in order to introduce the Finslerian metric $h_{\lambda\kappa}(x,y)$ in the form of (2.1).] By (2.1), the (y) field is embedded in the (x) field, so that in this sense, (2.1) may be considered our unification process of the (x) and (y) fields. This unification process is supported by our convention that on the side of the gravitational field, we adhere to the dimension number 4 and treat or observe only those quantities with Greek indices alone such as $g_{\lambda\kappa}, F^\kappa_{\lambda\mu}$, etc., not those quantities with mixed indices or Latin indices alone such as $g_{\lambda i}$, $F^\kappa_{\lambda i}, h_{ij}$, C^i_{jk}, etc. [Therefore, our unification cannot be treated within the theory of vector bundles (cf. Ref. 8).] In the following, the pro-

[a] Postal address.

 0022-2488/85/050958-03$02.50

cess (2.1) will be called the e mapping. Of course, the intrinsic behavior of y in the (y) field is reflected in the (x) field by the e mapping (see Sec. III).

At this stage, we can consider several kinds of unifications of $\gamma_{\lambda\kappa}(x)$ and $h_{\lambda\kappa}(x,y)$: As the most simple and typical example, the following form will be proposed:

$$g_{\lambda\kappa}(x,y) = \gamma_{\lambda\kappa}(x) + h_{\lambda\kappa}(x,y), \tag{2.2}$$

where $g_{\lambda\kappa}$ denotes the unified Finsler metric of our unified field. Starting from (2.2), we can construct several kinds of metrical Finsler connections with respect to $g_{\lambda\kappa}$ (i.e., $Dg_{\lambda\kappa} = 0$) (see Ref. 9) and can clarify the spatial structure of the Finslerian gravitational field (see Sec. IV). It should be remarked that in the case of (2.2), the inverse of $g_{\lambda\kappa}$ may be given by, at least as the first-order approximation, $g^{\kappa\lambda} = \gamma^{\kappa\lambda} - h^{\kappa\lambda}$, and in this case, the indices should be raised or lowered by $\gamma^{\kappa\lambda}$ or $\gamma_{\lambda\kappa}$ in the practical calculations for physical problems. And it should be noted that the concrete form of $g_{\lambda\kappa}$ such as (2.2) cannot be directly obtained from the standpoint of the theory of vector bundles.[8] By the way, from a physical viewpoint, the quantity $h_{\lambda\kappa}(x,y)$ represents, in general, some microscopic effects caused by the internal variable y, so that $g_{\lambda\kappa}(x,y)$ itself embodies some microscopic features of the gravitational field. Correspondingly, the Finslerian structure dominated by $g_{\lambda\kappa}$ gives, for example, a certain kind of fluctuating or perturbating image at some greater microstage than in Einstein's sense. Therefore, some interesting physical functions of $h_{\lambda\kappa}$ will be found with respect to its microcharacter, which will be reserved for another occasion.

III. ON THE INTRINSIC BEHAVIOR OF THE INTERNAL VARIABLE.I

As the internal variable, the vector y shows its own intrinsic behavior, which is geometrically grasped by its own intrinsic connection or parallelism (i.e., δy^i) in the (y) field. The intrinsic behavior of y is represented, as the typical example, by the rotational property such as, in Asanov's K group,[10]

$$y^i = K^i_j y^j, \tag{3.1}$$

where K^i_j means the rotation matrix. Here K^i_j may be regarded, in the most general case, as a function of (x,y). Equation (3.1) can be "geometrized" as the intrinsic parallelism of y^i (i.e., the intrinsic connection of y^i) denoted by δy^i in the form

$$\delta y^i = dy^i + K^i_{j\mu} y^j\, dx^\mu + L^i_{jk} y^j\, dy^k \quad (=0), \tag{3.2}$$

where $K^i_{j\mu} = -k^i_m(\partial K^m_j/\partial x^\mu)$ and $L^i_{jk} = -k^i_m(\partial K^m_j/\partial y^k)$, k^i_m being the inverse of $(\delta^m_i - K^m_i)$. Equation (3.2) may be regarded as a "Finslerization" of the intrinsic behavior of y^i. Then, δy^i is embedded in the external (x) field by the e mapping as follows:

$$\delta y^\kappa \equiv e^\kappa_i \delta y^i = dy^\kappa + K^\kappa_{\lambda\mu}\, dx^\mu + L^\kappa_{\lambda\mu} y^\lambda\, dy^\mu$$
$$(\equiv N^\kappa_\mu\, dx^\mu + P^\kappa_\mu\, dy^\mu) \quad (=0), \tag{3.3}$$

where $K^\kappa_{\lambda\mu} = e^\kappa_i e^j_\lambda K^i_{j\mu} - (\partial e^\kappa_i/\partial x^\mu) e^i_\lambda$, $L^\kappa_{\lambda\mu} = e^\kappa_i e^j_\lambda e^k_\mu L^i_{jk}$, $N^\kappa_\mu = K^\kappa_{\lambda\mu} y^\lambda$, and $P^\kappa_\mu = \delta^\kappa_\mu + L^\kappa_{\lambda\mu} y^\lambda$. [In the most general case where K^i_j and e are functions of (x,y), $L^\kappa_{\lambda\mu}$ is given by $L^\kappa_{\lambda\mu} = e^\kappa_i e^j_\lambda e^k_\mu L^i_{jk} - (\partial e^\kappa_i/\partial y^\mu) e^i_\lambda$.] Here, $K^\kappa_{\lambda\mu}$ and $L^\kappa_{\lambda\mu}$ play the role of horizontal and vertical coefficients of connection and N^κ_μ is the nonlinear connection[7] for the connection δ. If δy^i is not geometrized from the inherent law of y^i such as (3.1), then the conditions $\delta y^i = 0$ and $\delta y^\kappa = 0$ do not hold good, in general. But in our case mentioned above, the inherent law of y^i is satisfied automatically as in (3.1), so that the conditions $\delta y^i = 0$ and then $\delta y^\kappa = 0$ hold good as in (3.2) and (3.3). The connection δ given by (3.2) is assumed to be metrical for h_{ij}, i.e., $\delta h_{ij} = 0$, so that the connection δ given by (3.3) becomes also metrical for $h_{\lambda\kappa}$, i.e., $\delta h_{\lambda\kappa} (\equiv e^i_\lambda e^j_\kappa \delta h_{ij}) = 0$, where the homogeneity conditions with respect to the vertical coefficients of connection are not assumed from a general standpoint, i.e., $L^i_{jk} y^j \neq 0$ and $L^\kappa_{\lambda\mu} y^\lambda \neq 0$ (cf. Refs. 7 and 9). Concerning the conditions $\delta y^\kappa = 0$ and $\delta h_{\lambda\kappa} = 0$, we had better assume, from the beginning, the absolute parallelism of e (i.e., $\delta \mathrm{e} = 0$) in (2.1). This is quite appropriate from a physical viewpoint (cf. Ref. 10). Therefore, the conditions $\delta y^\kappa = 0$ and $\delta h_{\lambda\kappa} = 0$ are compatible with each other without loss of generality and naturalness.

From the most general case mentioned above, we can consider some special cases as follows.

(i) If K^i_j is a function of x alone, as in usual cases (cf. Ref. 10), then $L^i_{jk} = 0$ in (3.2) and $L^\kappa_{\lambda\mu} = 0$ in (3.3).

(ii) If K^i_j is a function of y alone, then (3.2) becomes a Riemannian parallelism in the (y) field, where $K^i_{j\mu} = 0$ in (3.2) and (3.3).

(iii) If the (y) field is flat (i.e., Minkowskian), then K^i_j = const and (3.2) reduces to $\delta y^i = dy^i = 0$ (*i.e.*, $K^i_{j\mu} = 0$ and $L^i_{jk} = 0$), so that $K^\kappa_{\lambda\mu} = -(\partial e^\kappa_i/\partial x^\mu) e^i_\lambda$ and $L^\kappa_{\lambda\mu} = -(\partial e^\kappa_i/\partial y^\mu) e^i_\lambda$ in (3.3), because e becomes a function of (x,y) in this case.

(iv) If $K^i_j = K^i_j(x)$ and $K^\kappa_{\lambda\mu}(x)$ in this case is equal to the one-form linear connection[11] $\Lambda^\kappa_{\lambda\mu} = e^\kappa_i(\partial e^i_\lambda/\partial x^\mu)$, then $K^i_{j\mu} = 0$, so that in this case, K^i_j turns out to be constant.

IV. ON THE INTRINSIC BEHAVIOR OF THE INTERNAL VARIABLE.II

As mentioned above, δy^κ given by (3.3) reflects the intrinsic behavior of y^i in the external (x) field, so that the whole Finslerian structure at the stage of connection is also influenced by δy^κ. Under these situations, the metrical Finsler connection D with respect to $g_{\lambda\kappa}$ [such as (2.2)] (i.e., $Dg_{\lambda\kappa} = 0$) can be represented by, for an arbitrary vector X^κ,

$$DX^\kappa = dX^\kappa + \Gamma^\kappa_{\lambda\mu} X^\lambda\, dx^\mu + C^\kappa_{\lambda\mu} X^\lambda\, dy^\mu$$
$$= dX^\kappa + F^\kappa_{\lambda\mu} X^\lambda\, dx^\mu + \Delta^\kappa_{\lambda\mu} X^\lambda \delta y^\mu, \tag{4.1}$$

where $F^\kappa_{\lambda\mu} (\equiv \Gamma^\kappa_{\lambda\mu} - M^\nu_\mu C^\kappa_{\lambda\nu})$ and $\Delta^\kappa_{\lambda\mu} (\equiv Q^\nu_\mu C^\kappa_{\lambda\nu})$ denote the horizontal and vertical coefficients of connection, $M^\nu_\mu (\equiv Q^\nu_\lambda N^\lambda_\mu)$ being the nonlinear connection for the connection D. [Q^κ_λ is the inverse of P^λ_κ: see (3.3).] Of course, $Dy^\kappa \neq \delta y^\kappa$. The horizontal coefficient of connection $F^\kappa_{\lambda\mu}$ represents, therefore, the concept of unified gauge field[12] for the Finslerian gravitational field. And the base $[\partial/\partial x^\lambda - M^\nu_\lambda(\partial/\partial y^\nu),\ Q^\nu_\lambda(\partial/\partial y^\nu)]$ and the dual base $(dx^\kappa, \delta y^\kappa)$ can be set for the unified field. From (4.1), the covariant derivatives with respect to x and y can be defined and then three kinds of curvature tensors and five kinds of torsion tensors can also be introduced through the Ricci identities, but they are all omitted here for the sake of simplicity (cf. Ref. 7).

Since $Dy^\kappa \neq \delta y^\kappa$ in our case, the relations $Dg_{\lambda\kappa} = 0$ and $Dh_{\lambda\kappa} \neq 0$ and $\delta g_{\lambda\kappa} \neq 0$ and $\delta h_{\lambda\kappa} = 0$ hold. (The metrical conditions $\delta h_{ij} = 0$ are assumed under $\delta y^i = 0$, so that $\delta h_{\lambda\kappa} = 0$ are assumed under the absolute parallelism of **e**, i.e., $\delta \mathbf{e} = 0$). That is to say, two different kinds of metrical Finsler connections $Dg_{\lambda\kappa} = 0$ and $\delta h_{\lambda\kappa} = 0$ are introduced owing to the difference of Dy^κ and δy^κ. In order to obtain the relation of Dy and δy, we reconsider the relations $Dg_{\lambda\kappa} = 0$ and $\delta g_{\lambda\kappa} \neq 0$ as follows: The connection D is a metrical connection for $g_{\lambda\kappa}$ derived from the nonmetrical one δ. Then, by use of Kawaguchi's theorem[13] which makes a nonmetrical connection metrical, the desiring relation can be obtained, with neglect of arbitrariness, as follows:

$$Dy^\kappa = \delta y^\kappa + \tfrac{1}{2} g^{\kappa\nu}(\delta g_{\nu\lambda})y^\lambda, \tag{4.2}$$

by which the following relations can be obtained from (4.1) and (4.2) (cf. Ref. 14):

$$\begin{aligned} \Gamma^{\kappa}_{\lambda\mu} &= K^{\kappa}_{\lambda\mu} + \tfrac{1}{2} g^{\kappa\nu}\left(\frac{\partial g_{\nu\lambda}}{\partial x^\mu} - K^{\iota}_{\nu\mu} g_{\iota\lambda} - K^{\iota}_{\lambda\mu} g_{\nu\iota}\right), \\ C^{\kappa}_{\lambda\mu} &= L^{\kappa}_{\lambda\mu} + \tfrac{1}{2} g^{\kappa\nu}\left(\frac{\partial g_{\nu\lambda}}{\partial y^\mu} - L^{\iota}_{\nu\mu} g_{\iota\lambda} - L^{\iota}_{\lambda\mu} g_{\nu\iota}\right). \end{aligned} \tag{4.3}$$

Therefore, from (4.3), the relations between $(F^{\kappa}_{\lambda\mu}, \Delta^{\kappa}_{\lambda\mu})$ and $(K^{\kappa}_{\lambda\mu}, L^{\kappa}_{\lambda\mu})$ can be obtained by inserting (4.3) into the definitions of $F^{\kappa}_{\lambda\mu}$ and $\Delta^{\kappa}_{\lambda\mu}$, which are, however, omitted here for simplicity's sake. As is understood from the above, the intrinsic behavior of y^i (i.e., δy^i) or y^κ (i.e., δy^κ) represented by $(K^{i}_{j\mu}, L^{i}_{jk})$ (3.2) or $(K^{\kappa}_{\lambda\mu}, L^{\kappa}_{\lambda\mu})$ (3.3) is absorbed into $(\Gamma^{\kappa}_{\lambda\mu}, C^{\kappa}_{\lambda\mu})$ or $(F^{\kappa}_{\lambda\mu}, \Delta^{\kappa}_{\lambda\mu})$ (4.1) by means of the relations (4.2) or (4.3).

Thus, the spatial structure of our Finslerian gravitational field, especially the connection structure, has been completely clarified by taking account of the intrinsic behavior of the internal variable y. In the future, some other essential unifications of $\gamma_{\lambda\kappa}$ and h_{ij} and some other interesting examples of δy^i or δy^κ should be investigated.

[1] A. Einstein, Ann. Phys. **49**, 769 (1916).
[2] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[3] See, for example, F. Hehl, P. van der Heyde, G. Kerlick, and J. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[4] D. Gregorash and G. Papini, Nuovo Cimento B **55**, 37 (1980).
[5] H. Yukawa, Phys. Rev. **77**, 219 (1950).
[6] S. Ikeda. Lett. Nuovo Cimento **21**, 165 (1978).
[7] See, for example, M. Matsumoto, *The Theory of Finsler Connections* (Okayama Univ., Okayama, Japan, 1970); "Foundations of Finsler Geometry and Special Finsler Spaces" (unpublished, 1977).
[8] R. Miron, in *Proceedings of the 2nd National Seminar on Finsler Spaces* (Braşov Univ., Braşov, Rumania 1983), pp. 147–188.
[9] R. Miron and M. Hashiguchi, Rep. Fac. Sci. Kagoshima Univ. **12**, 21 (1979).
[10] G. S. Asanov, Ann. Phys. **34**, 169 (1977).
[11] M. Matsumoto and H. Shimada, Tensor **32**, 161 (1978).
[12] See, for example, D. Ivanenko, in *Relativity, Quanta, and Cosmology*, edited by M. Pantaleo (Jonson, New York, 1979), pp. 295–354.
[13] A. Kawaguchi, Akad. Wetensch. Amsterdam Proc. **40**, 596 (1937).
[14] S. Ikeda, J. Math. Phys. **22**, 2831 (1981).

A summation method for the Rayleigh–Schrödinger series for the anharmonic oscillator

Leonid Friedlander
Kremenchygskaja 5, Apt. 228, Moscow, U.S.S.R.

(Received 1 October 1984; accepted for publication 28 December 1984)

We approximate the energy levels of the anharmonic oscillator with any coupling constant by eigenvalues $\lambda_j(g,T)$ of the operator $-d^2/dx^2 + x^2 + gV_T(x)$ with $V_T(x) = x^4$ when $|x| \leqslant T$ and $V_T(x) = T^4$ when $|x| > T$. The functions $\lambda_j(g,T)$ are holomorphic with respect to g in a neighborhood of the non-negative half-axis. The conformal transformation maps this neighborhood onto the unit circle of the complex plane. It gives the summation method for the Rayleigh–Schrödinger series for every $g \geqslant 0$.

I. INTRODUCTION

In this paper we consider the spectrum of the operator

$$A(g) = \frac{-d^2}{dx^2} + x^2 + gx^4, \quad g \geqslant 0,$$

in $L^2(-\infty, +\infty)$. It is well known (see, for example, Refs. 1 and 2) that the spectrum of such an operator is discrete and simple. Let $0 < \mu_0(g) < \mu_1(g) < \cdots$ be the eigenvalues of $A(g)$. In the case $g = 0$ we have the operator of an harmonic oscillator, so $\mu_j(0) = 2j + 1$. One can formally write the Rayleigh–Schrödinger series[3,4]

$$\mu_j(g) = 2j + 1 + \alpha_j^{(1)} g + \alpha_j^{(2)} g^2 + \cdots . \tag{1.1}$$

This series, however, converges nowhere except $g = 0$ (see Refs. 5 and 6). It is an asymptotic series[4] and one can sum it by the Borel method[7,8] or by the Padé approximant method.[9] Recently a new rather heuristic approach to the summation problem for perturbation theory's divergent series was suggested by Turbiner.[10] We intend to describe an alternative summation method for (1.1). We think it is rather simple and it can be applied in many cases, not only for the anharmonic oscillator.

We approximate $A(g)$ by the operator

$$B(g,T) = \frac{-d^2}{dx^2} + x^2 + gV_T(x),$$

with

$$V_T(x) = \begin{cases} x^4, & \text{if } |x| < T, \\ T^4, & \text{if } |x| \geqslant T. \end{cases}$$

Let $0 < \lambda_0(g,T) < \lambda_1(g,T) < \cdots$ be the eigenvalues of $B(g,T)$.

Proposition 1: $\lim_{T\to\infty} \lambda_j(g,T) = \mu_j(g)$ for every fixed $g > 0$. Moreover,

$$0 \leqslant \mu_j(g) - \lambda_j(g,T) \leqslant Cg\,\mu_j^{3/4}(g) T^{15/4} e^{-(T^{3/2} - T)/2}, \tag{1.2}$$

for sufficiently large T.

The linear operator family $B(g,T)$ is regular so the Rayleigh–Schrödinger series

$$\lambda_j(g,T) = 2j + 1 + \beta_j^{(1)} g + \beta_j^{(2)} g^2 + \cdots \tag{1.3}$$

converges in a circle $|g| < r_T$. The radius of this circle decreases when $T \to \infty$, so one cannot directly use (1.3) for approximate calculation of $\mu_j(g)$ with any precision. However the function $\lambda_j(g,T)$ is holomorphic in a domain bigger then $|g| < r_T$.

Proposition 2: The function $\lambda_j(g,T)$ is holomorphic in the domain

$$\Omega_T = \{g : |g| < 2T^{-4}\} \cup \{g : \operatorname{Re} g > 0,\ |\operatorname{Im} g| < CT^{-4} (\operatorname{Re} g)^{1/3}\},$$

when $T \geqslant T_0 = T_0(j)$; the constant C depends upon j only.

The domain Ω_T can be mapped conformally onto the circle $|\zeta| < 1$ in such a way that $\zeta(0) = 0$ and $\zeta((0, +\infty)) = (0,1)$. After substituting $g_T(\zeta)$ into (1.3) and transformating this series into the power series with respect to ζ, one has

$$\lambda_j(g_T(\zeta), T) = 2j + 1 + \gamma_j^{(1)}(T)\zeta + \gamma_j^{(2)}(T)\zeta^2 + \cdots . \tag{1.4}$$

This series converges in the circle $|\zeta| < 1$. After the inverse substitution $\zeta = \zeta_T(g)$ into its sum one can obtain the value of $\lambda_j(g,T)$ for every $g \in \Omega_T$, particularly for every $g \geqslant 0$. To obtain explicit formulas it is more convenient to map the half-strip

$$\Pi_T = \{g : \operatorname{Re} g \geqslant -C_1 T^{-4},\ |\operatorname{Im} g| \leqslant C_2 T^{-4}\} \subset \Omega_T$$

onto the circle $|\zeta| < 1$ by

$$\zeta = F(g) = \frac{\sinh((\pi/2C_2)T^4 g + \pi C_1/2C_2) - \sinh(\pi C_1/2C_2)}{\sinh((\pi/2C_2)T^4 g + \pi C_1/2C_2) + \sinh(\pi C_1/2C_2)}.$$

The F maps the half-axis $(0, +\infty)$ onto the interval $(0,1)$. The inverse mapping is

$$g = \frac{2C_2}{\pi T^4} \operatorname{arcsinh}\left[\frac{1+\zeta}{1-\zeta} \sinh\left(\frac{\pi C_1}{2C_2}\right)\right] - C_1 T^{-4}.$$

II. PROOF OF PROPOSITION 1

First of all, the potentials $x^2 + gV_T(x)$ increase when T increases; so by Courant minimax formulas for eigenvalues, $\lambda_j(g,T)$ increases with respect to T and $\lambda_j(g,T) \leqslant \mu_j(g)$. Thus,

$$\lambda_j(g,T) \nearrow \mu_j^*(g) \leqslant \mu_j(g), \quad \text{when } T \to \infty . \tag{2.1}$$

Let us introduce some notation: $Q(\Lambda, g, T)$ is the orthoprojector in $L_2(-\infty, +\infty)$ onto the subspace spanned by eigenfunctions of $B(g,T)$ with eigenvalues $\leqslant \Lambda$,

 0022-2488/85/050961-04$02.50

$$N(\Lambda, g,T) = \mathrm{Tr}\, Q(\Lambda, g,T) = \#\{ j{:}\lambda_j(g,T) \leqslant \Lambda \},$$

and

$$\widetilde{N}(\Lambda, g) = \#\{ j{:}\mu_j(g) \leqslant \Lambda \}.$$

By (2.1)

$$\widetilde{N}(\Lambda, g) \leqslant N(\Lambda, g,T). \tag{2.2}$$

Suppose we are able to obtain the estimate

$$\|[A(g) - B(g,T)]Q(\Lambda, g,T)\| \leqslant \gamma(g,\Lambda,T), \tag{2.3}$$

with $\gamma(g,\Lambda,T) \to 0$ when $T \to \infty$; g and Λ are fixed. Then for every $u \in Q(\Lambda, g,T)L_2(-\infty, +\infty)$ one has

$$|(A(g)u,u)| = |B(g,T)u,u) + ((A(g) - B(g,T))u,u)| \leqslant (\Lambda + \gamma(g,\Lambda,T))(u,u),$$

and by the Glazman variational lemma[11]

$$\widetilde{N}(\Lambda + \gamma(g,\Lambda,T),g) \geqslant \mathrm{Tr}\, Q(\Lambda, g,T) = N(\Lambda, g,T). \tag{2.4}$$

By (2.2) and (2.4),

$$\lim_{T\to\infty} N(\Lambda, g,T) = \widetilde{N}(\Lambda, g),$$

$$\lambda_j(g,T) \nearrow \mu_j(g), \quad \text{when } T \to \infty.$$

Moreover,

$$\mu_j(g) - \lambda_j(g,T) \leqslant \gamma(g,\lambda_j(g,T), T). \tag{2.5}$$

Now our aim is to obtain (2.3). Suppose that $gT^2 > \Lambda$. Let $\varphi(x)$ be the eigenfunction of $B(g,T)$ with an eigenvalue $\lambda \leqslant \Lambda$, $\|\varphi\| = 1$, and $\varphi(x) > 0$ if x is sufficiently large; let x_0 be the positive root of the equation

$$gx^4 + x^2 = \lambda$$

and $x_1 \leqslant x_0$ be the nearest to x_0 local maximum of φ (we will use general properties of solutions of the Sturm–Liouville equation and Sturm comparison theorems; see, for example, Refs. 1 and 2). The function $\varphi(x)$ is decreasing in the interval $(x_1, +\infty)$. Obviously $x_1 < T^{1/2}$. We have $x^2 + gV_T(x) - \lambda \geqslant T$ when $x > T^{1/2}$; so $\varphi(x)$ is majorated on the half-axis $(T^{1/2}, +\infty)$ by the decreasing solution of the equation $\psi'' - T\psi = 0$. Thus,

$$\varphi(x) \leqslant \varphi(T^{1/2})\exp(-T^{1/2}(x - T^{1/2}))$$

$$\leqslant \varphi(x_1)\exp(T^{1/2}(x - T^{1/2})), \quad x > T^{1/2}. \tag{2.6}$$

When $x \in (x_1 - \pi/(2\Lambda^{1/2}), x_1 + \pi/(2\Lambda^{1/2}))$ the $\varphi(x)$ can be estimated below by the solution of the problem

$$\psi'' + \Lambda\psi = 0, \quad \psi'(x_1) = 0, \quad \psi(x_1) = \varphi(x_1),$$

because $\lambda - (x^2 + gV_T(x)) < \Lambda$. So

$$\varphi(x) \geqslant \varphi(x_1)\cos(\Lambda^{1/2}(x - x_1))$$

on this interval. Thus,

$$1 = \int_{-\infty}^{+\infty} \varphi^2(x)dx \geqslant \int_{x_1 - \pi/(2\Lambda^{1/2})}^{x_1 + \pi/(2\Lambda^{1/2})} \varphi^2(x)dx$$

$$\geqslant \Lambda^{-1/2}\varphi^2(x_1) \int_{-\pi/2}^{\pi/2} \cos^2 x\, dx = \frac{\pi}{2}\Lambda^{-1/2}\varphi^2(x_1).$$

Hence

$$\varphi^2(x_1) \leqslant 2\pi^{-1}\Lambda^{1/2}.$$

Therefore

$$\|[A(g) - B(g,T)]\varphi\|^2 = g^2 \int_{|x|>T} (x^4 - T^4)^2 \varphi^2(x)dx$$

$$\leqslant 4g^2\pi^{-1}\Lambda^{1/2} \int_T^\infty x^8 \exp(-T^{1/2}(x - T^{1/2}))dx$$

$$\leqslant C_1^2\, g^2\Lambda^{1/2}T^{15/2} \exp(-(T^{3/2} - T)).$$

Finally,

$$\|[A(g) - B(g,T)]Q(\Lambda, g,T)\| \leqslant N^{1/2}(\Lambda, g,T)C_1 g\Lambda^{1/4}T^{15/4} \exp(-(T^{3/2} - T)/2)$$

$$\leqslant C_2 g\Lambda^{3/4}T^{15/4} \exp(-(T^{3/2} - T)/2). \qquad \blacksquare$$

III. PROOF OF PROPOSITION 2

The potentials $x^2 + gV_T(x)$ are even so all eigenfunctions are even or odd. The restrictions of theirs on the half-axis $[0, +\infty)$ are eigenfunctions of the Sturm–Liouville operator on $[0, +\infty)$ with Neumann or Dirichlet boundary condition correspondingly. Even eigenfunctions correspond to eigenvalues with even indices and odd eigenfunctions correspond to eigenvalues with odd indices. Our aim is to estimate the radius of convergence of $\lambda_j(g + h,T)$ with respect to h when $g \geqslant 0$. To apply the perturbation theory we should like to know lower estimates for $\lambda_{2j} - \lambda_{2j-2}$ and $\lambda_{2j+1} - \lambda_{2j-1}$ (the eigenvalues with even and odd indexes correspond to different operators on the positive half-axis, so we can investigate them separately). Let $\varphi_\lambda(x)$ be the solution of

$$-\varphi_\lambda'' + x^2\varphi_\lambda + gV_T(x)\varphi_\lambda = \lambda\varphi_\lambda \tag{3.1}$$

on $[0, +\infty)$ normalized by the following conditions:

(i) $\displaystyle\int_0^\infty \varphi_\lambda^2(x)dx = 1,$

(ii) $\varphi_\lambda(x) > 0$, if x is sufficiently large.

Such a solution exists and it is unique. The family $\varphi_\lambda(x)$ is pointwise C^1 with respect to λ. Let $\psi_\lambda = (d/d\lambda)\varphi_\lambda$. Then ψ_λ satisfies the equation

$$-\psi_\lambda'' + K\psi_\lambda = \lambda\psi_\lambda + \varphi_\lambda, \tag{3.2}$$

with $K(x) = x^2 + gV_T(x)$, and

$$\int_0^\infty \varphi_\lambda\psi_\lambda\, dx = 0.$$

Let $K(\tau) > \lambda$. Then the function

$$\tilde{\psi}_\lambda(x) = \varphi_\lambda(x) \int_\tau^x \varphi_\lambda^{-2}(y)dy \int_y^\infty \varphi_\lambda^2(z)dz$$

satisfies Eq. (3.2) on the half-axis $[\tau, +\infty)$. One can continue this function to the interval $[0,\tau)$ as a solution of (2.2) with conditions $\tilde{\psi}_\lambda(\tau - o) = \tilde{\psi}_\lambda(\tau + 0)$, $\tilde{\psi}_\lambda'(\tau - 0) = \tilde{\psi}_\lambda'(\tau + 0)$. By the Sturm comparison theorem,

$$\varphi_\lambda(z) \leqslant \varphi_\lambda(y)\exp(-(K(y) - \lambda)^{1/2}(z - y));$$

so

$$\int_y^\infty \varphi_\lambda^2(z)dz \leqslant \frac{\varphi_\lambda^2(y)}{2\sqrt{K(y) - \lambda}}$$

and

$$0<\tilde{\psi}_\lambda(x)\leqslant\varphi_\lambda(x)\int_T^x \frac{dy}{2\sqrt{K(y)-\lambda}}\,.$$

Therefore $\tilde{\psi}_\lambda(x)\in L_2(0,+\infty)$ and

$$\psi_\lambda(x)=\tilde{\psi}_\lambda(x)-\varphi_\lambda(x)\int_0^\infty \varphi_\lambda(x)\tilde{\psi}_\lambda(x)dx\,.$$

Now the C^1 dependence $\varphi_\lambda(x)$ upon λ is obvious.

Consider λ from the interval $[\lambda_{2j-1},\lambda_{2j}]$. Let $\sigma(\lambda)$ be the smallest zero of $\varphi_\lambda(x)$. According to the oscillatory theorem,[1,2] $\varphi_\lambda(x)$ has j zeros. The function $\sigma(\lambda)$ increases monotonically in the interval $[\lambda_{2j-1},\lambda_{2j}]$. Really, $\varphi_\lambda(x+\sigma)$ is the jth eigenfunction for the operator

$$\frac{-d^2}{dx^2}+K(x+\sigma) \tag{3.3}$$

on the positive half-axis with Dirichlet condition. The family of potentials $K(x+\sigma)$ increases with respect to σ, so $\lambda(\sigma)$ [and $\sigma(\lambda)$] increases by the Courant minimax principle. Let $\sigma(\lambda_{2j})=\sigma_0$. To estimate $\lambda_{2j}-\lambda_{2j-1}$ we will derive (i) the lower estimate for σ_0, and (ii) the lower estimate for $d\lambda/d\sigma$.

The first step is very simple. The function $\varphi_{\lambda_{2j}}(x)$ satisfies the Neumann condition, so by the Sturm comparison theorem

$$|\varphi_{\lambda_{2j}}(x)|\geqslant|\varphi_{\lambda_{2j}}(0)|\cos(\lambda_{2j}^{1/2}x)\,,$$
$$0\leqslant x\leqslant(\pi/2)\lambda_{2j}^{-1/2}\,.$$

Hence,

$$\sigma_0\geqslant(\pi/2)\lambda_{2j}^{-1/2}\,. \tag{3.4}$$

For the second step remember that $\varphi_\lambda(x+\sigma)$ is the eigenfunction for (3.3). By the Rayleigh formula,

$$\frac{d\lambda}{d\sigma}=\frac{\int_0^\infty[\partial K(x+\sigma)/\partial\sigma]\varphi_\lambda^2(x+\sigma)dx}{\int_0^\infty\varphi_\lambda^2(x+\sigma)dx}$$
$$=\frac{\int_\sigma^\infty(2x+gV'_T(x))\varphi_\lambda^2(x)dx}{\int_\sigma^\infty\varphi_\lambda^2(x)dx}\,. \tag{3.5}$$

Let $x_1=x_1(\lambda)$ be the smallest positive local extremum of $\varphi_\lambda(x)$. It is very simple to derive from the Sturm comparison theorem that

$$\int_0^{\sigma(\lambda)}\varphi_\lambda^2(x)dx\leqslant\int_{\sigma(\lambda)}^{x_1(\lambda)}\varphi_\lambda^2(x)dx\leqslant\int_{x_1(\lambda)}^\infty\varphi_\lambda^2(x)dx\,, \tag{3.6}$$

so

$$\int_0^{\sigma(\lambda)}\varphi_\lambda^2(x)dx\leqslant\frac{1}{3}\,.$$

Now let us remember the asymptotics for $\mu_k(g)$. Let $\psi_k(x)$ be the eigenfunction of $A(g)$. After the well-known scaling transformation $x=g^{-1/6}y$,

$$\frac{-d^2\psi_k}{dy^2}+y^4\psi_k+g^{-2/3}y^2\psi_k=\mu_k(g)g^{-1/3}\psi_k\,.$$

Therefore

$$\mu_k(g)\sim g^{1/3}\nu_k\,,\quad g\to+\infty\,, \tag{3.7}$$

where the ν_k are eigenvalues of the operator $-d^2/dy^2+y^4$. We choose T_0 from the conditions

(i) $\left(\frac{T_0}{2}\right)^2+g\left(\frac{T_0}{2}\right)^4>\mu_{2j}(g)$ for every $g\geqslant0$;

(ii) $T_0>(\frac{4}{3})^{1/4}(\log 9)^{1/2}$.

The asymptotics (3.7) allow us to fulfill the first condition. The function $\varphi_\lambda(x)$ decreases on the half-axis $[T/2,+\infty)$. When $x\geqslant T$, it is majorated by

$$\varphi_\lambda(T)\exp(-(T^2+gT^4-\lambda)^{1/2}(x-T))$$

and on the interval $[T/2,T]$ it is estimated below by the same function. Hence

$$\int_T^\infty\varphi_\lambda^2(x)dx$$
$$<\frac{\int_T^\infty\varphi_\lambda^2(x)dx}{\int_{T/2}^T\varphi_\lambda^2(x)dx}<\left[\exp(T\sqrt{T^2+gT^4-\lambda})-1\right]^{-1}$$
$$\leqslant\left[\exp(T\sqrt{\tfrac{3}{4}T^2+\tfrac{15}{16}gT^4})-1\right]^{-1}<\tfrac{1}{8}\,,$$

because $\lambda\leqslant\lambda_{2j}(g,T)\leqslant\mu_{2j}(g)$. Let $w(g)=(\pi/2)\mu_{2j}^{-1/2}(g)$. Clearly $w(g)<x_1(\lambda)$ and by (3.6)

$$\int_0^{w(g)}\varphi_\lambda^2(x)dx<\frac{1}{2}\,.$$

Thus,

$$\int_\sigma^\infty(2x+gV'_T(x))\varphi_\lambda^2(x)dx$$
$$>\int_{\max(\sigma,w)}^T(2x+4gx^3)\varphi_\lambda^2(x)dx$$
$$\geqslant C_1g\mu_{2j}^{-3/2}(g)\left(1-\int_T^\infty\varphi_\lambda^2(x)dx-\int_0^{\max(\sigma,w)}\varphi_\lambda^2(x)dx\right)$$
$$\geqslant C_2g^{1/2}\,.$$

Finally

$$\frac{d\lambda}{d\sigma}\geqslant C_3g^{1/2}\,.$$

Taking into account (2.4) and (2.7),

$$\lambda_{2j}-\lambda_{2j-1}\geqslant(\pi/2)C_3g^{1/2}\lambda_{2j}^{-1/2}\geqslant C_4g^{1/3}\,.$$

Let $d_j(g)=\min(\lambda_j-\lambda_{j-2},\lambda_{j+2}-\lambda_j)$. Then

$$d_j(g)\geqslant C_5g^{1/3}\,. \tag{3.8}$$

The constant C_5 depends upon j only. Consider the operator

$$B(g+h,T)=B(g,T)+hV_T(x)$$

as a perturbation of $B(g,T)$. The estimate (3.8) and $\|V_T(x)\|=T^4$ imply[3] the Rayleigh–Schrödinger series for $\lambda_j(g+h,T)$ with respect to h to be convergent in the circle $|h|\leqslant C_6T^{-4}g^{1/3}$. When $g=0$ it converges in the circle $|h|\leqslant 2T^{-4}$ because $d_j(0)=4$. ■

Remark: One can approximate x^4 not by $V_T(x)$ but by some other function, for example, by[10]

$$V_\alpha(x)=x^4e^{-\alpha x^2}\,,\quad \alpha\to0\,.$$

The only difficulty for such an approximation is that $x^2+gV_\alpha(x)$ does not increase monotonically and it invalidates our argument for existence of the function $\lambda(\sigma)$ in the proof of Proposition 2. However, this function exists. If it does not the Sturm–Liouville operator (3.3) on the half-axis with Dirichlet condition has for some σ two eigenfunctions which correspond to different eigenvalues but have the same number of zeros.

[1]E. G. Titchmarsh, *Eigenfunction Expansions Associated with Second-order Differential Equations* (Clarendon, Oxford, 1946).
[2]F. A. Berezin and M. A. Shubin, *The Schrödinger Equation* (Moscow State U. P., Moscow, 1983) (in Russian).
[3]T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
[4]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. IV. Analysis of Operators* (Academic, New York, 1978).
[5]F. Dyson, Phys. Rev. **85**, 631 (1952).
[6]C. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[7]S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[8]B. Hirsbrunner and J. Loeffel, Helv. Phys. Acta **48**, 546 (1975).
[9]J. Loeffel, A. Martin, R. Simon, and A. S. Wightman, Phys. Lett. B **30**, 656 (1979).
[10]A. V. Turbiner, "Quantum mechanics: a new approach to the old perturbation theory," preprint ITEF, Moscow, 1982 (in Russian).
[11]M. A. Shubin, *Pseudo-differential Operators and Spectral Theory* (Nauka, Moscow, 1978) (in Russian).

On the propagation of the operator average in truncated space

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Tokyo, 153, Japan

(Received 7 September 1984; accepted for publication 11 January 1985)

The propagation of the operator average is described in truncated space with some quantum number(s) being fixed. It is first shown that the propagation coefficient satisfies an analog of the Chapman–Kolmogorov equation. Next, particle-hole symmetry is incorporated into the propagation of the operator average. It yields an expression that facilitates evaluation of many-body trace. Fermion and boson systems are treated alike.

I. INTRODUCTION

Trace (or average) of an operator in the truncated space of n particles being distributed over definite N orbits has been studied in the analysis of atomic and nuclear spectroscopy.[1–7] The trace of interest is defined on the wave functions with some quantum number(s) λ other than n being fixed.

The operator average in the truncated space propagates in some cases.[1] Propagation here implies that the n-body trace (or average) of an operator is expressed as a linear combination of the first few-body traces of the same kind which are called input traces. The proportional coefficient that relates the n-body trace to each input trace is called the propagation coefficient.[1]

Realization of propagation of operator average relies on λ, though little attention has been given to boson cases.[2] Some examples of λ 's yielding the propagation in a fermion system are isospin,[3,4] seniority,[1,4] and quantum numbers specifying[5] U(4). Each of them is surely associated with a chain of the group. It is, however, unclear if conversely there would be a chain of the group in case the operator average propagates.

The purpose of the present paper is to describe properties of propagation of the λ-fixed operator average. Fermion and boson systems are treated alike throughout the work. Neither the group theoretical premise nor explicit λ is required. We first show that the propagation coefficient satisfies an analog of the Chapman–Kolmogorov equation. It implies that the propagation coefficient is akin to the Green function or propagator. The equation is rewritten as a difference equation which can be illustrated by a branching diagram with weighted paths. Next, we incorporate particle-hole symmetry of the operator into the propagation of the operator average. French[3] treated it for fermions as a trace network problem. Making use of relations among many-body operators, we combine the global nature of particle-hole symmetry with the local nature of the propagation of average, and obtain a simple result which does not require us to solve the trace network problem at all. Particle-hole symmetry is defined also in a boson system so as to treat fermion and boson systems alike. The present manipulation reduces the number of input traces and, therefore, facilitates evaluation of many-body traces.

Section II concerns the definition of the propagation of the operator average. In Sec. III, properties of the propagation coefficients are described in connection with relations for many-body operators. In Sec. IV, the operator satisfying particle-hole symmetry is expressed as a sum of mutually independent many-body operators. The result is used in Sec. V to combine the propagation of the λ-fixed operator average with particle-hole symmetry.

II. PROPAGATION OF THE λ-FIXED TRACE

The symbol $\langle\langle\ \rangle\rangle$ denotes the trace

$$\langle\langle V_k\rangle\rangle^{n\lambda}=\sum_{\mu}\langle n\lambda\mu|V_k|n\lambda\mu\rangle=\mathrm{Tr}(\rho_{n\lambda}V_k), \tag{1}$$

where V_k stands for a k-body operator. The n-body state is specified by quantum numbers λ and μ. We define $\rho_{n\lambda}$ by

$$\rho_{n\lambda}=\sum_{\mu}|n\lambda\mu\rangle\,\langle n\lambda\mu|. \tag{2}$$

The propagation of the average implies that

$$\frac{\langle\langle V_k\rangle\rangle^{n\lambda}}{d(n\lambda)}=\sum_{\lambda'}\frac{Z(n\lambda,k\lambda')\langle\langle V_k\rangle\rangle^{k\lambda'}}{d(k\lambda')}, \tag{3}$$

where $d(n\lambda)$ indicates the dimension of the space with n and λ being fixed. The factor $Z(n\lambda,k\lambda')$ is called the propagation coefficient. The left-hand side (lhs) of (3) indicates the average of V_k.

The k-body operator V_k is expressed as

$$V_k=\sum\langle k\lambda'\mu'|V_k|k\lambda''\mu''\rangle A^+(k\lambda'\mu')A(k\lambda''\mu''), \tag{4}$$

where the sum is over repeated Greek indices. The operator $A^+(k\lambda'\mu')$, made of fermion (or boson) creation operators of order k, creates the orthonormalized k-body state.[6] Its conjugate is denoted as $A(k\lambda'\mu')$. These are called state operators. The relation (3) is rewritten as

$$\langle\langle A^+(k\lambda'\mu')A(k\lambda''\mu'')\rangle\rangle^{n\lambda}/d(n\lambda)=\delta(\lambda',\lambda'')\delta(\mu',\mu'')Z(n\lambda,k\lambda')/d(k\lambda'). \tag{5}$$

For the case of $\lambda=T$ (isospin) and $\mu=$ the other quantum numbers including T_z, for example, the propagation coefficient is expressed as[4]

$$Z(nT,kT')=\left\{\binom{T+n/2+1}{T'+k/2+1}\binom{n/2-T}{k/2-T'}-\binom{T+n/2+1}{k/2-T'}\binom{n/2-T}{T'+k/2+1}\right\}\times(2T'+1)/(2T+1). \tag{6}$$

It satisfies

$$Z(nT,kT')=Z(n+2T\,0,kT')\times\sum_m\binom{n/2-T}{m}\left\{(2T'+1)\binom{T+n/2}{m}\right\}^{-1}, \tag{7}$$

 0022-2488/85/050965-05$02.50

where m runs from $k/2 - T'$ to $k/2 + T'$. To deduce (7) from (6), we have only to use (4.1) of Ref. 8. The relation (6) is valid also for $t = \frac{1}{2}$ bosons.

III. PROPERTIES OF PROPAGATION COEFFICIENTS IN CONNECTION WITH RELATIONS FOR MANY-BODY OPERATORS

We start from the following premise, in place of (5) itself:

$$\langle\langle A^+(n-1\,\lambda'\mu')A(n-1\,\lambda''\mu'')\rangle\rangle^{n\lambda}$$
$$= \delta(\lambda',\lambda'')\delta(\mu',\mu'')\times(\mu' - \text{independent factor}) \quad (8)$$
$$= \delta(\lambda',\lambda'')\delta(\mu',\mu'')Z(n\lambda,n-1\,\lambda')d(n\lambda)\times[d(n-1\,\lambda')]^{-1}, \quad (9)$$

which is (5) for $k = n-1$. We regard (9) as the defining relation of the propagation coefficients. We discuss (9) itself further in Appendix A.

By the induction method we prove that the relation (9) leads to (5). The relation (5) for $n = k+1$ reads (9). We assume (5) to be true for $n = n_0$ and derive it for $n = n_0 + 1$. We have the identity,[6] valid for fermions and bosons alike,

$$\sum \langle lv\sigma|A^+(k\lambda'\mu')A(k\lambda''\mu'')|lv'\sigma'\rangle \times A^+(lv\sigma)A(lv'\sigma') = \binom{\tilde{n}-k}{l-k}A^+(k\lambda'\mu')A(k\lambda''\mu''), \quad (10)$$

where the sum is taken over repeated Greek indices. The symbol $\tilde{n}$ indicates the number operator. Using (9) and (10) yields

$$(n_0+1-k)\langle\langle A^+(k\lambda'\mu')A(k\lambda''\mu'')\rangle\rangle^{n_0+1,\nu} = \sum_{\nu'} Z(n_0+1\,\nu,n_0\nu')\langle\langle A^+(k\lambda'\mu')A(k\lambda''\mu'')\rangle\rangle^{n_0\nu'} \times\frac{d(n_0+1\,\nu)}{d(n_0\nu')}. \quad (11)$$

Applying (5) for $n = n_0$ to the right-hand side (rhs), we obtain (5) for $n = n_0 + 1$ together with the result

$$(n_0+1-k)Z(n_0+1\,\nu,k\lambda') = \sum_{\nu'} Z(n_0+1\,\nu,n_0\nu')Z(n_0\nu',k\lambda'). \quad (12)$$

Changing notations in the last relation gives

$$\sum_{\nu} Z(n\lambda,n-1\,\nu)Z(n-1\,\nu,k\lambda') = (n-k)Z(n\lambda,k\nu'). \quad (12')$$

It can be regarded as the basic difference equation for the propagation coefficient once the explicit form of $Z(n\lambda,n-1\,\nu)$ is known. Actually, it, combined with (A2) and (A3) in Appendix A, yields the difference equation (31) of Ref. 4. From (12') we obtain, as easily proved by induction,

$$\sum_{\nu} Z(n\lambda,l\nu)Z(l\nu,k\lambda') = \binom{n-k}{l-k}Z(n\lambda,k\lambda'). \quad (13)$$

Applying (10) to the k-body operator in (5) also yields (13). The binomial coefficient in (13) is ascribed to normalization of the propagation coefficient, and will vanish if we use $Z(n\lambda,k\lambda')[\binom{n}{k}]^{-1}$ as the normalized propagation coefficient. The relation (13) is akin to the Chapman–Kolmogorov equation characteristic of Markov chains. The quantum numbers λ and n in the former correspond to space and time in the latter, respectively. The propagation coefficient may well be said to be an analog of the Green function or the propagator in this sense. French[3] noticed a similarity between the propagation coefficient and the Green function in a sense different from (13). The relation (12') in case $k = n-2$ reads

$$\sum_{\nu} Z(n\lambda,n-1\,\nu)Z(n-1\,\nu,n-2\,\lambda') = 2Z(n\lambda,n-2\,\lambda'). \quad (14)$$

Conversely, we easily get (12') from (14). A branching diagram with the path being weighted by $Z(n\lambda,n-1\,\lambda')$ is available to express (14), as used in Ref. 4. The relation (13) for $l = k+1$ reads

$$\sum_{\nu} Z(n\lambda,k+1\,\nu)Z(k+1\,\nu,k\lambda') = (n-k)Z(n\lambda,k\lambda'). \quad (15)$$

It is also a difference equation for the propagation coefficient. Let us multiply $(-1)^l$ on both sides of (13) and sum over l from k to an arbitrary integer $s(<n)$. Then, we get

$$\sum_{l=k}^{s}\sum_{\nu} Z(n\lambda,l\nu)Z(l\nu,k\lambda')(-1)^{l-s} = \binom{n-k-1}{s-k}Z(n\lambda,k\lambda'). \quad (16)$$

The last relation for $s = n$ is the orthogonality relation

$$\sum_{l=k}^{n}\sum_{\nu} Z(n\lambda,l\nu)Z(l\nu,k\lambda')(-1)^{l-k} = \delta(n,k)\delta(\lambda,\lambda'). \quad (17)$$

For $\lambda = T$, the orthogonality relation was given in (A.4) of Ref. 4.

Let O be a general operator which is expanded as

$$O = \sum_{k=0}^{u} V_k, \quad (18)$$

where u is called the maximum particle rank of O. From (3) it follows that

$$\frac{\langle\langle O\rangle\rangle^{n\lambda}}{d(n\lambda)} = \sum_{k=0}^{u}\sum_{\lambda'}\frac{Z(n\lambda,k\lambda')\langle\langle V_k\rangle\rangle^{k\lambda'}}{d(k\lambda')}. \quad (19)$$

Its inversion is expressed by using (17) as

$$\frac{\langle\langle V_k\rangle\rangle^{k\lambda'}}{d(k\lambda')} = \sum_{l=0}^{k}\sum_{\nu}\frac{Z(k\lambda',l\nu)\langle\langle O\rangle\rangle^{l\nu}(-1)^{k-l}}{d(l\nu)}. \quad (20)$$

We substitute (20) in (19) and, subsequently, use (16). Then, we get

$$\langle\langle O\rangle\rangle^{n\lambda}/d(n\lambda) = \sum_{l}\binom{n-l-1}{u-l}(-1)^{u-l}\sum_{\nu}\frac{Z(n\lambda,l\nu)\langle\langle O\rangle\rangle^{l\nu}}{d(l\nu)}. \quad (21)$$

It was derived by French[7] in a different way.

The k-hole operator $\tilde{V}_k$ is defined by

$$\tilde{V}_k = B(k)\sum\langle k\lambda'\mu'|V_k|k\lambda''\mu''\rangle A(k\lambda''\mu'')A^+(k\lambda'\mu'), \quad (22)$$

where $B(k)$ is the sign function given by

$$B(k) = 1 \text{ for fermions and } (-1)^k \text{ for bosons.} \tag{23}$$

The sign function $F(k)$ is defined by

$$F(k) = (-1)^k \text{ for fermions and 1 for bosons.} \tag{24}$$

The k-hole operator differs from the k-particle operator of (4) only in the ordering of A^+ and A. The term "hole" here implies antinormal ordering of them and, therefore, is used also for a boson system. The sign $B(k)$ in (22) is beyond a matter of convention. Without it, any of (30) and (51) given later would not be valid for bosons.

We commute a pair of operators in (22) using[6,9]

$$\begin{aligned}A(k'\lambda'\mu')A^+(k\lambda\mu) &= \sum \langle k'\lambda'\mu'|A^+(l'\nu'\sigma')A(l\nu\sigma)|k\lambda\mu\rangle \\ &\quad\times A^+(l\nu\sigma)A(l'\nu'\sigma')F(kk'+k-l),\end{aligned} \tag{25}$$

where the sum is taken over l and repeated Greek indices. The sign factor F, defined by (24), is characteristic of fermions. We summarize the resultant expression from (22) as

$$\widetilde{V}_k = B(k)\sum_{l=0}^{k} F(l)V_{k,l}, \tag{26}$$

where $V_{k,l}$ indicates the l-body operator generated from V_k by $(k-l)$ times contractions[2,6] and is given by

$$\begin{aligned}V_{k,l} &= \sum \langle k\lambda\mu|V_k|k\lambda'\mu'\rangle \\ &\quad\times\langle k\lambda'\mu'|A^+(l\nu'\sigma')A(l\nu\sigma)|k\lambda\nu\rangle \\ &\quad\times A^+(l\nu\sigma)A(l\nu'\sigma').\end{aligned} \tag{27}$$

The operator $V_{k,k}$ reads V_k. The l-body trace of $V_{k,l}$ is associated with the k-body trace of V_k as

$$\langle\langle V_{k,l}\rangle\rangle^{l\nu} = \sum_{\lambda} Z(k\lambda,l\nu)\langle\langle V_k\rangle\rangle^{k\lambda}. \tag{28}$$

To get it, we use (5) and (27). From (27), it follows that[2]

$$V'_{l,m} \equiv V_{(k,l),m} = \binom{k-m}{l-m}V_{k,m}. \tag{29}$$

Using both (26) and (29), we see that $\widetilde{V}''_k = V_k$ for the case $V''_k = \widetilde{V}_k$. The relation (13) is obtained again from m-body traces of the operators on both sides of (29). From (22), it follows that

$$(\widetilde{V_k V_l}) = \widetilde{V}_l\widetilde{V}_k, \tag{30}$$

which is checked by using (25).

For fermions, the following symmetry holds under suitable choice of quantum numbers:

$$\begin{aligned}&\langle l\nu\sigma|A^+(k\lambda\mu)A(k\lambda'\mu')|l\nu'\sigma'\rangle \\ &= \langle N-l\,\nu'\sigma'|A(k\lambda'\mu')A^+(k\lambda\mu)|N-l\,\nu\sigma\rangle \qquad (31)\\ &= \langle N-k\,\lambda\mu|A^+(N-l\,\nu\sigma) \\ &\quad\times A(N-l\,\nu'\sigma')|N-k\,\lambda'\mu'\rangle. \qquad (32)\end{aligned}$$

The last relation results from (31). Using (31) and (32) with $k=0$, we get

$$d(n\lambda) = Z(N0,n\lambda). \tag{33}$$

The symmetry (31) implies

$$\langle N-n\,\lambda\mu|\widetilde{V}_k|N-n\,\lambda'\mu'\rangle = \langle n\lambda'\mu'|V_k|n\lambda\mu\rangle. \tag{34}$$

Applying (26) to $\widetilde{V}_k$ in the matrix element on the lhs, we get

$$\sum_{l=0}^{k}(-1)^l\langle\langle V_{k,l}\rangle\rangle^{N-n,\lambda} = \langle\langle V_k\rangle\rangle^{n\lambda}. \tag{35}$$

Using (3) and (28), we express both sides in terms of the k-body trace of V_k. Then, we obtain a new relation for the propagation coefficients:

$$\sum_{l\nu}\frac{(-1)^l Z(n\lambda,l\nu)Z(k\lambda',l\nu)}{d(l\nu)} = \frac{Z(N-n\,\lambda,k\lambda')}{d(k\lambda')}. \tag{36}$$

The relations (31)–(36) are characteristic of a fermion system.

IV. A GENERAL FORM OF OPERATOR SATISFYING PARTICLE-HOLE SYMMETRY

Here, we discuss the expansion (18) in detail, imposing particle hole symmetry on O as

$$\widetilde{O} = (-1)^u O, \tag{37}$$

i.e.,

$$\sum_k \widetilde{V}_k = (-1)^u\sum_k V_k. \tag{38}$$

The sign $(-1)^u$ in the above expressions is due to the fact that $\widetilde{V}_u$ involves $(-1)^u V_u$, as seen in (26).

Substituting (26) in (38) yields

$$V_l = (-1)^u F(l)\sum_{k=l}^{u} B(k)V_{k,l}. \tag{39}$$

This, for $l = u-2p-1$, reads

$$V_{u-2p-1} = F(1)\sum_{k=0}^{2p}\frac{1}{2}B(k)V_{u-k,u-2p-1}. \tag{40}$$

The relation (39) for $u-l$ being even does not provide any new information, since it is shown to be deduced from (40).

Let us derive from (40)

$$V_{u-2p-1} = F(1)\sum_{k=0}^{p}\frac{1}{2}C_{2k+1}V_{u-2p+2k,u-2p-1}, \tag{41}$$

where C_{2k+1} is a definite number determined by

$$\sum_{k=0}^{n}\binom{2n+1}{2k+1}C_{2k+1} = 2C_1 - C_{2n+1}, \tag{42}$$

with $C_1 = 1$.

Proof: We prove (41) by induction on p. In case $p=0$ it reads (40) with $p=0$. Supposing (41) with $p<q$, let us derive (41) with $p=q+1$. From (41) with $r\leqslant q$ it follows that

$$\begin{aligned}&V_{u-2r-1,u-2q-3} \\ &= F(1)\sum_{k=0}^{r}\frac{1}{2}C_{2k+1}\binom{2k-2r+2q+3}{2k+1} \\ &\quad\times V_{u-2r+2k,u-2q-3}.\end{aligned} \tag{43}$$

To get the rhs, the relation (29) is used. The rhs of (40) for $p=q+1$ is split into two parts according to whether k is even or odd:

$$\begin{aligned}V_{u-2q-3} &= F(1)\sum_{r=0}^{q+1}\frac{1}{2}V_{u-2r,u-2q-3} \\ &\quad-\sum_{r=0}^{q}\frac{1}{2}V_{u-2r-1,2q-3}.\end{aligned} \tag{44}$$

We put (43) in the second term on the rhs. Applying the defining relation (42) to the resultant expression, we get (41) with $p = q + 1$. Q.E.D.

To determine the explicit form of C_{2k+1}, we solve (42) for $n = 1,2,\ldots,k$. Putting $n = 1$ in (42) yields $C_1 = -2C_3$ from which we get $C_3 = -\frac{1}{2}$. From (42) with $n = 2$, it follows that $3C_1 + 10C_3 = -2C_5$. Using the known values of C_1 and C_3, we get $C_5 = 1$. Repeating this way, we get $C_7 = -\frac{17}{4}$, $C_9 = 31$, $C_{11} = -\frac{691}{2}$, $C_{13} = 5461$, etc. The relation (41) implies that the terms $\{V_{u-1}, V_{u-3},\ldots\}$ are uniquely determined by $\{V_u, V_{u-2},\ldots\}$ as a consequence of the symmetry (37).

The expansion (41) is substituted into (18) to yield

$$O = \sum_m R(V_{u-2m}), \tag{45}$$

where $m = 0,1,\ldots,[u/2]$, which denotes the largest integer contained in $u/2$, and

$$R(V_k) \equiv V_k + F(1) \sum_l \frac{1}{2} C_{2l+1} V_{k,k-2l-1}. \tag{46}$$

The index l runs over $0,1,\ldots,[(k-1)/2]$. The operator $R(V_k)$ is generated from V_k. Its particle rank is indefinite ($\leqslant k$).

Let us show that the operator $R(V_k)$ has the definite particle-hole symmetry:

$$\widetilde{R}(V_k) = (-1)^k R(V_k). \tag{47}$$

Proof: The operator $\widetilde{R}(V_k)$ is transformed, by using (26) and (29), as

$$\widetilde{R}(V_k) = B(k) \sum_{m=0}^{k} F(m) V_{k,m} - B(k) \sum_l C_{2l+1} \sum_{m=0}^{k-2l-1} \frac{1}{2} F(m) \binom{k-m}{2l+1} V_{k,m}, \tag{48}$$

where the sum of l is taken over $0,1,\ldots,[(k-1)/2]$. Changing the order of the summation over l and m yields

$$\widetilde{R}(V_k) = B(k) \sum_{m=0}^{k} F(m) V_{k,m} \left\{ 1 - \sum_l \frac{1}{2} C_{2l+1} \binom{k-m}{2l+1} \right\}, \tag{49}$$

where l runs over $0,1,\ldots,[(k-m-1)/2]$. In case $m = k$, the sum over l seen on the rhs reads 0. In case $m > k$, the summation over l is easily done by using

$$\sum_k \binom{b}{2k+1} C_{2k+1} \{1 + \delta(b, 2k+1)\} = 2C_1, \tag{50}$$

where $b > 0$, and k runs over possible values of k, i.e., $0,1,\ldots,[(b-1)/2]$. For b being odd the relation (50) reads as the defining relation (42). We see that the rhs of (49) vanishes for $k - m$ being nonzero even. Replacing the argument m in (49) by a new argument$\equiv (k - m - 1)/2$, we obtain (47) from (49). Q.E.D.

The operator O expressed in the form of (45) satisfies the symmetry (37) because of (47). It is now concluded that the expression (45) together with (46) is the necessary and sufficient condition for O to satisfy (37).

In Appendix B, the decomposition (45) is compared with the unitary scalar decomposition discussed in Refs. 2, 6, and 7.

V. TRACE PROPAGATION COMBINED WITH PARTICLE-HOLE SYMMETRY

For fermions where (34) is valid, the n-hole and the n-particle traces agree with each other in magnitude, if the operator satisfies the symmetry (37). This property is not reflected in (19). Equating (19) to the same expression with n in it being replaced by $N - n$ produces coupled linear equations that provide mutual dependence among input traces. Selection of independent input traces requires us to solve the equations numerically. French[3] called this the trace network problem. Here, we solve the problem without recourse to the trace network problem. The present manipulation relies on the result in the last section and is applied to a boson system as well as a fermion system.

Let us apply (3) and (28) to an n-body trace of O which is expanded as (45). We then obtain

$$\frac{\langle\langle O \rangle\rangle^{n\lambda}}{d(n\lambda)} = \sum_{m\lambda'} \frac{X(n\lambda, u-2m\,\lambda')\langle\langle V_{u-2m}\rangle\rangle^{u-2m,\lambda'}}{d(u-2m,\lambda')}, \tag{51}$$

where m runs over $0,1,\ldots,[u/2]$. The coefficient X is defined by

$$X(n\lambda, k\lambda') \equiv Z(n\lambda, k\lambda') + F(1) d(k\lambda') \sum_l C_{2l+1} \times \sum_\sigma \frac{Z(n\lambda, k-2l-1\,\sigma) Z(k\lambda', k-2l-1\,\sigma)}{2d(k-2l-1\,\sigma)}, \tag{52}$$

where l runs over $0,1,\ldots,[(k-1)/2]$.

For fermions where (34) is valid, the coefficient (52) is symmetric (antisymmetric) in n-$N/2$ for k being even (odd). In case it is a polynomial in n, it is expressed as a polynomial in $n - N/2$ of degree k as (13) suggests. For example, we get, using (6),

$$X(nT, 3\ 1/2) = (n - N/2)\{2(n - N/2)^2 - 8T(T+1) - 3N^2/2 + 4\}/24. \tag{53}$$

For a boson system, the coefficient (52) is a function in $n + N/2$, as suggested by (4.12) of Ref. 2. Notice that the coefficient (52) survives even if $n < k$ and that it does not satisfy (13).

The input traces in (51) are

$$\langle\langle V_u \rangle\rangle^{u\lambda},\ \langle\langle V_{u-2}\rangle\rangle^{u-2\lambda'},\ldots,\langle\langle V_0\rangle\rangle^0 \text{ or } \langle\langle V_1 \rangle\rangle^1, \tag{54}$$

which are a subset of the input traces in (19). The number of input traces is much reduced in (51). It implies an advantage to (51) in the trace calculation. The propagation coefficient $Z(n\lambda, k\lambda')$ for a fermion system becomes very large as n approaches N. Since the n-body average is not large, there occurs remarkable cancellation among terms on the rhs of (19). It will cause a problem in numerical calculation. Such a situation will be avoided by using (51) because the coefficient (52) is symmetric in n-$N/2$.

Among input traces (54), the trace $\langle\langle V_u \rangle\rangle^{u\lambda}$ is most troublesome in actual calculation. We can avoid it by replacing the input trace by new inputs $\{\langle\langle V_k \rangle\rangle^{k\lambda};\ u - k = \text{odd}\}$ which are excluded from (54). Let us consider, for example, the T-fixed trace with $u = 4$. Using (41) together with (28) yields

$$\langle\langle V_3\rangle\rangle^{3T} = F(1) \sum_{T'} \frac{1}{2} Z(4T',3T)\langle\langle V_4\rangle\rangle^{4T'} \tag{55}$$

and

$$\langle\langle V_1\rangle\rangle^{1,1/2} = \Big\{2 \sum_{T'} Z(2T',1\ 1/2)\langle\langle V_2\rangle\rangle^{2T'} - \sum_{T'} Z(4T',1\ 1/2)\langle\langle V_4\rangle\rangle^{4T'}\Big\} \frac{F(1)}{4}. \tag{56}$$

Solving them, we can express $\langle\langle V_4\rangle\rangle^{4T}$ in terms of $\langle\langle V_3\rangle\rangle^{3T'}$, where $T' = \frac{1}{2}$ and $\frac{3}{2}$, and $\langle\langle V_1\rangle\rangle^{1,1/2}$.

It is straightforward to extend (51) to a general operator which breaks particle-hole symmetry (37). We have only to decompose the operator O' as

$$O' = (O' + \widetilde{O}')/2 + (O' - \widetilde{O}')/2 \tag{57}$$

so that each term on the rhs should satisfy (37). In the usual cases, the operator O' has the form of an operator product. The relation (30) is then conveniently used to get $\widetilde{O}'$ from O'.

APPENDIX A

The relation (9) is expressed in terms of coefficients of fractional parentage (cfp's) (see Ref. 6) as

$$\begin{aligned} n \sum_{\mu\nu\sigma} &\langle n-1,\lambda'\mu' + 1,\nu,\sigma|n,\lambda\mu\rangle \\ &\times \langle n-1,\lambda''\mu'' + 1,\nu,\sigma|n\lambda\mu\rangle \\ &= \delta(\lambda',\lambda'')\delta(\mu',\mu'')Z(n\lambda,n-1\,\lambda')d(n\lambda) \\ &\times [d(n-1\,\lambda')]^{-1}. \end{aligned} \tag{A1}$$

It is crucial that the coefficient Z be independent of μ'. This type of relation is realized in case cfp is factorized[10] into a few parts as a result of a chain of the group. If there is a chain of the group attached to λ and μ, the propagation coefficient can be evaluated from cfp or a Clebsch–Gordan coefficient characteristic of a chain of the group. For the case of $\lambda = T$, we easily evaluate[4] the lhs of (A1) and get

$$Z(nT,n-1\ T-1/2) = T(n+2T+2)/(2T+1) \tag{A2}$$

and

$$Z(nT,n-1\ T+1/2) = (T+1)(n-2T)/(2T+1), \tag{A3}$$

which is valid for $t = \frac{1}{2}$ fermions and bosons alike. Each of them being divided by n is just a ratio of dimensions of the representation of symmetric group.

We get the same relation as (13) if we start from

$$\Big\langle n\lambda\mu\Big|\sum_{\mu'} A^{+}(n-1\,\lambda'\mu')A(n-1\,\lambda'\mu')\Big|n\lambda''\mu''\Big\rangle = \delta(\lambda,\lambda'')\delta(\mu,\mu'')Z(n\lambda,n-1\,\lambda') \tag{A4}$$

as a premise instead of (9). For a fermion system where (32) is valid, both (9) and (A4) coincide with each other.

APPENDIX B

We compare the decomposition (45) with the unitary scalar decomposition.[2,6,7]

An arbitrary k-body operator V_k can be expanded as

$$V_k = \sum_{m=0}^{k} \frac{1}{2} R(V_{k,m})\{1 + \delta(k,m)\}, \tag{B1}$$

as easily be checked by using (46). The operator $(V_k + \widetilde{V}_k)/2$ is expressed as a sum of even m terms on the rhs, and fulfills particle-hole symmetry as is expected. The unitary scalar decomposition implies

$$V_k = \sum_m \binom{\hat{n}-m}{k-m} U_m, \tag{B2}$$

where U_m is the m-body operator which satisfies (47) as $R(V_{k,m})$ in (B1) does. While the number operator $\hat{n}$ appears in (B2), it does not in (B1). The operator $R(V_{k,m})$ in (B1) has indefinite particle rank ($\leqslant m$), while the particle rank of U_m in (B2) is the definite value m. The relation (46) in case $V_k = U_k$ reads $R(U_k) = (-1)^k U_k$, as $U_{k,p}$ for $p < k$ always vanishes. The operators $\{U_m\}$ and $\{V_{k,m}\}$ are related to each other through (2.21) and (2.22) of Ref. 2.

[1] J. B. French, *Nuclear Structure*, edited by A. Hossain, Harum-ar-Raschid, and M. Islam (North-Holland, Amsterdam, 1967), p. 85.
[2] M. Nomura, Ann. Phys. **91**, 83 (1975).
[3] J. B. French, *Isospin in Nuclear Physics*, edited by D. H. Wilkinson (North-Holland, Amsterdam, 1969), p. 264.
[4] M. Nomura, Prog. Theor. Phys. **51**, 489 (1974).
[5] C. Quesne, *Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1980), p. 257.
[6] M. Nomura, Prog. Theor. Phys. **59**, 451 (1978).
[7] F. S. Chang, J. B. French, and T. H. Thio, Ann. Phys. **66**, 137 (1971).
[8] H. W. Gould, "Combinatorial Identities," unpublished note, West Virginia University, 1972.
[9] M. Nomura, J. Math. Phys. **26**, 732 (1985).
[10] G. Racah, Phys. Rev. **76**, 1352 (1949).

Strong approximation of time evolution operators for a finite system of oscillators with nonlinear coupling

Helmut Kröger
Département de Physique, Université Laval, Québec, P.Q., Canada G1K 7P4

(Received 14 September 1984; accepted for publication 21 December 1984)

We consider a system in three space dimensions consisting of a finite number of oscillators with a nonlinear interaction. Using projectors on N-particle subspaces of the Fock space, we show that the time evolution operator is strongly approximatable by exponentials of self-adjoint finite-rank operators (finite-dimensional Hermitian matrices), which can easily be calculated in the corresponding eigenrepresentation.

I. INTRODUCTION

It has been suggested in Refs. 1–4 for nonrelativistic N-body potential scattering to approximate strongly the wave operators and weakly the S matrix by exponentials of bounded self-adjoint or even finite-rank self-adjoint operators. In the latter case the approximate wave operators and the approximate S matrix can be calculated easily using the eigenrepresentations of the finite-dimensional Hermitian matrices corresponding to the full and asymptotic Hamiltonian, respectively.

This approach has been tested successfully in the two-body system for a variety of short-range nucleon-nucleon potentials plus the long-range Coulomb potential.[2] It has been applied to the three-body charged particle process $d + p \rightarrow p + p + n$ (see Ref. 2).

Let us recall some of the basic features of the approach.

(i) As long as the wave operators exist, the interaction can be rather arbitrary, in particular it works for every real coupling constant.

(ii) The strong approximation of the wave operators and the weak approximation of the S matrix is based on strong resolvent convergence of the approximated full and asymptotic Hamiltonian, respectively. That is a rather "weak" condition and it can be generalized to functions of operators.

(iii) In contrast to stationary (multichannel) scattering theory, where the scattering process is expressed in terms of Green's functions which contain singularities and which correspond to the physical picture of multiple vertices and free propagation, our approach is by construction free of singularities.

The properties (i)–(iii) encourage us to try to extend the approach to field theoretic models. As a first step in this direction we want to consider in this paper the time evolution operator for a system consisting of a finite number of oscillators with a nonlinear coupling. The purpose of this paper is twofold. First, for a finite system of oscillators with a nonlinear coupling, which cannot be solved via normal coordinates as in the linear coupling case, an approximation scheme is provided. Second, the ground work is laid for a generalization to field theoretical models.

The paper is organized as follows. In Sec. II we discuss the approximation in detail for an anharmonic oscillator in one space dimension (i.e., one oscillator with a nonlinear self-interaction). The generalization to a finite number of oscillators in three space dimensions, given in Sec. III is then very straightforward.

II. THE ANHARMONIC OSCILLATOR IN ONE SPACE DIMENSION

The harmonic oscillator Hamiltonian in one space dimension is given by

$$H^0 = p^2/2m + (m\omega^2/2)x^2 = \omega(a^+a + \tfrac{1}{2}), \tag{2.1}$$

where

$$a = (m\omega/2)^{1/2}x + [i/(2m\omega)^{1/2}]p, \tag{2.2}$$

$$a^+ = (m\omega/2)^{1/2}x - [i/(2m\omega)^{1/2}]p, \tag{2.3}$$

$$[a,a^+] = 1. \tag{2.4}$$

We consider all the operators to act in the Fock space $\mathcal{F}$, which is a Hilbert space. The vacuum state $|0\rangle$ is defined by

$$a|0\rangle = 0. \tag{2.5}$$

The n-particle states

$$|n\rangle = (a^{+n}/\sqrt{n!})|0\rangle \tag{2.6}$$

form a complete orthonormal basis. The particle number operator N is given by

$$N = a^+a \tag{2.7}$$

and yields

$$N|n\rangle = n|n\rangle. \tag{2.8}$$

We consider the interaction Hamiltonian

$$H^{\text{int}} = \exp(-N) \sum_{i,j=0}^{G} h_{ij}^{\text{int}} a^{+i} a^{j} \exp(-N) \tag{2.9}$$

and a total Hamiltonian

$$H = H^0 + H^{\text{int}}. \tag{2.10}$$

Because we want to deal with self-adjoint operators, we require the complex $G \times G$ matrix h^{int} to be symmetric. The exponential operator has been introduced to generate a fall-off behavior, such that H^{int} becomes bounded.

Proposition 1: H^0 with the domain $D(H^0) \subset \mathcal{F}$ is a self-adjoint operator; H^{int} has the domain $D(H^{\text{int}}) = \mathcal{F}$ and is self-adjoint and bounded; and H has the domain $D(H) = D(H^0)$ and is self-adjoint.

Proof: Obviously H^0 is symmetric, and

$$D(H^0) = \left\{ \psi | \psi \in \mathcal{F}, |\psi\rangle = \sum_{n=0}^{\infty} \psi_n |n\rangle, \sum_{n=0}^{\infty} n^2 |\psi_n|^2 < \infty \right\} \tag{2.11}$$

is dense in $\mathcal{F}$. For each $\psi \in \mathcal{F}$ there are $\psi_\pm \in \mathcal{F}$ given by

$$|\psi_\pm\rangle = \sum_{n=0}^{\infty} \frac{\psi_n}{\omega(n + \frac{1}{2}) \pm i} |n\rangle. \tag{2.12}$$

 0022-2488/85/050970-03$02.50

Hence, one has

$$(H^0 \pm i)|\psi_\pm\rangle = |\psi\rangle, \tag{2.13}$$

which implies

$$\text{Range}(H^0 \pm i) = \mathscr{F}. \tag{2.14}$$

Then H^0 is self-adjoint by Theorem 5.21 in Weidmann's book.[5]

Next we claim that for each $i,j \in \mathbb{N}$ there is a polynomial

$$R^{(i+j)}(x) = \sum_{k=0}^{i+j} r_k x^k, \tag{2.15}$$

of degree $i+j$ with real coefficients r_k, such that for each $\psi \in D(N^{(i+j)/2})$, it holds that

$$\|a^{+i}a^j\psi\| \leqslant \|R^{(i+j)}(N^{1/2})\psi\|. \tag{2.16}$$

Application of the shift operator $a^{+i}a^j$ on $|k\rangle$ yields

$$a^{+i}a^j|k\rangle = \begin{cases} (k(k-1)\cdots(k-j+1)(k-j+1) \\ \quad \cdots(k-j+i))^{1/2}|k+i-j\rangle, & \text{if } k-j \geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

Hence

$$\begin{aligned} \|a^{+i}a^j\psi\|^2 &= \sum_{k=j}^{\infty} |(k(k-1)\cdots(k-j+1)(k-j+1) \\ &\quad \cdots(k-j+i))^{1/2}\psi_{k+i-j}|^2 \\ &= \sum_{n=i}^{\infty} |(n+j-i)(n+j-i-1) \\ &\quad \cdots(n-i+1)(n-i+1)(n-i+2)\cdots(n)\psi_n^2|, \end{aligned} \tag{2.18}$$

which establishes Eq. (2.16).

Thus one can find a polynomial $R^{(G)}(x)$ with real coefficients such that

$$\left\|\sum_{i,j=0}^{G} h_{ij}^{\text{int}} a^{+i}a^j\psi\right\| \leqslant \|R^{(G)}(N^{1/2})\psi\|, \tag{2.19}$$

for each $\psi \in D(N^{G/2})$.

We have that H^0 is self-adjoint, N is self-adjoint and non-negative, hence $\exp(-N)$ is bounded. Moreover $\exp(-N)$ maps $\mathscr{F}$ into $D(N^{k/2})$ for an arbitrary $k \in \mathbb{N}$. Thus we can estimate

$$\|H^{\text{int}}\psi\| \leqslant \|\exp(-N)\| \|R^{(G)}(N^{1/2})\exp(-N)\psi\|, \tag{2.20}$$

for all $\psi \in \mathscr{F}$. Obviously $R^{(G)}(N^{1/2})\exp(-N)$ is bounded. Hence is H^{int} a bounded operator with $D(H^{\text{int}}) = \mathscr{F}$. Because H^{int} is also symmetric, we can conclude that H^{int} is self-adjoint.[6] For H^{int} to be bounded means H^{int} is relatively H^0 bound with a H^0 bound 0. Thus Rellich–Kato's Theorem[7] implies that H is self-adjoint and $D(H) = D(H^0)$.

Now we introduce finite-dimensional approximations. Let P_n denote the orthogonal projector on the subspace generated by $\{|0\rangle,|1\rangle,\ldots,|n\rangle\}$. We define

$$H_n^0 = P_n H^0 P_n, \tag{2.21}$$

$$H_n^{\text{int}} = P_n H^{\text{int}} P_n, \tag{2.22}$$

$$H_n = P_n H P_n. \tag{2.23}$$

Proposition 2: $H_n^0, H_n^{\text{int}}, H_n$ are self-adjoint, finite-rank operators having the domain $\mathscr{F}$.

Proof: H^0, H^{int}, H are self-adjoint due to Proposition 1, P_n is self-adjoint and of finite rank with $D(P_n) = \mathscr{F}$, Range $(P_n) \subset D(H^0)$ for all $n \in \mathbb{N}$. Hence $H_n^0, H_n^{\text{int}}, H_n$ are symmetric, bounded and $D(H_n^0) = D(H_n^{\text{int}}) = D(H_n) = \mathscr{F}$, thus $H_n^0, H_n^{\text{int}}, H_n$ are self-adjoint.[6]

Now we introduce the time evolution operators for $t \in \mathbb{R}$

$$U^0(t) = \exp(iH^0 t), \tag{2.24}$$

$$U(t) = \exp(iHt), \tag{2.25}$$

$$U_n^0(t) = \exp(iH_n^0 t), \tag{2.26}$$

$$U_n(t) = \exp(iH_n t), \tag{2.27}$$

which are well defined, because H^0, H, H_n^0, H_n are self-adjoint.

Theorem 1: For every $t \in \mathbb{R}$ $U_n^0(t)$, $U_n(t)$ tend strongly to $U^0(t)$, $U(t)$, respectively, if n tends to infinity.

Proof: For any self-adjoint operator X, being densely defined in a Hilbert space, it holds that[8]

$$X = X^+ = \overline{X}. \tag{2.28}$$

For a closable operator Y, defined in a Hilbert space, the domain $D(Y)$ is a core of $\overline{Y}$ (see Ref. 9). Hence $D(X)$ is a core of X. This can be applied to H^0, H, i.e., $D(H^0)$ is a core of $H^0, D(H)$ is a core of H. Now we claim for every $\psi \in D(H^0) = D(H)$,

$$H_n^0\psi \to H^0\psi, \tag{2.29}$$

$$H_n\psi \to H\psi. \tag{2.30}$$

One has

$$|\psi\rangle = \sum_{n=0}^{\infty} \psi_n |n\rangle, \tag{2.31}$$

$$H^0|\psi\rangle = \sum_{n=0}^{\infty} \omega\left(n+\frac{1}{2}\right)\psi_n|n\rangle, \tag{2.32}$$

and $H^0|\psi\rangle$ is an element of $\mathscr{F}$, which means

$$\sum_{n=0}^{\infty} \left|\omega\left(n+\frac{1}{2}\right)\psi_n\right|^2 < \infty. \tag{2.33}$$

That implies

$$\|H^0(P_n - 1)\psi\|^2 = \sum_{k=n+1}^{\infty} \left|\omega\left(k+\frac{1}{2}\right)\psi_k\right|^2 \to 0, \tag{2.34}$$

if n tends to infinity. Because P_n tends strongly to 1, one also has

$$(P_n - 1)H^0\psi \to 0, \tag{2.35}$$

if n tends to infinity. Equations (2.34) and (2.35) imply Eq. (2.29).

Due to Proposition 1, H^{int} is bounded, P_n tends strongly to 1, $\|P_n\| = 1$, thus one concludes[10] that $P_n H^{\text{int}} P_n$ tends strongly to H^{int}, i.e., for every $\psi \in \mathscr{F}$,

$$P_n H^{\text{int}} P_n \psi \to H^{\text{int}}\psi. \tag{2.36}$$

Equations (2.29) and (2.36) imply Eq. (2.30). From H^0, H_n^0, H, H_n being self-adjoint, $D(H^0) = D(H)$ being a core of H^0, H, and Eqs. (2.29) and (2.30) we conclude[11] that H_n^0, H_n tend to H^0, H, respectively, in the sense of strong resolvent convergence. From that we conclude[12] that $U_n^0(t)$, $U_n(t)$ tend strongly to $U^0(t)$, $U(t)$, respectively, which proves the claim.

We want to conclude this section with a remark on the usefulness of the approximation $U_n(t)$. We have that $U^0(t)$ and hence $U_n^0(t)$ are already diagonal in the Fock space $\mathscr{F}$. But $U(t)$, $U_n(t)$ in general are not diagonal in $\mathscr{F}$. However, $U_n(t)$ can easily be calculated by diagonalizing the finite-dimensional Hermitian matrix corresponding to H_n and expressing $U_n(t)$ in the eigenrepresentation of H_n.

III. A FINITE SYSTEM OF OSCILLATORS IN THREE SPACE DIMENSIONS

In this section we want to generalize the model and the results of Sec. II to a finite system, of oscillators with nonlinear coupling in three space dimensions. Because all the proofs can be carried over essentially from Sec. II but require only more tedious writing, we will omit them here.

Let $a_{\rho,\nu}, a^+_{\rho,\nu}$ denote the annihilation and creation operators, respectively, for particles corresponding to the frequencies $\omega_{\rho,\nu}$, where $\rho = 1,2,3$ counts the space dimension and $\nu = 1,2,\ldots, \nu_m$ counts the oscillators in each space dimension. One has

$$[a_{\rho,\nu}, a^+_{\tau,\mu}] = \delta_{\rho,\tau}\delta_{\nu,\mu}, \tag{3.1}$$

$$[a_{\rho,\nu}, a_{\tau,\mu}] = [a^+_{\rho,\nu}, a^+_{\tau,\mu}] = 0. \tag{3.2}$$

We introduce a new counting $I = 1,\ldots,I_m$, $I_m = 3\nu_m$, where I corresponds one to one to a pair (ρ,ν). Hence Eqs. (3.1) and (3.2) read

$$[a_I, a_J^+] = \delta_{I,J}, \tag{3.3}$$

$$[a_I, a_J] = [a_I^+, a_J^+] = 0. \tag{3.4}$$

We put

$$H^0 = \sum_{I=1}^{I_m} \omega_I \left(a_I^+ a_I + \frac{1}{2}\right). \tag{3.5}$$

The Fock space $\mathscr{F}$ is a Hilbert space, with the vacuum $|0\rangle$ defined by

$$a_I|0\rangle = 0, \quad I = 1,\ldots,I_m, \tag{3.6}$$

and spanned by

$$|n_1,\ldots,n_{I_m}\rangle = \frac{(a_1^+)^{n_1}\cdots(a_{I_m}^+)^{n_{I_m}}}{(n_1!\cdots n_{I_m}!)^{1/2}}|0\rangle. \tag{3.7}$$

The particle number operator N is given by

$$N = \sum_{I=1}^{I_m} a_I^+ a_I, \tag{3.8}$$

and yields

$$N|n_1,\ldots,n_{I_m}\rangle = (n_1 + \cdots + n_{I_m})|n_1,\ldots,n_{I_m}\rangle. \tag{3.9}$$

We consider the interaction Hamiltonian

$$H^{\text{int}} = \exp(-N) \sum_{i,j=1}^{G} \sum_{I,J=1}^{I_m} h^{\text{int}}_{iIjJ} a_I^{+\,i} a_J^j \exp(-N) \tag{3.10}$$

and the total Hamiltonian

$$H = H^0 + H^{\text{int}}. \tag{3.11}$$

Again we assume the matrix h^{int} to be symmetric.

Proposition 3: Proposition 1 holds true, if H^0, H^{int}, H, and $\mathscr{F}$ are substituted by definitions given in this section.

Let P_n denote the orthogonal projector on the subspace generated by all states characterized by a particle number less or equal to n. Usually this subspace is denoted by $\mathscr{F}_0 \oplus \mathscr{F}_1 \oplus \cdots \oplus \mathscr{F}_n$. We define $H_n^0, H_n^{\text{int}}, H_n$ by Eqs. (2.21)–(2.23), where the definitions of $H^0, H^{\text{int}}, H, P_n$ are taken from this section. Then the following proposition holds.

Proposition 4: Proposition 2 holds true, if $H_n^0, H_n^{\text{int}}, H_n$, and $\mathscr{F}$ are substituted by the definitions of this section.

We define $U^0(t)$, $U(t)$, $U_n^0(t)$, and $U_n(t)$ by Eqs. (2.24)–(2.27) with the definitions of H^0, H, H_n^0, and H_n taken from this section. Then the following theorem holds.

Theorem 2: Theorem 1 holds if $U^0(t)$, $U(t)$, $U_n^0(t)$ and $U_n(t)$ are substituted by the definitions of this section.

IV. CONCLUSION

We have shown for a model consisting of a finite number of oscillators with a nonlinear coupling in three space dimensions that the time evolution operator corresponding to the full Hamiltonian can be approximated with arbitrary accuracy by a time evolution operator corresponding to an approximate full Hamiltonian. Due to the finite rank of the latter Hamiltonian, the corresponding time evolution operator can be calculated exactly, that means in numerical calculations within the accuracy of the diagonalization of a finite-dimensional Hermitian matrix and the calculation of the functions $\exp(ix)$ for real x. The model interaction studied here can be generalized to include unbounded operators in the form of a polynomial in the creation and annihilation operators.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge helpful discussions with W. Hengartner.

This work was supported by the National Science and Engineering Research Council of Canada.

[1]H. Kröger, J. Math. Phys. **24**, 1509 (1983).

[2]H. Kröger, Phys. Lett. B **135**, 1 (1984).

[3]H. Kröger, J. Math. Phys. **25**, 1875 (1984).

[4]H. Kröger, J. Math. Phys. **26**, 139 (1985).

[5]J. Weidmann, *Linear Operators in Hilbert Spaces* (Springer, New York 1980).

[6]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics, Lecture Notes and Supplements in Physics* (Benjamin, Reading, MA, 1977), p. 53, Remark.

[7]See Theorem 5.28, p. 111 in Ref. 5.

[8]See Theorem 5.3, in Ref. 5.

[9]See Remark on p. 88 in Ref. 5.

[10]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York 1976), p. 151, Lemma 3.8.

[11]See Theorem 9.16 in Ref. 5.

[12]See Theorem 9.18 in Ref. 5.

Lie algebras for systems with mixed spectra. I. The scattering Pöschl–Teller potential

Alejandro Frank
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, 04510 Mexico, D. F., Mexico
Kurt Bernardo Wolf[a)]
Departamento de Física, Centro de Investigación y Estudios Avanzados, Instituto Politécnico Nacional, 07000 Mexico, D. F., Mexico

(Received 23 March 1984; accepted for publication 4 October 1984)

Starting from an N-body quantum space, we consider the Lie-algebraic framework where the Pöschl–Teller Hamiltonian, $-\frac{1}{2}\partial_\chi^2 + c\,\mathrm{sech}^2\chi + s\,\mathrm{csch}^2\chi$, is the single sp(2,$R$) Casimir operator. The spectrum of this system is *mixed*: it contains a finite number of negative-energy bound states and a positive-energy continuum of free states; it is identified with the Clebsch–Gordan series of the $\mathscr{D}^+\times\mathscr{D}^-$ representation coupling. The wave functions are the sp(2,R) Clebsch–Gordan coefficients of that coupling in the parabolic basis. Using only Lie-algebraic techniques, we find the asymptotic behavior of these wave functions; for the special pure-trough potential ($s=0$) we derive thus the transmission and reflection amplitudes of the scattering matrix.

I. INTRODUCTION

Symmetry methods involving dynamical algebras have been long used to study the eigenstates and spectra of Schrödinger equations for certain one-dimensional potentials.[1–4] Notable among them are the hydrogen atom[1] (bound states and scattering states), the harmonic oscillator,[3] and the radial oscillator[5] (bound states only); among the systems with continuous spectra we know the free-fall (or linear) potential, the free particle, and the repulsive oscillator,[4] and the latter two in their welled versions. Here the symmetry method builds the dynamical algebra, and —in all but the first case, where it is the *pseudo*-Coulomb system which enjoys the algebra—the Hamiltonian is an element of this Lie algebra, which is[4,6] sp(2,R) = so(2,1) = sl(2,R) = su(1,1), and which we refer to as the two-dimensional real symplectic algebra sp(2,R). (We note that Lie *algebras* are involved, rather than Lie *groups*, as it is often stated.) Symmetries have a longer history, of course, since the angular properties of any central potential Hamiltonian and the rigid rotator acquainted physicists with group theory in the first place.[7] We are here concerned with *dynamical* algebras, i.e., those whose representations correspond with the whole energy spectrum of the system.

It is important to state that the only spectra that have been obtained from a dynamical algebra of which the Hamiltonian is an element are equally spaced spectra[8,9] if discrete (with a lower bound if realistic), a lower-bound continuum, or a double non-lower-bound continuum. These cases correspond to the Hamiltonian being on the elliptic, parabolic, or hyperbolic subalgebras of sp(2,R).[6]

Next, the Hamiltonian may be a simple function of one or more of the generators, the spectrum now being that function of the integers or subset thereof. This construction may be made for the bound hydrogen atom states, for its scattering states separately,[1] and has been used recently for the Morse[10] and Pöschl–Teller[11] potentials, among others, by Alhassid, Gürsey, and Iachello.

The Morse potential[12] is very well-known for its role in molecular physics while the Pöschl–Teller potential[13] emerges in connection with diverse physical systems, such as completely integrable many-body systems in one dimension,[14] the solitary wave solutions to the Korteweg–de Vries equation,[15] and in the Hartree mean field equation of many-body systems interacting through a δ force[16] among others. The Pöschl–Teller Schrödinger equation also stems from the Klein–Gordon equation on a space of constant curvature, with an appropriate set of separating variables, the D'Alembertian being the Laplace–Beltrami operator on a sphere or hyperboloid.[17]

The Pöschl–Teller potential has two free parameters:

$$V^{cs}(\chi) = c\,\mathrm{sech}^2\chi + s\,\mathrm{csch}^2\chi\,. \tag{1.1}$$

See Fig. 1. There is a $\sim s/\chi^2$ core at the origin plus a trough $\sim \mathrm{sech}^2\chi$. When $0 < -s/c < 1$, the two may combine to a potential with a core ($s>0$) and a trough ($c>0$). This trough may capture one or more quantum bound states when $\sqrt{2s+\frac{1}{4}} < \sqrt{-2c+\frac{1}{4}}$, which will be part of the spectrum of the Pöschl–Teller quantum Hamiltonian $\mathbb{H}^{\mathrm{PT}}$. The number of bound states is the integer part of the difference between $\sqrt{2s+\frac{1}{4}}$ and $\sqrt{-2c+\frac{1}{4}}$.

Alhassid,[18] Gürsey,[19] and Iachello[11] used the algebra

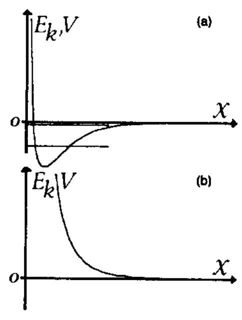


FIG. 1. (a) The Pöschl–Teller potential with a core and trough, exhibiting two bound states and the continuum. (b) A Pöschl–Teller potential where the trough parameter is smaller than the core parameter; it has only a continuum of positive-energy states.

[a)] On sabbatical leave from Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, 04510 Mexico, D. F., Mexico.

 0022-2488/85/050973-11$02.50

so(3) with one subalgebra generator for full representation l and row m classification. The Pöschl–Teller equation is then found to be the square of that generator and thus the bound-state spectrum is accounted for, being $\sim -m^2$ over the multiplet. This potential also has a continuum of positive-energy scattering states, and the Weyl analytic continuation is used to turn the algebra into so(2,1), where the positive continuous energy eigenvalue is the square of the eigenvalue of a noncompact generator of the algebra. They also investigate a more general version of the Pöschl–Teller potential, which is obtained from a representation of the direct sum algebra su(1,1) ⊕ su(1,1) realized by the symmetric top system in which one of the Euler angles is made hyperbolic.[20] They are thus able to show that the Pöschl–Teller Hamiltonian emerges as essentially the Casimir operator of the algebra and that it has mixed spectrum, including the bound and scattering states of the potential,[20] a result also found by Basu and Wolf.[21]

In this article we shall reexamine the Pöschl–Teller potential, showing that the Clebsch–Gordan *series*[22] of sp(2,R) yields the spectrum of the system, while the eigenstates turn out to be the sp(2,R) Clebsch–Gordan *coefficients* in the parabolic chain of Basu and Wolf[21] for a lower- and an upper-bound sp(2,R) discrete series representation, coupling into a finite sum of discrete series plus an integral over continuous-series representation. The energy values are determined by the coupled-sp(2,R) representations, while the potential parameters in (1.1) are determined by the two-factor sp(2,R) representations. The action of the raising and lowering operators in the conjugate so(2,2) algebra allow us to relate potentials (1.1) with *different values* of *the potential parameters* s and c for eigenstates of the *same energy*. In particular they can be made to relate a given potential with the free-particle potential $V^{00}(\chi) = 0$, the eigenstates of the two systems then being related through an algebra with shift operators, thus allowing a derivation of the reflection and transmission coefficients of the S matrix by purely algebraic means. These will be functions of the potential parameters and the energy of the state.

The mixed-spectrum character of the Pöschl–Teller potential makes it attractive for nuclear physics models of scattering. It is shown in Sec. II and III that this potential arises in an N-particle space out of the quadratic operators in position and momentum, forming an oscillator sp(2N,R) algebra, which contains sp(2,R) through the maximal subalgebra sp(2,R) ⊕ so(n,m). In this reduction, the representations of the two summands are conjugate. We further decompose so(n,m) ⊃ so(n) ⊕ so(m), each direct summand algebra having a conjugate sp(2,R), which provide the Pöschl–Teller potential parameters s and c with restrictions to discrete values. In Sec. IV we use the so(2,2) algebra generators to raise and lower[2,23] these values: the dimensions n and m are not crucially important for the structure of the system, and so(2,2) has most of the general features, plus some particularly useful ones. In this way we find the reflection and transmission coefficients[24] and the scattering matrix for this potential. The closing section offers some conclusions as to the place of the system treated here within the general systems whose spectrum is given by the Clebsch–Gordan series for sp(2,R), which may include the Coulomb system in the proper representation coupling class.

II. THE OSCILLATOR REALIZATION OF sp(2n,R) ⊃ sp(2,R) ⊕ so(n)

We consider the Schrödinger realization of the quantum operators of position and momentum in an n-dimensional Euclidean space R^n,

$$\mathbb{Q}_a f(\mathbf{x}) := x_a f(\mathbf{x}), \quad \mathbb{P}_a f(\mathbf{x}) := -i\frac{\partial f(\mathbf{x})}{\partial x_a}, \quad a = 1,2,\ldots,n. \tag{2.1a}$$

They are self-adjoint in a common invariant domain dense in $\mathscr{L}^2(R^n)$, and satisfy the well-known Heisenberg commutation relations

$$[\mathbb{Q}_a, \mathbb{P}_b] = i\delta_{a,b}\mathbf{1}, \tag{2.1b}$$

where $\mathbf{1}$ is the unit operator.[25] Next, we build all bilinear self-adjoint operators in $\mathbb{Q}_a$ and $\mathbb{P}_b$, denoting them as

$$\mathbb{J}^1_{ab} := \tfrac{1}{4}(\mathbb{P}_a\mathbb{P}_b - \mathbb{Q}_a\mathbb{Q}_b), \tag{2.2a}$$

$$\mathbb{J}^2_{ab} := \tfrac{1}{4}\mathbb{N}_{ab} := \tfrac{1}{4}(\mathbb{Q}_a\mathbb{P}_b + \mathbb{Q}_b\mathbb{P}_a - i\delta_{a,b}\mathbf{1}), \tag{2.2b}$$

$$\mathbb{J}^0_{ab} := \tfrac{1}{4}(\mathbb{P}_a\mathbb{P}_b + \mathbb{Q}_a\mathbb{Q}_b), \tag{2.2c}$$

$$\mathbb{M}_{ab} := \mathbb{Q}_a\,\mathbb{P}_b - \mathbb{Q}_b\,\mathbb{P}_a, \quad a,b = 1,2,\ldots,n. \tag{2.2d}$$

This set of operators closes under commutation, with the commutation relations defining the $2n$-dimensional real symplectic algebra[2] sp(2n,R). Since $\mathbb{J}^k_{ab} = \mathbb{J}^k_{ba}$ and $\mathbb{M}_{ab} = -\mathbb{M}_{ba}$, there are $2n^2 + n$ operators in the set and they are self-adjoint in $\mathscr{L}^2(R^n)$. On this space, they yield the *oscillator*[3] (or *metaplectic*[26]) representation of sp(2n,R). On this space this representation is not irreducible, since the inversion commutes with the set (2.2), and decomposes into two irreducible representations, one in the subspace of even functions and one in the subspace of odd functions.

Now we construct the linear combinations

$$\mathbb{J}^k := \sum_{a=1}^{n} \mathbb{J}^k_{aa}, \quad k = 1,2,0. \tag{2.3}$$

These three operators generate an algebra sp(2,R) which commutes with the operators $\mathbb{M}_{ab}$ in (2.2d). The latter commute among themselves and generate the n-dimensional orthogonal algebra so(n). We thus consider the algebra chain

$$sp(2n,R) \supset \mathrm{sp}(2,R) \oplus \mathrm{so}(n), \tag{2.4}$$

where the subalgebra is maximal in the parent algebra. The two direct summands in the subalgebra are, moreover, *conjugate*, i.e., within the oscillator representation of sp(2n,R), the representation of one direct summand determines the representation of the other. Indeed, the Casimir operator of sp(2,R) is[6]

$$\mathbb{C}^{\mathrm{sp}} := (\mathbb{J}^1)^2 + (\mathbb{J}^2)^2 - (\mathbb{J}^0)^2, \tag{2.5a}$$

while the second-order Casimir operator of so(n) is [2]

$$\mathbb{C}^{\mathrm{so}} := \frac{1}{2}\sum_{a,b=1}^{n} (\mathbb{M}_{ab})^2, \tag{2.5b}$$

and all higher-order Casimir operators of the latter are zero since the algebra is realized on the n-sphere[27] S^{n-1}. The second-order Casimir operator (2.5b) is the Laplace–Beltrami operator on that $(n-1)$-dimensional space S^{n-1}, with constant curvature related to the radius of the sphere.[28] One may show directly replacing (2.2) that the two operators (2.5) are related by

$$\mathbb{C}^{\rm sp} = -\tfrac{1}{4}\mathbb{C}^{\rm so} + \tfrac{1}{16}n(4-n)\,. \tag{2.6}$$

The eigenvalues of $\mathbb{C}^{\rm so}$ on S^{n-1} (i.e., the spectrum of the D'Alembertian) are given by

$$c^{\rm so} = l(l+n-2), \quad l = 0,1,2,..., \tag{2.7}$$

and thus through (2.6), the eigenvalues of $\mathbb{C}^{\rm sp}$ are

$$c^{\rm sp} = k(1-k)\,, \tag{2.8a}$$

$$k = \tfrac{1}{2}(l + \tfrac{1}{2}n) = \tfrac{1}{4}n, \tfrac{1}{4}n + \tfrac{1}{2}, \tfrac{1}{4}n + 1, ..., \tag{2.8b}$$

where k is referrred to as the *Bargmann* sp(2,R) representation index.[6] The representations of sp(2,R) present in the decomposition (2.4) are thus the lower-bound discrete-series representations $\mathscr{D}_k^+$. The parity of the so(n) representation l on S^{n-1} is well known to be $(-1)^l$. It follows that on the irreducible subspace of even functions, the oscillator realization (2.2) decomposes into the direct sum of sp(2,R) ⊕ so(n) representations $(k,l) = (\tfrac{1}{4}n,0) + (\tfrac{1}{4}n+1,2) + (\tfrac{1}{4}n+2,4) + \cdots$, while in the subspace of odd functions it is $(k,l) = (\tfrac{1}{4}n+\tfrac{1}{2},1) + (\tfrac{1}{4}n+\tfrac{3}{2},3) + (\tfrac{1}{4}n+\tfrac{5}{2},5) + \cdots$.

In the case $n = 1$, the generatorless algebra "so(1)" is replaced by the inversion operator with eigenvalues $+1$ and -1 on the two-point space S^0. The former goes with $k = \tfrac{1}{4}$ and the latter with $k = \tfrac{3}{4}$. This "so(1)" also effects the "algebra reduction" of so(2) to eigenvalues $m = \pm l$ of the latter's single generator, the sign being the "so(1)" eigenlabel. In the general-n case, we need not concern ourselves with the representation row labeling.

Regarding the subalgebra reduction of sp(2,R), the better-known[6] chain involves the compact subalgebra with generator $\mathbb{J}^0$. This operator is the n-dimensional harmonic oscillator Hamiltonian with angular momentum l, whose spectrum is lower bounded by k and is linearly spaced by integers.[2] In this work we shall use the *parabolic* subalgebra generator[29]

$$\mathbb{J}^- := \mathbb{J}^0 - \mathbb{J}^1 = \frac{1}{2}\sum_{a=1}^{n} \mathbb{Q}_a\mathbb{Q}_a\,. \tag{2.9}$$

This has been implicitly used whenever sp(2,R) is realized in terms of up-to-second-order differential operators, but has not often appeared as an abstract subalgebra chain in the physics literature. The operator $\mathbb{J}^-$ is noncompact and its spectrum therefore continuous; in the discrete-series representations $\mathscr{D}_k^+$, it is positive, as seen here, and simple. In the continuous-series representations of the next section, it is still simple but both positive *and* negative.[30]

The algebra sp(2,R) also has *negative* discrete-series representations, denoted by $\mathscr{D}_k^-$. These may be obtained from the positive-series operators (2.2) and (2.3) on $\mathscr{L}^2(R^n)$ through the mapping[6]

$$A\!:\{\mathbb{J}^1,\mathbb{J}^2,\mathbb{J}^0\} \mapsto \{-\mathbb{J}^1,\mathbb{J}^2,-\mathbb{J}^0\}\,. \tag{2.10}$$

This is an automorphism of sp(2,R), so the Casimir operator eigenvalues (2.8) are unaffected. It is involutive, but not within the group generated by it, *i.e.*, it is outer. It inverts the harmonic-oscillator spectrum of $\mathbb{J}^0$ to negative values, so the eigenvalues of the latter are upper bound by $-k$. The spectrum of $A\mathbb{J}^-$ is now the negative half-axis.

III. COUPLING AND REDUCTION IN sp(2N,R) ⊃ sp(2,R) ⊕ so(n,m)

We now consider the following Euclidean spaces: R^n, R^m, and R^N, $N = n + m$, where the two first spaces are disjoint subspaces of the latter, arranged so that $x_a \in R^n$ for $1 \leqslant a \leqslant n$ and $x_a \in R^m$ for $n+1 \leqslant a \leqslant n+m = N$. On the $\mathscr{L}^2(R^N)$ space of functions $f(\mathbf{x})$ we may build the oscillator representation of the symplectic algebra sp(2N,R), which has been presented in the last section and given in (2.2), letting all index ranges grow to N. We reproduce the structure for R^n and R_m placing their oscillator algebras sp(2n,R) and sp(2m,R) as subalgebras of sp(2N,R). Each of the former two will be decomposed as sp(2n,R) ⊃ $\mathrm{sp}_{(n)}(2,R) \oplus \mathrm{so}(n)$ and sp(2m,R) ⊃ $\mathrm{sp}_{(m)}(2,R) \oplus \mathrm{so}(m)$, where the generators of the first factors will be labeled as $\mathbb{J}^k_{(n)}$ and $\mathbb{J}^k_{(m)}$ for $k = 1,2,0$, built as in (2.3) with the appropriate summation index range. Now, if we follow the same procedure with sp(2N,R), we are *coupling*[21,31] the representations of $\mathrm{sp}_{(n)}(2,R)$ and $\mathrm{sp}_{(m)}(2,R)$ to a representation of $\mathrm{sp}_{(N)}(2,R)$. If the two factor representations belong to the $\mathscr{D}^+$ series, their product[21,22,31,32] will be reducible in terms of irreducible representations of the latter also belonging to the $\mathscr{D}^+$ series. If the former are given by their Bargmann indices $k_{(n)}$ and $k_{(m)}$, the Clebsch–Gordan series will contain the $\mathrm{sp}_{(N)}(2,R)$ representations $k_{(N)} = k_{(n)} + k_{(m)}, k_{(n)} + k_{(m)} + 1,...$ and its Casimir operator would have eigenvalues $k_{(N)}(1-k_{(N)})$ with $k_{(N)}$ on the series. These facts may be easily seen in the compact subalgebra reductions, where $\mathbb{J}^0_{(N)}$ is the sum of the harmonic oscillators $\mathbb{J}^0_{(n)}$ and $\mathbb{J}^0_{(m)}$ with the consequent sum of their discrete, lower-bound spectra to a "radial" discrete, lower-bound spectrum which, were we to follow this coupling, would lead to the constraining Pöschl–Teller potential (of the *first* type) $V(\chi) = c\sec^2\chi + s\csc^2\chi$, $0 < \chi < \pi/2$.

Our interest in this paper lies in the *scattering* Pöschl–Teller potential (i.e., of the *second* kind):

$$V^{cs}(\chi) = c\,\mathrm{sech}^2\chi + s\,\mathrm{csch}^2\chi\,, \tag{3.1}$$

which, according to the values of the two parameters, c and s, will have a lower-bound continuum of scattering states, with the possibility of a finite number of bound states.

To achieve this, we couple the two sp(2,R) subalgebras in sp(2n,R) and sp(2m,R) to the $\mathrm{sp}_{(N)}(2,R)$ algebra in sp(2N,R) through essentially the *difference* of the generators, following the linear combinations[2,31,32]

$$\mathbb{J}^1_{(N)} = \mathbb{J}^1_{(n)} - \mathbb{J}^1_{(m)}\,, \tag{3.2a}$$

$$\mathbf{J}^2_{(N)} = \mathbf{J}^2_{(n)} + \mathbf{J}^2_{(m)}, \quad (3.2b)$$
$$\mathbf{J}^0_{(N)} = \mathbf{J}^0_{(n)} - \mathbf{J}^0_{(m)}. \quad (3.2c)$$

This corresponds to coupling one $\mathscr{D}^+_{k(n)}$ with one $\mathscr{D}^-_{k(m)}$ irreducible representation. The so-algebra commuting with this $\mathrm{sp}_{(N)}(2,R)$ is the pseudo-orthogonal algebra so(n,m) whose set of generators is the union of the so(n) generators $\mathbf{M}_{ab}$, $a,b = 1,...,n$ (communting nontrivially with $\mathbf{J}^k_{(m)}$), the so(m) generators $\mathbf{M}_{ab}$, $a,b = n+1,...,n+m$ (commuting nontrivially with $\mathbf{J}^k_{(n)}$), and the "cross" noncompact boost generators $\mathbf{N}_{ab}$, $a = 1,...,n$, $b = n+1,...,n+m$ in (2.2b). We thus work with the subalgebra chain

$$\mathrm{sp}(2N,R) \supset \mathrm{sp}(2,R) \oplus \mathrm{so}(n,m). \quad (3.3)$$

The second-order Casimir operator of this so(n,m) may be expressed in the following form, in terms of the three constituent sets of generators:

$$\mathbb{C}^{\mathrm{so}(n,m)} = \mathbb{C}^{\mathrm{so}(n)} + \mathbb{C}^{\mathrm{so}(m)} - \sum_{a=1}^{n} \sum_{b=n+1}^{n+m} (\mathbf{N}_{ab})^2, \quad (3.4)$$

while that of $\mathrm{sp}_{(N)}(2,R)$ is given by (2.5a) in terms of (3.2). The two direct summand algebras in (3.3), sp$(2,R)$ and so(n,m) are again conjugate in sp$(2N,R)$, and their Casimir operators are related as in (2.6), with N replacing n; the eigenvalues relate accordingly.

The spectrum of the so(n,m) Casimir on the (n,m) hyperboloid H^{N-1} may be written as in (2.7), but with a different range of values of l. Through (2.8) we conclude that the conjugate $\mathrm{sp}_{(N)}(2,R)$ representation is labeled by $k = (1+i\lambda)/2$ and thus belongs to the *continuous* (nonexceptional[6]) representation series $\mathscr{C}^\epsilon_c$ with Casimir eigenvalue $c = (1+\lambda^2)/4 \geqslant \frac{1}{4}$ and multivaluation index $\epsilon = 0, \frac{1}{2}$ resolved in the $\mathscr{L}^2(R^N)$ subspaces of even and odd functions.

When $n = 1 = m$, both "so(1)" Casimir operators are zero and $\mathbb{C}^{\mathrm{so}(1,1)}$ is the square of a single boost generator $\mathbf{N}_{12}$, with a negative sign, i.e., $c^{\mathrm{so}(1,1)} = -\lambda^2$ corresponding to $l = i\lambda$, $\lambda \geqslant 0$.

For general $n,m > 1$, the coupling of $\mathscr{D}^+_{k_{(n)}}$ and $\mathscr{D}^-_{k_{(m)}}$ representations of sp$(2,R)$ has the following Clebsch–Gordan series[21,22,31,32]:

$$\mathscr{D}^+_{k_{(n)}} \dot{\times} \mathscr{D}^-_{k_{(m)}} = \sum_{|k_{(n)} - k_{(m)}| \geqslant k_{(N)} > \frac{1}{2}} \mathscr{D}^{\mathrm{sgn}[k_{(n)} - k_{(m)}]}_{k_{(N)}} + \int_{\kappa=0}^{\infty} \mathscr{C}^{k_{(n)} - k_{(m)} \bmod 1}_{\frac{1}{4} + \kappa \frac{2}{4}}, \quad (3.5)$$

i.e., for $k_{(n)} > k_{(m)}$, a direct finite sum of lower-bound discrete-series representations $\mathscr{D}^+_{k(N)}$ from $k_{(N)} = k_{(n)} - k_{(m)}$ in integer steps down to (but not including) $\frac{1}{2}$, *plus* a direct integral over all nonexceptional continuous representation series with the appropriate multivaluation index $\epsilon = 0$ or $\epsilon = \frac{1}{2}$, according to the total space inversion parity. The discrete part of the spectrum is absent if $k_{(n)} - k_{(m)} \leqslant \frac{1}{2}$. The Casimir operator of $\mathrm{sp}_{(N)}(2,R)$ has thus the *mixed* spectrum

$$c^{\mathrm{sp}_{(N)}(2,R)} = \begin{cases} k(1-k) < \frac{1}{4}, & k = k_{\min}, k_{\min} - 1, ..., > \frac{1}{2}, \quad k_{\min} = k_{(n)} - k_{(m)}, \\ \frac{1}{4}(1+\kappa^2) \geqslant \frac{1}{4}, & \kappa \geqslant 0. \end{cases} \quad (3.6)$$

This is the form of the spectra fitting into our coupling scheme: mixed spectra with a continuum of positive energy and a finite number of bound states with a characteristic quadratically downward-increasing separation for negative energy.

The previous statements are basis independent. In order to see how (3.6) becomes the spectrum of the scattering Pöschl–Teller potential Schrödinger Hamiltonian, we introduce the appropriate coordinates in R^N. These are (n,m)-bipolar-hyperbolic coordinates x_j $(\sigma,\rho,\chi,\{v_k\},\{\omega_k\})$:

$$x_j =: v\nu_j, \quad j = 1,2,...,n; \quad \gamma \geqslant 0, \quad \sum_j (\nu_j)^2 = 1, \quad (3.7a)$$

$$x_j =: s\omega_j, \quad j = n+1,...,n+m; \quad s \geqslant 0, \quad \sum_j (\omega_j)^2 = 1; \quad (3.7b)$$

$$\sigma = +1 \quad (r > s) \quad \begin{cases} r =: \rho\cosh\chi, & \rho,\chi \geqslant 0, \\ s =: \rho\sinh\chi, \end{cases} \quad (3.8a)$$

$$\sigma = -1 \quad (r < s) \quad \begin{cases} r =: \rho\sinh\chi, \\ s =: \rho\cosh\chi. \end{cases} \quad (3.8b)$$

The $\{\nu_j\}_{j=1}^n$ and $\{\omega_k\}_{k=n+1}^{n+m}$ are coordinates on the spheres S^{n-1} and S^{m-1}, and the Casimir operators of so(n) and so(m) are second-order differential operators in them, while r and s do not enter their expression. The so(n,m) Casimir operator will be a differential operator in all variables but ρ. One should be careful to note that these coordinates are *not global*, i.e., two *charts*, labeled by $\sigma = \pm 1$, are needed to cover the $r,s \geqslant 0$ quadrant by ρ, $\chi \geqslant 0$, (2.8a) *and* (3.8b). The so(n,m) Casimir operator will have two forms, one in each chart.[21,31]

We now detail the forms of the six operators and their eigenvalues, whose eigenfunctions are—once appropriately normalized—the $\mathscr{D}^+_{k(n)} \dot{\times} \mathscr{D}^-_{k(m)}$ Clebsch–Gordan coefficients.[21]

$k_{(n)}$: $\mathbb{C}^{\mathrm{sp}(n)(2,R)}$ has eigenvalue $k_{(n)}(1-k_{(n)})$. This fixes the $\mathbf{v}$ dependence of the eigenfunction to be an so(n) harmonic with angular momentum $l_{(n)} = 2k_{(n)} - \frac{1}{2}n$.

$k_{(m)}$: $\mathbb{C}^{\mathrm{sp}(m)(2,R)}$ has eigenvalue $k_{(m)}(1-k_{(m)})$. The $\boldsymbol{\omega}$ dependence is that of an so(m) harmonic $l_{(m)} = 2k_{(m)} - \frac{1}{2}m$.

$k_{(N)}$: $\mathbb{C}^{\mathrm{sp}(N)(2,R)}$ has eigenvalue $c^{\mathrm{sp}(N)(2,R)}$ given by the Clebsch–Gordan series (3.6), expressible through (2.6), (3.4), and (3.7) as the differential operator on $(\sigma,\chi,\mathbf{v},\boldsymbol{\omega})$,

$$
\begin{aligned}
\mathbb{C}^{\mathrm{sp}(N)(2,R)} &= \tfrac{1}{16}N(4-N) - \tfrac{1}{4}\left[\mathbb{C}^{\mathrm{so}(n)} + \mathbb{C}^{\mathrm{so}(m)} + \sum_{a=1}^{n}\sum_{b=n+1}^{n+m}\left(x_a\frac{\partial}{\partial x_b} + x_b\frac{\partial}{\partial x_a}\right)\right] \\
&= \frac{-1}{4}\left[N\left(N-\frac{1}{4}\right) + \left(1-\frac{s^2}{r^2}\right)\mathbb{C}^{\mathrm{so}(n)} + \left(1-\frac{r^2}{s^2}\right)\mathbb{C}^{\mathrm{so}(m)}\right. \\
&\quad \left. + \left(r\frac{\partial}{\partial s} + s\frac{\partial}{\partial r}\right)^2 + \left((n-1)\frac{s}{r} + (m-1)\frac{r}{s}\right)\left(r\frac{\partial}{\partial s} + s\frac{\partial}{\partial r}\right)\right] \\
&= \frac{-1}{4}\left[N\left(N-\frac{1}{4}\right) + \frac{\partial^2}{\partial\chi^2} + \left((n-1)\begin{Bmatrix}\tanh\chi\\ \coth\chi\end{Bmatrix} + (m-1)\begin{Bmatrix}\coth\chi\\ \tanh\chi\end{Bmatrix}\right)\frac{\partial}{\partial\chi}\right. \\
&\quad \left. + \begin{Bmatrix}\operatorname{sech}^2\chi\\ -\operatorname{csch}^2\chi\end{Bmatrix}\mathbb{C}^{\mathrm{so}(n)} + \begin{Bmatrix}-\operatorname{csch}^2\chi\\ \operatorname{sech}^2\chi\end{Bmatrix}\mathbb{C}^{\mathrm{so}(m)}\right], \quad \text{for} \quad \begin{Bmatrix}\sigma=+1\\ \sigma=-1\end{Bmatrix}.
\end{aligned}
\tag{3.9}
$$

Note that this operator has one form on each chart. Now we come to the parabolic subalgebra "row" labels.

r: $\mathbb{J}_{(n)}^{-} = \frac{1}{2}r^2$ with nonnegative eigenvalue.

s: $\mathbb{J}_{(m)}^{-} = \frac{1}{2}s^2$ with nonnegative eigenvalue.

(σ,ρ): $\mathbb{J}_{(N)}^{-} = -\mathbb{J}_{(n)}^{-} - \mathbb{J}_{(m)}^{-} = \frac{1}{2}(r^2 - s^2) = \frac{1}{2}\sigma\rho^2$.

This eigenvalue is fixed by r and s, which determine the chart σ on which $\mathbb{C}^{\mathrm{sp}(n)(2,R)}$ lies, and $\rho\geqslant 0$.

The normalized eigenfunctions of the first three operators, valuated at the eigenvalues of the last three, are the numerical Clebsch–Gordan coefficients. Only two of the latter three are independent. We may fix (σ,ρ) and, say, r, to determine s. We decide to fix σ and ρ, and let the coefficient be a function of the single free coordinate χ. Then, the Clebsch–Gordan coefficients are obtained as functions of χ satisfying the differential eigenfunction equation (3.9) with (3.6) for its spectrum.[21] It is particularly important to fix σ since this places us on a single chart σ, which we choose hereafter to be $\sigma = +1$. (Choosing $\sigma = -1$ only exchanges n and m.)

Clebsch–Gordan coefficients, even for noncompact algebras, are best known when reduced with respect to a compact subalgebra,[7,21,32] so that the row indices are integers m_1, m_2, and $m = m_1 + m_2$, for example. These satisfy three-term recursion relations—a second-order difference equation—which stem from the coupled Casimir operator. Their proper summation for normalization is a rather difficult problem. In the noncompact parabolic subalgebra basis, the row labels are continuous and the eigenfunctions of (3.9) satisfy an ordinary second-order differential equation when the so(n) and so(m) eigenfunction subspaces are taken. We anticipate that the solutions of (3.9) are ${}_2F_1$ Gauss hypergeometric functions,[21] while the elliptic or hyperbolic subalgebras lead to ${}_3F_2$ functions of unit argument.[31,32]

The original $\mathscr{L}^2(R^N)$ eigenfunctions of the Casimir operator are orthogonal under a maximal set of commuting operators under the measure

$$d^N\mathbf{x} = \rho^{N-1}\,d\rho\,\Omega_{nm}(\chi)d\chi\,d^{n-1}\mathbf{v}\,d^{m-1}\boldsymbol{\omega}, \tag{3.10a}$$

$$\Omega_{nm}(\chi) = \sigma\sinh^{n-1}\chi\cosh^{m-1}\chi. \tag{3.10b}$$

The integration on $\mathbf{v}$ and $\boldsymbol{\omega}$ leads to orthogonality in the so(n) and so(m) representation labels $k_{(n)}$ and $k_{(m)}$, and row labels which are absent from the sp(2,R) coefficient. Definite $\mathbb{J}_{(N)}^{-}$ eigenfunctions restrict to a definite (σ,ρ) value, on the χ half-line, the operator (3.9) is symmetric with respect to the measure $\Omega_{nm}(\chi)d\chi$. By similarity we may transform (3.9) to an operator symmetric with respect to $d\chi$, containing thus no first-order derivative terms:

$$
\begin{aligned}
\widetilde{\mathbb{C}}^{\mathrm{sp}} &:= \Omega^{1/2}\mathbb{C}^{\mathrm{sp}(N)(2,R)}\Omega^{-1/2}\big|_{k_{(n)},k_{(m)}} \\
&= \tfrac{1}{16}N(4-N) - \tfrac{1}{4}[\partial_\chi^2 - \{(2k_{(n)}-1)^2 - \tfrac{1}{4}\}\operatorname{sech}^2\chi \\
&\quad + \{(2k_{(m)}-1)^2 - \tfrac{1}{4}\}\operatorname{csch}^2\chi - \tfrac{1}{4}(N-2)^2].
\end{aligned}
\tag{3.10c}
$$

To obtain the usual $-\frac{1}{2}\partial_\chi^2 + V(\chi)$ form of Schrödinger equations, we define

$$\mathbb{H}^{\mathrm{PT}} = 2\widetilde{\mathbb{C}}^{\mathrm{sp}} - \tfrac{1}{2} = -\tfrac{1}{2}\partial_\chi^2 + V^{cs}(\chi), \tag{3.11a}$$

where $V^{cs}(\chi)$ is the scattering Pöschl–Teller potential (3.1) with parameters

$$c = -\tfrac{1}{2}[(2k_{(n)}-1)^2 - \tfrac{1}{4}] < \tfrac{1}{8}, \tag{3.11b}$$

$$s = \tfrac{1}{2}[(2k_{(m)}-1)^2 - \tfrac{1}{4}] \geqslant -\tfrac{1}{8}, \tag{3.11c}$$

and spectrum

$$E_k = 2k_{(N)}(1-k_{(N)}) - \tfrac{1}{2} = -\tfrac{1}{2}(2k_{(N)}-1)^2, \tag{3.11d}$$

where the range of k in the Clebsch–Gordan series (3.5) and (3.6) yields negative-energy bound states for couplings to the discrete series: $k = k_{\min}, k_{\min}-1,\ldots > \frac{1}{2}$, $k_{\min} = k_{(n)} - k_{(m)}$. The continuum of positive-energy scattering states appears for couplings to the continuous series for $k = \frac{1}{2}(1+i\kappa)$, $\kappa > 0$. The multivaluation index is determined by $k_{(n)} - k_{(m)}$ mod 1; $\epsilon = 0$ allows the $\kappa = 0$ value, while $\epsilon = \frac{1}{2}$ excludes it since the representation belongs to the exceptional type and is not square-integrable.[29]

Some remarks about the allowed values of the Pöschl–Teller parameters c and s, and the proper spectrum of $\mathbb{H}^{\mathrm{PT}}$ follow.

The coefficient s in (3.11b) multiplies the $\operatorname{csch}^2\chi$ term of the potential, which is singular as $\sim\chi^{-2}$ at the origin. The coefficient s represents thus a *core* parameter. There may be three cases.

(a) The core may be a singular, negative *well* $(0 > s \geqslant -\frac{1}{8})$ for $\frac{1}{4} < k_{(m)} < \frac{3}{4}$. Among the representations of $\mathrm{sp}_{(m)}(2,R)$ contained in the oscillator representation of sp(2N,R), only $k_{(m)} = \frac{1}{2}$ leads to a potential with an attractive well $s = -\frac{1}{8}$ at the origin. This is what we would call a *weak*[8] well for the purposes of investigating the conditions under which the Pöschl–Teller Hamiltonian has a unique spectrum.

(b) When $k_{(m)} = \frac{1}{4}$ or $\frac{3}{4}$, the core parameter s is zero.[24] These values of $k_{(m)}$ are allowed within sp(2N,R) [see (2.8b)] only when $l=0$, $m=1$, i.e., when the reduction of so$(n,1)$ is the canonical one to so(n), and (3.7) and (3.8) are spherical coordinates. For $n=2$, this is the chain sp$(6,R) \supset$ sp$(2,R) \oplus$ so$(2,1)$, and the algebra so$(2,1)$ is doubly at play. Using only so$(2,1)$ with its Casimir operator and generators, thus, we cannot get a nonzero core parameter. We do get, however, the two parity values which allow the system to be extended to the full line $\chi \in R$.

(c) Finally, when $s>0$ we get true core potentials. We must note the interval $(0,\frac{1}{4}) \cup (\frac{3}{4},1)$ for which $0<s<\frac{3}{8}$. It is in fact a *weak* core,[8] and the problem is having *two* values of $k_{(m)}$ leading to a single potential with fixed s; the interval, luckily, does not include values allowed by sp$(2N,R)$. For $k_{(m)} \geqslant 1$ the core is *strong*[33] ($\frac{3}{8} \leqslant s = \frac{3}{8}, 1, \frac{15}{8}, 3, \ldots$, for $k = 1, \frac{5}{4}, \frac{3}{2}, \frac{7}{4}, \ldots$) and the square-integrable solutions of the Schrödinger equation must be zero at the origin.

The second parameter c is factor to a $\operatorname{sech}^2 \chi$ term in the Pöschl–Teller potential. It is a trough parameter for $c<0$, and a smooth bump for $0<c \leqslant \frac{1}{8}$. There are also three cases: (a) a $c=\frac{1}{8}$ bump for $k_{(n)}=\frac{1}{2}$, (b) a $c=0$ zero potential for $k_{(n)} = \frac{1}{4}$ and $\frac{3}{4}$ allowing the system to be extended to $\chi \in R$, and (c) troughs of coefficient $c = -\frac{3}{8}, -1, -\frac{15}{8}, -3, \ldots$ for $k_{(n)} = 1, \frac{5}{4}, \frac{3}{2}, \frac{7}{4}, \ldots$.

The core and trough parameters combine to a Pöschl–Teller potential with a trough when $0< -s/c <1$ (i.e., $k_{(n)} > k_{(m)}$) at the position $\chi_{\min} = \operatorname{arctanh}(-s/c)^{1/4}$ and of depth $V(\chi_{\min}) = -(s - \sqrt{-c})^2$. This potential is able to hold bound states for $\frac{1}{2} < k_{\min} = k_{(n)} - k_{(m)}$ or $\sqrt{-2c+\frac{1}{4}} - \sqrt{2s+\frac{1}{4}} > 1$, and the number of these states is the integer part of $k_{\min}$. The bound state energies are given by (3.11) for $k_{(N)} = k_{\min}, k_{\min}-1, \ldots > \frac{1}{2}$, each corresponding to a discrete-series term in the Clebsch–Gordan series. For all sp$(2N,R)$-allowed Pöschl–Teller potential parameters, there is a continuum extending over positive energies, and corresponding to the continuous-series representations in the integral in the Clebsch–Gordan series for $k=\frac{1}{2}(1+i\kappa)$, $\kappa \in R^+$. It should finally be remarked again that for n even, $k_{(n)}$ may be fixed to be integer or half-integer, while for n odd, $k_{(n)}$ may only be a quarter-integer.

The Pöschl–Teller bound and free wave functions

$$\mathbb{H}^{\mathrm{PT}} \psi_k = E_k \psi_k \tag{3.12}$$

may be obtained directly, in normalized form, from the work of Basu and Wolf on sp$(2,R)$. They relate through[34]

$$C_> \begin{pmatrix} k_1 & k_2 & ; & k \\ +1,r & -1,r & ; & +1,\rho \end{pmatrix} = \delta(\tfrac{1}{2} r^2 - \tfrac{1}{2} s^2 - \tfrac{1}{2}\rho^2)(\rho^2 \cosh\chi \sinh\chi)^{-1/2} \psi_k^{(k_1,k_2)}(\chi), \tag{3.13}$$

where $k_1 = k_{(n)}$ and $k_2 = k_{(m)}$ determine the c and s potential parameters, $-\frac{1}{2}(2k-1)^2$ is the energy [see (3.11), setting $k=k_{(N)}$], and $\tanh\chi = s/r$ [$s/r<1$ on the $\sigma=+1$ chart] is the position in the Pöschl–Teller Schrödinger equation.

The bound states have the following wave functions normalized on $[0,\infty)$:

$$\psi_k^{(k_1,k_2)}(\chi) = c_k^{k_1k_2} (\cosh\chi)^{-2k_1+3/2} (\sinh\chi)^{2k_2-1/2} F\left[\begin{matrix} -k_1+k_2+k, -k_1+k_2-k+1 \\ 2k_2 \end{matrix}; -\sinh^2\chi\right], \tag{3.14a}$$

$$c_k^{k_1k_2} = \frac{1}{\Gamma(2k_2)} \left[\frac{(2k-1)\Gamma(k_1+k_2-k)\Gamma(k_1+k_2+k-1)}{\Gamma(k_1-k_2+k)\Gamma(k_1-k_2-k+1)}\right]^{1/2}, \tag{3.14b}$$

where $F\left[{a,b \atop c};z\right]$ is the ${}_2F_1$ Gauss hypergeometric function. These also may be put in terms of Jacobi polynomials of degree k_1-k_2-k and argument $2\operatorname{sech}^2\chi - 1$. The scattering states are described by the wave functions

$$\psi_k^{(k_1,k_2)}(\chi) = c_k^{k_1k_2}(\cosh\chi)^{2k_1-1/2}(\sinh\chi)^{2k_2-1/2} \times F\left[\begin{matrix} k_1+k_2-k, k_1+k_2+k-1 \\ 2k_2 \end{matrix}; -\sinh^2\chi\right], \tag{3.15a}$$

$$c_k^{k_1k_2} = [1/\pi\Gamma(2k_2)]\,[\tfrac{1}{2}\kappa \sinh\pi\kappa \times \Gamma(k_1+k_2-k)\Gamma(-k_1+k_2+k) \times \Gamma(k_1+k_2+k-1)\Gamma(-k_1+k_2-k+1)]^{1/2}. \tag{3.15b}$$

We note the symmetry relations

$$\psi_k^{(1-k_1,k_2)}(\chi) = \psi_k^{(k_1,k_2)}(\chi), \quad \text{for } k_1 \in R, \tag{3.15c}$$

$$\psi_k^{(k_1,1-k_2)}(\chi) = \psi_k^{(k_1,k_2)}(\chi), \quad \text{for } 2k_2 \text{ integer}, \tag{3.15d}$$

for the scattering states. This is an invariance transformation for the potential parameter c in (3.11b), obviously. Not so for s, however, as we shall see in the next section. In particular, from the coupling of two oscillator represention to the continuous-series representations[35] we obtain

$$\psi_k^{(3/4,3/4)}(\chi) = \psi_k^{(1/4,3/4)}(\chi) = \sqrt{2/\pi}\,\sin\kappa\chi, \tag{3.16a}$$

$$\psi_k^{(3/4,1/4)}(\chi) = \psi_k^{(1/4,1/4)}(\chi) = \sqrt{2/\pi}\,\cos\kappa\chi, \tag{3.16b}$$

normalized on $[0,\infty)$.

The asymptotic behavior of the scattering states under $\chi \to \infty$ corresponds to the oscillatory behavior of the Clebsch–Gordan coefficient (3.13) in the neighborhood of the cone $r=s$, where ρ becomes vanishingly small. Out of the tabulated asymptotic properties of the hypergeometric function, one may find, for $k=\frac{1}{2}(1+i\kappa)$,

$$\psi_k^{(k_1,k_2)}(\chi) \underset{\chi\to\infty}{\sim} \alpha_k^{k_1k_2} e^{i\kappa\chi} + \beta_k^{k_1k_2} e^{-i\kappa\chi}, \tag{3.17a}$$

$$\alpha_k^{k_1k_2} = \frac{2^{-i\kappa}}{\sqrt{2\pi}} \left[\frac{\Gamma(i\kappa)}{\Gamma(-i\kappa)} \times \frac{\Gamma(k_1+k_2-k)\Gamma(-k_1+k_2-k+1)}{\Gamma(k_1+k_2+k-1)\Gamma(-k_1+k_2+k)}\right]^{1/2}, \tag{3.17b}$$

$$\beta_k^{k_1k_2} = \alpha_{1-k}^{k_1k_2} = (\alpha_k^{k_1k_2})^* = \beta_k^{1-k_1,k_2}. \tag{3.17c}$$

In the next section we shall rederive these asymptotic coefficients out of pure Lie-algebraic considerations. In particular, we shall use them to find the reflection and transmission amplitudes of the pure trough potential $V^{c0}(\chi)$.

IV. so(2,2) SHIFT OPERATORS FOR THE SCATTERING STATES

In Sec. II we started with the parent algebra sp($2N,R$) in its oscillator representation (2.2) in order to provide an underlying N-particle phase space. There, the sp($2,R$) subalgebra in the chain (3.3) has for Casimir operator (in bipolar-hyperbolic coordinates) the Pöschl–Teller equation. The price we pay is to be able to account only for certain values for the potential coefficients c and s. In particular, we do *not* obtain (for $n,m > 1$) the null potential $V^{00}(\chi)$.

We may do away with this restriction in one important case:

$$\mathrm{sp}(8,R) \supset \mathrm{sp}(2,R) \oplus \mathrm{so}(2,2). \tag{4.1a}$$

There, the dimensional accident occurs that[2]

$$\mathrm{so}(2,2) = \mathrm{sp}_a(2,R) \oplus \mathrm{sp}_b(2,R). \tag{4.1b}$$

In fact, it also allows us to present the Pöschl–Teller equation as the *Klein–Gordon* equation[17] $(\Delta - \mu^2)\psi = 0$ with Δ being the Laplace–Beltrami operator on the three-dimensional surface of the (2,2) hyperboloid H^3 [(3.7) and (3.8] $\rho = \text{const}$, $\chi \in [0,\infty)$, θ, $\phi \in S^1$,

$$x_1 = \rho \cosh\chi \cos\theta, \quad x_3 = \rho \sinh\chi \cos\phi,$$
$$x_2 = \rho \cosh\chi \sin\theta, \quad x_4 = \rho \sinh\chi \sin\phi. \tag{4.2}$$

The eigenvalues μ may be interpreted as the masses allowed in such a model.

The so(2,2) algebra in the decomposition (4.1b) may be written explicitly in terms of the generators (2.2) as

$$\mathbb{K}_a^0 = \tfrac{1}{2}(\mathbb{M}_{12} + \mathbb{M}_{34}) = -(i/2)(\partial_\theta + \partial_\phi),$$
$$\mathbb{K}_b^0 = \tfrac{1}{2}(-\mathbb{M}_{12} + \mathbb{M}_{34}) = (i/2)(\partial_\theta - \partial_\phi), \tag{4.3a}$$
$$\mathbb{K}_a^1 = \tfrac{1}{2}(\mathbb{N}_{23} + \mathbb{N}_{14}), \quad \mathbb{K}_b^1 = \tfrac{1}{2}(-\mathbb{N}_{23} + \mathbb{N}_{14}), \tag{4.3b}$$
$$\mathbb{K}_a^2 = \tfrac{1}{2}(-\mathbb{N}_{13} + \mathbb{N}_{24}), \quad \mathbb{K}_b^2 = \tfrac{1}{2}(-\mathbb{N}_{13} - \mathbb{N}_{24}). \tag{4.3c}$$

We note that on H^3 the two Casimir operators of the two sp($2,R$)'s in (4.1b) are equal to each other and related to the sp($2,R$) Casimir in (4.1a) through

$$\mathbb{C}^{\mathrm{sp}_c} = (\mathbb{K}_c^1)^2 + (\mathbb{K}_c^2)^2 - (\mathbb{K}_c^0)^2$$
$$= -\tfrac{1}{4}\mathbb{C}^{\mathrm{so}(2,2)} = \mathbb{C}^{\mathrm{sp}(2,R)}, \quad c = a,b. \tag{4.4}$$

This means we have a "square" (k,k) representation of so(2,2) corresponding to the degenerate representation with Casimir eigenvalue $l(l+2)$, $k = \tfrac{1}{2}l + 1$, as before. The so(2,2) representation basis elements, classified through their eigenvalues M_a and M_b under $\mathbb{K}_a^0$ and $\mathbb{K}_b^0$, and k under $\mathbb{C}^{\mathrm{sp}_a} = \mathbb{C}^{\mathrm{sp}_b}$, may be written as functions over H^3 through

$$\varphi_k^{m_a m_b}(\theta,\phi,\chi) = e^{i(m_1\theta + m_2\phi)}\psi_k^{(k_1,k_2)}(\chi), \tag{4.5}$$

where m_1 and m_2 are the eigenvalues under $\mathbb{M}_{12}$ and $\mathbb{M}_{34}$,

$$m_a - m_b = m_1 = 2k_1 - 1, \quad m_a + m_b = m_2 = 2k_2 - 1 \tag{4.6}$$

(on H^3, m_1 and m_2 can be all and only integers), and $\psi_k^{(k_1,k_2)}(\chi)$ is the normalized Pöschl–Teller wave function (3.14) and (3.15) of energy $-\tfrac{1}{2}(2k-1)^2$.

The four so(2,2) shift operators are

$$\mathbb{K}_a^\updownarrow = \mathbb{K}_a^1 \pm i\mathbb{K}_a^2$$
$$= \tfrac{1}{2}e^{\pm i(\theta+\phi)}[\mp\partial_\chi + \tanh\chi(-i\partial_\theta \pm \tfrac{1}{2}) + \coth\chi(-i\partial_\phi \pm \tfrac{1}{2})], \tag{4.7a}$$
$$\mathbb{K}_b^\updownarrow = \mathbb{K}_b^1 \pm i\mathbb{K}_b^2$$
$$= \tfrac{1}{2}e^{\pm i(-\theta+\phi)}[\mp\partial_\chi + \tanh\chi(i\partial_\theta \pm \tfrac{1}{2}) + \coth\chi(-i\partial_\phi \pm \tfrac{1}{2})]. \tag{4.7b}$$

Shift operators, independently of their realization, have the well-known constants and action[6] on any normalized eigenbasis of $\mathbb{C}^{\mathrm{sp}_c}$ and $\mathbb{K}^0$ given by

$$\mathbb{K}^\updownarrow \varphi_k^m = [(k \pm m)(1 - k \pm m)]^{1/2}\varphi_k^{m\pm 1}. \tag{4.8}$$

The shifts produced by (4.7) on a fixed Pöschl-Teller potential generates the elements of its *multiplet*:

$$\mathbb{K}_a^\updownarrow : (m_1,m_2) \mapsto (m_1 \pm 1, m_2 \pm 1),$$
$$\mathbb{K}_b^\updownarrow : (m_1,m_2) \mapsto (m_1 \mp 1, m_2 \pm 1), \tag{4.9a}$$
$$\mathbb{K}_a^\updownarrow : (k_1,k_2) \mapsto (k_1 \pm \tfrac{1}{2}, k_2 \pm \tfrac{1}{2}),$$
$$\mathbb{K}_b^\updownarrow : (k_1,k_2) \mapsto (k_1 \mp \tfrac{1}{2}, k_2 \pm \tfrac{1}{2}). \tag{4.9b}$$

We noted in (3.15) that reflection in k_1 through $\tfrac{1}{2}$, $\psi_k^{(k_1,k_2)}(\chi) = \psi_k^{(1-k_1,k_2)}(\chi)$ holds for all k_1 due to hypergeometric function identities. This is as expected, since m_1 and $-m_1$ provide the same $\mathrm{sech}^2\chi$-potential coefficient c in (3.11b). When $2k_2$ is an integer (i.e., when m_2 is an integer, but *only* then) it holds that k_2 may be reflected through $\tfrac{1}{2}$, corresponding to the same $\mathrm{csch}^2\chi$-potential core coefficient s in (3.11c).

We thus arrive at the point of view of so(2,2) as the "dynamical potential group" (named in Ref. 11, Sec. 7) of the Pöschl–Teller system, where the irreducible representation basis elements are the wave functions, with the same energy, of *different potentials*. We may thus speak of the multiplet $\{m_1,m_2\}$ of potentials for a given energy level E_k.

For the bound states of a given energy $E_k = -\tfrac{1}{2}(2k-1) < 0$ allowed for $k = k_1 - k_2$, $k_1 - k_2 - 1$, $\ldots > \tfrac{1}{2}$, m_a and m_b range over $k, k+1, k+2, \ldots$. See Fig. 2(a), where the axes are drawn for m_1 and m_2. The set of dots constitute the so(2,2) multiplet; each dot in the first quadrant is associated to a given potential, the second quadrant being a reflection through the m_2 axis of the first. The $m_1 = 0$, $m_2 = 2k$ potential has a minimal (negative) core $s = -\tfrac{1}{8}$; shifting along the lattice boundary with $\mathbb{K}_a^\uparrow$ we increase both the repulsive core over positive values of s, and the trough parameter c to ever more negative values. The lowest allowed state in the first potential remains the lowest allowed one in all potentials of the "boundary" of the lattice in Fig. 2(a). Moving into the lattice we deepen the trough and thicken—to a lower degree—the core. Our energy E_k eigenstate will have more bound states below it, one for every nested layer we cross.

For the scattering states of a given energy $E_k = -\tfrac{1}{2}(2k-1)^2 > 0$, $k = \tfrac{1}{2}(1 + i\kappa)$, $\kappa \geqslant 0$, only the integer representations $\mathscr{C}^0_{k(1-k)}$ of the conjugate sp($2,R$) are allowed. The so(2,2) multiplet is shown in Fig. 2(b). Points in the first quadrant (including the axes) represent Pöschl–Teller potentials. The multiplet may be traversed on diagonals by means

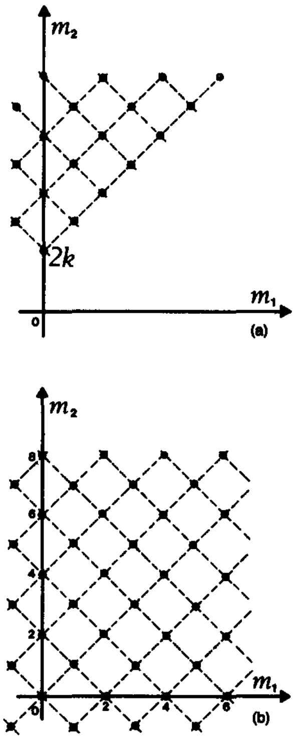


FIG. 2. (a) The so(2,2) multiplet of Pöschl–Teller potentials corresponding to a (bound) energy level $E_k < 0$. (b) The so(2,2) multiplet corresponding to a (free) energy level $E_k > 0$.

of the shift operators (4.7). Potentials along the m_2 axis ($m_1 = 0$) have a $c = \frac{1}{8}$ bump and all others a $c < 0$ trough, while those along the m_1 axis ($m_2 = 0$) have a $s = -\frac{1}{8}$ weak attractive core, and all others a true $s > 0$ core. Potentials lying above the diagonal in the first quadrant have a trough deep enough to hold bound states.

This is the situation on the (2,2) hyperboloid H^3. We would like to be able to include other Pöschl–Teller potentials with different values of the c and s parameters, in particular the null potential $V^{00}(\chi)$.

The coefficient c of the $\operatorname{sech}^2 \chi$ well, we saw, vanishes for $m_1 = \pm \frac{1}{2}$ ($k_1 = \frac{1}{4}, \frac{3}{4}$) and the coefficient s of the $\operatorname{csch}^2 \chi$ core for $m_2 = \pm \frac{1}{2}$ ($k_2 = \frac{1}{4}, \frac{3}{4}$). But note, only integer m's are allowed on H^3. This situation may be remedied for the well and managed for the barrier in the following way.

The three-dimensional hyperboloid space $H^3(\theta, \phi, \chi)$ projected on $\phi = \phi_0, \phi_0 + \pi$ is a two-dimensional space: a one-sheeted hyperboloid; the coordinate which circles it is θ. The $\theta = \theta_0, \theta_0 + \pi$ subspace, on the other hand, is two-sheeted hyperboloid circled by ϕ. We may *cover* the original H^3 hyperboloid n times in θ to a space of constant curvature (θ, ϕ, χ), where $\theta \in [0, 2\pi n)$, $\phi \in [0, 2\pi]$, $\chi \in [0, \infty)$, with the proper identifications, including $\theta \equiv \theta \bmod 2\pi n$. In the *double covered* three-hyperboloid $\overline{H^3}$, $\mathbf{M}_{12} = -i\,\partial_\theta$ may have integer as well as *half*-integer eigenvalues m_1. The zero-trough Pöschl–Teller system $m_1 = \pm \frac{1}{2}$ may be thus placed in the same multiplet with other potentials with half-integer m_1's and integer m_2's. Note that we may *not* do the same covering using ϕ, so the barrierless potential $V^{00}(\chi)$ *cannot* be realized on $\overline{H^3}$, nor partake in a multiplet belonging to a self-adjoint representation of so(2,2).

In Ref. 24 we took up the barrierless case through working with so(2,1) instead of so(2,2). In that case, only the one-sheeted two-hyperboloid space $H^2(\theta, \chi)$ is used, the algebra so(2) generated by $\mathbf{M}_{34}$ leaves an inversion "so(1)" group which in turn allows the unfolding of χ to the full line, where only coreless Pöschl–Teller potentials are allowed. The shift operators (4.7) become a single pair which lead to the scattering matrix, as detailed there.

In the full so(2,2) case followed here we note that up to this point we are consistent in having $+m_1$ and $-m_1$ half-integer (i.e., k_1 and $1 - k_1$ quarter-integer) describing the same potential constant $c = -\frac{1}{2}[m_1^2 - \frac{1}{4}]$ and the same eigenfunctions (3.15c). The question of allowing the core parameter s to vanish is more delicate. On a pedestrian level it would seem that one could "analytically continue"[11] the $\mathbf{M}_{34}$ eigenvalue m_2 to half-integer values, in spite of the fact that the hyperboloid coordinate ϕ allows *no* covering. In fact, we may do just that provided we realize that the resulting representation of the algebra so(2,2) will no longer be self-adjoint,[36] since, as we shall see below, the action of the shift operators does *not* leave the space of square-integrable wave functions invariant. Such representations of the algebra are *not* integrable to representations of the group. The results of Basu and Wolf[21] on sp(2,R) Clebsch–Gordan coefficients which provide the explicit eigenfunctions (3.14) and (3.15) continue to be valid for any real $k_2 > 0$ ($m_2 > -1$) since they were built out of the algebra, not the group.

The continuous-spectrum wave functions for the null Pöschl–Teller potential ($c = 0$, $s = 0$) may be found from (3.16a) and (3.16b) and are, as expected,

$$\varphi^{1/2,0}(\theta, \phi, \chi) = \sqrt{2/\pi}\, e^{i(\theta + \phi)/2} \sin \kappa\chi\,, \tag{4.10a}$$

$$\varphi_k^{-1/2,0}(\theta, \phi, \chi) = \sqrt{2/\pi}\, e^{i(-\theta + \varphi)/2} \cos \kappa\chi\,, \tag{4.10b}$$

and similar ones for $m_1 = -\frac{1}{2}$. Both (4.10a) and (4.10b) are solutions to the same, null potential $V^{00}(\chi)$ but they are obviously not the same. If χ were extended to the full real line, they would be the odd and even solutions of the free Schödinger equation for energy $\frac{1}{2}\kappa^2$. If we repeatedly apply $\mathbf{K}_a^{\uparrow}$ to $\varphi_k^{1/2,0}$ we find through (4.8) $\varphi_k^{3/2,0}$, $\varphi_k^{5/2,0}$, etc., which coincide with the functions on the hyperboloid built with $\psi_k^{(5/4,5/4)}$, $\psi_k^{(7/4,7/4)}$, etc. See Fig. 3. Applying $\mathbf{K}_a^{\downarrow}$ to $\varphi_k^{1/2,0}$ we obtain ($\kappa/2$ times) $\varphi_k^{-1/2,0}$. If we apply $\mathbf{K}_a^{\downarrow}$ to $\varphi_k^{-1/2,0}$, generating $\varphi_k^{-3/2,0}$, $\varphi_k^{-5/2,0}$,..., we find as expected that these functions are obtained out of $\psi_k^{(-1/4,-1/4)}$, $\psi_k^{(-3/4,-3/4)}$,... . The surprising

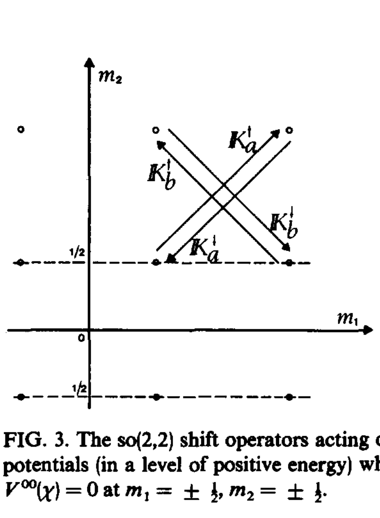


FIG. 3. The so(2,2) shift operators acting on a multiplet of Pöschl–Teller potentials (in a level of positive energy) which includes the null potential $V^{00}(\chi) = 0$ at $m_1 = \pm \frac{1}{2}$, $m_2 = \pm \frac{1}{2}$.

element in this construction is that while $\psi_k^{(k_1,k_1)}$ and $\psi_k^{(1-k_1,1-k_1)}$ are solutions to the *same* potential, they are the *two independent* solutions. Recall that (3.15b) holds only for integer $2k_2$. For $m_2 > \frac{1}{2}$, $\psi_k^{(k_1,k_2)}(\chi) \underset{\chi\to\infty}{\to} 0$ is the "good" solution, while its companion $m_2' = -m_2 < -\frac{1}{2}$, $\psi_k^{(k_1,1-k_2)}(\chi) \underset{\chi\to\infty}{\to} \infty$ is the "bad" solution to the same potential. For $m_2' < -1$ ($k_2 < 0$) it is not even square-integrable.

Bases for algebra representations built out of "good" and "bad" (i.e., non-square-integrable) functions are known in other contexts[36,37] in which the wave functions to raise and lower are, for instance, the Bessel functions $J_m(x)$.

We make use now of the so(2,2) shift operators in order to obtain algebraic relations between the asymptotic expansion coefficients $\alpha_k^{k_1k_2}$ and $\beta_k^{k_1k_2}$ in (3.17) for different values of k_1 and k_2, in particular, to relate them to those of (4.10). This will lead to the reflection and transmission coefficient of the coreless Pöschl–Teller potentials.

To this end we examine the asymptotic form of the shift operators (4.7) and recall that, as $\chi \to +\infty$, $\tanh\chi \to 1$ and $\coth\chi \to 1$. We define

$$\mathbb{K}_a^{\updownarrow(\infty)} := \lim_{\chi\to\infty} \mathbb{K}_a^{\updownarrow} = \tfrac{1}{2}e^{i(\theta+\phi)}[\mp\partial_\chi - i(\partial_\theta + \partial_\phi) \pm 1], \tag{4.11a}$$

$$\mathbb{K}_b^{\updownarrow(\infty)} := \lim_{\chi\to\infty} \mathbb{K}_b^{\updownarrow} = \tfrac{1}{2}e^{i(-\theta+\phi)}[\mp\partial_\chi - i(-\partial_\theta + \partial_\phi) \pm 1]. \tag{4.11b}$$

Next, we propose the asymptotic form of the so(2,2) basis wave functions (4.5) to be

$$\lim_{\chi\to\infty} \varphi_k^{m_am_b}(\theta,\phi,\chi) = e^{i(m_1\theta + m_2\phi)}(A_k^{m_am_b}e^{i\kappa\chi} + B_k^{m_am_b}e^{-i\kappa\chi}). \tag{4.12}$$

Now, consider (4.8) with the appropriate labels m_a or m_b, and its limit as $\chi \to \infty$. The left-hand side entails applying (4.11) to (4.12), while the right-hand side retains the square root factor and (4.12), with the replacement $m_a \mapsto m_a + 1$ or $m_b \mapsto m_b + 1$. This yields the following recursion relations between neighboring coefficients through the phase factor:

$$\begin{aligned} F(k,m) &= [F(1-k,m)]^{-1} = [(1-k+m)/(k+m)]^{1/2} \\ &= \exp[-\tfrac{1}{2}\arg\{k+m\}], \end{aligned} \tag{4.13}$$

viz.

$$\begin{aligned} A_k^{m_a+1,m_b} &= F(k,m_a)A_k^{m_am_b}, \\ B_k^{m_a+1,m_b} &= F(1-k,m_a)B_k^{m_am_b}, \end{aligned} \tag{4.14a}$$

$$\begin{aligned} A_k^{m_a,m_b+1} &= F(1-k,m_b)A_k^{m_am_b}, \\ B_k^{m_a,m_b+1} &= F(k,m_b)B_k^{m_am_b}, \end{aligned} \tag{4.14b}$$

Once the appropriate replacements are made, namely $m_a = k_1 + k_2 - 1$, $m_b = -k_1 + k_2$, it may be seen that the coefficients in (3.17b) and (3.17c) obey (4.14). Finally, the asymptotic coefficients in (3.17) are found when we start the recurrence (4.14) from the $V^{00}(\chi)$ eigenfunctions (4.10), where

$$\begin{aligned} A_k^{1/2,0} &= -i\sqrt{2\pi} = -B_k^{1/2,0}, \\ A_k^{0,-1/2} &= 1/\sqrt{2\pi} = B_k^{0,-1/2}. \end{aligned} \tag{4.15}$$

We may thus identify the α's and A's, and the β's and B's. In this way we derive the asymptotic behavior, the "good" as well as the "bad" solutions to the Pöschl–Teller potential using only algebraic techniques.

Scattering *through* a $\operatorname{csch}^2\chi$ core or well does not make sense since the $\chi > 0$ and $\chi < 0$ regions of the Schrödinger equation are uncoupled. We may remain with $\chi > 0$, but we cannot in general impose an arbitrary asymptotic behavior on the wave functions so as to make them "incoming" or "outgoing" without receiving a linear combination of "good" and "bad" solutions. We can speak about scattering for the coreless class of potentials, however. These, we saw, lie at $m_2 = \pm\frac{1}{2}$, and for them it is the full real line which makes sense in the scattering process. From (3.16) or (4.5)–(4.10) it is obvious that the basis functions $(m_1,m_2) = (\pm\frac{1}{2},\frac{1}{2})$ are odd in χ and $(\pm\frac{1}{2},-\frac{1}{2})$ are even. Use of the χ parity changing shift operators (4.7) shows that all $m_2 = \frac{1}{2}$ ($k_2 = \frac{3}{4}$) basis functions are odd, and all $m_2 = -\frac{1}{2}$ ($k_2 = \frac{1}{4}$) ones are even. Since we may cover the H^3 hyperboloid any number of times, this is also true for real m_1 (or k_1).

For a fixed, real trough parameter we may build the general wave function for the system with energy E_k as a linear combination:

$$\begin{aligned} &\psi_k^{(k_1)}(\chi) := \sigma_1\psi_k^{(k_1,1/4)}(\chi) + \sigma_3\psi_k^{(k_1,3/4)}(\chi), \\ &|\sigma_1|^2 + |\sigma_3|^2 = 1. \end{aligned} \tag{4.16}$$

The asymptotic behavior of this at $\chi \to \pm\infty$ is obtained from (3.17) and the parity of the two summands. Denoting $\alpha^1 := \alpha_k^{k_1,1/4}$, $\beta^1 := \beta_k^{k_1,1/4}$, $\alpha^3 := \alpha_k^{k_1,3/4}$, and $\beta^3 := \beta_k^{k_1,3/4}$, we have

$$\psi_k^{(k_1)}(\chi) \underset{\chi\to+\infty}{\sim} (\sigma_1\alpha^1 + \sigma_3\alpha^3)e^{i\kappa\chi} + (\sigma_1\beta^1 + \sigma_3\beta^3)e^{-i\kappa\chi}, \tag{4.17a}$$

$$\psi_k^{(k_1)}(\chi) \underset{\chi\to-\infty}{\sim} (\sigma_1\beta^1 - \sigma_3\beta^3)e^{i\kappa\chi} + (\sigma_1\alpha^1 - \sigma_3\alpha^3)e^{-i\kappa\chi}. \tag{4.17b}$$

Now, a *scattering* state for that potential and energy is a wavefunction which represents a flux of particles incoming far from the right ($ae^{-i\kappa\chi}$), part of the wave reflecting back ($be^{i\kappa\chi}$), and part of it transmitting towards the left ($ce^{i\kappa\chi}$), i.e.,

$$\psi_k^s(\chi) \underset{\chi\to+\infty}{\sim} be^{i\kappa\chi} + ae^{-i\kappa\chi}, \tag{4.18a}$$

$$\psi_k^s(\chi) \underset{\chi\to-\infty}{\sim} ce^{i\kappa\chi}. \tag{4.18b}$$

When this behavior at $\chi \to -\infty$ is imposed on the general solution (4.16), it implies $\sigma_1\beta^1 = \sigma_3\beta^3$ and this in turn fixes the coefficients a, b, and c. The *transmission* T and *reflection* R amplitudes[38] are then found to be

$$T:=\frac{c}{a}=\frac{1}{2}\left(\frac{\alpha^1}{(\alpha^1)^*}-\frac{\alpha^3}{(\alpha^3)^*}\right)=iC(k_1,\kappa)\sinh\pi\kappa\,, \tag{4.19a}$$

$$R:=\frac{b}{a}=\frac{1}{2}\left(\frac{\alpha^1}{(\alpha^1)^*}+\frac{\alpha^3}{(\alpha^3)^*}\right)=-C(k_1,\kappa)\cos 2\pi k_1, \tag{4.19b}$$

where

$$C(k_1,\kappa)=\frac{1}{\pi}\frac{\Gamma(i\kappa)}{\Gamma(-i\kappa)}\Gamma\left(2k_1-\frac{1}{2}-i\kappa\right)\times\Gamma\left(-2k_1+\frac{3}{2}-i\kappa\right). \tag{4.19c}$$

These results agree—as expected—with those of Ref. 39, once we replace $k\mapsto\kappa$ and $j\mapsto m_1-\frac{1}{2}=2k_1-\frac{3}{2}$. They are here obtained as a consequence of the form of the asymptotic coefficients (3.17).

V. CONCLUSION

We have worked with the $2N$-dimensional real symplectic algebra sp($2N,R$) so as to have an N-particle configuration space in its oscillator representation. This may be reduced with respect to its sp($2,R$) $\oplus$ so(n,m) subalgebras. Their conjugate Casimir operator is then the system's Hamiltonian with a Pöschl–Teller potential. Seen as a sp($2,R$) Casimir operator, the spectrum of this Hamiltonian becomes the Clebsch–Gordan series of sp($2,R$) which has a mixed spectrum. Seen as an so(n,m) Casimir operator, the Pöschl–Teller Schrödinger Hamiltonian becomes the Laplace–Beltrami operator in the so(n) $\oplus$ so(m) reduction, i.e., a Klein–Gordon equation on a space with constant curvature. In either case we worked specifically on so(2,2) which has already the essence of the properties on any more general so(n,m).

We devoted little space to mention the reduction sp($2N,R$) $\supset$ sp($2,R$) $\oplus$ so(N), where instead of a hyperboloid we have a sphere. This is the trigonometric Pöschl–Teller potential of the first type, which contains only bound states and no continuum. The constraining of the sphere—a point rotor—yields the familiar quadratically increasing eigenvalues (2.7) associated with angular momentum. In so(4), in particular, we have the rigid rotator system which belongs to the canonical reduction so(4) $\supset$ so(3) $\supset$ so(2), while the trigonometric Pöschl–Teller potential belongs to so(4) = so(3) $\oplus$ so(3) $\supset$ so(2) $\oplus$ so(2).

Beyond so(4), we have so(3,1), which is real, semisimple, and has not been treated explicitly here, but which can be shown to correspond to Hamiltonians built as Casimir operators with a spectrum given by the Clebsch–Gordan series $\mathscr{D}\dot{\times}\mathscr{C}$, which decompose[21,22,31,32] into an *infinite* set of quadratically decreasing values (for the full discrete series), plus a continuum of positive-energy "scattering" states. We did not pursue this line further since, as shown in the work of Basu and Wolf, the Hamiltonians—which are indeed of the Pöschl–Teller type but with a strong attractive core well—have the additional feature of being multichart operators [i.e., the index σ in (3.8) can no longer be fixed by a parabolic subalgebra representation] and both charts must be coupled properly.[21] Multichart operators with these kind of wells and non-lower-bound spectra do not make for attractive physical models. The *inverse* (reciprocal) of the spectrum, however, shifted by $\frac{1}{4}$, is the *full* spectrum of the hydrogen atom system, so it may well be that a good description of this system will lead to the $\mathscr{D}\times\mathscr{C}$ Clebsch–Gordan series and coefficients in the parabolic basis.

We worked here with the general case so(n,m) which leads to the $\mathscr{D}^+\dot{\times}\mathscr{D}^-$ coupling to a mixed spectrum with a finite number of bound states. It must be mentioned that one may also effect the reduction so(2,2)$\supset$so(1,1)$\oplus$so(1,1), as a particular case for so(n,m) $\supset$ so(p,q) $\oplus$ so(r,s) ($p+r=n$, $q+s=m$) leading to the last sp($2,R$) representation coupling,[21,22,31,32] $\mathscr{C}\times\mathscr{C}$ which has the same reduction as $\mathscr{D}\times\mathscr{C}$, doubled by parity, and possibly containing *one* exceptional representation. Unfortunately, the resulting Pöschl–Teller Hamiltonian is a three-chart operator, and no attractive physical interpretation can be attached to it. Beyond these reductions, one has the nonsubgroup reductions of Winternitz and collaborators[40] which probably lead to the periodic potentials studied in Ref. 11. We intend to pursue their inclusion in this scheme to provide a unified sp($2,R$)-based description of mixed and other spectra. In any case, we hope to have made the point that Pöschl–Teller systems are quite general with a clear-cut geometric interpretation.

ACKNOWLEDGMENTS

We would like to thank Y. Alhassid, O. Castaños, J. Lomnitz, M. Moshinsky, J. Plebański, and L. F. Urrutia for their interest and helpful comments.

This work was supported in part by CONACyT project PCCBCEU-020061.

[1]See, e.g., M. J. Engelfield, *Group Theory and the Coulomb Problem* (Wiley, New York, 1972).

[2]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).

[3]M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1968).

[4]K. B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979), Part IV.

[5]M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 1634 (1972).

[6]V. Bargmann, Ann. Math. **48**, 568 (1947).

[7]E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).

[8]K. B. Wolf, Kinam **3**, 323 (1981).

[9]B. Mielnik, preprint, CINVESTAV, Instituto Politécnico Nacional, Mexico, 1984.

[10]M. Berrondo and A. Palma, J. Phys. A **13**, 773 (1980); in *Proceedings of the IX International Colloquium on Group Theoretical Methods in Physics, Lecture Notes in Physics* Vol. 135 (Springer, Heidelberg, 1980), p. 2; Y. Alhassid, F. Iachello, and F. Gürsey, Chem. Phys. Lett. **99**, 27 (1983); R. Montemayor and L. Urrutia, Am J. Phys. **51**, 641 (1983).

[11](a) Y. Alhassid, F. Gürsey, and F. Iachello, Phys. Rev. Lett. **50**, 873 (1983); (b) Ann. Phys. (N. Y.) **148**, 346 (1983).

[12]P. M. Morse, Phys. Rev. **34**, 57 (1929).

[13]G. Pöschl and E. Teller, Z. Phys. **83**, 143 (1933).

[14]F. Calogero, Lett. Nuovo Cimento **13**, 411 (1975); M. A. Olshanetsky and A. M. Perelomov, Phys. Rep. **94**, 313 (1983).

[15]P. Lax, Commun. Pure Appl. Math. **27**, 97 (1968); H. Segur, *Topics in Ocean Physics* (Corso, Italy, 1982).

[16]F. Calogero and A. Degasperis, Phys. Rev. A **11**, 265 (1975); B. Yoon and J. W. Negele, Phys. Rev. A **16**, 1451 (1977).

[17]J. Plebański (unpublished).

[18]Y. Alhassid, in *Proceedings of the VII Oaxtepec Symposium on Nuclear Physics*, 1984, Instituto de Física, Universidad Nacional Autónoma de México, p. 1.

[19]F. Gürsey, in *Proceedings of the XI International Colloquium on Group Theoretical Methods in Physics, Lecture Notes in Physics* Vol. **180** (Springer, Heidelberg, 1982), p. 106.
[20]Y. Alhassid, in *Bosons in Nuclei* (World Scientific, Singapore, 1983).
[21]D. Basu and K. B. Wolf, J. Math. Phys. **24**, 478 (1983).
[22]L. Pukański, Trans. Am. Math. Soc. **100**, 116 (1961).
[23]L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951), Sec. 4.8.
[24]A. Frank and K. B. Wolf, Phys. Rev. Lett. **52**, 1737 (1984).
[25]See, e.g., G. Fano, *Mathematical Methods of Quantum Mechanics* (McGraw–Hill, New York, 1971).
[26]A. Weil, Acta Math. **11**, 143 (1963).
[27]R. L. Anderson and K. B. Wolf, J. Math. Phys. **11**, 3176 (1970).
[28]W. Miller, Jr., *Symmetry and Separation of Variables, Encyclopedia of Mathematics*, Vol. 4 (Addison–Wesley, Reading, MA, 1977).
[29]D. Basu and K. B. Wolf, J. Math. Phys. **23**, 189 (1982).
[30]Reference 6, Appendix.
[31]N. Mukunda and B. Radhakrishnan, J. Math. Phys. **15**, 1320, 1322, 1643, 1656 (1974).
[32]W. J. Holman and L. C. Biedenharn, Ann. Phys. (N. Y). **39**, 1 (1966); **47**, 205 (1968).
[33]An analysis similar to that of Ref. 8 on singular potentials with csch^2 terms has not yet been performed for the strongly attractive core cases.
[34]Reference 22, Eqs. (6.19) and (6.20).
[35]Reference 22, Eqs. (4.8) and (4.11).
[36]W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).
[37]K. B. Wolf, in *Latin American School of Physics, 1980, Proceedings*, AIP *Conference Proceedings*, Vol. 71 (AIP, New York, 1981), p. 1.
[38]See, e.g., E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970).
[39]Reference 11(b), Eqs. (6.39) and (6.40).
[40]P. Winternitz, I. Lukac, and Ya. Smorodinskii, Sov. J. Nucl. Phys. **7**, 139 (1968); N. Macfadyen and P. Winternitz, J. Math. Phys. **12**, 281 (1971); P. Winternitz, in *Proceedings of the V International Colloquium on Group Theoretical Methods in Physics* (Academic, New York, 1977), p. 549.

The semi-classical expansion for a charged particle on a curved space background

David Elworthy
Mathematics Institute, University of Warwick, Coventry CV4 7AL, England
Aubrey Truman
Mathematics Department, University College of Swansea, Singleton Park, Swansea, SA2 8PP, Wales
Keith Watling
Mathematics Institute, University of Warwick, Coventry CV4 7AL, England

(Received 17 July 1984; accepted for publication 12 October 1984)

We give the semi-classical expansion, with remainder to any order in $\hbar$, for the wave function of a nonrelativistic quantum particle in a classical external magnetic field on a curved space background. The basic assumption is of a "no caustics condition" on the underlying classical mechanics, at least up to the time in question. The gauge invariance of the result is emphasized together with a discussion of the geometric meaning of the classical mechanical quantities involved.

I. INTRODUCTION

An extension of the results of Truman,[1] and Elworthy–Truman[2] to the case of a particle in an external magnetic field was announced by the first two authors at the CIRM–Marseilles conference on "Stochastic Processes in Quantum Theory and Statistical Physics" in June 1982. This gave a rigorous proof, plus estimate of the remainder, of the convergence to the WKB term as $\hbar \to 0$ of the wave function for our nonrelativistic quantum particle. More recently, Watling[3] observed that the results of Ref. 2 extend rather easily to give the semi-classical expansion for the wave function, with remainder term, to any power of $\hbar$. This program is carried out here and with the inclusion of an external magnetic field. In principle it should be possible to obtain the corresponding expansion for the propagator as a corollary of these results, as in the case of the diffusion equation.[3,4] However we do not go into that here.

The basic hypothesis we have to make is a "no caustics condition" for the classical mechanics, as in Ref. 2. This is discussed in Sec. VI where it is shown to hold for all sufficiently small times for a wide class of vector and scalar potentials, in particular for a compact state space.

The results are put in a geometric context, and their gauge invariance emphasized, in Sec. VII. This also highlights the "topological triviality" we have had to assume for our gauge theory and initial wave function. A similar assumption was also intrinsic to Ref. 2. In general the appreciation of gauge invariance makes the arguments in this article clearer and simpler than the very similar ones in Ref. 3.

For a general survey of the relationships between Schrödinger equations and classical mechanics, with an extensive bibliography we refer the reader to Albeverio and Arede.[5]

II. THE PROBLEM

Suppose that our initial wave function $\psi_0: M \to \mathbb{C}$ has the form

$$\psi_0(x) = \exp\{i(e/\hbar)S_0(x)\}\,T_0(x) \tag{1}$$

for S_0 and T_0 real-valued functions on M [but see Remarks (i) at the end of Sec. IV]. This form is invariant under the group $\mathscr{G}$ of local gauge transformations which we suppose to act on our wave functions ψ by

$$\psi \to \exp\{i(e/\hbar)\theta(\cdot)\}\,\psi(\cdot) \tag{2}$$

for $\theta: M \to \mathbb{R}$. Various degrees of differentiability will be required of S_0 and T_0 as we proceed. Our Schrödinger equation will be

$$\frac{\partial \psi_t}{\partial t} = \tfrac{1}{2} i\hbar \Delta \psi_t + (i\hbar)^{-1} V \psi_t$$

for $V: M \to \mathbb{R}$, where Δ denotes the gauge-invariant Laplacian

$$\Delta = (\nabla - i(e/\hbar)A)^2, \tag{3}$$

i.e.,

$$\Delta\psi = \Delta_M \psi - 2i(e/\hbar)\langle A, \nabla\psi\rangle - i(e/\hbar)(\operatorname{div} A)\psi - (e^2/\hbar^2)|A|^2\psi \tag{4}$$

for Δ_M the Laplace—Beltrami operator on M

$$\Delta_M \psi = \operatorname{div} \nabla\psi$$

and A the vector potential (units: $[A] = Q^{-1}MLT^{-1}$), treated as a vector field on M so that $\langle A, \nabla\psi\rangle$ for example stands for

$$x \to \langle A(x), \nabla\psi(x)\rangle_x$$

using the Riemannian metric on M.

We seek an asymptotic expansion of ψ_t in $\hbar$ and in particular the approximate behavior of ψ_t when $\hbar$ is small compared to the other quantities involved, together with an understanding of when the approximations are valid.

III. ASSOCIATED CLASSICAL MECHANICS

We shall use the convention that # denotes a raising or lowering of indices: if ϕ is a one-form on M then $\phi^{\#}$ is the associated vector field, and if X is a vector field then $X^{\#}$ is the one-form

$$X^{\#}(v) = \langle X(x), v\rangle_x$$

for v a tangent vector to M at x.

Define the momentum Z as the one-form on M

$$Z = e\,dS_0 - eA^{\#} \tag{5}$$

so that if ψ_0 is nonzero and "Im" denotes the imaginary part

0022-2488/85/050984-07$02.50

$$Z^{\#} = \operatorname{Im}\{\psi_0^{-1}(\hbar\nabla - ieA)\psi_0\}$$
$$= (2i)^{-1}|\psi_0|^{-2}\{\bar{\psi}_0(\hbar\nabla - ieA)\psi_0 + \psi_0\,\overline{(\hbar\nabla - ieA)\psi_0}\}, \quad (6)$$

where the bar denotes complex conjugation. (Thus $|\psi_0|^2 Z^{\#}$ is the "current.")

Let R denote the two-form on M, the curvature form,

$$R = -e^{-1}\,dZ = dA^{\#} \quad (7)$$

and consider the classical equations of motion for $x_s = \phi_s(a)$ on M, each $a \in M$:

$$\ddot{\Phi}_t(a) + eR(\dot{\Phi}_t(a), -)^{\#} + \nabla V(\Phi_t(a)) = 0, \quad (8)$$

$$\Phi_0(a) = a,\ \dot{\Phi}_0(a) = Z(a)^{\#}. \quad (9)$$

For $M = \mathbf{R}^3$ this reduces to the Lorentz force equation

$$\ddot{\Phi}_t + e\dot{\Phi}_t \wedge B + \nabla V = 0, \quad (10)$$

where $B = \operatorname{curl} A$ is the magnetic induction.

As in Ref. 2 we have to assume that for each $a \in M$ and some $0 < t < \tau$ a solution $\Phi_t(a)$ exists for $0 < t < \tau$ and that no caustics develop up to that time, i.e.,

No caustics condition: For $0 < t < \tau$ the map $\Phi_t : M \to M$ is a diffeomorphism of M onto itself.

We show in Sec. VII below that such a τ exists whenever M is compact (and S_0, A, and V sufficiently differentiable) and for noncompact but complete M given bounds on S_0, A, and V and their derivatives.

With this assumption from now on, define

$$\check{S}(x,t) = \tfrac{1}{2}\int_0^t |\dot{\Phi}_s \circ \Phi_t^{-1}(x)|^2 ds - \int_0^t V(\Phi_s \circ \Phi_t^{-1}(x))ds \quad (11)$$

for $x \in M$ and $0 < t < \tau$, and set

$$S(x,t) = \check{S}(x,t) - \int_0^t Z(\dot{\Phi}_s \circ \Phi_t^{-1}(x))ds, \quad (12)$$

the Hamilton–Jacobi principal function.

Lemma 3: S satisfies the Hamilton–Jacobi equation

$$\tfrac{1}{2}|\nabla S + Z^{\#}|^2 + V + \frac{\partial S}{\partial t} = 0 \quad (13)$$

on $M \times [0,\tau]$. Also

$$\dot{\Phi}_t(a) = \nabla S(\Phi_t(a),t) + Z^{\#}(\Phi_t(a)),$$
$$a \in M,\ 0 < t < \tau. \quad (14)$$

Proof: For fixed $0 < t < \tau$ set $\Theta_s = \Phi_s \circ \Phi_t^{-1} : M \to M$ and write $\dot{\Theta}_s$ for $(\partial/\partial s)\Theta_s$. Then

$$S(x,t) = \tfrac{1}{2}\int_0^t |\dot{\Theta}_s(x)|^2 ds - \int_0^t V(\Theta_s(x))ds + eS_0(\Phi_t^{-1}(x)) - eS_0(x) + e\int_0^t \langle A(\Theta_s(x)), \dot{\Theta}_s(x)\rangle ds. \quad (15)$$

Therefore, if $L^* : T_y M \to T_x M$ denotes the adjoint of any given linear map $L : T_x M \to T_y M$ of tangent spaces to M and if $Tf : TM \to TM$ denotes the induced (derivative) map of any smooth $f : M \to M$,

$$\nabla S(x,t) = e(T\Phi_t^{-1})^* \nabla S_0(\Phi_t^{-1}(x)) - e\nabla S_0(x) + \int_0^t \frac{D}{\partial s}[T\Theta_s]^*[\dot{\Theta}_s(x) + eA(\Theta_s(x))]ds - \int_0^t [T\Theta_s]^*[\nabla V(\Theta_s(x))]ds + e\int_0^t [T\Theta_s]^*(\nabla A)^* \dot{\Theta}_s(x) ds.$$

On integration by parts the first integral becomes

$$[T\Theta_s]^*(\dot{\Theta}_s(x) + eA(\Theta_s(x)))|_0^t - \int_0^t [T\Theta_s]^*(\ddot{\Theta}_s(x) + e\nabla A(\dot{\Theta}_s(x)))ds.$$

However, our equation of motion gives

$$\ddot{\Theta}_s(x) + e(\nabla A - (\nabla A)^*)\dot{\Theta}_s(x) + \nabla V(\Theta_s(x)) = 0, \quad (16)$$

with

$$\dot{\Theta}_0(x) = e\nabla S_0(\Phi_t^{-1}(x)) - eA(\Phi_t^{-1}(x)) \quad (17)$$

and (14) follows.

On the other hand, differentiating Eq. (15) for t:

$$\frac{\partial S}{\partial t}(x,t) = e\,dS_0(\dot{\Phi}_t^{-1}(x)) + \tfrac{1}{2}|\dot{\Theta}_t(x)|^2 + \int_0^t \Big\langle \frac{D}{\partial s} T\Phi_s(\Phi_t^{-1}(x)),\ \dot{\Theta}_s(x) + eA(\Theta_s(x))\Big\rangle ds + \int_0^t \langle e\nabla A(T\Phi_s \circ [\Phi_t^{-1}(x)]), \dot{\Theta}_s(x)\rangle ds - V(x) - \int_0^t dV(T\Phi_s(\dot{\Phi}_t^{-1}(x)))ds + \langle eA(x), \dot{\Theta}_t(x)\rangle.$$

The first integral, when integrated by parts, becomes

$$\langle T\Phi_t(\dot{\Phi}_t^{-1}(x)), \dot{\Theta}_t(x) + eA(\Theta_t(x))\rangle - \langle \dot{\Phi}_t^{-1}(x), \dot{\Theta}_0(x) + eA(\Theta_0(x))\rangle - \int_0^t \langle T\Phi_s(\dot{\Phi}_t^{-1}(x)), \ddot{\Theta}_s(x) + eA(\Theta_s(x))\rangle ds$$

and (13) follows using (14), the equations of motion, and the equality

$$T\Phi_t \circ \dot{\Phi}_t^{-1} + \dot{\Phi}_t \circ \Phi_t^{-1} = 0. \quad (18)$$

□

Now set

$$\phi(x,t) = |\det T_x \Phi_t^{-1}|,$$

the Jacobian determinant of Φ_t^{-1} at x where by "determinant" we mean the determinant of the matrix of $T_x \Phi_t^{-1} : T_x M \to T_y M$, for $y = \Phi_t^{-1}(x)$, obtained by using orthonormal bases for $T_x M$ and $T_y M$. Then by general principles, or the proof in Ref. 2, by (14) we have the continuity equation

$$\frac{\partial \phi}{\partial t}(x,t) + \operatorname{div}(\phi(x,t)[\nabla S(x,t) + Z(x)^{\#}]) = 0. \quad (19)$$

IV. THE WKB TERM

First we need the analog of Lemma 6C of Ref. 2.

Lemma 4A: (i) For C^2 maps $f : M \to \mathbf{C}$ and $\theta : M \to \mathbf{R}$

$$\Big(\nabla - i\frac{e}{\hbar}A\Big)^2\Big(\exp\Big\{i\frac{e}{\hbar}\theta\Big\}f\Big) = \exp\Big\{i\frac{e}{\hbar}\theta\Big\}\Big(\nabla - i\frac{e}{\hbar}(A - \nabla\theta)\Big)^2 f. \quad (20)$$

(ii) For a C^1 map $\theta:M\to\mathbb{C}$

$$\frac{\partial}{\partial t}[\phi^{1/2}\theta\circ\Phi_t^{-1}] = -\tfrac{1}{2}(\mathrm{div}(\nabla S + Z^{\#}))\phi^{1/2}\theta\circ\Phi_t^{-1} - \langle\nabla(\phi^{1/2}\theta\circ\Phi_t^{-1}),\nabla S + Z^{\#}\rangle. \tag{21}$$

Proof: Equation (20) is just the expression of the invariance of our operator Δ under gauge transformations. Equation (21) comes directly from the continuity equation (19) together with Eqs. (14) and (18). $\square$

Now define $W_0(t):L^2(M;\mathbb{C})\to L^2(M;\mathbb{C})$ by

$$[W_0(t)\theta](x) = \exp\{(i/\hbar)S(x,t)\}\phi^{1/2}(x,t)\theta(\Phi_t^{-1}(x))\psi(x)|\psi(x)^{-1}| \tag{22}$$

for $\theta\in L^2$ and $x\in M$.

Then

$$[W_0(t)\theta](x) = \exp\left\{\frac{i}{\hbar}\check{S}(x,t)\right\}\phi^{1/2}(x,t) \times\square^t\left\{\theta(\Phi_t^{-1}(x))\frac{\psi(\Phi_t^{-1}(x))}{|\psi(\Phi_t^{-1}(x))|}\right\}, \tag{23}$$

where $\square^t:\mathbb{C}\to\mathbb{C}$ is the "parallel translation operator"

$$\square^t(z) = z\exp\left\{i\frac{e}{\hbar}\int_0^t A^{\#}(\dot{\Phi}_s\circ\Phi_t^{-1}(x))ds\right\}. \tag{24}$$

See Sec. VII below.

Next let $U_t:L^2(M;\mathbb{C})\to L^2(M;\mathbb{C})$ be the unitary one-parameter group corresponding to our quantum mechanical Hamiltonian $H = -\frac{1}{2}\hbar^2\Delta + V$, i.e.,

$$U_t = \exp(-it\bar{H}/\hbar)$$

for a suitable self-adjoint extension $\bar{H}$ of H. Such an extension will exist and be uniquely determined under mild conditions on M, A, and V; see Ref. 6.

We write $U_t\psi_0 = \psi_t$.

Define $W(t,s):L^2(M;\mathbb{C})\to L^2(M;\mathbb{C}),\quad s,t\in\mathbb{R}$

by

$$W(t,s) = W_0(s)^{-1}U(s-t)W_0(t) \tag{25}$$

making commutative the diagram:

$$\begin{tikzcd} L^2(M;\mathbb{C}) \arrow[r, "W_0(t)"] \arrow[d, "W(t,s)"'] & L^2(M;\mathbb{C}) \arrow[d, "U(s-t)"] \\ L^2(M;\mathbb{C}) \arrow[r, "W_0(s)"'] & L^2(M;\mathbb{C}). \end{tikzcd} \tag{26}$$

Proposition 4: The family $\{W(t,s)\}$ is a (time inhomogeneous) semigroup of unitary operators:

$$W(t,s)W(u,t) = W(u,s),\quad 0\leqslant s,t,u\leqslant\tau. \tag{27}$$

For $\theta:M\to\mathbb{C}$ a C^2 function of compact support

$$\begin{aligned}\lim_{s\to0^+}s^{-1}(W(t+s,t)\theta-\theta) &= -\tfrac{1}{2}i\hbar\phi^{-1/2}(\Phi_t(\cdot),t)(\Delta_M(\phi^{1/2}\theta\circ\Phi_t^{-1}))\circ\Phi_t \\ &= -\tfrac{1}{2}i\hbar\Delta_t\theta,\ \text{say}\end{aligned} \tag{28}$$

with convergence in $L^2(M;\mathbb{C})$ for each $t\in[0,\tau)$.

Proof: It is clear that we have a unitary semigroup. For (28) we proceed as in Ref. 2. Using Eq. (21):

$$\begin{aligned}\frac{\partial}{\partial t}(W_0(t)\theta) &= \left(\frac{i}{\hbar}\frac{\partial S}{\partial t} - \tfrac{1}{2}\,\mathrm{div}(\nabla S + Z^{\#})\right)W_0(t)\theta \\ &\quad - \exp\left(\frac{i}{\hbar}S\right)\langle\nabla(\phi^{1/2}\theta\circ\Phi_t^{-1}), \\ &\quad \nabla S + Z^{\#}\rangle\psi|\psi|^{-1} \\ &= \left(-\tfrac{1}{2}\frac{i}{\hbar}|\nabla S + Z^{\#}|^2 - \frac{i}{\hbar}V\right. \\ &\quad \left.-\tfrac{1}{2}\mathrm{div}(\nabla S + Z^{\#})\right)W_0(t)\theta \\ &\quad - \exp\left(\frac{i}{\hbar}S\right)\langle\nabla(\phi^{1/2}\theta\circ\Phi_t^{-1}), \\ &\quad \nabla S + Z^{\#}\rangle\psi|\psi|^{-1}.\end{aligned}$$

This derivative can be interpreted as an L^2 derivative by the joint continuity of $\Phi_t(x)$ on $[0,\tau)\times N$, the compactness of the support of θ and the dominated convergence theorem applied to

$$\int_0^1\frac{\partial W_0(t+\lambda s)}{\partial(t+\lambda s)}d\lambda = s^{-1}(W_0(t+s) - W_0(t))$$

as $s\to0^+$.

Working in $L^2(M;\mathbb{C})$, Eqs. (22) and (20) give

$$\begin{aligned}&\frac{\partial}{\partial s}U(-s)W_0(t)\theta\Big|_{s=0} \\ &= -(\tfrac{1}{2}i\hbar\Delta)W_0(t)\theta + (i\hbar)^{-1}VW_0(t)\theta \\ &= -\tfrac{1}{2}i\hbar\exp\left(\frac{i}{\hbar}S\right)\left[\left(\nabla + \frac{i}{\hbar}(\nabla S + Z^{\#})\right)^2\right. \\ &\quad \left.\times(\phi^{1/2}\theta\circ\Phi_t^{-1})\right]\psi|\psi|^{-1} + (i\hbar)^{-1}VW_0(t)\theta\end{aligned}$$

since $\psi/|\psi| = \exp\{i(e/\hbar)S_0\}$ and

$$\nabla - i(e/\hbar)(A - e^{-1}\nabla S - \nabla S_0) = \nabla + (i/\hbar)(\nabla S + Z^{\#})$$

by definition (5).

The result follows on writing out the operator

$$(\nabla + (i/\hbar)(\nabla S + Z^{\#}))^2$$

as in Eq. (4) and using

$$\begin{aligned}&\lim_{s\to0^+}s^{-1}\{W(t+s,t)\theta - \theta\} \\ &= W_0(t)^{-1}\left(\frac{\partial}{\partial t}W_0(t)\theta + \frac{\partial}{\partial s}U(-s)W_0(t)\theta\right)_{s=0}.\end{aligned}$$ $\square$

Let $\Phi_t^*:L^2(M;\mathbb{C})\to L^2(M;\mathbb{C})$ be the unitary map induced by the classical flow Φ_t, $0\leqslant t\leqslant\tau$:

$$\Phi_t^*(\theta)(x) = \phi^{1/2}(x,t)\theta(\Phi_t^{-1}(x)). \tag{29}$$

Note that, just as in Ref. 2, the operator Δ_t of the previous proposition is the transport of Δ_M by the classical flow, i.e.,

$$\Delta_t = \Phi_t^{*-1}\Delta_M\Phi_t^*. \tag{30}$$

Let Δ_M^0 stand for the restriction of Δ_M to the space C_0^∞ of C^∞ functions on M with compact support. It is well known[6,7] that Δ_M^0, considered as an unbounded operator on $L^2(M,\mathbb{C})$ is essentially self-adjoint when M is complete, and so has a unique closure $\bar{\Delta}_M$ with domain $\mathscr{D}(\bar{\Delta}_M)$ say.

Since Δ_t is unitarily equivalent to Δ_M by (30) we can restrict it to C_0^∞ and take its closure $\bar{\Delta}_t$ say, obtaining a self-adjoint operator

$$\bar{\Delta}_t = \Phi_t^{*-1}\bar{\Delta}_M\Phi_t^* \tag{31}$$

with domain $\mathscr{D}(\bar{\Delta}_t) = \{\theta\in L^2(M;\mathbb{C}):\Phi_t^*(\theta)\in\mathscr{D}(\bar{\Delta}_M)\}$. Our next lemma encases some technicalities extending the previous proposition. It will be used to give precise conditions under which our semiclassical approximations are valid.

Lemma 4B: Assume M is complete and the flow Φ is C^3.

(i) $\mathscr{D}(\bar{\Delta}_t)$ consists of those θ in $L^2(M;\mathbb{C})$ for which $\Delta_t\theta$, computed in the sense of distributions, lie in $L^2(M;\mathbb{C})$.

(ii) If $\theta\in\mathscr{D}(\bar{\Delta}_s)$ for $0\leqslant s\leqslant t$ then $\bar{\Delta}_s\theta$ is measurable in $s\in[0,t]$.

(iii) If also $||\bar{\Delta}_s\theta||_{L^2}$ is integrable over $[0,t]$ then

$$W(t,0)\theta = \theta - \tfrac{1}{2}i\hbar\int_0^t W(s,0)\bar{\Delta}_s\theta\, ds. \tag{32}$$

Proof: Since $\bar{\Delta}_M$ is self-adjoint we have $\bar{\Delta}_M = (\Delta_M^0)^*$. It follows immediately from the definitions of adjoint and distributional derivatives that $\mathscr{D}(\bar{\Delta}_M) = \{\theta\in L^2(M;\mathbb{C}): \Delta_M\theta\in L^2(M;\mathbb{C})\}$ and (i) follows in turn.

Since weak measurability implies measurability for separable spaces,[8] and Φ_t^* is strongly continuous in t, to prove (ii) it is enough to show the measurability of $s\mapsto\langle\bar{\Delta}_M(\Phi_s^*)^{-1}\theta,\psi\rangle_{L^2}$ for arbitrary ψ in $L^2(M;\mathbb{C})$. For this take $\psi_i\to\psi$ with $\psi_i\in\mathscr{D}(\bar{\Delta}_M)$ for each i. Then, for $0\leqslant s\leqslant t$,

$$\langle\bar{\Delta}_M(\Phi_s^*)^{-1}\theta,\psi\rangle = \lim_{i\to\infty}\langle(\Phi_s^*)^{-1}\theta,\Delta\psi_i\rangle$$

which is a pointwise limit of continuous functions, and therefore measurable.

For (iii) assume first that $\theta\in C_0^2$. Then by the previous proposition

$$\begin{aligned}
&\lim_{s\to 0^+} s^{-1}[W(t+s,0)\theta - W(t,0)\theta]\\
&\quad = \lim_{s\to 0^+} s^{-1}[W(t,0)W(t+s,t)\theta - W(t,0)\theta]\\
&\quad = -\tfrac{1}{2}(i/\hbar)W(t,0)\Delta_t\theta
\end{aligned}$$

which is continuous in t into $L^2(M;\mathbb{C})$. It follows by general principles (e.g., Ref. 8, Chap. IX, Sec. 3) that $W(t,0)\theta$ is differentiable in t into L^2 and so (32) holds as required.

For general θ let

$$D_t = \left\{\theta\in L^2(M;\mathbb{C}):\theta\in\mathscr{D}(\bar{\Delta}_s) \text{ for } 0\leqslant s\leqslant t \text{ with } \int_0^t ||\bar{\Delta}_s\theta||_{L^2}ds<\infty\right\}. \tag{33}$$

Define

$$||\theta||_t = \int_0^t ||\bar{\Delta}_s\theta||_{L^2}ds + ||\theta||_{L^2}, \quad \theta\in D_t.$$

Since both sides of Eq. (32) are continuous on D_t with its norm $||\ ||_t$ it suffices to show that C_0^2 is dense in D_t. For this let $W=\mathscr{D}(\bar{\Delta}_M)$ made into separable Hilbert space by the norm

$$||\theta||_W = ||\bar{\Delta}_M\theta||_{L^2} + ||\theta||_{L^2},$$

and let $L^{1,1}(W)$ denote the space of absolutely continuous paths

$$\sigma:[0,t]\to W$$

with norm $||\sigma||_{1,1} = ||\sigma(0,1)||_W + \int_0^t ||\dot{\sigma}(s)||_W ds. \quad (34)$

Absolute continuity means that $\sigma(s) = \sigma(0) + \int_0^s\dot{\sigma}(r)dr$, where the derivative $\dot{\sigma}(r)\in W$ is assumed to exist for almost all points r in $[0,t]$ and determine a measurable function of r. For paths in $\mathbb{R}$ it is well known, e.g., see Ref. 9, that this holds if σ is differentiable at all points of $[0,t]$ with integrable derivative. The same is therefore true by the Hahn–Banach theorem for paths σ in W. Since

$$||\theta_t||_t = \int_0^t ||\bar{\Delta}_M\Phi_s^*\theta||_{L^2}ds + ||\Phi_s^*\theta||_{L^2},$$

it follows that $\theta\in D_t$ if and only if $s\mapsto\Phi_s^*$ lies in $L^{1,1}([0,t],W)$, and furthermore

$$\begin{aligned}
\Phi^*:D_t&\to L^{1,1}([0,t],W)\\
\theta&\mapsto[s\mapsto\Phi_s^*\theta]
\end{aligned} \tag{35}$$

is an isometry. Now C_0^∞ is dense in W since $\bar{\Delta}_M$ is the closure of Δ_M^0 and any absolutely continuous path in W can be approximated in $||\ ||_{1,1}$ by a piecewise linear path, and then by such a path whose vertices lie in C_0^∞. Therefore paths with values in C_0^∞ are dense in $L^{1,1}$. Since the inverse image in D_t by Φ_0^* of such a path is in C_0^2 it follows that C_0^2 is dense in D_t, as required. □

Theorem 4 (WKB approximation): Assume that A, V, and S_0, are C^4 real-valued functions on the complete manifold M, and that A and V are such that there is a unitary one-parameter group $\{U_t:t\in\mathbb{R}\}$ on $L^2(M;\mathbb{C})$ describing the evolution determined by the Schrödinger equation

$$\frac{\partial}{\partial t}\psi_t = \tfrac{1}{2}i\hbar\Delta\psi_t + (i\hbar)^{-1}V\psi_t$$

as discussed above.

Let $\psi_0(x) = T_0(x)\exp\{i(e/\hbar)S_0(x)\}$ for $T_0:M\to\mathbf{R}$ a C^2 map. Assume the no caustics condition of Sec. III for the associated classical mechanics. Then $\psi_t = U_t\psi_0$ has the WKB approximation

$$\begin{aligned}
\psi_t(x) &\approx \exp\{(i/\hbar)\check{S}(x,t)\}\phi^{1/2}(x,t)\square\{\psi_0(\Phi_t^{-1}(x))\} && (36)\\
&= \exp\{(i/\hbar)S(x,t) + (ie/\hbar)S_0(x)\}\\
&\quad\times\phi^{1/2}(x,t)T_0(\Phi_t^{-1}(x)) && (37)
\end{aligned}$$

for

$$\begin{aligned}
S(x,t) &= eS_0(\Phi_t^{-1}(x)) + \tfrac{1}{2}\int_0^t|\dot{\Phi}_s\circ\Phi_t^{-1}(x)|^2ds\\
&\quad - \int_0^t V(\Phi_s\circ\Phi_t^{-1}(x))ds\\
&\quad + e\int_0^t A(\dot{\Phi}_s\circ\Phi_t^{-1}(x))ds - eS_0(x),
\end{aligned}$$

where Φ_s is the classical flow coming from the Lorentz force equations (8) and (9). The approximation holds in the sense that

$$\begin{aligned}
&||\psi_t(x) - \exp\left\{\frac{i}{\hbar}\check{S}(x,t)\right\}\phi^{1/2}(x,t)\square\{\psi_0(\Phi_t^{-1}(x))\}||_{L^2}\\
&\quad\leqslant\tfrac{1}{2}\hbar\int_0^t||\Delta_sT_0||_{L^2}ds
\end{aligned} \tag{38}$$

for Δ_s as given in Proposition 4, whenever $\Delta_sT_0\in L^2(M;\mathbb{C})$ for $0\leqslant s\leqslant t$ and the integral is finite. If M is not complete the result still holds for ψ_0 of compact support.

[Note that

$$\|\Delta_s T_0\|_{L^2} = \|\Delta_M(\phi^{1/2}(-,s)T_0\circ\Phi_s^{-1}(\cdot)\|_{L^2} \tag{39}$$

which is "independent of $\hbar$."]

Proof: The C^4 assumptions ensure that Φ_s is C^3 on M. According to the previous lemma if the right-hand side of (38) exists and is finite then

$$W(t,0)T_0 = T_0 - \tfrac{1}{2}i\hbar\int_0^t W(s,0)\bar{\Delta}_s T_0\,ds. \tag{40}$$

Therefore,

$$\|W(t,0)T_0 - T_0\|_{L^2} \leqslant \tfrac{1}{2}\hbar\int_0^t \|\bar{\Delta}_s T_0\|_{L^2}ds.$$

But

$$\begin{aligned}\|W(t,0)T_0 - T_0\|_{L^2} &= \|W_0(t)T_0 - U(t)W_0(0)T_0\|_{L^2}\\ &= \|W_0(t)T_0 - \psi_t\|_{L^2}\end{aligned}$$

and so (38) holds. For noncomplete M Eq. (40) still holds for T_0 of compact support, by direct reference to Lemma 4B. □

Remarks: (i) On occasion in the proofs we have written terms like Im log ψ_0 to mean $(e/\hbar)S_0$ even if ψ_0 itself vanishes. This is safe since we have assumed S_0 defined throughout M. We have taken T_0 real valued: an assumption would be T_0 non-negative so that $T_0 = |\psi_0|$. In fact the proof works equally well for T_0 complex valued.

(ii) Note that the WKB approximation depends on the vector potential (through the initial condition on the classical flow). In particular the magnetic field shows its presence at the WKB level even if the magnetic induction B vanishes.

V. SEMICLASSICAL EXPANSION

Watling's derivation[3] of a precise expression for the semiclassical expansion extends immediately to the case of a charged particle in a magnetic field.

Theorem 5: (Semiclassical expansion): Suppose that all the assumptions of Theorem 4 hold. Then, for $m = 1,2,3,\ldots$, if

$$\Delta_{s_p}\Delta_{s_{p-1}}\cdots\Delta_{s_1}T_0\in L^2(M;\mathbb{C}),\quad 0<s_1<s_2<\ldots<s_p<t,$$

with

$$\int_0^t\int_{s_1}^t\cdots\int_{s_{p-1}}^t \|\Delta_{s_p}\cdots\Delta_{s_1}T_0\|_{L^2}ds_p\cdots ds_1 < \infty$$

for $p = 1,2,\ldots,m+1$, we have

$$\begin{aligned}\psi_t(x) = \exp&\left\{\frac{i}{\hbar}\check{S}(x,t)\right\}\phi^{1/2}(x,t)\Box^t\{\psi_0(\Phi_t^{-1}(x))\}\\ &+\left[\sum_{k=1}^m A_k(\Phi_t^{-1}(x),t)(i\hbar)^k\right.\\ &\left.+R_m(\Phi_t^{-1}(x),t)(i\hbar)^{m+1}\right]\\ &\times\exp\{(ie/\hbar)S_0(\Phi_t^{-1}(x))\},\end{aligned} \tag{41}$$

$$A_k(x,t) = 2^{-k}\int_0^t\int_{s_1}^t\cdots\int_{s_{k-1}}^t \Delta_{s_k}\cdots\Delta_{s_1}(T_0)(x)ds_k\cdots ds_1$$

and

$$\begin{aligned}R_m(x,t) = \frac{1}{2^{m+1}}&\int_0^t\int_{s_1}^t\cdots\int_{s_m}^t W(s_{m+1},t)\\ &\times\Delta_{s_{m+1}}\cdots\Delta_{s_1}(T_0)(x)ds_{m+1}\cdots ds_1.\end{aligned}$$

The result still holds for M incomplete provided T_0 is in the space C_0^{2m+2} of C^{2m+2} functions of compact support and S_0, A, V are of class C^{2m+4}. Note that

$$\|R_m(-,t)\|_{L^2} \leqslant \frac{1}{2^{m+1}}\int_0^t\int_{s_1}^t\cdots\int_{s_m}^t \|\Delta_{s_{m+1}}\cdots\Delta_{s_1}T_0\|_{L^2}ds_{m+1}\cdots ds_1,$$

which is "independent of $\hbar$."

Proof: First note that for $0<s,t<\tau$

$$W(s,t)\theta = \theta - \tfrac{1}{2}i\hbar\int_t^s W(r,t)\bar{\Delta}_r\theta\,dr,$$

whenever $\theta\in D_r$ for all r between t and s and $\int_s^t\|\bar{\Delta}_r\theta\|_{L^2}dr<\infty$. This follows exactly as for the case $t=0$ in Lemma 4B. Next let D_t^m be the space of all $\theta\in L^2(M;\mathbb{C})$ with

$$\Delta_{s_p}\Delta_{s_{p-1}}\cdots\Delta_{s_1}\theta\in L^2,\quad 0<s_1<s_2<\ldots<s_p<t$$

for $p = 1,2,\ldots,m+1$ (where the derivatives are in the sense of distributions on M). Then for $\theta\in D_t^2$

$$\begin{aligned}W(t,0)\theta &= \theta - \tfrac{1}{2}i\hbar\int_0^t W(s,0)\bar{\Delta}_s\theta\,ds\\ &= \theta - \tfrac{1}{2}i\hbar W(t,0)\int_0^t W(s,t)\bar{\Delta}_s\theta\,ds\\ &= \theta - \tfrac{1}{2}i\hbar W(t,0)\int_0^t\bar{\Delta}_s\theta\,ds\\ &\quad+(-\tfrac{1}{2}i\hbar)^2W(t,0)\int_0^t\int_t^s W(r,t)\bar{\Delta}_r\bar{\Delta}_s\theta\,dr\,ds\end{aligned}$$

provided the integrals exist (the last one as an iterated integral). By induction, for $\theta\in D_t^m$

$$\begin{aligned}W(t,0)\theta &= \theta - W(t,0)\{\tfrac{1}{2}i\hbar\int_0^t\bar{\Delta}_s\theta\,ds+\cdots\\ &\quad+(\tfrac{1}{2}i\hbar)^m\int_0^t\int_{s_1}^t\cdots\int_{s_{m-1}}^t\bar{\Delta}_{s_m}\cdots\bar{\Delta}_{s_1}\theta\,ds_m\cdots ds_1\\ &\quad+(\tfrac{1}{2}i\hbar)^{m+1}\int_0^t\int_{s_1}^t\cdots\int_{s_m}^t\\ &\quad\times W(s_{m+1},t)\bar{\Delta}_{s_{m+1}}\cdots\bar{\Delta}_{s_1}\theta\,ds_{m+1}\cdots ds_1\}\end{aligned}$$

provided the integrals exist [in $L^2(M;\mathbb{C})$ as always].

The result follows on operating on both sides of this equation by $U(t)W_0(0)$ and taking $\theta = T_0$. When θ is in C_0^{2n} and the flow Φ_t is C^{2n+1} then the integrals exist with continuous integrands, and completeness of M is not required in the proof. □

VI. THE NO CAUSTICS PROPERTY

A. The derivative of the classical flows

Let $R^{\#}:TM\to TM$ denote the family of skew-symmetric linear maps $T_xM\to T_xM$, determined by the curvature form R of the magnetic field:

$$R^{\#} = \nabla A - (\nabla A)^* = e^{-1}((\nabla Z^{\#})^* - (\nabla Z^{\#}))$$

so that if $v\in T_xM$

$$R^{\#}(v) = R(v,-)^{\#} = v\wedge B(x)\in T_xM$$

for $M = \mathbf{R}^3$.

For fixed $a \in M$ define

$$O_t : T_{\Phi_t(a)} M \to T_{\Phi_t(a)} M, \quad t \geq 0$$

by

$$\frac{D}{\partial t} O_t = \tfrac{1}{2} e O_t R^{\#} \tag{42}$$

[i.e. $(D/\partial t) O_t(v_t) = \frac{1}{2} e O_t R^{\#}(v_t)$ whenever $\{v_t : t \geq 0\}$ is a parallel vector field along $\{\Phi_t(a) : t \geq 0\}$], with $O_0 = I$.

Then each O_t is orthogonal since $R^{\#}$ is skew symmetric on each tangent space.

Let R_M denote the curvature tensor of our Riemannian manifold M, using Kobayashi and Nomizu's sign convention.[10]

Lemma 6A: Set $K_t = O_t \circ T\Phi_t : T_a M \to T_{\Phi_t(a)} M$, $t \geq 0$. Then K_t satisfies

$$\frac{D^2}{\partial t} K_t v + P_t K_t v = 0, \quad t \geq 0, v \in T_a M$$

$$K_0 = I, \ \frac{D}{\partial t} K_t \bigg|_{t=0} = Q, \tag{43}$$

where

$$P_t : T_{\Phi_t(a)} M \to T_{\Phi_t(a)} M \quad \text{and} \quad Q : T_a M \to T_a M$$

are given by $P_t = O_t \tilde{P}_t O_t^{-1}$ for

$$\begin{aligned}\tilde{P}_t w = {} & \nabla^2 V(w) - R_M[\dot{\Phi}_t(a), w]\dot{\Phi}_t(a) \\ & + e\nabla R^{\#}(w)\dot{\Phi}_t(a) - \tfrac{1}{2} e \nabla R^{\#}(\dot{\Phi}_t(a)) w \\ & - \tfrac{1}{4} e^2 R^{\#} R^{\#}(w), \quad w \in T_{\Phi_t(a)} M\end{aligned}$$

and

$$Q(v) = \tfrac{1}{2}(\nabla Z^{\#}(v) + (\nabla Z^{\#})^*(v)), \quad v \in T_a M. \tag{44}$$

Proof: Differentiation of the equations of motion (8) and (9) as in Sec. 3F of Ref. 2 yields

$$\begin{aligned}\frac{D^2}{\partial t^2} T\Phi_t(v) & - R_M[\dot{\Phi}_t(a), T\Phi_t(v)]\dot{\Phi}_t(a) \\ & + \nabla^2 V(T\Phi_t(v)) + eR^{\#}\left(\frac{D}{\partial t} T\Phi_t(v)\right) \\ & + e\nabla R^{\#}(T\Phi_t(v))\dot{\Phi}_t(a) = 0,\end{aligned} \tag{45}$$

with

$$T\Phi_0(v) = v \text{ and } \frac{D}{\partial t} T\Phi_t(v) \bigg|_{t=0} = \nabla Z^{\#}(v)$$

for all $v \in T_a M$.

Also, using the definition (42) of O_t

$$\frac{D}{\partial t} K_t(v) = \tfrac{1}{2} e O_t R^{\#} \circ T\Phi_t(v) + O_t \frac{D}{\partial t}(T\Phi_t(v))$$

and so

$$\begin{aligned}\frac{D^2}{\partial t^2} K_t(v) = {} & \tfrac{1}{4} e^2 O_t R^{\#} R^{\#} \circ T\Phi_t(v) \\ & + \tfrac{1}{2} e O_t \nabla R^{\#}(\dot{\Phi}_t(a)) \circ T\Phi_t(v) \\ & + e O_t R^{\#} \frac{D}{\partial t}(T\Phi_t(v)) + O_t \frac{D^2}{\partial t^2}(T\Phi_t(v))\end{aligned} \tag{46}$$

with

$$\begin{aligned}\frac{D}{\partial t} K_t(v) \bigg|_{t=0} & = \tfrac{1}{2} e R^{\#} v + \nabla Z^{\#}(v) \\ & = \tfrac{1}{2}(\nabla Z^{\#}(v) + (\nabla Z^{\#})^*(v))\end{aligned}$$

and (43) follows on elimination of $(D^2/\partial t^2) T\Phi_t(v)$ from (45) via (46). □

B. A criterion for Φ_t to be a diffeomorphism

Theorem 6B: For fixed $\tau_1 > 0$ let $\rho^2 > 0$ be an upper bound over (t,a) in $[0,\tau_1] \times M$ of the operator norm of $\tilde{P}_t : T_{\Phi_t(a)} M \to T_{\Phi_t(a)} M$, given by (44). Let λ, Λ be the lower and upper bounds of the set of the eigenvalues of the operators $\frac{1}{2}(\nabla Z^{\#} + (\nabla Z^{\#})^*)$ on $T_a M$, $a \in M$. Assume λ and Λ are finite.

Let $t = \tau_2$ be the least positive solution of the equation

$$(1 + \lambda t)(2 + \lambda t + \Lambda(t - \rho^{-1} \sinh \rho t) - \cosh \rho t) = 0.$$

Then, if M is complete, $\Phi_t : M \to M$ is a diffeomorphism for $0 \leq t \leq \min(\tau_1, \tau_2)$.

Proof: The proof is immediate from the previous Lemma and Lemmas 3H and 3I of Ref. 2. □

Corollary[11]*:* If M is compact or more generally if M is complete and $\nabla V, \nabla^2 V, Z, \nabla Z, \nabla R^{\#}$, and R_M are all uniformly bounded on M then there exists $\tau > 0$ such that $\Phi_t : M \to M$ is defined and a diffeomorphism for $0 \leq t \leq \tau$. □

VII. GAUGE INVARIANCE: FIBER BUNDLE FORMULATION

A. The Dirac quantization condition and the underlying fiber bundles

The conventional idea is that the gauge invariance of our Schrödinger equation has a geometric formulation with a principal bundle coming from the classical magnetic field and an associated Hermitian complex line bundle whose sections are the wave functions under consideration. The group G of the bundle is either $\mathbb{R}$ or U(1); the latter being needed to take into account the possibility of magnetic monopoles. The coupling of the charged particle with the magnetic field is expressed by the representation of G into U(1) which determines the associated bundle. The simplest interpretation of our formulas (2) and (3) is that this representation is

$$\alpha \mapsto \exp\{i(e/\hbar)\alpha\}, \quad G = \mathbb{R},$$

or

$$\exp i\theta \mapsto \exp\{i(e/\hbar)\theta\}, \quad G = \mathrm{U}(1) = S^1.$$

There are two immediate objections to this. The most serious is that for $G = \mathrm{U}(1)$ we do not have a representation unless $e/\hbar$ is an integral multiple of 2π. The second is that $e/\hbar$ has dimensions $[e\hbar^{-1}] = QM^{-1}L^{-2}T$ and one can only take exponentials of dimensionless quantities.

Both problems are got over by introducting the magnetic charge g, dimensions $[g] = Q^{-1}ML^2T^{-1}$, and letting the representations be

$$\alpha \mapsto \exp\{i(e/\hbar) g\alpha\}, \quad G = \mathbb{R},$$

or

$$\exp\{i\theta\} \mapsto \exp\{i(e/\hbar) g\theta\}, \quad G = \mathrm{U}(1).$$

The Dirac quantization condition[12,13] ensures that if there is a magnetic monopole in the universe then eg must be an integral multiple of $2\pi\hbar$.

The "classical" principal bundle then has the connection represented by $g^{-1}A$ while that of the associated bundle is still represented by $e\hbar^{-1}A$, both having dimension L^{-1}. None of this makes any essential difference to our formulas apart from the fact that the gauge transformation in Eq. (2) should really be writter

$$\psi \mapsto \exp\{i(eg/\hbar)\theta(\cdot)\}\psi(\cdot)$$

and it would have been more sensible, geometrically, to work in terms of $g^{-1}A, g^{-1}S_0, g^{-1}R^{\#}$, etc., rather than $A, S_0, R^{\#},\ldots$. To avoid confusion we shall from now on assume we are working in units in which $g = 1$.

B. Assumptions on the bundle and initial wave function

Let $\pi{:}L\to M$ be our associated line bundle. In writing (1) we have assumed that we have chosen a trivialization of π. In fact a basic assumption was that S_0 should be defined on the whole of M. When $T_0 = |\psi_0|$ this implies in particular that the "angular part" $\psi_0|\psi_0|^{-1}$ of our initial wave function extends to a section of the S^1 bundle inside L and so gives a trivialization of π, even if T_0 vanishes at some points of M. In this trivialization S_0 is represented as being identically zero: it has been "gauged away." Thus our basic assumption of the existence of S_0 can be replaced by the gauge invariant assumption that $\psi_0|\psi_0|^{-1}$ has an extension over all M into the S^1 bundle. This way the wave function $\psi_0(z) = |z|e^{i\theta/\hbar}$ on $M = \mathbb{R}^2 - \{0\} = \mathbb{C} - \{0\}$ is allowed (but was not in Ref. 2 even for zero magnetism) while $\psi_0(z) = |z|e^{i\theta/\hbar}$ on $M = \mathbb{R}^2 = \mathbb{C}$ is not covered even for zero magnetism.

The reason we need such an assumption, even with no magnetism, is to have Z defined, Eq. (5), so that the classical mechanics, Eq. (9), makes sense for all intial points of M.

C. The geometry of the proof

Let $L^2(M;\pi)$ denote the space of L^2 sections of π. Throughout the proof we identified this, the space of wave functions, with $L^2(M;\mathbb{C})$ by using our fixed trivialization. In fact the Hamiltonian H is an operator with domain in $L^2(M;\pi)$ and $\{U_t{:}t\in\mathbb{R}\}$ acts on $L^2(M;\pi)$. However $W_0(t)$, given in Eq. (23) should be considered as a map

$$W_0(t){:}L^2(M;\mathbb{C})\to L^2(M;\pi),$$

with (26) replaced by

$$\begin{tikzcd} L^2(M;\mathbb{C}) \arrow[r] \arrow[d, "W(t,s)"'] & L^2(M;\pi) \arrow[d, "U(s-t)"] \\ L^2(M;\mathbb{C}) \arrow[r] & L^2(M;\pi). \end{tikzcd}$$

Note that Φ_t^{*}, Δ_M and therefore Δ_t, all act on $L^2(M;\mathbb{C})$.

The connection on π is represented in our trivialization by the one-form $-ie\hbar^{-1}A^{\#}$ and in the trivialization determined by $\psi_0|\psi_0|^{-1}$, or its extension, when $|\psi_0| = T_0$, it is represented by $\hbar^{-1}Z$. This determines the initial velocity $Z^{\#}$ for the classical flow, Eq. (9).

The parallel translation operator $\square^t$ of Eq. (24) maps $\pi^{-1}(\Phi_t^{-1}(x))$ to $\pi^{-1}(x)$ along the path $\{\Phi_s\Phi_t^{-1}(x){:}0\leqslant s\leqslant t\}$. Our expression (36) for the WKB approximation is geometric and manifestly gauge invariant. In it the "external" phase change coming from $\check{S}$ is separated from the "internal" change coming from the parallel translation. The same is true for the semiclassical expansion (41).

ACKNOWLEDGMENTS

Conversations with J. Rawnsley and C. J. Isham helped our understanding of some of the geometric aspects. This research is part of a program supported by S. E. R. C. grant GR/C/13644. The typing was by Peta McAllister.

[1] A. Truman, J. Math. Phys. **18**, 2308 (1977).
[2] K. D. Elworthy and A. Truman, J. Math. Phys. **22**, 2144 (1981).
[3] K. Watling, unpublished manuscripts. See also the preliminary report "Semi-classical Expansions for Degenerate Diffusions," and Warwick Ph. D. thesis in preparation, 1983.
[4] K. D. Elworthy and A. Truman, "The diffusion equation and classical mechanics: an elementary formula," in *Stochastic Processes in Quantum Physics* edited by S. Albeverio *et al.* (Springer-Verlag, Berlin, 1982), Lecture Notes in Physics 173. pp. 136–146.
[5] S. Albeverio and T. Arede, "The relationships between quantum mechanics and classical mechanics: a survey of some mathematical aspects." Preprint: Zentrum für interdisziplinäre Forschung, Universität Bielefeld, West Germany, 1983. Also A. Weinstein has brought our attention to the articles by H. Hogreve, J. Potthoff, and R. Schrader, Comm. Math. Phys. **91**, 573 (1983), especially Sec. 6, and its sequel by R. Schrader and M. E. Taylor, Commun. Math. Phys. **92**, 555 (1984).
[6] (a) P. R. Chernoff, J. Funct. Anal. **12**, 401 (1973). (b). P. R. Chernoff, Pacific J. Math. **72**, 361 (1977).
[7] (a) M. P. Gaffney, Ann. Math. (NY) **60**, 140 (1954). (b) R. S. Strichartz, J. Funct. Anal. **52**, 48 (1983),
[8] K. Yosida, *Functional Analysis*, Grundlehren der math. Wissenschaftan, 123 (Springer-Verlag, Berlin, 1968), 2nd ed.
[9] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966).
[10] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (John Wiley–Interscience, New York, (1963), Vol. 1.
[11] For the case $M = \mathbb{R}^n$ see also D. Fujiwara, Duke Math. J. **47**, 559 (1980).
[12] P. A. M. Dirac, Proc. R. Soc. London **133**, 60 (1931).
[13] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).

Hamiltonian representation for helically symmetric magnetic fields

Leaf Turner[a)]
Department of Physics, University of Wisconsin, Madison, Wisconsin 53706

(Received 10 May 1984; accepted for publication 4 January 1985)

It is proved that one of the two potentials used in a standard representation of helically symmetric magnetic fields is a Hamiltonian that generates the magnetic lines of force. It is proved also that if the other potential is restricted to be proportional to this Hamiltonian, then the radial components of current density and magnetic field are proportional. This restriction applies, for example, to constant-λ, helically symmetric, force-free magnetic fields [$\mathbf{j}(\mathbf{r}) = \lambda\,\mathbf{B}(\mathbf{r})$] relevant to both fusion physics and astrophysics.

I. INTRODUCTION

First we shall introduce the representation for a helically symmetric magnetic field. Next we shall both define a momentum, canonically conjugate to the radial coordinate, and use it to verify that one of the two potentials of the representation is a Hamiltonian that generates the magnetic lines of force. Finally we shall prove that if the other potential is proportional to this Hamiltonian, then the radial components of current density and magnetic field are also proportional. Thus the formalism of Hamiltonian mechanics can be applied to elucidate the structure of constant-λ, force-free [$\mathbf{j}(\mathbf{r}) = \lambda\,\mathbf{B}(\mathbf{r})$], helically symmetric magnetic fields relevant to both fusion physics[1] and astrophysics.[2]

II. POTENTIAL REPRESENTATION FOR A HELICALLY SYMMETRIC MAGNETIC FIELD

A helically symmetric magnetic field depending on only r and $m\theta + kz$ can be represented in terms of two potentials ψ_1 and ψ_2 as follows[3]:

$$\mathbf{B}(r,\chi) = \psi_1(r,\chi)\mathbf{e} - \nabla\psi_2(r,\chi)\times\mathbf{e}\,, \tag{1}$$

where

$$\chi \equiv m\theta + kz\,, \tag{2a}$$

$$\mathbf{e} \equiv (kr\hat{\theta} - m\hat{z})/(k^2r^2 + m^2)\,. \tag{2b}$$

Using the properties of $\mathbf{e}$ that

$$\nabla\times\mathbf{e} = -[2km/(k^2r^2 + m^2)]\mathbf{e}\,, \tag{3a}$$

$$\mathbf{e}\cdot\nabla f(r,\chi) = 0\,, \tag{3b}$$

for any scalar function $f(r,\chi)$, one can verify easily that the two terms on the right-hand side of Eq. (1) are each solenoidal. Thus the two potentials ψ_1 and ψ_2 can be prescribed independently to specify an arbitrary magnetic field possessing helical symmetry. To determine ψ_1 and ψ_2 for a specific magnetic field, one must take the scalar and vector products of Eq. (1) with $\mathbf{e}$ and utilize Eqs. (2b) and (3b) to obtain

$$\psi_1(r,\chi) = krB_\theta(r,\chi) - mB_z(r,\chi)\,, \tag{4a}$$

$$\nabla\psi_2(r,\chi) = -\hat{r}[krB_z(r,\chi) + mB_\theta(r,\chi)] + (m\hat{\theta} + kr\hat{z})B_r(r,\chi)\,. \tag{4b}$$

a) Permanent address: Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545.

III. DEMONSTRATION OF HAMILTONIAN NATURE OF ψ_2

We now shall demonstrate that ψ_2 is a Hamiltonian that generates the magnetic lines of force.

Let us define a momentum p, canonically conjugate to r, by[4]

$$p \equiv \int_0^\chi rB_z(r,\chi')d\chi'\,. \tag{5}$$

Note that

$$p = p(r,\chi)\,, \tag{6a}$$

$$\chi = \chi(r,p)\,. \tag{6b}$$

In particular, note that Eq. (5) yields

$$\frac{\partial p(r,\chi)}{\partial\chi} = rB_z(r,\chi)\,. \tag{7}$$

We wish to determine the Hamiltonian $\widetilde{H}(r,p)$ that generates the magnetic lines of force according to Hamilton's equations

$$\frac{dr}{dz} = \frac{\partial\widetilde{H}(r,p)}{\partial p}\,, \tag{8a}$$

$$\frac{dp}{dz} = -\frac{\partial\widetilde{H}(r,p)}{\partial r}\,. \tag{8b}$$

Note that the z coordinate plays the role of the time coordinate in the standard Hamilton's equations of motion. The left-hand sides of Eqs. (8) are, respectively, the rate of change of r and of $p(r,\chi)$ with z as one moves along a given line of force, i.e.,

$$\frac{dr}{dz} = \frac{B_r}{B_z}\,, \tag{9a}$$

$$r\frac{d\theta}{dz} = \frac{B_\theta}{B_z}\,. \tag{9b}$$

According to Eqs. (6), we can define

$$H(r,\chi(r,p)) \equiv \widetilde{H}(r,p)\,. \tag{10}$$

Using Eqs. (6), (9a), and (10), we rewrite Eq. (8a) as

$$\frac{\partial H(r,\chi(r,p))}{\partial p} = \frac{\partial\chi(r,p)}{\partial p}\frac{\partial H(r,\chi)}{\partial\chi} = \frac{B_r}{B_z}\,. \tag{11}$$

Inserting Eq. (7) into Eq. (11), we obtain the implication of the first of Hamilton's equations, Eq. (8a),

 0022-2488/85/050991-03$02.50

$$\frac{\partial H(r,\chi)}{\partial \chi} = rB_r(r,\chi), \tag{12}$$

or equivalently,

$$\frac{1}{r}\frac{\partial H(r,\chi)}{\partial \theta} = mB_r(r,\chi), \tag{13a}$$

$$\frac{\partial H(r,\chi)}{\partial z} = krB_r(r,\chi), \tag{13b}$$

where we used the helical symmetry property, Eq. (2a).

We now derive the corresponding implication of the second of Hamilton's equations, Eq. (8b). Using Eq. (6a), we observe

$$\frac{dp}{dz} = \frac{d\chi}{dz}\frac{\partial p(r,\chi)}{\partial \chi} + \frac{dr}{dz}\frac{\partial p(r,\chi)}{\partial r}.$$

Application of Eqs. (2a), (7), and (9) yields

$$\frac{dp}{dz} = \left(m\frac{d\theta}{dz} + k\right)rB_z + \frac{B_r}{B_z}\frac{\partial p(r,\chi)}{\partial r} = mB_\theta + krB_z + \frac{B_r}{B_z}\frac{\partial p(r,\chi)}{\partial r} \tag{14}$$

as the left-hand side of Eq. (8b). Examining the right-hand side of Eq. (8b), we observe

$$\begin{aligned} -\frac{\partial \widetilde{H}(r,p)}{\partial r} &= -\frac{dH(r,\chi(r,p))}{dr} \\ &= -\frac{\partial H(r,\chi)}{\partial r} - \frac{\partial \chi(r,p)}{\partial r}\frac{\partial H(r,\chi)}{\partial \chi} \\ &= -\frac{\partial H(r,\chi)}{\partial r} + \left[\frac{\partial p(r,\chi)/\partial r}{\partial p(r,\chi)/\partial \chi}\right]\frac{\partial H(r,\chi)}{\partial \chi}, \end{aligned} \tag{15}$$

where we have used Eq. (10) and have employed a standard identity relating partial derivatives. Inserting Eqs. (7) and (12) into Eq. (15) yields

$$-\frac{\partial \widetilde{H}(r,p)}{\partial r} = -\frac{\partial H(r,\chi)}{\partial r} + \frac{B_r}{B_z}\frac{\partial p(r,\chi)}{\partial r} \tag{16}$$

as the right-hand side of Eq. (8b). Equating the right-hand sides of Eq. (14) and Eq. (16), we finally obtain the implication of the second of Hamilton's equations, Eq. (8b),

$$\frac{\partial H}{\partial r} = -\left[mB_\theta(r,\chi) + krB_z(r,\chi)\right]. \tag{17}$$

Comparing Eqs. (13) and (17) with Eq. (4b), we conclude that H and ψ_2 can be identified as

$$H(r,\chi) = \psi_2(r,\chi). \tag{18}$$

Therefore

$$\mathbf{B}(r,\chi) = \psi_1(r,\chi)\mathbf{e} - \nabla H(r,\chi)\times\mathbf{e} \tag{19}$$

defines an arbitrary, helically symmetric magnetic field in which the potential H is a Hamiltonian that generates the lines of force and in which ψ_1 is a measure of the magnetic field strength along the direction of symmetry.

Before proceeding to the particular case in which ψ_1 is restricted to be proportional to H, we wish to evaluate $\nabla\psi_1$ and ψ_2 for the general case. Utilizing Eq. (4a), we note

$$\nabla\psi_1 = \hat{r}\left[k\frac{\partial}{\partial r}(rB_\theta) - m\frac{\partial B_z}{\partial r}\right] + \left(\hat{\theta}\frac{\partial}{\partial \theta} + \hat{z}r\frac{\partial}{\partial z}\right)\left(kB_\theta - \frac{m}{r}B_z\right). \tag{20}$$

Since the current density $\mathbf{j}(r,\chi)$ is merely the curl of the magnetic field, the first bracket can be rewritten

$$k\frac{\partial}{\partial r}(rB_\theta) - m\frac{\partial B_z}{\partial r} = krj_z + mj_\theta + \left(k\frac{\partial}{\partial \theta} - m\frac{\partial}{\partial z}\right)B_r.$$

By virtue of the helical symmetry, we thus obtain

$$\hat{r}\cdot\nabla\psi_1 = krj_z + mj_\theta.$$

The remaining portion of Eq. (20) can be evaluated similarly. We thereby obtain

$$\nabla\psi_1(r,\chi) = \hat{r}\left[krj_z(r,\chi) + mj_\theta(r,\chi)\right] - (m\hat{\theta} + kr\hat{z})j_r(r,\chi). \tag{21}$$

Comparing Eqs. (4b) with (21) and noting Eq. (4a), we obtain the corollary

$$\psi_2(r,\chi) = H(r,\chi) = mA_z(r,\chi) - krA_\theta(r,\chi), \tag{22}$$

where $\mathbf{A}(r,\chi)$ is the associated magnetic vector potential. Note that Eq. (22) is gauge invariant.

IV. APPLICATION TO FORCE-FREE MAGNETIC FIELDS

We now restrict our attention to the particular case

$$\psi_1(r,\chi) = -\lambda H(r,\chi). \tag{23}$$

Equations (4b), (21), and (18) yield

$$j_r(r,\chi) = \lambda B_r(r,\chi), \tag{24a}$$

$$krj_z(r,\chi) + mj_\theta(r,\chi) = \lambda\left[krB_z(r,\chi) + mB_\theta(r,\chi)\right]. \tag{24b}$$

Of course, Eq. (24b) is merely a consequence of Eq. (24a), the helical symmetry, and the solenoidal nature of $\mathbf{j}$ and $\mathbf{B}$.

Because of their success in accounting for the gross features of turbulently relaxed magnetic field configurations, constant-λ, force-free magnetic fields [fields satisfying $\mathbf{j}(\mathbf{r}) = \lambda\,\mathbf{B}(\mathbf{r})$] have been relevant to fusion physicists studying the reversed-field pinch and spheromak configurations as well as to astrophysicists. According to Eqs. (19), (23), and (24a), a force-free field possessing helical symmetry can be represented as[5]

$$\mathbf{B}(r,\chi) = -\lambda H(r,\chi)\mathbf{e} - \nabla H(r,\chi)\times\mathbf{e}. \tag{25}$$

As a result, the elaborate machinery of Hamiltonian mechanics can be invoked to elucidate further the properties of such fields.[6] Hamiltonians have long provided a tool for studying magnetic island structures.[4] We believe that Hamiltonian formalism also can be a tool for studying the "dynamo" mechanism operant in the sustainment of magnetic field structures observed both in laboratory experiments (reversed-field pinches and spheromaks) and in astrophysical entities (earth, sun, etc.).

ACKNOWLEDGMENTS

I wish to thank both Dr. Harry Dreicer and Dr. Richard A. Gerwin of Los Alamos as well as the University of Wisconsin administration for permitting me the privilege of working and wintering in Madison. In addition, I am grateful to the University of Wisconsin's Department of Physics, to its Plasma Group, and, in particular, to Professor Stewart C. Prager and Professor Keith R. Symon for warm hospitality that enhanced the quality of my visit.

This work was performed under the auspices of the U. S. Department of Energy.

[1]J. B. Taylor, Phys. Rev. Lett. **33**, 1139 (1974); M. N. Rosenbluth and M. N. Bussac, Nucl. Fusion **19**, 489 (1979).
[2]R. Lüst and A. Schlüter, Z. Astrophys. **34**, 263 (1954); S. Chandrasekhar and L. Woltjer, Proc. Natl. Acad. Sci. U.S.A. **44**, 285 (1958).
[3]For example, see D. Lortz, Plasma Phys. **10**, 967 (1968).
[4]D. W. Kerst, J. Nucl. Energy, Part C **4**, 253 (1962).
[5]For the more general case in which ψ_1 equals an arbitrary function of the Hamiltonian, say $\psi_1[H(r,\chi)]$, Eq. (24) remains valid, but with a spatially dependent λ defined by $-d\psi_1/dH$.
[6]J. R. Cary and R. G. Littlejohn, Ann. Phys. **151**, 1 (1983).

Ladder and cross terms in second-order distorted Born approximation

Y. Q. Jin and J. A. Kong
Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 20 September 1984; accepted for publication 28 December 1984)

In the strong fluctuation theory for a bounded layer of random discrete scatterers, the second moments of the fields in the second-order distorted Born approximation are obtained for copolarized and cross-polarized fields. The backscattering cross sections per unit area are calculated by including the mutual coherence of the fields due to the coincidental ray paths, and that due to the opposite ray paths, corresponding to the ladder and cross terms in the Feynman diagramatic representation. It is proved that the contributions from ladder and cross terms for the copolarized backscattering cross sections are the same, while the contributions for the cross-polarized backscattering cross sections are of the same order. The bistatic scattering coefficients in the second-order approximation for both the ladder and cross terms are also obtained. The contributions from the cross terms explain the enhancement in the backscattering direction.

I. INTRODUCTION

The strong fluctuation theory[1,2] for a bounded layer of random discrete scatterers has been studied in Ref. 3 by taking into account the singularity of the dyadic Green's functions to calculate for the second moments of the fields with the first-order distorted Born approximation. By including high-order copolarized and cross-polarized second moments, and making use of the Fourier transform of the mean dyadic Green's function and mean unperturbed fields, we shall calculate the second-order copolarized and cross-polarized backscattering cross sections per unit area, and the bistatic scattering coefficients. The second-order backscattering are caused by the coherent waves with the coincidental and opposite ray paths. In the Feynman diagram representation of the Bethe–Salpeter equation, which governs the second moments of the fields, they are expressed, respectively, by the ladder and cross terms. Kuga and Ishimaru[4] have observed the enhancement phenomenon in the backscattering direction of a wave incident upon a layer of latex microspheres. Tsang and Ishimaru[5] explained this enhancement by using the model of point scatterers embedded in a half-space medium, and showed that the enhancement is due to the cross terms' contribution in the backscattering direction. In this paper, we employ the model of a bounded layer of finite-sized discrete scatterers to calculate the ladder and cross terms for backscattering and bistatic scattering and discuss the enhancement in the backscattering direction. In the backscattering direction the ladder and cross terms will be shown to contribute equally to the copolarizations, and give rise to contributions of the same order for the cross polarizations. It also explains the discrepancies between the wave theory and the radiative transfer theory, which is based upon the ladder approximation.

II. SECOND MOMENTS IN THE SECOND-ORDER DISTORTED BORN APPROXIMATION

For a bounded layer of random discrete scatters (Fig. 1), we have

$$\nabla\times\nabla\times\mathbf{E}_1(\mathbf{r})-\omega^2\mu\epsilon(\mathbf{r})\mathbf{E}_1(\mathbf{r})=0, \tag{2.1}$$

where the subscript 1 denotes the fields in the bounded random medium. Introducing an auxiliary permittivity ϵ_g as in Refs. 1–3, we have

$$\nabla\times\nabla\times\mathbf{E}_1(\mathbf{r})-k_g^2\mathbf{E}(\mathbf{r})=k_0^2\{[\epsilon(\mathbf{r})-\epsilon_g]/\epsilon_0\}\mathbf{E}_1(\mathbf{r}) \tag{2.2}$$

and

$$\mathbf{E}_1(\mathbf{r})=\mathbf{E}_1^0(\mathbf{r})+\int d\mathbf{r}_1\,\mathbf{G}_g(\mathbf{r},\mathbf{r}_1)\cdot k_0^2\,\frac{\epsilon(\mathbf{r}_1)-\epsilon_g}{\epsilon_0}\,\mathbf{E}_1(\mathbf{r}_1). \tag{2.3}$$

We let

$$\xi(\mathbf{r})\equiv3(\epsilon_g/\epsilon_0)[\epsilon(\mathbf{r})-\epsilon_g]/[\epsilon(\mathbf{r})+2\epsilon_g] \tag{2.4}$$

and

$$\mathbf{F}_1(\mathbf{r})\equiv\{[\epsilon(\mathbf{r})+2\epsilon_g]/3\epsilon_g\}\mathbf{E}_1(\mathbf{r}) \tag{2.5}$$

and note that

$$\mathbf{G}_g(\mathbf{r},\mathbf{r}_1)=\mathrm{PV}\,\mathbf{G}_g(\mathbf{r},\mathbf{r}_1)-(\mathrm{l}/3k_g^2)\delta(\mathbf{r}-\mathbf{r}_1), \tag{2.6}$$

where PV denotes the principal value. The principal volume is assumed to be a sphere. Substituting (2.4)–(2.6) into (2.3), we obtain the equation

$$\mathbf{F}_1(\mathbf{r})=\mathbf{E}_1^0(\mathbf{r})+k_0^2\int d\mathbf{r}_1\,\mathrm{PV}\,\mathbf{G}_g(\mathbf{r},\mathbf{r}_1)\cdot\xi(\mathbf{r}_1)\mathbf{F}_1(\mathbf{r}_1), \tag{2.7}$$

where the superscript 0 denotes the unperturbed incident

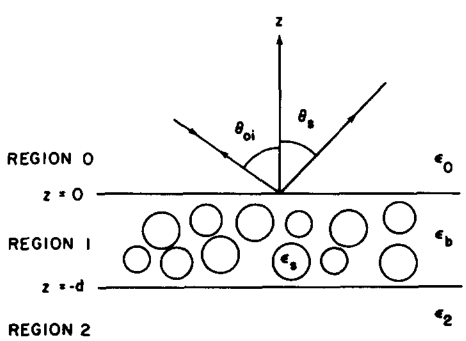


FIG. 1. Geometry of the problem.

 0022-2488/85/050994-18$02.50

wave. Averaging (2.7), we obtain the Dyson equation for $\mathbf{F}_1(\mathbf{r})$:

$$\langle \mathbf{F}_1(\mathbf{r})\rangle = \mathbf{E}_1^0(\mathbf{r}) + k_0^2 \int d\mathbf{r}_1 \mathrm{PV}\, \mathsf{G}_g(\mathbf{r},\mathbf{r}_1)\cdot \mathsf{M}(\mathbf{r}_1,\mathbf{r}_1')\langle \mathbf{F}_1(\mathbf{r}_1')\rangle, \tag{2.8}$$

where the mass operator $\mathsf{M}(\mathbf{r}_1,\mathbf{r}_1')$ is defined as

$$\mathsf{M}(\mathbf{r}_1,\mathbf{r}_1')\langle \mathbf{F}_1(\mathbf{r}_1')\rangle = \langle \xi(\mathbf{r}_1)\mathbf{F}_1(\mathbf{r}_1)\rangle, \tag{2.9}$$

where ϵ_g is determined from $\langle \xi(\mathbf{r})\rangle = 0$.

Thus, from (2.2), the fields observed in region 0 are

$$\mathbf{E}_0(\mathbf{r}) = \mathbf{E}_0^0(\mathbf{r}) + \int d\mathbf{r}_1\, \mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \xi(\mathbf{r}_1)\mathbf{F}_1(\mathbf{r}_1), \tag{2.10}$$

where the effective propagation constant $k_1 = k_{\mathrm{eff}}$ in $\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)$ is calculated according to Refs. 1–3. By carrying out the iteration of $\mathbf{F}_1(\mathbf{r}_1)$ in (2.7), we find the expansion of $\mathbf{E}_0(\mathbf{r})$ as follows:

$$\begin{aligned}\mathbf{E}_0(\mathbf{r}) &= \mathbf{E}_0^0(\mathbf{r}) + k_0^2 \int d\mathbf{r}_1\, \mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \xi(\mathbf{r}_1)\mathbf{E}_1^0(\mathbf{r}_1)\\ &\quad + k_0^4 \int d\mathbf{r}_1\, d\mathbf{r}_2\, \mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \xi(\mathbf{r}_1)\\ &\quad \times \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)\cdot \xi(\mathbf{r}_2)\mathbf{E}_1^0(\mathbf{r}_2)\\ &\quad + \cdots\\ &= \mathbf{E}_0^0(\mathbf{r}) + \sum_{n=1}^{\infty} \mathbf{E}_{0s}^{(n)}(\mathbf{r}),\end{aligned} \tag{2.11}$$

where the superscript (n) denotes the iteration order, the subscript s denotes the scattered fields, and $\mathbf{r}_1,\ldots,\mathbf{r}_n$ are all in the region 1. In (2.11),

$$\begin{aligned}\mathbf{E}_{0s}^{(n)}(\mathbf{r}) &= k_0^{2n} \int d\mathbf{r}_1 \cdots d\mathbf{r}_n\, \mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\\ &\quad \cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)\cdots \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_{n-1},\mathbf{r}_n)\\ &\quad \cdot \xi(\mathbf{r}_1)\cdots \xi(\mathbf{r}_n)\mathbf{E}_1^0(\mathbf{r}_n).\end{aligned} \tag{2.12}$$

We can prove that the terms of $2\,\mathrm{Re}[\mathbf{E}_0^{0*}(\mathbf{r})\Sigma_{n=1}^{\infty}\langle \mathbf{E}_0^{(n)}(\mathbf{r})\rangle]$ are negligible with $\langle \Sigma_{n,m=1}^{\infty} \mathbf{E}_0^{(n)}\mathbf{E}_0^{(m)*}\rangle$ in a similar way as in Ref. 6. Thus from (2.11) we may obtain the second moment of the fields in the region 0 with the first-order distorted Born approximation $\langle |\mathbf{E}_{0s}^{(1)}(\mathbf{r})|^2\rangle_{\alpha\beta}$ as shown in Ref. 3. It can be proved that $\langle \mathbf{F}_1(\mathbf{r})\rangle = \langle \mathbf{E}_1(\mathbf{r})\rangle$. The unperturbed mean field is denoted as $\mathbf{F}_{1m}(\mathbf{r}) \equiv \mathbf{E}_1^0(\mathbf{r})$. Now carrying out further to the *second order*, we find

$$\begin{aligned}&\langle |\mathbf{E}_{0s}^{(2)}(\mathbf{r})|^2\rangle_{\alpha\beta}\\ &\quad = k_0^8 \int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4 \langle \xi(\mathbf{r}_1)\xi(\mathbf{r}_2)\xi^*(\mathbf{r}_3)\xi^*(\mathbf{r}_4)\rangle\\ &\quad \cdot [\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathbf{F}_{1m}(\mathbf{r}_2)]\\ &\quad \times [\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_3)\cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_3,\mathbf{r}_4)\cdot \mathbf{F}_{1m}(\mathbf{r}_4)]_{\alpha\beta}^*,\end{aligned} \tag{2.13}$$

where α, β denote the polarization h or v. The superscript (2) denotes the second order.

Defining $\Gamma(\xi_1 \cdots \xi_n)$ as an n-point cumulant function and denoting $\xi_i = \xi(\mathbf{r}_i)$, $i = 1,2,3,4$, we have the cumulant expansion as follows:

$$\begin{aligned}&\langle \xi_1\xi_2\xi_3^*\xi_4^*\rangle\\ &\quad = \Gamma(\xi_1)\Gamma(\xi_2\xi_3^*\xi_4^*) + \Gamma(\xi_2)\Gamma(\xi_1\xi_3^*\xi_4^*)\\ &\qquad + \Gamma(\xi_3^*)\Gamma(\xi_1\xi_2\xi_4^*) + \Gamma(\xi_4^*)\Gamma(\xi_1\xi_2\xi_3^*)\\ &\qquad + \Gamma(\xi_1\xi_2)\Gamma(\xi_3^*\xi_4^*) + \Gamma(\xi_1\xi_3^*)\Gamma(\xi_2\xi_4^*)\\ &\qquad + \Gamma(\xi_1\xi_4^*)\Gamma(\xi_2\xi_3^*) + \Gamma(\xi_1\xi_2\xi_3^*\xi_4^*).\end{aligned} \tag{2.14}$$

Noting that $\Gamma(\xi_i) = 0$ and the correlation functions have peak values at

(1) $\mathbf{r}_1 = \mathbf{r}_2$ and $\mathbf{r}_3 = \mathbf{r}_4$ for $\Gamma(\xi_1\xi_2)\Gamma(\xi_3^*\xi_4^*)$,

which leads to the integrations with $\mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)$ and $\mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_3,\mathbf{r}_4)$ to be zero, and at

(2) $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}_3 = \mathbf{r}_4$ for $\Gamma(\xi_1\xi_2\xi_3^*\xi_4^*)$,

which leads to the integrations with $\mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)$ and $\mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_3,\mathbf{r}_4)$ to be zero, and at

(3) $\mathbf{r}_1 = \mathbf{r}_3$ and $\mathbf{r}_2 = \mathbf{r}_4$ for $\Gamma(\xi_1\xi_3^*)\Gamma(\xi_2\xi_4^*)$,

(4) $\mathbf{r}_1 = \mathbf{r}_4$ and $\mathbf{r}_2 = \mathbf{r}_4$ for $\Gamma(\xi_1\xi_4^*)\Gamma(\xi_2\xi_3^*)$.

Combining the above results (1)–(4), we obtain

$$\begin{aligned}&\langle |\mathbf{E}_{0s}^{(2)}(\mathbf{r})|^2\rangle_{\alpha\beta}\\ &\quad = (4\pi)^2 k_0^8 W^2\\ &\quad \times \Big\{ \int\!\!\int d\mathbf{r}_1\, d\mathbf{r}_2 [\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathbf{F}_{1m}(\mathbf{r}_2)]\\ &\quad \times [\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathbf{F}_{1m}(\mathbf{r}_2)]_{\alpha\beta}^*\\ &\quad + \int d\mathbf{r}_1\, d\mathbf{r}_2 [\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)(\cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathbf{F}_{1m}(\mathbf{r}_2)]\\ &\quad \times [\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_2)\cdot \mathrm{PV}\, \mathsf{G}_{g11}(\mathbf{r}_2,\mathbf{r}_1)\cdot \mathbf{F}_{1m}(\mathbf{r}_1)]_{\alpha\beta}^* \Big\},\end{aligned} \tag{2.15}$$

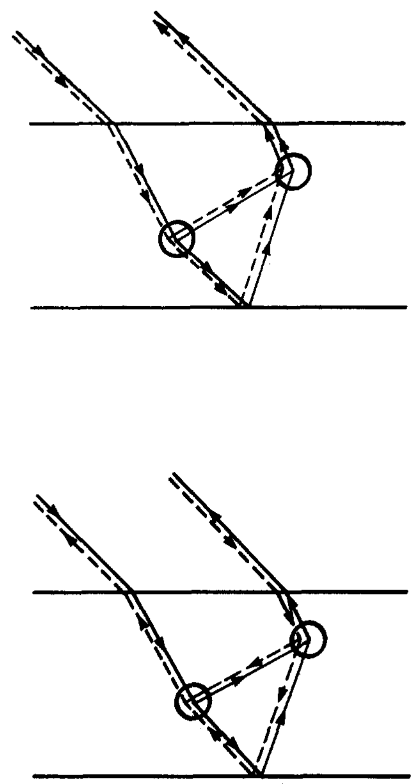


FIG. 2. Second-order scattering figure.

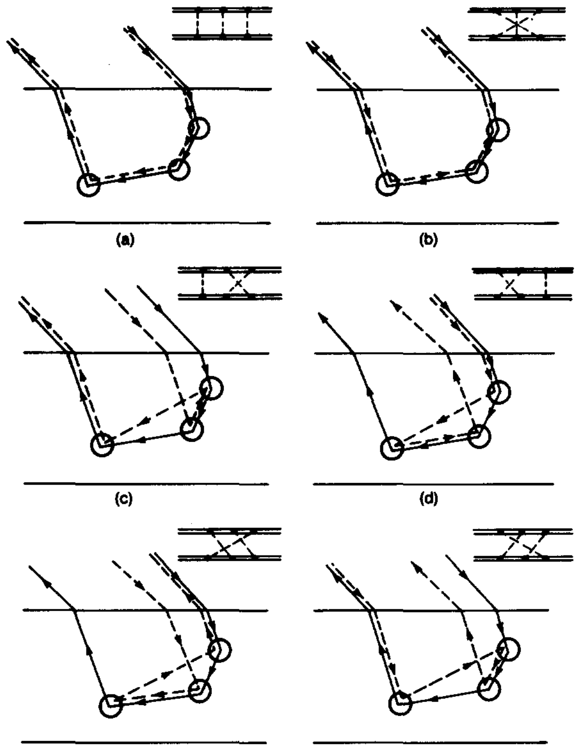


FIG. 3. Third-order scattering figure.

where W is defined as in Refs. 1–3

$$W = \int_0^\infty dr\, r^2 C_\xi(|\mathbf{r}|). \tag{2.16}$$

The correlation functions of random discrete scatterers $C_\xi(|\mathbf{r}|)$ are derived as shown in Ref. 3, which are determined by the scatterers' sizes, fractional volumes, and dielectric permittivities. It is noted that the first integration of (2.15) account for the contributions due to the waves with coincident ray paths which have been taken into account in the RT theory or MRT theory and are derived from the ladder-approximated Bethe–Salpeter equation. The second integration is due to the waves with the opposite ray paths which have not been included in the ladder approximation (Fig. 2). Furthermore, we may proceed to higher order, and see which kinds of cross terms give significant contributions:

$$\begin{aligned}\langle|\mathbf{E}_{0s}^{(3)}(\mathbf{r})|^2\rangle &= k_0^{12}\int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4\, d\mathbf{r}_5\, d\mathbf{r}_6 \langle \xi_1\xi_2\xi_3\xi_4^*\xi_5^*\xi_6^*\rangle \\ &\times \{[\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2) \\ &\cdot \mathrm{PV}\, \mathrm{G}_{g11}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)] \\ &\cdot [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_4)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_4,\mathbf{r}_5) \\ &\cdot \mathrm{PV}\, \mathrm{G}_{g11}(\mathbf{r}_5,\mathbf{r}_6)\cdot \mathbf{F}_{1m}(\mathbf{r}_6)]^*\}. \end{aligned} \tag{2.17}$$

Expanding $\langle \xi_1\xi_2\xi_3\xi_4^*\xi_5^*\xi_6^*\rangle$ into cumulant clusters, and following the discussion of the second order, we only need consider the following clusters:

$$\begin{aligned}\langle \xi_1\xi_2\xi_3\xi_4^*\xi_5^*\xi_6^*\rangle &= \Gamma(1,4)\Gamma(2,5)\Gamma(3,6) + \Gamma(1,6)\Gamma(2,5)\Gamma(3,4) \\ &+ \Gamma(1,4)\Gamma(2,6)\Gamma(3,5) + \Gamma(1,5)\Gamma(2,4)\Gamma(3,6) \\ &+ \Gamma(1,5)\Gamma(2,6)\Gamma(3,4) + \Gamma(1,6)\Gamma(2,4)\Gamma(3,5), \end{aligned} \tag{2.18}$$

where $\Gamma(1,4) \equiv \Gamma(\xi_1,\xi_4^*)$ and so on.

Considering that the correlation function $\langle \xi_i\xi_j^*\rangle$ has peak value at $\mathbf{r}_i = \mathbf{r}_j$, we obtain

$$\begin{aligned}\langle|\mathbf{E}_{0s}^{(3)}(\mathbf{r})|^2\rangle_{\alpha\beta} &= (4\pi W)^3 k_0^{12}\Big\{\int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 |\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)|^2 \\ &+ \int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)] \\ &\times [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_3)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_3,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_1)\cdot \mathbf{F}_{1m}(\mathbf{r}_1)]^* \\ &+ \int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)] \\ &\times [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_3)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_3,\mathbf{r}_2)\cdot \mathbf{F}_{1m}(\mathbf{r}_2)]^* \\ &+ \int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)] \\ &\times [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)]^* \\ &+ \int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)] \\ &\times [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_3)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_3,\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathbf{F}_{1m}(\mathbf{r}_2)]^* \\ &+ \int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_1)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_1,\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathbf{F}_{1m}(\mathbf{r}_3)] \\ &\times [\mathrm{G}_{01}(\mathbf{r},\mathbf{r}_2)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_2,\mathbf{r}_3)\cdot \mathrm{PV}\, \mathrm{G}_{11g}(\mathbf{r}_3,\mathbf{r}_1)\cdot \mathbf{F}_{1m}(\mathbf{r}_1)]^*\Big\}_{\alpha\beta}, \end{aligned} \tag{2.19}$$

where in the backscattering direction the first integration corresponds to the scattering with *coincidental* ray paths [Fig. 3(a)], the second one corresponds to that with *quite-opposite* ray paths [Fig. 3(b)], and the other four integrations correspond to those part-coincidental and part-opposite ray paths [Figs. 3(c)–3(f)]. From Fig. 3, we may conclude that the significant contributions are from the first and second integrations, and there would be no constructive coherences from the other four terms.

III. THE FEYNMAN DIAGRAM

Corresponding to (2.1), we have the equation for the random dyadic Green's function as

$$\nabla\times\nabla\times\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1) - \omega^2\mu\epsilon(\mathbf{r})\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1) = \mathbf{I}\delta(\mathbf{r}-\mathbf{r}_1). \quad (3.1)$$

Introducing ϵ_g as before, we find

$$\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1) = \mathbf{G}_g(\mathbf{r},\mathbf{r}_1) + \int d\mathbf{r}_1'\ \mathbf{G}_g(\mathbf{r},\mathbf{r}_1')\cdot k_0^2\,\frac{\epsilon(\mathbf{r}_1')-\epsilon_g}{\epsilon_0}\,\mathbf{G}_{11}(\mathbf{r}_1',\mathbf{r}_1). \quad (3.2)$$

Defining the new dyadic Green's function

$$\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1) \equiv \{[\epsilon(\mathbf{r}) + 2\epsilon_g]/3\epsilon_g\}\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1) \quad (3.3)$$

and by using (2.4) and (2.6), we obtain from (3.2) and (3.3)

$$\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1) = \mathbf{G}_g(\mathbf{r},\mathbf{r}_1) + k_0^2\int d\mathbf{r}_1'\ \mathrm{PV}\ \mathbf{G}_g(\mathbf{r},\mathbf{r}_1')\cdot\xi(\mathbf{r}_1')\mathbf{G}_{11}(\mathbf{r}_1',\mathbf{r}_1). \quad (3.4)$$

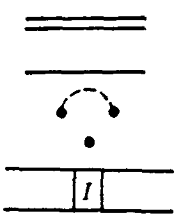
Averaging (3.4), we obtain the Dyson equation of the dyadic Green's function

$$\langle\mathbf{G}_{11}(\mathbf{r},\mathbf{r}_1)\rangle = \mathbf{G}_g(\mathbf{r},\mathbf{r}_1) + k_0^2\int d\mathbf{r}_1'\ \mathrm{PV}\ \mathbf{G}_g(\mathbf{r},\mathbf{r}_1')\cdot\mathbf{M}(\mathbf{r}_1',\mathbf{r}_1)\langle\mathbf{G}_{11}(\mathbf{r}_1',\mathbf{r}_1)\rangle, \quad (3.5)$$

which is corresponding to (2.8), where the mass operator is denoted as

$$\mathbf{M}(\mathbf{r}_1',\mathbf{r}_1)\langle\mathbf{G}_{11}(\mathbf{r}_1',\mathbf{r}_1)\rangle = \langle\xi(\mathbf{r}_1')\mathbf{G}_{11}(\mathbf{r}_1',\mathbf{r}_1)\rangle. \quad (3.6)$$

In the Feynman diagram,[7] we define



$\equiv\langle\mathbf{G}(\mathbf{r},\mathbf{r}_1)\rangle$,

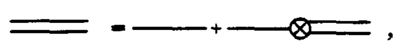
$\equiv\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}_1)$,

$\equiv\langle\xi(\mathbf{r})\xi^*(\mathbf{r}_1)\rangle$,

$\equiv$vertex over which integration is implied,

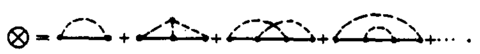
$\equiv\langle\mathbf{G}(\mathbf{r},\mathbf{r})\mathbf{G}^*(\mathbf{p},\mathbf{p}_1)\rangle$.

Thus, the Dyson equation may be expressed as



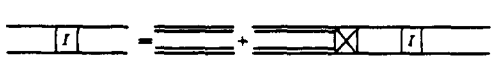
(3.7)

where the mass operator $\otimes$ is



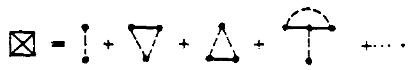
(3.8)

The Bethe–Salpeter equation may be written as



(3.9)

where the intensity operator is

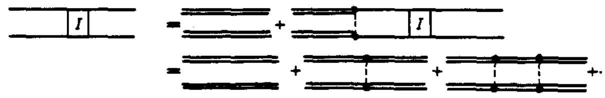


(3.10)

Under the ladder approximation

(3.11)

the Bethe-Salpeter equation is written as

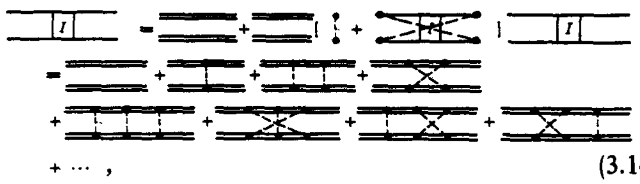


(3.12)

where there are no cross terms. To include cross terms the renormalized intensity operator should be used. According to Ref. 6, the intensity operator is renormalized as

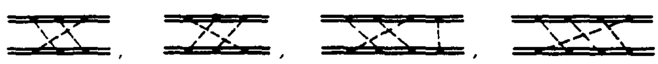
(3.13)

then we have the Bethe–Salpeter equation as



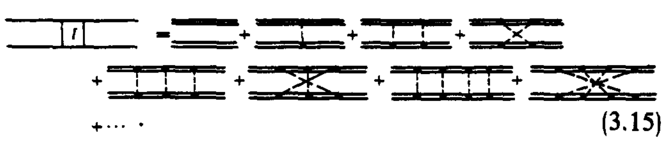
(3.14)

We may see that the cross terms appear in the above renormalized Bethe–Salpeter equation, although some terms such as



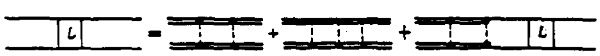
are not included in (3.14).

From the discussion in the last section we only need to keep the ladder and cross terms to explain the constructive interferences in the backscattering direction. So, we should have



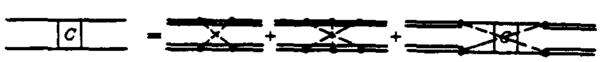
(3.15)

Defining



(3.16)

and



(3.17)

we have

(3.18)

The first and second terms on the right side of (3.18) correspond to the zeroth- and first-order (also a ladder) approximation, which correspond to nonscattering and single scattering of mean fields. The ladder and cross terms at second- and higher-order approximations will correspond to double and higher scattering of the mean fields. In the non-backscattering direction (3.17) is negligible.

IV. THE SECOND-ORDER BACKSCATTERING CROSS SECTIONS FOR THE LADDER AND CROSS TERMS

Now we calculate the second-order backscattering cross sections for the ladder and cross terms. By making use of the Fourier transform of the mean dyadic Green's function and the mean unperturbed field of the two-layer model, we have

$$\mathsf{G}_{01}(\mathbf{r},\mathbf{r}_1)=\int d\,\mathbf{k}_\rho\ \mathsf{g}_{01}^{>}(\mathbf{k}_\rho,z,z_1)\exp[i\mathbf{k}_\rho\cdot(\mathbf{r}_\rho-\mathbf{r}_{1\rho})],\tag{4.1}$$

$$\mathrm{PV}\ \mathsf{G}_{g11}^{\gtrless}(\mathbf{r}_1,\mathbf{r}_2)=\int d\,\mathbf{k}'_\rho\ \mathsf{g}_{11}^{\gtrless}(\mathbf{k}'_\rho,z_1,z_2)\exp[i\mathbf{k}'_\rho\cdot(\mathbf{r}_{1\rho}-\mathbf{r}_{2\rho})],\tag{4.2}$$

$$\mathbf{F}_{1m}(\mathbf{r})=\mathbf{F}_{1m}(z)\exp(i\mathbf{k}_{\rho i}\cdot\mathbf{r}_\rho),\tag{4.3}$$

where $\mathsf{g}_{01}^{>}$, $\mathsf{g}_{11}^{\gtrless}$ and $\mathbf{F}_{1m}(z)$ are listed in Appendix A, i denotes the incident direction, ρ the transversal direction, $>$ for $z_1>z_2$ in $\mathsf{g}_{11}^{>}(\mathbf{k}_\rho,z_1,z_2)$, and vice versa. Substituting (4.1)–(4.3) into (2.13), we obtain the second moment of field in the second order:

$$\begin{aligned}\langle|\mathbf{E}_{0s}^{(2)}(\mathbf{r})|^2\rangle_{\alpha\beta}
&=(2\pi)^4k_0^8(4\pi W)^2\int d\,\mathbf{k}_\rho\int d\,\mathbf{k}'_\rho\int_{-d}^{0}dz_1\int_{-d}^{0}dz_2\\
&\quad\times\{|\mathsf{g}_{01}^{>}(\mathbf{k}_\rho,z,z_1)\cdot\mathsf{g}_{11}^{\gtrless}(\mathbf{k}'_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)|^2\\
&\quad+[\mathsf{g}_{01}^{>}(\mathbf{k}_\rho,z,z_1)\cdot\mathsf{g}_{11}^{\gtrless}(\mathbf{k}'_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)]\\
&\quad\cdot[\mathsf{g}_{01}^{>}(\mathbf{k}_\rho,z,z_2)\cdot\mathsf{g}_{11}^{\gtrless}(\mathbf{k}_{\rho i}-\mathbf{k}'_\rho+\mathbf{k}_\rho,z_2,z_1)\\
&\quad\cdot\mathbf{F}_{1m}(z_1)]^*\}_{\alpha\beta}.\end{aligned}\tag{4.4}$$

In the backscattering direction $\mathbf{k}=-\mathbf{k}_{0i}$, $\mathbf{k}_\rho=\mathbf{k}_{\rho i}$ we rewrite $\mathbf{k}'_\rho$ of (4.4) as $\mathbf{k}_\rho$, and find the second-order backscattering cross section per unit area:

$$\begin{aligned}\sigma_{\alpha\beta}^{(2)}(\mathbf{k}_{0i},-\mathbf{k}_{0i})
&=4\pi k_0^2\cos^2\theta_{0i}\langle|\mathbf{E}_{0s}^{(2)}(\mathbf{k}_{\rho i},z)|^2\rangle/E_0^2\\
&=(2k_0)^{10}\pi^7W^2\cos^2\theta_{0i}\int d\,\mathbf{k}_\rho\int_{-d}^{0}dz_1\int_{-d}^{0}dz_2\\
&\quad\times\{|\mathsf{g}_{01}^{>}(-\mathbf{k}_{\rho i},z,z_1)\cdot\mathsf{g}_{11}^{\gtrless}(\mathbf{k}_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)|^2_{\alpha\beta}\\
&\quad+[\mathsf{g}_{(01)}^{>}(-\mathbf{k}_{\rho i},z,z_1)\cdot\mathsf{g}_{11}^{\gtrless}(\mathbf{k}_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)]\\
&\quad\times[\mathsf{g}_{01}^{>}(-\mathbf{k}_{\rho i},z,z_2)\cdot\mathsf{g}_{11}^{\gtrless}(-\mathbf{k}_\rho,z_2,z_1)\cdot\mathbf{F}_{1m}(z_1)]^*_{\alpha\beta}\}\\
&=(2k_0)^{10}\pi^7W^2\cos^2\theta_{0i}\ [\text{the first term}\\
&\qquad+\text{the second term}].\end{aligned}\tag{4.5}$$

Substituting the expressions of

$$\mathsf{g}_{01}^{>}(-\mathbf{k}_{\rho i},z,z_1)=(1)_{hi}+(2)_{hi}+(3)_{vi}+(4)_{vi},$$
$$\mathsf{g}_{11}^{\gtrless}(\mathbf{k}_\rho,z_1,z_2)=(5)_h+(6)_h+(7)_h+(8)_h+(9)_v+(10)_v+(11)_v+(12)_v,$$
$$\mathbf{F}_{1m}(z)=(13)_{hi}+(14)_{hi}+(15)_{vi}+(16)_{vi}$$

in Appendix A into (4.5), we sort out the same phase terms and find the following expansions which consist of 16 terms:

$$\begin{aligned}&[\mathsf{g}_{01}^{>}(-\mathbf{k}_{\rho i},z,z_1)\cdot\mathsf{g}_{11}^{\gtrless}(\mathbf{k}_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)]_{\alpha\beta}\\
&=[(1)_{hi}+(2)_{hi}+(3)_{vi}+(4)_{vi}]^{>}\cdot\{[(5)_h+(9)_v]\\
&\quad+[(6)_h+(10)_v]+[(7)_h+(11)_v]+[(8)_h+(12)_v]\}^{\gtrless}\\
&\quad\cdot[(13)_{hi}+(14)_{hi}+(15)_{vi}+(16)_{vi}]\\
&=\sum_{s,s'}\sum_{p,p'}[A^{\gtrless}_{\alpha\beta}(\phi)]_{s,s',p,p'}\exp(i\kappa_{s,s'}z_2)\\
&\quad\times\exp(i\kappa_{p,p'}z_1)\exp(ik_{0zi}z),\end{aligned}\tag{4.6}$$

where

$$\begin{aligned}\kappa_{s,s'}&=sk_{1zi}+s'k_{1z},\\ \kappa_{p,p'}&=pk_{1zi}+p'k_{1z},\end{aligned}\tag{4.7}$$

$s,s',p,p'=1$ or -1. Here, $\phi\equiv\phi'-\phi_i$, where ϕ', ϕ_i are the azimuthal angles of $\mathbf{k}_\rho$, $\mathbf{k}_{\rho i}$. The $A^{\gtrless}_{\alpha\beta}(\phi)_{s,s',p,p'}$ ($\alpha\beta=hh,vv,hv,vh$) are listed in Appendix B. The terms of (i) $i=1,\dots,16$ are shown in Appendix A. Thus we have

$$\begin{aligned}\text{the first term of (4.5)}&=\int d\,\mathbf{k}_\rho\int_{-d}^{0}dz_1\left\{\int_{-d}^{z_1}dz_2\left[\sum_{s,s'}\sum_{p,p'}A^{>}_{\alpha\beta}(\phi)_{ss'pp'}\exp(i\kappa_{ss'}z_2)\exp(i\kappa_{pp'}z_1)\right]\right.\\
&\quad\times\left[\sum_{t,t'}\sum_{q,q'}A^{>}_{\alpha\beta}(\phi)_{tt'qq'}\exp(i\kappa_{tt'}z_2)\exp(i\kappa_{qq'}z_1)\right]^*_{\alpha\beta}\\
&\quad+\int_{z_1}^{0}dz_2\left[\sum_{s,s'}\sum_{p,p'}A^{<}_{\alpha\beta}(\phi)_{ss'pp'}\exp(i\kappa_{ss'}z_2)\exp(i\kappa_{pp'}z_1)\right]\\
&\quad\left.\times\left[\sum_{t,t'}\sum_{q,q'}A^{<}_{\alpha\beta}(\phi)_{tt'qq'}\exp(i\kappa_{tt'}z_2)\exp(i\kappa_{qq'}z_1)\right]^*_{\alpha\beta}\right\},\end{aligned}\tag{4.8}$$

where $t,t',q,q'=1$ or -1 as s,s',p,p'. In low loss medium we only need to consider $t=s,t'=s',q=p,q'=p'$, since after integrating over z_2 there are only imaginary parts of k_{1zi} and k_{1z} in the denominators which would give most significant contributions. It follows that

$$\text{the first term of (4.5)}=\int d\,\mathbf{k}_\rho\sum_{s,s'}\sum_{p,p'}[|A^{>}_{\alpha\beta}(\phi)|^2_{ss'pp'}M_{ss'pp'}+|A^{<}_{\alpha\beta}(\phi)|^2_{ss'pp'}N_{ss'pp'}],\tag{4.9a}$$

$$\text{the second term of (4.5)}=\int d\,\mathbf{k}_\rho\sum_{s,s'}\sum_{p,p'}[A^{>}_{\alpha\beta}(\phi)_{ss'pp'}A^{<*}_{\alpha\beta}(\phi+\pi)_{pp'ss'}M_{ss'pp'}+A^{<}_{\alpha\beta}(\phi)_{ss'pp'}A^{>*}_{\alpha\beta}(\phi+\pi)_{pp'ss'}N_{ss'pp'}],\tag{4.9b}$$

where

$$M_{ss'pp'} = \frac{1}{4\kappa''_{ss'}}\left[\frac{1-\exp(2(\kappa''_{ss'}+\kappa''_{pp'})d)}{\kappa''_{ss'}+\kappa''_{pp'}} - \frac{\exp(2\kappa''_{ss'}d)(1-\exp(2\kappa''_{pp'}d))}{\kappa''_{pp'}}\right], \tag{4.10}$$

$$N_{ss'pp'} = \frac{1}{4\kappa''_{ss'}}\left[\frac{1-\exp(2\kappa''_{pp'}d)}{\kappa''_{pp'}} - \frac{1-\exp(2(\kappa''_{ss'}+\kappa''_{pp'})d)}{\kappa''_{ss'}+\kappa''_{pp'}}\right] \tag{4.11}$$

are the integrations of exponential terms over z_1,z_2 in (4.7).

The first summation, (4.8), corresponds to the second-order ladder term (coincidental ray paths), the second one (4.9), corresponds to the cross term (opposite ray paths). From Appendix B, we may see that

$$A^{<*}_{\alpha\alpha}(\phi+\pi)_{pp'ss'} = A^{>*}_{\alpha\alpha}(\phi)_{ss'pp'}, \tag{4.12a}$$

$$A^{>*}_{\alpha\alpha}(\phi+\pi)_{pp'ss'} = A^{<*}_{\alpha\alpha}(\phi)_{ss'pp'}, \tag{4.12b}$$

where $\alpha\alpha = hh$ or vv. Thus, comparing (4.8),(4.9), we conclude that the *ladder* and *cross terms* for *copolarized* waves have the same significant contributions. That is

$$\sigma^{+}_{\alpha\alpha} = \sigma^{-}_{\alpha\alpha}, \tag{4.13}$$

where $\pm$ denote the ladder and cross (or coincidental and opposite) terms, respectively. From Figs. 3(a) and 3(b) we also see physically that the copolarized ladder and cross terms should be the same. However, from A_{hv} or A_{vh} in Appendix B for cross-polarized waves, we do not have the equality as in (4.12a) and (4.12b). Therefore, the ladder and cross terms are generally not equal, but have the same order.

Carrying out (4.8) and (4.9), sorting out the three types of $1/(k''_{1zi}+k''_{1z})^2$, $1/(k''^2_{1zi}-k''^2_{1z})$, $1/(k''_{1zi}-k''_{1z})^2$, we obtain the following results:

$$\begin{aligned}\sigma^{(2)}_{hh} &= \sigma^{(2)+}_{hh} + \sigma^{(2)-}_{hh} = 2\sigma^{(2)+}_{hh}\\ &= \frac{1}{8}W^2k_0^{10}\cos^2\theta_{0i}\frac{T_hT'_h}{|D_{2i}|^4}\\ &\quad\times\int_0^{\pi/2} d\theta\, k_0^2\sin\theta\cos\theta\frac{1}{|k_{1zi}k_{1z}|^2}\\ &\quad\times\left[\frac{1}{(k''_{1zi}+k''_{1z})^2}N_{1h} + \frac{1}{(k''^2_{1zi}-k''^2_{1z})}N_{2h} + \frac{1}{(k''_{1zi}-k''_{1z})^2}N_{3h}\right],\end{aligned} \tag{4.14}$$

where

$$\begin{aligned}N_{1h} &= C_{11}\cdot\left[|R_{12i}|^4e^{-4k''_{1zi}d}\cdot H_1 + H_2\right]\\ &\quad - C_{12}\cdot 2|R_{12i}|^2e^{-4k''_{1zi}d}\cdot H_3\\ &\quad + C_{13}\cdot 2|R_{12i}|^4e^{-4k''_{1zi}d}\cdot H_4,\end{aligned} \tag{4.15a}$$

$$\begin{aligned}N_{2h} &= C_{21}\cdot\left[1+|R_{12i}|^4e^{-4k''_{1zi}d}\right]\cdot H_4\\ &\quad + C_{22}\cdot\left[1+|R_{12i}|^4e^{-4k''_{1zi}d}\right]\cdot H_3\\ &\quad - C_{23}\cdot 2|R_{12i}|^2e^{-4k''_{1zi}d}\cdot[H_1+H_2],\end{aligned} \tag{4.15b}$$

$$\begin{aligned}N_{3h} &= C_{31}\cdot\left[|R_{12i}|^4e^{-4k''_{1zi}d}\cdot H_2 + H_1\right]\\ &\quad - C_{32}\cdot 2|R_{12i}|^2e^{-4k''_{1zi}d}\cdot H_4\\ &\quad + C_{33}\cdot 2|R_{12i}|^2e^{-4k''_{1zi}d}\cdot H_3,\end{aligned} \tag{4.15c}$$

where the coefficients C_{mn}, $m,n = 1,2,3$, are given in Appendix C. The functions H_n, $n = 1,2,3,4$, are given in Appendix D. And

$$T_h = |X_{10i}|^2, \quad T'_h = |X_{01i}|^2. \tag{4.16}$$

It can be obtained[8] that the results from the modified radiative transfer theory (MRT) based upon the ladder approximation is half the value of the wave approach (4.14), since the cross terms $\sigma^{(2)-}_{hh}$ were not included in the ladder approximation.

For the vertical polarization, we obtain

$$\begin{aligned}\sigma^{(2)}_{vv} &= \sigma^{(2)+}_{vv} + \sigma^{(2)-}_{vv} = 2\sigma^{(2)+}_{vv}\\ &= \frac{1}{8}W^2k_0^{10}\cos^2\theta_{0i}\frac{T_vT'_v}{|F_{2i}|^4}\\ &\quad\times\int_0^{\pi/2} d\theta\, k_0^2\sin\theta\cos\theta\frac{1}{|k_{1zi}k_{1z}|^2}\\ &\quad\times\left[\frac{1}{(k''_{1zi}+k''_{1z})^2}N_{1v} + \frac{1}{(k''^2_{1zi}-k''^2_{1z})}N_{2v} + \frac{1}{(k''_{1zi}-k''_{1z})^2}N_{3v}\right],\end{aligned} \tag{4.17}$$

where

$$\begin{aligned}N_{1v} &= C_{11}\cdot\left[|S_{12i}|^4e^{-4k''_{1zi}d}\cdot V_1 + V_2\right]\\ &\quad - C_{12}\cdot 2|S_{12i}|^2e^{-4k''_{1zi}d}\cdot V_3\\ &\quad + C_{13}\cdot 2|S_{12i}|^2e^{-4k''_{1zi}d}\cdot V_4,\end{aligned} \tag{4.18a}$$

$$\begin{aligned}N_{2v} &= C_{21}\cdot\left[1+|S_{12i}|^4e^{-4k''_{1zi}d}\right]\cdot U_4\\ &\quad + C_{22}\cdot\left[1+|S_{12i}|^4e^{-4k''_{1zi}d}\right]\cdot U_3\\ &\quad - C_{23}\cdot 2|S_{12i}|^4e^{-4k''_{1zi}d}\cdot[U_2+U_1]\end{aligned} \tag{4.18b}$$

$$\begin{aligned}N_{3v} &= C_{31}\cdot\left[|S_{12i}|^4e^{-4k''_{1zi}d}\cdot V_2 + V_1\right]\\ &\quad - C_{32}\cdot 2|S_{12i}|^2e^{-4k''_{1zi}d}\cdot V_4\\ &\quad + C_{33}\cdot 2|S_{12i}|^2e^{-4k''_{1zi}d}\cdot V_3,\end{aligned} \tag{4.18c}$$

where V_n and U_n, $n = 1,2,3,4$, are given in Appendix D:

$$T_v = |(k_0/k_1)Y_{01i}|^2, \quad T'_v = |(k_1/k_0)Y_{10i}|^2. \tag{4.19}$$

For the cross-polarized waves we may have

$$\begin{aligned}\sigma^{(2)+}_{hv} &= \frac{1}{8}W^2k_0^{10}\cos^2\theta_{0i}\frac{T_hT_v}{|D_{2i}F_{2i}|^2}\\ &\quad\times\int_0^{\pi/2} d\theta\, k_0^2\sin\theta\cos\theta\frac{1}{|k_{1zi}k_{1z}|^2}\\ &\quad\times\left[\frac{1}{(k''_{1zi}+k''_{1z})^2}M_1^+ + \frac{1}{(k''^2_{1zi}-k''^2_{1z})}M_2^+ + \frac{1}{(k''_{1zi}-k''_{1z})^2}M_3^+\right],\end{aligned} \tag{4.20a}$$

$$\sigma_{hv}^{(2)-} = \frac{1}{8} W^2 k_0^{10} \cos^2\theta_{0i} \frac{T_h T_v}{|D_{2i} F_{2i}|^2}$$
$$\times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta \cos\theta \frac{1}{|k_{1zi} k_{1z}|^2}$$
$$\times \left[\frac{1}{(k''_{1zi} + k''_{1z})^2} M_1^- + \frac{1}{(k''^2_{1zi} - k''^2_{1z})} M_2^- + \frac{1}{(k''_{1zi} - k''_{1z})^2} M_3^-\right], \qquad (4.20b)$$

where

$$M_1^+ = C_{11}\cdot\tfrac{1}{2}\big[|R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\cdot W_1^+ + W_2^+\big] - C_{12}\cdot\tfrac{1}{2}\big[|R_{12i}|^2 + |S_{12i}|^2\big] e^{-4k''_{1z}d}\cdot W_3^+ + C_{13}\cdot\tfrac{1}{2}\big[|R_{12i}|^2 + |S_{12i}|^2\big] e^{-4k''_{1z}d}\cdot W_4^+, \qquad (4.21a)$$

$$M_1^- = C_{11}\cdot\tfrac{1}{2}\big[|R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\cdot W_1^- + W_2^-\big] - C_{12}\cdot\tfrac{1}{2}\big[-R_{12i} S^*_{12i} - R^*_{12i} S_{12i}\big] e^{-4k''_{1z}d}\cdot W_3^- + C_{13}\cdot\tfrac{1}{2}\big[-R_{12i} S^*_{12i} - R^*_{12i} S_{12i}\big] e^{-4k''_{1z}d}\cdot W_4^-; \qquad (4.21b)$$

$$M_2^+ = C_{21}\cdot\tfrac{1}{2}\big[1 + |R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\big]\cdot W_4^+ + C_{22}\cdot\tfrac{1}{2}\big[1 + |R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\big]\cdot W_3^+ - C_{23}\cdot\tfrac{1}{2}\big[|R_{12i}|^2 + |S_{12i}|^2\big] e^{-4k''_{1z}d}\cdot\big[W_1^+ + W_2^+\big], \qquad (4.22a)$$

$$M_2^- = C_{21}\cdot\tfrac{1}{2}\big[1 + |R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\big]\cdot W_4^+ + C_{22}\cdot\tfrac{1}{2}\big[1 + |R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\big]\cdot W_3^+ - C_{23}\cdot\tfrac{1}{2}\big[-R_{12i} S^*_{12i} - R^*_{12i} S_{12i}\big] e^{-4k''_{1z}d}\cdot\big[W_1^+ + W_2^+\big]; \qquad (4.22b)$$

$$M_3^+ = C_{31}\cdot\tfrac{1}{2}\big[|R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\cdot W_2^+ + W_1^+\big] - C_{32}\cdot\tfrac{1}{2}\big[|R_{12i}|^2 + |S_{12i}|^2\big] e^{-4k''_{1z}d}\cdot W_4^+ + C_{33}\cdot\tfrac{1}{2}\big[|R_{12i}|^2 + |S_{12i}|^2\big] e^{-4k''_{1z}d}\cdot W_3^+, \qquad (4.23a)$$

$$M_3^- = C_{31}\cdot\tfrac{1}{2}\big[|R_{12i} S_{12i}|^2 e^{-4k''_{1z}d}\cdot W_2^- + W_1^-\big] - C_{12}\cdot\tfrac{1}{2}\big[-R_{12i} S^*_{12i} - R^*_{12i} S_{12i}\big] e^{-4k''_{1z}d}\cdot W_4^- + C_{33}\cdot\tfrac{1}{2}\big[-R_{12i} S^*_{12i} - R^*_{12i} S_{12i}\big] e^{-4k''_{1z}d}\cdot W_3^-; \qquad (4.23b)$$

where

$$W_1^\pm = \left|\frac{k_{1zi}}{k_1}\right|^2 \left|\frac{R_{12}}{D_2} + \frac{S_{12} k_{1z}^2}{F_2 k_1^2}\right|^2 \pm 4|k_\rho k_{\rho i}|^2 \left|\frac{S_{12} k_{1z}}{k_1^3}\right|^2, \qquad (4.24a)$$

$$W_2^\pm = \left|\frac{k_{1zi}}{k_1}\right|^2 \left|\frac{R_{10}}{D_2} + \frac{S_{10} k_{1z}^2}{F_2 k_1^2}\right|^2 \pm 4|k_\rho k_{\rho i}|^2 \left|\frac{S_{10} k_{1z}}{k_1^3}\right|^2, \qquad (4.24b)$$

$$W_3^\pm = \left|\frac{k_{1zi}}{k_1}\right|^2 \left|\frac{R_{10} R_{12}}{D_2} - \frac{S_{10} S_{12} k_{1z}^2}{F_2 k_1^2}\right|^2 \pm 4|k_\rho k_{\rho i}|^2 \left|\frac{S_{10} S_{12} k_{1z}}{k_1^3}\right|^2, \qquad (4.24c)$$

$$W_4^\pm = \left|\frac{k_{1zi}}{k_1}\right|^2 \left|\frac{1}{D_2} - \frac{k_{1z}^2}{F_2 k_1^2}\right|^2 \pm 4|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}}{k_1^3}\right|^2. \qquad (4.24d)$$

Comparing M_n^+ and M_n^-, we may see that σ_{hv}^+ and σ_{hv}^- are the same order, the difference between them is only that $|R_{12i}|^2 + |S_{12i}|^2$ in M_n^+ replaced by $-2\,\mathrm{Re}[R_{12i} S^*_{12i}]$ in M_n^-.

It can be seen that

$$\sigma_{hv}^{(2)} = \sigma_{hv}^{(2)+} + \sigma_{hv}^{(2)-}, \qquad (4.25)$$

where as $d \to \infty$,

$$\sigma_{hv}^{(2)+} = \tfrac{1}{8} W^2 k_0^{10} \cos^2\theta_{0i} T_h T_v \times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \frac{1}{|k_{1zi} k_{1z}|^2} \times \left\{\frac{1}{(k''_{1zi} + k''_{1z})^2}\left[\frac{1}{8}\left|\frac{k_{1zi}}{k_1}\right|^2 \left|R_{10} + \frac{S_{10} k_{1z}^2}{k_1^2}\right|^2 + \frac{1}{2}|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}}{k_1}\frac{S_{10}}{k_1^2}\right|^2\right] + \frac{1}{k''_{1zi} + k''_{1z}}\frac{1}{k''_{1zi}}\left[\frac{1}{8}\left|\frac{k_{1zi}}{k_1}\right|^2 \left|1 - \frac{k_{1z}^2}{k_1^2}\right|^2 + \frac{1}{2}|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}}{k_1}\frac{1}{k_1^2}\right|^2\right]\right\}, \qquad (4.26a)$$

$$\sigma_{hv}^{(2)-} = \tfrac{1}{8} W^2 k_0^{10} \cos^2\theta_{0i} T_h T_v \times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \frac{1}{|k_{1zi} k_{1z}|^2} \times \left\{\frac{1}{(k''_{1zi} + k''_{1z})^2}\left[\frac{1}{8}\left|\frac{k_{1zi}}{k_1}\right|^2 \left|R_{10} + \frac{S_{10} k_{1z}^2}{k_1^2}\right|^2 - \frac{1}{2}|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}}{k_1}\frac{S_{10}}{k_1^2}\right|^2\right] + \frac{1}{k''_{1zi} + k''_{1z}}\frac{1}{k''_{1zi}}\left[\frac{1}{8}\left|\frac{k_{1zi}}{k_1}\right|^2 \left|1 - \frac{k_{1z}^2}{k_1^2}\right|^2 + \frac{1}{2}|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}}{k_1}\frac{1}{k_1^2}\right|^2\right]\right\}, \qquad (4.26b)$$

while

$$M_{h1}^{\pm} = \frac{1}{8}\left|\frac{k_{1zi}}{k_1}\right|^2 \left|\frac{R_{10}}{D_2} + \frac{S_{10}k_{1z}^2}{F_2 k_1^2}\right|^2 \pm \frac{1}{2}|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}S_{10}}{k_1^3}\right|^2,$$

$$M_{h2}^{\pm} = \frac{k''_{1zi} - k''_{1z}}{k''_{1zi}}\left[\frac{1}{8}\left|\frac{k_{1zi}}{k_1}\right|^2 \left|1 - \frac{k_{1z}^2}{k_1^2}\right|^2 + \frac{1}{2}|k_\rho k_{\rho i}|^2 \left|\frac{k_{1z}}{k_1^3}\right|^2\right].$$

$$M_{h3}^{\pm} = 0.$$

The $\sigma_{hv}^{(2)+}$ of (4.26a) has been obtained in the RT results,[9] while $\sigma_{hv}^{(2)-}$ has not been included.

Note that

$$\int_{-\infty}^{\infty} d\bar{\mathbf{k}}_\rho = \int_0^{\infty} d\theta\, k_0^2 \sin\theta\cos\theta \int_0^{2\pi} d\phi = \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \int_0^{2\pi} d\phi, \tag{4.27}$$

since the observation point is in the far field, only the radiating mode contributing to the integral needs to be considered, and $d\bar{k}_\rho = d\theta_1\, d\phi\, k_1^2 \sin\theta_1\cos\theta_1 = d\theta\, d\phi\, k_0^2 \sin\theta\cos\theta$. So we *truncate* the integral as $\int_0^{\pi/2} d\theta$, where θ,θ_1 are defined in region 0 and region 1, respectively. The effective propagation constant $k_1 = \omega\sqrt{\mu\epsilon_{\rm eff}}$ is calculated according to the strong fluctuation approach,[1–3] where

$$\epsilon_{\rm eff} = \epsilon_g + i\tfrac{2}{3} k_0^2 k_g \epsilon_0 \int_0^{\infty} dr\, r^2 C_\xi(r).$$

We also obtain $\sigma_{vh}^{(2)}$ by carrying out (4.8) and (4.9) in terms of $A_{vh}^{\gtrless}(\phi)$ of Appendix D:

$$\sigma_{vh}^{(2)\pm} = \frac{1}{8} W^2 \frac{k_0^{10}}{k_1^2}\cos^2\theta_{0i} \frac{T'_v T'_h}{|F_{2i}D_{2i}|^2} \times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \times \frac{1}{|k_{1zi}k_{1z}|^2}\left[\frac{1}{(k''_{1zi} + k''_{1z})^2} M_1^{\pm} + \frac{1}{(k''^2_{1zi} - k''^2_{1z})} M_2^{\pm} + \frac{1}{(k''_{1zi} - k''_{1z})^2} M_3^{\pm}\right]. \tag{4.28}$$

It is noted that the difference between (4.28) and (4.20a),(4.20b) is only in the coefficients of T_h, T_v, and T'_h, T'_v, since in the second-order approximation with the spherical scatterers model, the top boundary plays a unique role to effect the difference in the result.

V. BISTATIC SCATTERING AND ENHANCEMENT IN BACKSCATTERING DIRECTION

In the bistatic scattering direction, the constructive interference of scattered fields is dominated by coincidental ray paths. So we temporarily omit the second term in (4.4) to calculate the bistatic scattering coefficient. Mathematically, since the bistatic direction $\mathbf{k}'_\rho$ in $\mathbf{g}_{11}^{\gtrless}(\mathbf{k}'_\rho, z_1, z_2)$ is not equal to $\pm(\mathbf{k}_{\rho i} - \mathbf{k}'_\rho + \mathbf{k}_\rho)$ in $\mathbf{g}_{11}^{\gtrless}(\mathbf{k}_{\rho i} - \mathbf{k}'_\rho + \mathbf{k}_\rho, z_2, z_1)$, the expansion of the second multiplications on the right side of (4.4) will give rise to denominators with $i(k'_{1zi} + k'_{1zs}) + k''_{1z}$, where k'_{1zi}, k'_{1zs} are the real parts of k_{1zi}, k_{1zs}, respectively. As long as $k'_{1zi} + k'_{1zs} \gg k''_{1z}$, the second term is much less than the first term. Thus, in the low loss medium and nonbackscattering direction, we obtain

$$\gamma_{hh}^{(2)}(\theta_{0s},\theta_{0i}) = 4\pi k_0^2 \cos^2\theta_{0i} \langle|\mathbf{E}_{0s}^{(2)}(\mathbf{k}_{\rho s}, z)|^2\rangle / E_0^2 = (2k_0)^{10}\pi^7 W^2 \cos^2\theta_{0i} \int d\mathbf{k}_\rho \int_{-d}^{0} dz_1 \int_{-d}^{0} dz_2 \times |\mathbf{g}_{01}^{>}(-\mathbf{k}_{\rho s}, z, z_1)\cdot\mathbf{g}_{11}^{\gtrless}(\mathbf{k}_\rho, z_1, z_2)\cdot\mathbf{F}_{1m}(z_2)|_{hh}^2, \tag{5.1}$$

with

$$[\mathbf{g}_{01}^{>}(-\mathbf{k}_{\rho s}, z, z_1)\cdot\mathbf{g}_{11}^{\gtrless}(\mathbf{k}_\rho, z_1, z_2)\cdot\mathbf{F}_{1m}(z_2)]_{hh} = \sum_{s,s'}\sum_{p,p'} [D_{hh}^{\gtrless}(\phi)]_{ss'pp'} \exp(i\kappa_{ss'} z_2) \times \exp(i\kappa_{pp'} z_1)\exp(ik_{0zi} z), \tag{5.2}$$

where

$$\kappa_{ss'} \equiv s k_{1zi} + s' k_{1z}, \qquad \kappa_{pp'} \equiv p k_{1zs} + p' k_{1z}, \tag{5.3}$$

which are different from (4.7). In (5.2) $\phi \equiv \phi' - \phi_i$ and $\phi_s = \phi_i$, where ϕ', ϕ_i, ϕ_s are the azimuthal angles of $\mathbf{k}_\rho, \mathbf{k}_{\rho i}, \mathbf{k}_{\rho s}$. The $D_{hh}^{\gtrless}(\phi)_{ss'pp'}$ are listed in Appendix E. Substituting (5.2) into (5.1), we find

$$\gamma_{hh}^{(2)}(\theta_{0s},\theta_{0i}) = (2k_0)^{10}\pi^7 W^2 \cos^2\theta_{0i} \int_0^{\pi/2} k_0^2 \sin\cos\theta \int_0^{2\pi} d\phi = \sum_{s,s'}\sum_{p,p'} [|D_{hh}^{>}(\phi)|_{ss'pp'}^2 M_{ss'pp'}^s + |D_{hh}^{<}(\phi)|_{ss'pp'}^2 N_{ss'pp'}^s], \tag{5.4}$$

where the definitions of $M_{ss'pp'}^s$ and $N_{ss'pp'}^s$ are the same as $M_{ss'pp'}$ and $N_{ss'pp'}$ of (4.10) and (4.11), but $\kappa''_{ss'}$ and $\kappa''_{pp'}$ follow the definitions of (5.3).

After careful algebra, we obtain

$$\gamma_{hh}^{(2)}(\theta_{0s},\theta_{0i}) = \frac{1}{16} W^2 k_0^{10} \cos^2\theta_{0i} \frac{|X_{10s}X_{01i}|^2}{|D_{2i}D_{2s}|^2} \times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \frac{1}{|k_{1zs}k_{1z}|^2} \times \left[\frac{1}{(k''_{1zi} + k''_{1z})(k''_{1zs} + k''_{1z})} N_{1h}^s + \frac{1}{(k''_{1zi} - k''_{1z})(k''_{1zs} + k''_{1z})} N_{2h}^{s(1)} + \frac{1}{(k''_{1zi} + k''_{1z})(k''_{1zs} - k''_{1z})} N_{2h}^{s(2)} + \frac{1}{(k''_{1zi} - k''_{1z})(k''_{1zs} - k''_{1z})} N_{3h}^s\right], \tag{5.5}$$

where

$$N^{s}_{1h} = C^{s}_{11}\cdot\left[\,|R_{12i}R_{12s}|^2 e^{-2k''_{1zi}d}e^{-2k''_{1zs}d}\cdot H_1 + H_2\right] - C^{s}_{12}\cdot 2|R_{12i}|^2 e^{-4k''_{1zi}d}\cdot H_3 + C^{s}_{13}\cdot 2|R_{12s}|^2 e^{-4k''_{1zs}d}\cdot H_4, \tag{5.6a}$$

$$N^{s(1)}_{2h} = C^{s(1)}_{21}\cdot|R_{12i}R_{12s}|^2 e^{-2(k''_{1zi}+k''_{1zs})d}\cdot H_4 + C^{s(1)}_{22}\cdot H_3 - C^{s(1)}_{23}\cdot\left[\,|R_{12s}|^2 e^{-2(k''_{1zi}+k''_{1zs})d}\cdot H_1 + |R_{12i}|^2 e^{-4k''_{1zi}d}\cdot H_2\right], \tag{5.6b}$$

$$N^{s(2)}_{2h} = C^{s(2)}_{21}\cdot H_4 + C^{s(2)}_{22}\cdot|R_{12i}R_{12s}|^2 e^{-2(k''_{1zi}+k''_{1zs})d}\cdot H_3 - C^{s(2)}_{23}\cdot\left[\,|R_{12i}|^2 e^{-2(k''_{1zi}+k''_{1zs})d}\cdot H_1 + |R_{12s}|^2 e^{-4k''_{1zs}d}\cdot H_2\right], \tag{5.6c}$$

$$N^{s}_{3h} = C^{s}_{31}\cdot\left[\,|R_{12i}R_{12s}|^2 e^{-2k''_{1zi}d}e^{-2k''_{1zs}d}\cdot H_2 + H_1\right] - C^{s}_{32}\cdot 2|R_{12i}|^2 e^{-4k''_{1zi}d}\cdot H_4 + C^{s}_{33}\cdot 2|R_{12s}|^2 e^{-4k''_{1zs}d}\cdot H_3, \tag{5.6d}$$

where C^{s}_{mn}, $m,n = 1,2,3$, are given in Appendix C. The functions H_n, $n = 1,2,3,4$, are the same as given in Appendix D.

As $\theta_{0s}\to\theta_{0i}$, it can be seen that $C^{s}_{mn}\to C_{mn}$, and $C^{s(1)}_{2n} + C^{s(2)}_{2n}\to C_{2n}$, $m = 1,3$, $n = 1,2,3$, then we have $\gamma^{(2)}_{hh}\to\sigma^{(2)+}_{hh}$. In the backscattering direction there is an additional contribution from the cross term. It explains the backscattering enhancement phenomena observed in experiment.[4,5] In the distorted Born approximation, the backscattering cross section is

$$\sigma_{hh} = \sigma^{(1)}_{hh} + 2\sigma^{(2)}_{hh}. \tag{5.7a}$$

In the RT or MRT approaches, it is

$$\sigma^{+}_{hh} = \sigma^{(1)}_{hh} + \sigma^{(2)+}_{hh}. \tag{5.7b}$$

Thus, the enhancement in dB can be expressed as

$$\begin{aligned} &10\times(\log\sigma_{hh} - \log\sigma^{+}_{hh}) \\ &\quad = 10\times\log\left(1 + \sigma^{(2)+}_{hh}/(\sigma^{(1)}_{hh} + \sigma^{(2)+}_{hh})\right) \\ &\quad \approx 4.3\left[\sigma^{(2)+}_{hh}/(\sigma^{(1)}_{hh} + \sigma^{(2)+}_{hh})\right]\mathrm{dB}. \end{aligned} \tag{5.8}$$

If $\sigma^{(2)+}_{hh}/\sigma^{(1)}_{hh}\approx 0.1$ or 0.01, the enhancement would be 0.43 dB, or 0.043 dB, which seem to be consistent with observations in the experiments.[4] Actually, the enhancement in the backscattering direction characterizes the significance of high-order scattering, and cross terms should be taken into account in the backscattering direction. Since we neglected the second term in (4.4), there will be a sharp peak with no angular width in the backscattering direction. As the bistatic direction is so close to the backscattering direction that $k'_{1zi} + k'_{1zs}\approx k''_{1z}$, there should be a gradually increasing enhancement until the maximum in the backscattering direction. The angular width is proportional to k''_1/k'_1.

To see this further, we carry out the second term of (4.4) to calculate $\gamma^{(2)-}_{hh}$, which is due to the coherence of the fields with the opposite ray paths, and gradually up to the maximum in the backscattering direction. From (4.4) we have

$$\begin{aligned} &\gamma^{(2)-}_{hh}(\theta_{0s},\theta_{0i}) \\ &\quad = (2k_0)^{10}\pi^7 W^2\cos^2\theta_{0i}\int d\mathbf{k}_\rho\int_{-d}^{0}dz_1\int_{-d}^{0}dz_2 \\ &\quad\times\left[\mathbf{g}^{>}_{01}(-\mathbf{k}_{\rho s},z,z_1)\cdot\mathbf{g}^{\gtrless}_{11}(\mathbf{k}_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)\right] \\ &\quad\times\left[\mathbf{g}^{>}_{01}(-\mathbf{k}_{\rho s},z,z_2)\cdot\mathbf{g}^{\gtrless}_{11}(\mathbf{k}_{\rho i}-\mathbf{k}_\rho+\mathbf{k}_{\rho s},z_2,z_1) \cdot\mathbf{F}_{1m}(z_1)\right]^{*}_{hh}. \end{aligned} \tag{5.9}$$

As before, $\phi\equiv\phi' - \phi_i$ and $\phi_i = \phi_s$, where ϕ',ϕ_i,ϕ_s are the azimuthal angles of $\mathbf{k}_\rho,\mathbf{k}_{\rho i},\mathbf{k}_{\rho s}$. Since in the vicinity of the backscattering direction, $\mathbf{k}_{\rho i} + \mathbf{k}_{\rho s}\approx 0$, the azimuthal angle ϕ'_1 of $\mathbf{k}_{\rho i} - \mathbf{k}_\rho + \mathbf{k}_{\rho s}$ is assumed to equal $\phi' + \pi$. Expanding the multiplications in (5.9), we obtain

$$\begin{aligned} &\left[\mathbf{g}^{>}_{01}(-\mathbf{k}_{\rho s},z,z_1)\cdot\mathbf{g}^{\gtrless}_{11}(\mathbf{k}_\rho,z_1,z_2)\cdot\mathbf{F}_{1m}(z_2)\right]_{hh} \\ &\quad = \sum_{s,s'}\sum_{p,p'}\left[D^{\gtrless}_{hh}(\phi)\right]_{ss'pp'}\exp(i(sk_{1zi} + s'k_{1z})z_2) \\ &\quad\times\exp(i(pk_{1zs} + p'k_{1z})z_1)\exp(ik_{0zi}z), \end{aligned} \tag{5.10a}$$

$$\begin{aligned} &\left[\mathbf{g}^{>}_{01}(-\mathbf{k}_{\rho s},z,z_1)\cdot\mathbf{g}^{\gtrless}_{11}(\mathbf{k}_{\rho i}-\mathbf{k}_\rho+\mathbf{k}_{\rho s},z_2,z_1)\cdot\mathbf{F}_{1m}(z_1)\right]_{hh} \\ &\quad = \sum_{s,s'}\sum_{p,p'}\left[D^{\gtrless}_{hh}(\phi+\pi)\right]_{ss'pp'} \\ &\quad\times\exp(i[sk_{1zi} + s'(k_{1z} + k_{1zi} + k_{1zs})]z_2) \\ &\quad\times\exp(i[pk_{1zs} + p'(k_{1z} + k_{1zi} + k_{1zs})]z_1) \\ &\quad\times\exp(ik_{0zi}z). \end{aligned} \tag{5.10b}$$

Let $\Delta\equiv k_{1zi} + k_{1zi}$, and Δ',Δ'' denote the real and imaginary parts of Δ. Close to the backscattering direction we may assume $\Delta'' = k''_{1zi} + k''_{1zs}\approx 0$. Substituting (5.10a) and (5.10b) into (5.9), we obtain

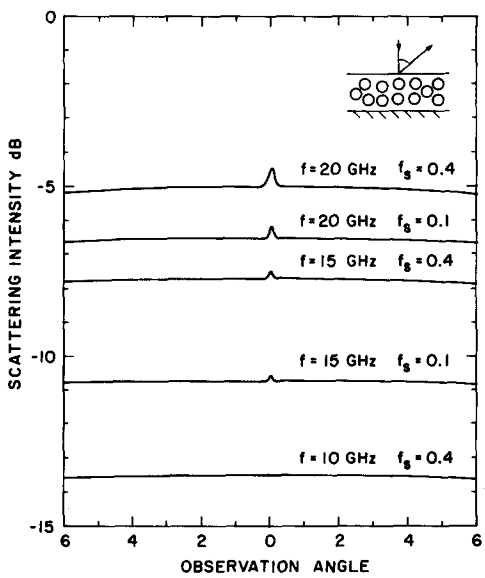


FIG. 4. Scattering intensity versus the observation angle. Scatterers radius = 0.7 mm, layer depth = 100 cm, $\epsilon_s = (3 + i0.001)\epsilon_0$, $\epsilon_b = 1.5\epsilon_0$, and $\epsilon_2 = 4\epsilon_0$.

$$
\begin{aligned}
\gamma_{hh}^{(2)-}(\theta_{0s},\theta_{0i}) &= (2k_0)^{10}\pi^7 W^2 \cos^2\theta_{0i} \\
&\times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \int_0^{2\pi} d\phi \int_{-d}^{0} dz_1 \\
&\times \sum_{s,s'}\sum_{p,p'} \Big[\int_{-d}^{z_1} dz_2\, D^{>}_{hh}(\phi)_{ss'pp'} D^{<*}_{hh}(\phi+\pi)_{pp'ss'} \\
&\times \exp(-2[sk''_{1zi} + s'(k''_{1z} + i\Delta')]z_2) \\
&\times \exp(-2[pk''_{1zs} + p'(k''_{1z} + i\Delta')]z_1) \\
&+ \int_{z_1}^{0} dz_2\, D^{<}_{hh}(\phi)_{ss'pp'} D^{>*}_{hh}(\phi+\pi)_{pp'ss'} \\
&\times \exp(-2[sk''_{1zi} + s'(k''_{1z} + i\Delta')]z_2) \\
&\times \exp(-2[pk''_{1zs} + p'(k''_{1z} + i\Delta')]z_1) \Big]. \qquad (5.11)
\end{aligned}
$$

From Appendix E, it can be seen that

$$
\begin{aligned}
D^{<*}_{hh}(\phi+\pi)_{pp'ss'} &= D^{>*}_{hh}(\phi)_{ss'pp'}, \\
D^{>*}_{hh}(\phi+\pi)_{pp'ss'} &= D^{<*}_{hh}(\phi)_{ss'pp'}.
\end{aligned}
\qquad (5.12)
$$

Thus, we have

$$
\begin{aligned}
\gamma_{hh}^{(2)-}(\theta_{0s},\theta_{0i}) &= (2k_0)^{10}\pi^7 W^2 \cos^2\theta_{0i} \\
&\times \int_0^{\pi/2} d\theta\, k_0^2 \sin\theta\cos\theta \int_0^{2\pi} d\phi \int_{-d}^{0} dz_1 \\
&\times \sum_{s,s'}\sum_{p,p'} \Big[\int_{-d}^{z_1} dz_2 |D^{>}_{hh}(\phi)_{ss'pp'}|^2 \\
&\times \exp(-2[sk''_{1zi} + s'(k''_{1z} + i\Delta')]z_2) \\
&\times \exp(-2[pk''_{1zs} + p'(k''_{1z} + i\Delta')]z_1) \\
&+ \int_{z_1}^{0} dz_2 |D^{<}_{hh}(\phi)_{ss'pp'}|^2 \\
&\times \exp(-2[sk''_{1zs} + s'(k''_{1z} + i\Delta')]z_2) \\
&\times \exp(-2[pk''_{1zi} + p'(k''_{1z} + i\Delta')]z_1) \Big]. \qquad (5.13)
\end{aligned}
$$

It is easily seen that (5.13) follows from (5.4) by replacing k''_{1z} in (5.5) by $k''_{1z} + i\Delta'$, and taking the real part. As $\Delta' = k'_{1zi} + k'_{1zs} \gg k''_{1z}$ (away from the backscattering direction), $\gamma_{hh}^{(2)-}$ will be negligible as the denominators contain $i(k'_{1zi} + k'_{1zs}) + k''_{1z}$. As $\theta_{0s} \to \theta_{0i}$ (close to the backscattering direction), Δ' is approaching zero, and $\gamma_{hh}^{(2)-}$ increases and reaches the maximum value $\sigma_{hh}^{(2)-}$, when $\Delta' = 0$. The above discussions describe the constructive interference due to the cross term near the backscattering direction.

Numerical results are given in Fig. 4 at frequencies 10, 15, and 20 GHz. It is seen that as frequency or fractional volume increase, the enhancement in the backscattering direction will be increased due to high-order scattering effects. An enlarged Fig. 5 shows the angular widths near the backscattering direction.

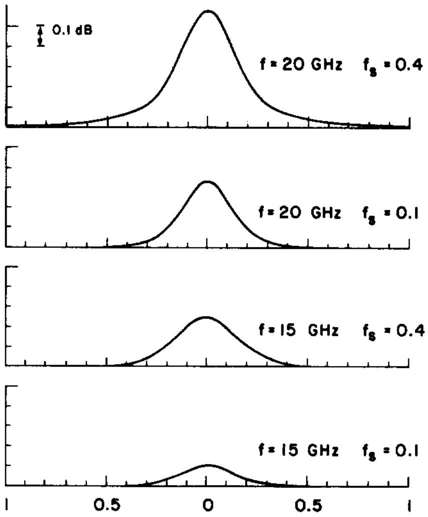


FIG. 5. Increasing enhancements near the backscattering direction. All parameters are the same as Fig. 4.

VI. CONCLUSIONS

The backscattering cross section per unit area in the second-order distorted Born approximation for copolarization and cross polarization, and the bistatic scattering coefficients are obtained. Contributions from both the coincidental and opposite ray paths in the Feynman diagram, where they are expressed by the ladder and cross terms, are calculated. The enhancements in the backscattering direction are explained with the phenomenon of the constructive interferences of waves with the quite opposite ray paths corresponding to the contributions from the cross terms. Meanwhile it also explains discrepancies between the wave approaches and the RT and MRT theory in the second-order solutions in the backscattering direction. For the copolarization both ladder and cross terms have exactly the same constributions, for the cross polarization they are of the same order.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Grant No. ECS-8203390, NASA Contract No. NAG5-270, and Office of Naval Research Contract No. N00014-83-K-0528.

APPENDIX A: FOURIER TRANSFORM OF THE MEAN DYADIC GREEN'S FUNCTIONS

We have

$$
\begin{aligned}
g^{>}_{01}(\mathbf{k}_\rho,z,z_1) &= \frac{i}{8\pi^2}\frac{1}{k_{1z}} \Big\{ \frac{X_{10}(k_\rho)}{D_2(k_\rho)} \hat{e}(k_{0z}) \big[R_{12}(k_\rho) e^{i2k_{1z}d} e^{ik_{1z}z_1} \hat{e}_1(-k_{1z}) + e^{-ik_{1z}z_1} \hat{e}_1(k_{1z}) \big] \\
&+ \frac{k_1}{k_0}\frac{Y_{10}(k_\rho)}{F_2(k_\rho)} \hat{h}(k_{0z}) \big[S_{12}(k_\rho) e^{i2k_{1z}d} e^{ik_{1z}z_1} \hat{h}_1(-k_{1z}) + e^{-ik_{1z}z_1} \hat{h}_1(k_{1z}) \big] \Big\} e^{ik_{0z}z} \\
&= (1)_h + (2)_h + (3)_v + (4)_v,
\end{aligned}
$$

$$\mathsf{g}_{\bar{1}1}^{>}(\mathbf{k}_\rho,z_1,z_2)=\frac{i}{8\pi^2}\frac{1}{k_{1z}}\Bigg\{\frac{1}{D_2(k_\rho)}\,[e^{ik_{1z}(z_1-z_2)}\hat{e}_1(k_{1z})\hat{e}_1(k_{1z})$$

$$+R_{10}(k_\rho)e^{-ik_{1z}(z_1+z_2)}\hat{e}(-k_{1z})\hat{e}_1(k_{1z})+R_{12}(k_\rho)e^{i2k_{1z}d}e^{ik_{1z}(z_1+z_2)}\hat{e}_1(k_{1z})\hat{e}_1(-k_{1z})$$

$$+R_{10}(k_\rho)R_{12}(k_\rho)e^{i2k_{1z}d}e^{-ik_{1z}(z_1-z_2)}\hat{e}_1(-k_{1z})\hat{e}_1(-k_{1z})]$$

$$+\frac{1}{F_2(k_\rho)}\,[e^{ik_{1z}(z_1-z_2)}\hat{h}_1(k_{1z})\hat{h}(k_{1z})+S_{10}(k_\rho)e^{-ik_{1z}(z_1+z_2)}\hat{h}_1(-k_{1z})\hat{h}_1(k_{1z})$$

$$+S_{12}(k_\rho)e^{i2k_{1z}d}e^{ik_{1z}(z_1+z_2)}\hat{h}_1(k_{1z})\hat{h}_1(-k_{1z})+S_{10}(k_\rho)S_{12}(k_\rho)e^{i2k_{1z}d}e^{-ik_{1z}(z_1-z_2)}\hat{h}_1(-k_{1z})\hat{h}_1(-k_{1z})]\Bigg\}$$

$$=[(5)_h+(6)_h+(7)_h+(8)_h+(9)_v+(10)_v+(11)_v+(12)_v]^{>},$$

$$\mathsf{g}_{\bar{1}1}^{<}(\mathbf{k}_\rho,z_1,z_2)=\frac{i}{8\pi^2}\frac{1}{k_{1z}}\Bigg\{\frac{1}{D_2(k_\rho)}\,[e^{-ik_{1z}(z_1-z_2)}\hat{e}_1(-k_{1z})\hat{e}_1(-k_{1z})$$

$$+R_{10}(k_\rho)e^{-ik_{1z}(z_1+z_2)}\hat{e}_1(-k_{1z})\hat{e}_1(k_{1z})+R_{12}(k_\rho)e^{i2k_{1z}d}e^{ik_{1z}(z_1+z_2)}\hat{e}_1(k_{1z})\hat{e}_1(-k_{1z})$$

$$+R_{10}(k_\rho)R_{12}(k_\rho)e^{i2k_{1z}d}e^{ik_{1z}(z_1-z_2)}\hat{e}_1(k_{1z})\hat{e}_1(k_{1z})]$$

$$+\frac{1}{F_2(k_\rho)}\,[e^{-ik_{1z}(z_1-z_2)}\hat{h}_1(-k_{1z})\hat{h}_1(-k_{1z})+S_{10}(k_\rho)e^{-ik_{1z}(z_1+z_2)}\hat{h}_1(-k_{1z})\hat{h}_1(k_{1z})$$

$$+S_{12}(k_\rho)e^{i2k_{1z}d}e^{ik_{1z}(z_1+z_2)}\hat{h}_1(k_{1z})\hat{h}_1(-k_{1z})+S_{10}(k_\rho)S_{12}(k_\rho)e^{i2k_{1z}d}e^{ik_{1z}(z_1-z_2)}\hat{h}_1(k_{1z})\hat{h}_1(k_{1z})]\Bigg\}$$

$$=[(5)_h+(6)_h+(7)_h+(8)_h+(9)_v+(10)_v+(11)_v+(12)_v]^{<}.$$

For the horizontal incident wave,

$$\mathbf{F}_{1m}(z_2)=E_0\frac{X_{01i}}{D_{2i}}\,[R_{12i}e^{i2k_{1z}d}e^{ik_{1z}z_2}+e^{-ik_{1z}z_2}]\,\hat{e}_{1i}=(13)_h+(14)_h.$$

For the vertical incident wave,

$$\mathbf{F}_{1m}(z_2)=E_0\frac{k_0Y_{01i}}{k_1F_{2i}}\,[S_{12}e^{i2k_{1z}d}e^{ik_{1z}z_2}\hat{h}_1(k_{1zi})+e^{-ik_{1z}z_2}\hat{h}_1(-k_{1zi})]=(15)_v+(16)_v.$$

APPENDIX B: FUNCTIONS $A^{\gtrless}_{\alpha\beta}(\phi)_{ss'pp'}$ FOR CALCULATIONS OF $\sigma^{(2)}_{\alpha\beta}$

For $z_1>z_2$,

$$A^{>}_{hh}(\phi)_{1,1,1,1}=(1)\cdot[(7)+(11)]^{>}\cdot(13)=\Gamma_{hh}R^2_{12i}e^{i4k_{1z}d}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi-\frac{k^2_{1z}S_{12}}{k^2_1F_2}\sin^2\phi\right],\tag{B1}$$

$$A^{>}_{hh}(\phi)_{1,1,-1,1}=(2)\cdot[(7)+(11)]^{>}\cdot(13)=\Gamma_{hh}R_{12i}e^{i2k_{1z}d}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi-\frac{k^2_{1z}S_{12}}{k^2_1F_2}\sin^2\phi\right],\tag{B2}$$

$$A^{>}_{hh}(\phi)_{1,1,1,-1}=(1)\cdot[(8)+(12)]^{>}\cdot(13)=\Gamma_{hh}R^2_{12i}e^{i4k_{1z}d}e^{i2k_{1z}d}\left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi+\frac{k^2_{1z}S_{10}S_{12}}{k^2_1F_2}\sin^2\phi\right],\tag{B3}$$

$$A^{>}_{hh}(\phi)_{1,-1,1,1}=(1)\cdot[(5)+(9)]^{>}\cdot(13)=\Gamma_{hh}R^2_{12i}e^{i4k_{1z}d}\left[\frac{1}{D_2}\cos^2\phi+\frac{k^2_{1z}}{k^2_1F_2}\sin^2\phi\right],\tag{B4}$$

$$A^{>}_{hh}(\phi)_{-1,1,1,1}=(1)\cdot[(7)+(11)]^{>}\cdot(14)=\Gamma_{hh}R_{12i}e^{i2k_{1z}d}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi-\frac{k^2_{1z}S_{12}}{k^2_1F_2}\sin^2\phi\right],\tag{B5}$$

$$A^{>}_{hh}(\phi)_{1,1,-1,-1}=(2)\cdot[(8)+(12)]^{>}\cdot(13)=\Gamma_{hh}R_{12i}e^{i2k_{1z}d}e^{i2k_{1z}d}\left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi+\frac{k^2_{1z}S_{10}S_{12}}{k^2_1F_2}\sin^2\phi\right],\tag{B6}$$

$$A^{>}_{hh}(\phi)_{-1,-1,1,1}=(1)\cdot[(5)+(9)]^{>}\cdot(14)=\Gamma_{hh}R_{12i}e^{i2k_{1z}d}\left[\frac{1}{D_2}\cos^2\phi+\frac{k^2_{1z}}{k^2_1F_2}\sin^2\phi\right],\tag{B7}$$

$$A^{>}_{hh}(\phi)_{-1,1,-1,1}=(2)\cdot[(7)+(11)]^{>}\cdot(14)=\Gamma_{hh}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi-\frac{k^2_{1z}S_{12}}{k^2_1F_2}\sin^2\phi\right],\tag{B8}$$

$$A^{>}_{hh}(\phi)_{-1,1,1,-1} = (1)\cdot[(8)+(12)]^{>}\cdot(14) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi \right], \tag{B9}$$

$$A^{>}_{hh}(\phi)_{1,-1,-1,1} = (2)\cdot[(5)+(9)]^{>}\cdot(13) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi \right], \tag{B10}$$

$$A^{>}_{hh}(\phi)_{1,-1,1,-1} = (1)\cdot[(6)+(10)]^{>}\cdot(13) = \Gamma_{hh} R_{12i}^2 e^{i4k_{1z}d} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi \right], \tag{B11}$$

$$A^{>}_{hh}(\phi)_{-1,-1,-1,-1} = (2)\cdot[(6)+(10)]^{>}\cdot(14) = \Gamma_{hh} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi \right], \tag{B12}$$

$$A^{>}_{hh}(\phi)_{1,-1,-1,-1} = (2)\cdot[(6)+(10)]^{>}\cdot(13) = \Gamma_{hh}\, R_{12i} e^{i2k_{1z}d} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi \right], \tag{B13}$$

$$A^{>}_{hh}(\phi)_{-1,-1,-1,1} = (2)\cdot[(5)+(9)]^{>}\cdot(14) = \Gamma_{hh} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi \right], \tag{B14}$$

$$A^{>}_{hh}(\phi)_{-1,1,-1,1} = (2)\cdot[(8)+(12)]^{>}\cdot(14) = \Gamma_{hh} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi \right], \tag{B15}$$

$$A^{>}_{hh}(\phi)_{-1,-1,1,-1} = (1)\cdot[(6)+(10)]^{>}\cdot(14) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi \right]. \tag{B16}$$

For $z_1 < z_2$,

$$A^{<}_{hh}(\phi)_{1,1,1,1} = \mathrm{A}^{>}_{hh}(\phi)_{1,1,1,1}, \tag{B1'}$$

$$A^{<}_{hh}(\phi)_{1,1,-1,1} = A^{>}_{hh}(\phi)_{1,1,-1,1}, \tag{B2'}$$

$$A^{<}_{hh}(\phi)_{1,1,1,-1} = (1)\cdot[(5)+(9)]^{<}\cdot(13) = \Gamma_{hh} R_{12i}^2 e^{i4k_{1z}d} \left[\frac{1}{D^2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B4}), \tag{B3'}$$

$$A^{<}_{hh}(\phi)_{1,-1,1,1} = (1)\cdot[(8)+(12)]^{<}\cdot(13) = \Gamma_{hh} R_{12i}^2 e^{i4k_{1z}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B3}), \tag{B4'}$$

$$A^{<}_{hh}(\phi)_{-1,1,1,1} = A^{>}_{hh}(\phi)_{-1,1,1,1}, \tag{B5'}$$

$$A^{<}_{hh}(\phi)_{1,1,-1,-1} = (2)\cdot[(5)+(9)]^{<}\cdot(13) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1 F_2}\sin^2\phi \right] = (\mathrm{B7}), \tag{B6'}$$

$$A^{<}_{hh}(\phi)_{-1,-1,1,1} = (1)\cdot[(8)+(12)]^{<}\cdot(14) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B6}), \tag{B7'}$$

$$A^{>}_{hh}(\phi)_{-1,1,-1,1} = A^{<}_{hh}(\phi)_{-1,1,-1,1}, \tag{B8'}$$

$$A^{<}_{hh}(\phi)_{-1,1,1,-1} = (1)\cdot[(5)+(9)]^{<}\cdot(14) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B10}), \tag{B9'}$$

$$A^{<}_{hh}(\phi)_{1,-1,-1,1} = (2)\cdot[(8)+(12)]^{<}\cdot(13) = \Gamma_{hh} R_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B9}), \tag{B10'}$$

$$A^{<}_{hh}(\phi)_{1,-1,1,-1} = A^{>}_{hh}(\phi)_{1,-1,1,-1}, \tag{B11'}$$

$$A^{<}_{hh}(\phi)_{-1,-1,-1,-1} = A^{>}_{hh}(\phi)_{-1,-1,-1,-1}, \tag{B12'}$$

$$A^{<}_{hh}(\phi)_{1,-1,-1,-1} = A^{>}_{hh}(\phi)_{1,-1,-1,-1}, \tag{B13'}$$

$$A^{<}_{hh}(\phi)_{-1,-1,-1,1} = (2)\cdot[(8)+(12)]^{<}\cdot(14) = \Gamma_{hh} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B15}), \tag{B14'}$$

$$A^{<}_{hh}(\phi)_{-1,1,-1,-1} = \Gamma_{hh} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi \right] = (\mathrm{B14}), \tag{B15'}$$

$$A^{<}_{hh}(\phi)_{-1,-1,1,-1} = A^{>}_{hh}(\phi)_{-1,-1,1,-1}, \tag{B16'}$$

where

$$\Gamma_{hh} \equiv \frac{1}{(8\pi^2)^2} \frac{X_{10i}}{D_{2i}} \frac{X_{01i}}{D_{2i}} \frac{1}{k_{1zi}k_{1z}}.$$

For $z_1 > z_2$,

$$A^{>}_{vv}(\phi)_{1,1,1,1} = (3)\cdot[(7)+(11)]^{>}\cdot(15) = \Gamma_{vv} S^{2}_{12i} e^{i4k_{1z}d} e^{i2k_{1z}d}\left[-\frac{R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{12}}{k^{4}_{1}F_2}\alpha^2_{-}(\phi)\right], \tag{B17}$$

$$A^{>}_{vv}(\phi)_{1,1,-1,1} = (4)\cdot[(7)+(11)]^{>}\cdot(15) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d}\left[\frac{R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{S_{12}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B18}$$

$$A^{>}_{vv}(\phi)_{1,1,1,-1} = (3)\cdot[(8)+(12)]^{>}\cdot(15) = \Gamma_{vv} S^{2}_{12i} e^{i4k_{1z}d} e^{i2k_{1z}d}\left[-\frac{R_{10}R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{S_{10}S_{12}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B19}$$

$$A^{>}_{vv}(\phi)_{1,-1,1,1} = (3)\cdot[(5)+(9)]^{>}\cdot(15) = \Gamma_{vv} S^{2}_{12i} e^{i4k_{1z}d}\left[-\frac{k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{1}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B20}$$

$$A^{>}_{vv}(\phi)_{-1,1,1,1} = (3)\cdot[(7)+(11)]^{>}\cdot(16) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d}\left[\frac{R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{S_{12}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B21}$$

$$A^{>}_{vv}(\phi)_{1,1,-1,-1} = (4)\cdot[(8)+(12)]^{>}\cdot(15) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d}\left[\frac{R_{10}R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{10}S_{12}}{k^{4}_{1}F_2}\alpha^2_{-}(\phi)\right], \tag{B22}$$

$$A^{>}_{vv}(\phi)_{-1,-1,1,1} = (3)\cdot[(5)+(9)]^{>}\cdot(16) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d}\left[\frac{k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{1}{F_2k^{4}_{1}}\alpha^2_{-}(\phi)\right], \tag{B23}$$

$$A^{>}_{vv}(\phi)_{-1,1,-1,1} = (4)\cdot[(7)+(11)]^{>}\cdot(16) = \Gamma_{vv} e^{i2k_{1z}d}\left[-\frac{R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{12}}{k^{4}_{1}F_2}\alpha^2_{+}(\phi)\right], \tag{B24}$$

$$A^{>}_{vv}(\phi)_{-1,1,1,-1} = (3)\cdot[(8)+(12)]^{>}\cdot(16) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d}\left[\frac{R_{10}R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{10}S_{12}}{k^{4}_{1}F_2}\alpha^2_{+}(\phi)\right], \tag{B25}$$

$$A^{>}_{vv}(\phi)_{1,-1,-1,1} = (4)\cdot[(5)+(9)]^{>}\cdot(15) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d}\left[\frac{k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{1}{k^{4}_{1}F_2}\alpha^2_{+}(\phi)\right], \tag{B26}$$

$$A^{>}_{vv}(\phi)_{1,-1,1,-1} = (3)\cdot[(6)+(10)]^{>}\cdot(15) = \Gamma_{vv} S^{2}_{12i} e^{i4k_{1z}d}\left[-\frac{R_{10}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{10}}{k^{4}_{1}F_2}\alpha^2_{+}(\phi)\right], \tag{B27}$$

$$A^{>}_{vv}(\phi)_{-1,-1,-1,-1} = (4)\cdot[(6)+(10)]^{>}\cdot(16) = \Gamma_{vv}\left[-\frac{R_{10}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{10}}{k^{4}_{1}F_2}\alpha^2_{-}(\phi)\right], \tag{B28}$$

$$A^{>}_{vv}(\phi)_{1,-1,-1,-1} = (4)\cdot[(6)+(10)]^{>}\cdot(15) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d}\left[\frac{R_{10}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{S_{10}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B29}$$

$$A^{>}_{vv}(\phi)_{-1,-1,-1,1} = (4)\cdot[(5)+(9)]^{>}\cdot(16) = \Gamma_{vv}\left[-\frac{k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{1}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B30}$$

$$A^{>}_{vv}(\phi)_{-1,1,-1,-1} = (4)\cdot[(8)+(12)]^{>}\cdot(16) = \Gamma_{vv} e^{i2k_{1z}d}\left[-\frac{R_{10}R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{S_{10}S_{12}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B31}$$

$$A^{>}_{vv}(\phi)_{-1,-1,1,-1} = (4)\cdot[(5)+(9)]^{>}\cdot(16) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d}\left[-\frac{R_{10}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{S_{10}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right], \tag{B32}$$

For $z_1 < z_2$,

$$A^{<}_{vv}(\phi)_{1,1,1,1} = A^{>}_{vv}(\phi)_{1,1,1,1}, \tag{B17$'$}$$

$$A^{<}_{vv}(\phi)_{1,1,-1,1} = A^{>}_{vv}(\phi)_{1,1,-1,1}, \tag{B18$'$}$$

$$A^{<}_{vv}(\phi)_{1,1,1,-1} = (3)\cdot[(5)+(9)]^{<}\cdot(15) = \Gamma_{vv} S^{2}_{12i} e^{i4k_{1z}d}\left[-\frac{k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{1}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right] = \text{(B20)}, \tag{B19$'$}$$

$$A^{>}_{vv}(\phi)_{1,-1,1,1} = (3)\cdot[(8)+(12)]^{<}\cdot(15) = \Gamma_{vv} S^{2}_{12i} e^{i4k_{1z}d} e^{i2k_{1z}d}\left[-\frac{R_{10}R_{12}k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi - \frac{S_{10}S_{12}}{k^{4}_{1}F_2}\alpha_{+}(\phi)\alpha_{-}(\phi)\right] = \text{(B19)}, \tag{B20$'$}$$

$$A^{<}_{vv}(\phi)_{-1,1,1,1} = A^{>}_{vv}(\phi)_{-1,1,1,1}, \tag{B21$'$}$$

$$A^{<}_{vv}(\phi)_{1,1,-1,-1} = (4)\cdot[(5)+(9)]^{<}\cdot(15) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d}\left[\frac{k^{2}_{1zi}}{D_2k^{2}_{1}}\sin^2\phi + \frac{1}{k^{4}_{1}F_2}\alpha^2_{-}(\phi)\right] = \text{(B23)}, \tag{B22$'$}$$

$$A^{<}_{vv}(\phi)_{-1,-1,1,1} = (3)\cdot[(8)+(12)]^{<}\cdot(16) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}k^2_{1zi}}{D_2 k_1^2} \sin^2\phi + \frac{S_{10}S_{12}}{k_1^4 F_2} \alpha^2_-(\phi) \right] = (\text{B22}), \quad (\text{B23}')$$

$$A^{<}_{vv}(\phi)_{-1,1,-1,1} = A^{>}_{vv}(\phi)_{-1,1,-1,1}, \quad (\text{B24}')$$

$$A^{<}_{vv}(\phi)_{-1,1,1,-1} = (3)\cdot[(5)+(9)]^{<}\cdot(16) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} \left[\frac{k^2_{1zi}}{D_2 k_1^2} \sin^2\phi + \frac{1}{k_1^4 F_2} \alpha^2_+(\phi) \right] = (\text{B26}), \quad (\text{B25}')$$

$$A^{<}_{vv}(\phi)_{1,-1,-1,1} = (4)\cdot[(8)+(12)]^{<}\cdot(15) = \Gamma_{vv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}k^2_{1zi}}{D_2 k_1^2} \sin^2\phi + \frac{S_{10}S_{12}}{k_1^4 F_2} \alpha^2_+(\phi) \right] = (\text{B25}), \quad (\text{B26}')$$

$$A^{<}_{vv}(\phi)_{1,-1,1,-1} = A^{>}_{vv}(\phi)_{1,-1,1,-1}, \quad (\text{B27}')$$

$$A^{<}_{vv}(\phi)_{-1,-1,-1,-1} = A^{>}_{vv}(\phi)_{-1,-1,-1,-1}, \quad (\text{B28}')$$

$$A^{<}_{vv}(\phi)_{1,-1,-1,-1} = A^{>}_{vv}(\phi)_{1,-1,-1,-1}, \quad (\text{B29}')$$

$$A^{<}_{vv}(\phi)_{-1,-1,-1,1} = (4)\cdot[(8)+(12)]^{<}\cdot(16) = \Gamma_{vv} e^{i2k_{1z}d} \left[-\frac{R_{10}R_{12}k^2_{1zi}}{D_2 k_1^2} \sin^2\phi - \frac{S_{10}S_{12}}{k_1^4 F_2} \alpha_+(\phi)\alpha_-(\phi) \right] = (\text{B31}), \quad (\text{B30}')$$

$$A^{<}_{vv}(\phi)_{-1,1,-1,-1} = \Gamma_{vv} \left[-\frac{k^2_{1zi}}{D_2 k_1^2} \sin^2\phi - \frac{1}{k_1^4 F_2} \alpha_+(\phi)\alpha_-(\phi) \right] = (\text{B30}), \quad (\text{B31}')$$

$$A^{<}_{vv}(\phi)_{-1,-1,1,-1} = A^{>}_{vv}(\phi)_{-1,-1,1,-1}, \quad (\text{B32}')$$

where

$$\Gamma_{vv} \equiv \frac{1}{(8\pi^2)^2} \frac{k_1 Y_{10i}}{k_0 F_{2i}} \frac{k_0 Y_{01i}}{k_1 F_{2i}} \frac{1}{k_{1zi} k_{1z}},$$

$$\alpha_\pm(\phi) \equiv k_{1zi} k_{1z} \cos\phi \pm k_\rho k_{\rho i}.$$

We observe that changes from $A^{\gtrless}_{hh}(\phi)_{s,s',p,p'}$ to $A^{\gtrless}_{vv}(\phi)_{s,s',p,p'}$ correspond to the following.

(1) $\cos^2\phi$ replaced by $\pm(k^2_{1zi}/k_1^2)\sin^2\phi$ since $\hat{e}_{\pm i}\cdot\hat{e}_\pm\,\hat{e}_\pm\cdot\hat{e}_{\pm i}$ replaced by $\hat{h}_{\pm i}\cdot\hat{e}_\pm\,\hat{e}_\pm\cdot\hat{h}_{\pm i}$.

(2) $\pm(k^2_{1z}/k_1^2)\sin^2\phi$ replaced by $(1/k_1^4)\alpha^2_\pm(\phi)$ or $-(1/k_1^4)\alpha_+(\phi)\alpha_-(\phi)$ since $\hat{e}_{\pm i}\cdot\hat{h}_\pm\,\hat{h}_\pm\cdot\hat{e}_{\pm i}$ is replaced by $\hat{h}_{\pm i}\cdot\hat{h}_\pm\,\hat{h}_\pm\cdot\hat{h}_{\pm i}$.

For $z_1 > z_2$,

$$A^{>}_{hv}(\phi)_{1,1,1,1} = (1)\cdot[(7)+(11)]^{>}\cdot(15) = \Gamma_{hv} R_{12i} S_{12i} e^{i4k_{1z}d} e^{i2k_{1z}d} \left[-\frac{k_{1zi}R_{12}}{k_1 D_2} \cos\phi\sin\phi - \frac{k_{1z}}{k_1} \frac{S_{12}}{k_1^2 F_2} \alpha_-(\phi)\sin\phi \right], \quad (\text{B33})$$

$$A^{>}_{hv}(\phi)_{1,1,-1,1} = (2)\cdot[(7)+(11)]^{>}\cdot(15) = \Gamma_{hv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[-\frac{k_{1zi}R_{12}}{k_1 D_2} \cos\phi\sin\phi - \frac{k_{1z}}{k_1} \frac{S_{12}}{k_1^2 F_2} \alpha_-(\phi)\sin\phi \right], \quad (\text{B34})$$

$$A^{>}_{hv}(\phi)_{1,1,1,-1} = (1)\cdot[(8)+(12)]^{>}\cdot(15) = \Gamma_{hv} R_{12i} S_{12i} e^{i4k_{1z}d} e^{i2k_{1z}d} \left[-\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2} \cos\phi\sin\phi + \frac{k_{1z}}{k_1} \frac{S_{10}S_{12}}{k_1^2 F_2} \alpha_-(\phi)\sin\phi \right], \quad (\text{B35})$$

$$A^{>}_{hv}(\phi)_{1,-1,1,1} = (1)\cdot[(5)+(9)]^{>}\cdot(15) = \Gamma_{hv} R_{12i} S_{12i} e^{i4k_{1z}d} \left[-\frac{k_{1zi}}{k_1 D_2} \cos\phi\sin\phi + \frac{k_{1z}}{k_1} \frac{1}{k_1^2 F_2} \alpha_+(\phi)\sin\phi \right], \quad (\text{B36})$$

$$A^{>}_{hv}(\phi)_{-1,1,1,1} = (1)\cdot[(7)+(11)]^{>}\cdot(16) = \Gamma_{hv} R_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[-\frac{k_{1zi}R_{12}}{k_1 D_2} \cos\phi\sin\phi + \frac{k_{1z}}{k_1} \frac{S_{12}}{k_1^2 F_2} \alpha_+(\phi)\sin\phi \right], \quad (\text{B37})$$

$$A^{>}_{hv}(\phi)_{1,1,-1,-1} = (2)\cdot[(8)+(12)]^{>}\cdot(15) = \Gamma_{hv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[-\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2} \cos\phi\sin\phi + \frac{k_{1z}}{k_1} \frac{S_{10}S_{12}}{k_1^2 F_2} \alpha_-(\phi)\sin\phi \right], \quad (\text{B38})$$

$$A^{>}_{hv}(\phi)_{-1,-1,1,1} = (1)\cdot[(5)+(9)]^{>}\cdot(16) = \Gamma_{hv} R_{12i} e^{i2k_{1z}d} \left[\frac{k_{1zi}}{k_1 D_2} \cos\phi\sin\phi - \frac{k_{1z}}{k_1} \frac{1}{k_1^2 F_2} \alpha_-(\phi)\sin\phi \right], \quad (\text{B39})$$

$$A^{>}_{hv}(\phi)_{-1,1,-1,1} = (2)\cdot[(7)+(11)]^{>}\cdot(16) = \Gamma_{hv} e^{i2k_{1z}d} \left[\frac{k_{1zi}R_{12}}{k_1 D_2} \cos\phi\sin\phi + \frac{k_{1z}}{k_1} \frac{S_{12}}{k_1^2 F_2} \alpha_+(\phi)\sin\phi \right], \quad (\text{B40})$$

$$A^{>}_{hv}(\phi)_{-1,1,1,-1} = (1)\cdot[(8)+(12)]^{>}\cdot(16) = \Gamma_{hv} R_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d} \left[\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2} \cos\phi\sin\phi - \frac{k_{1z}}{k_1} \frac{S_{10}S_{12}}{k_1^2 F_2} \alpha_+(\phi)\sin\phi \right], \quad (\text{B41})$$

$$A^{>}_{hv}(\phi)_{1,-1,-1,1} = (2)\cdot[(5)+(9)]^{>}\cdot(15) = \Gamma_{hv} S_{12i} e^{i2k_{1z}d}\left[-\frac{k_{1zi}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{1}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B42}$$

$$A^{>}_{hv}(\phi)_{1,-1,1,-1} = (1)\cdot[(6)+(10)]^{>}\cdot(15) = \Gamma_{hv} R_{12i} S_{12i} e^{i4k_{1z}d}\left[-\frac{k_{1zi}R_{10}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{S_{10}}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B43}$$

$$A^{>}_{hv}(\phi)_{-1,-1,-1,-1} = (2)\cdot[(6)+(10)]^{>}\cdot(16) = \Gamma_{hv}\left[-\frac{k_{1zi}R_{10}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{S_{10}}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B44}$$

$$A^{>}_{hv}(\phi)_{1,-1,-1,-1} = (2)\cdot[(6)+(10)]^{>}\cdot(15) = \Gamma_{hv} S_{12i} e^{i2k_{1z}d}\left[-\frac{k_{1zi}R_{10}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{S_{10}}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B45}$$

$$A^{>}_{hv}(\phi)_{-1,-1,-1,1} = (2)\cdot[(5)+(9)]^{>}\cdot(16) = \Gamma_{hv}\left[-\frac{k_{1zi}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{1}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B46}$$

$$A^{>}_{hv}(\phi)_{-1,1,-1,-1} = (2)\cdot[(8)+(12)]^{>}\cdot(16) = \Gamma_{hv} e^{i2k_{1z}d}\left[-\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{S_{10}S_{12}}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B47}$$

$$A^{>}_{hv}(\phi)_{-1,-1,1,-1} = (1)\cdot[(6)+(10)]^{>}\cdot(16) = \Gamma_{hv} R_{12i} e^{i2k_{1z}d}\left[\frac{k_{1zi}R_{10}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{S_{10}}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B48}$$

For $z_1 < z_2$,

$$A^{<}_{hv}(\phi)_{1,1,1,1} = A^{>}_{hv}(\phi)_{1,1,1,1}, \tag{B33'}$$

$$A^{<}_{hv}(\phi)_{1,1,-1,1} = A^{>}_{hv}(\phi)_{1,1,-1,1}, \tag{B34'}$$

$$A^{<}_{hv}(\phi)_{1,1,1,-1} = (1)\cdot[(5)+(9)]^{<}\cdot(15) = \Gamma_{hv} R_{12i} S_{12i} e^{i4k_{1z}d}\left[-\frac{k_{1zi}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{1}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B35'}$$

$$A^{<}_{hv}(\phi)_{1,-1,1,1} = (1)\cdot[(8)+(12)]^{<}\cdot(15) = \Gamma_{hv} R_{12i} S_{12i} e^{i4k_{1z}d} e^{i2k_{1z}d}\left[-\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{S_{10}S_{12}}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B36'}$$

$$A^{<}_{hv}(\phi)_{-1,1,1,1} = A^{>}_{hv}(\phi)_{-1,1,1,1}, \tag{B37'}$$

$$A^{<}_{hv}(\phi)_{1,1,-1,-1} = (2)\cdot[(5)+(9)]^{<}\cdot(15) = \Gamma_{hv} S_{12i} e^{i2k_{1z}d}\left[-\frac{k_{1zi}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{1}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B38'}$$

$$A^{<}_{hv}(\phi)_{-1,-1,1,1} = (1)\cdot[(8)+(12)]^{<}\cdot(16) = \Gamma_{hv} R_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d}\left[\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{S_{10}S_{12}}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B39'}$$

$$A^{<}_{hv}(\phi)_{-1,1-1,1} = A^{>}_{hv}(\phi)_{-1,1,-1,1}, \tag{B40'}$$

$$A^{<}_{hv}(\phi)_{-1,1,1,-1} = (1)\cdot[(5)+(9)]^{<}\cdot(16) = \Gamma_{hv} R_{12i} e^{i2k_{1z}d}\left[\frac{k_{1zi}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{1}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B41'}$$

$$A^{<}_{hv}(\phi)_{1,-1,-1,1} = (2)\cdot[(8)+(12)]^{<}\cdot(15) = \Gamma_{hv} S_{12i} e^{i2k_{1z}d} e^{i2k_{1z}d}\left[-\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2}\cos\phi\sin\phi + \frac{k_{1z}}{k_1}\frac{S_{10}S_{12}}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B42'}$$

$$A^{<}_{hv}(\phi)_{1,-1,1,-1} = A^{>}_{hv}(\phi)_{1,-1,1,-1}, \tag{B43'}$$

$$A^{<}_{hv}(\phi)_{-1,-1,-1,-1} = A^{>}_{hv}(\phi)_{-1,-1,-1,-1}, \tag{B44'}$$

$$A^{<}_{hv}(\phi)_{1,-1,-1,-1} = A^{>}_{hv}(\phi)_{1,-1,-1,-1}, \tag{B45'}$$

$$A^{<}_{hv}(\phi)_{-1,-1,-1,1} = (2)\cdot[(8)+(12)]^{<}\cdot(16) = \Gamma_{hv} e^{i2k_{1z}d}\left[\frac{k_{1zi}R_{10}R_{12}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{S_{10}S_{12}}{k_1^2 F_2}\alpha_-(\phi)\sin\phi\right], \tag{B46'}$$

$$A^{<}_{hv}(\phi)_{-1,1,-1,-1} = (2)\cdot[(5)+(9)]^{<}\cdot(16) = \Gamma_{hv}\left[\frac{k_{1zi}}{k_1 D_2}\cos\phi\sin\phi - \frac{k_{1z}}{k_1}\frac{1}{k_1^2 F_2}\alpha_+(\phi)\sin\phi\right], \tag{B47'}$$

$$A^{<}_{hv}(\phi)_{-1,-1,1,-1} = A^{>}_{hv}(\phi)_{-1,-1,1,-1}, \tag{B48'}$$

where

$$\Gamma_{hv} \equiv \frac{1}{(8\pi^2)^2}\frac{X_{10i}}{D_{2i}}\frac{k_0 Y_{01i}}{k_1 F_{2i}}\frac{1}{k_{1zi}k_{1z}}, \quad \alpha_\pm(\phi) \equiv [k_{1zi}k_{1z}\cos\phi \pm k_\rho k_{\rho i}].$$

For $A^{\gtrless}_{vh}(\phi)_{s,s',p,p'}$, we have

$$\Gamma_{vh} \equiv \frac{1}{(8\pi^2)^2} \frac{k_1 Y_{10i}}{k_0 F_{2i}} \frac{X_{01i}}{D_{2i}} \frac{1}{k_{1zi} k_{1z}}$$

and the following replacements:

(1) (B33),(B34),(B47),(B48) replaced by (B35),(B36),(B45),(B46), respectively. Therefore, we have (2).

(2) $\Gamma_{hv} R_{12i} S_{12i}$ is replaced by $-\Gamma_{vh} S_{12i} R_{12i}$.

(3) $\Gamma_{hv} R_{12i}$ or $\Gamma_{hv} S_{12i}$ are replaced by $\Gamma_{vh} S_{12i}$ or $\Gamma_{vh} R_{12i}$.

(4) Γ_{hv} replaced by $-\Gamma_{vh}$. (For those terms without R_{12i} and S_{12i}).

(5) In (B33), (B34), (B35), (B36), (B45), (B46), (B47), (B48), and their primed versions, which connect with the term of $1/(k''^2_{1zi} - k''^2_{1zi})$, $\alpha_\pm(\phi)$ is replaced by $\alpha_\mp(\phi)$.

APPENDIX C: COEFFICIENTS C_{mn} AND C^s_{mn} $(m,n = 1,2,3)$ FOR CALCULATIONS OF $\sigma^{(2)}_{\alpha\beta}$ AND $\gamma^{(2)}_{\alpha\beta}$

We have

$$C_{11} = \tfrac{1}{4}\left[1 - e^{-2(k''_{1zi}+k''_{1z})d}\right]^2,$$

$$C_{12} = \tfrac{1}{4}\left[2d(k''_{1zi}+k''_{1z})e^{-4k''_{1z}d} - \left[1 - e^{-2(k''_{1zi}+k''_{1z})d}\right]e^{2(k''_{1zi}-k''_{1z})d}\right],$$

$$C_{13} = \tfrac{1}{4}\left[2d(k''_{1zi}+k''_{1z}) - \left[1 - e^{-2(k''_{1zi}+k''_{1z})d}\right]\right],$$

$$C_{21} = (1/4k''_{1zi})\left[k''_{1zi} - k''_{1z} + (k''_{1zi}+k''_{1z})e^{-4k''_{1z}d} - 2k''_{1zi}e^{-2(k''_{1zi}+k''_{1z})d}\right],$$

$$C_{22} = (1/4k''_{1zi})\left[(k''_{1zi}-k''_{1z})e^{-4(k''_{1zi}+k''_{1z})d} + (k''_{1zi}+k''_{1z})e^{-4k''_{1z}d} - 2k''_{1zi}e^{-2(k''_{1zi}-k''_{1z})d}\right],$$

$$C_{23} = \tfrac{1}{4}\left[1 - e^{-2(k''_{1zi}+k''_{1z})d}\right]\left[1 - e^{2(k''_{1zi}-k''_{1z})d}\right],$$

$$C_{31} = \tfrac{1}{4}\left[e^{-2k''_{1zi}d} - e^{-2k''_{1z}d}\right]^2,$$

$$C_{32} = \tfrac{1}{4}\left[2d(k''_{1zi}-k''_{1z}) + 1 - e^{2(k''_{1zi}-k''_{1z})d}\right],$$

$$C_{33} = \tfrac{1}{4}e^{-4k''_{1z}d}\left[2d(k''_{1zi}-k''_{1z}) - 1 + e^{-2(k''_{1zi}-k''_{1z})d}\right],$$

$$C^s_{11} = \tfrac{1}{4}\left[1 - e^{-2(k''_{1zi}+k''_{1z})d}\right]\left[1 - e^{-2(k''_{1zs}+k''_{1z})d}\right],$$

$$C^s_{12} = \frac{1}{8}\left[\frac{k''_{1zs}+k''_{1z}}{k''_{1zi}-k''_{1zs}}\left(e^{2(k''_{1zi}-k''_{1z}d)} - e^{-2(k''_{1zi}-k''_{1zs})d}\right)e^{-4k''_{1z}d} + e^{-4k''_{1z}d}\left(1 - e^{2(k''_{1zs}+k''_{1z})d}\right) + \left(e^{2k''_{1zi}d} - e^{2k''_{1zs}d}\right)\right],$$

$$C^s_{13} = \frac{1}{8}\left[\frac{k''_{1zs}+k''_{1z}}{k''_{1zi}-k''_{1zs}}\left(e^{2(k''_{1zi}-k''_{1zs})d} - e^{-2(k''_{1zi}-k''_{1zs})d}\right) + e^{-2(k''_{1zi}+k''_{1z})d} + e^{-2(k''_{1zs}+k''_{1z})d} - e^{-2(k''_{1zi}-k''_{1zs})d} - 1\right],$$

$$C^{s(1)}_{21} = \left[1/4(k''_{1zi}+k''_{1zs})\right]\left[(k''_{1zi}-k''_{1z}) + (k''_{1zs}+k''_{1z})e^{-2(k''_{1zi}+k''_{1zs})d} - (k''_{1zi}+k''_{1zs})e^{-2(k''_{1zi}+k''_{1z})d}\right],$$

$$C^{s(2)}_{21} = \left[1/4(k''_{1zi}+k''_{1zs})\right]\left[(k''_{1zs}-k''_{1z}) + (k''_{1zi}+k''_{1z})e^{-2(k''_{1zi}+k''_{1zs})d} - (k''_{1zi}+k''_{1zs})e^{-2(k''_{1zs}+k''_{1z})d}\right],$$

$$C^{s(1)}_{22} = \left[1/4(k''_{1zi}+k''_{1zs})\right]\left[(k''_{1zi}+k''_{1z}) + (k''_{1zs}-k''_{1z})e^{-2(k''_{1zi}+k''_{1zs})d} - (k''_{1zi}+k''_{1zs})e^{-2(k''_{1zs}-k''_{1z})d}\right]e^{-4k''_{1z}d},$$

$$C^{s(2)}_{22} = \left[1/4(k''_{1zi}+k''_{1zs})\right]\left[(k''_{1zs}+k''_{1z}) + (k''_{1zi}-k''_{1z})e^{-2(k''_{1zi}+k''_{1zs})d} - (k''_{1zi}+k''_{1zs})e^{-2(k''_{1zs}-k''_{1z})d}\right]e^{-4k''_{1z}d},$$

$$C^{s(1)}_{23} = \tfrac{1}{4}\left(1 - e^{2(k''_{1zi}-k''_{1z})d}\right)\left(1 - e^{-2(k''_{1zs}+k''_{1z})d}\right),$$

$$C^{s(2)}_{23} = \tfrac{1}{4}\left(1 - e^{-2(k''_{1zi}+k''_{1z})d}\right)\left(1 - e^{2(k''_{1zs}-k''_{1z})d}\right),$$

$$C^s_{31} = \tfrac{1}{4}\left[e^{-2k''_{1zi}d} - e^{-2k''_{1z}d}\right]\left[e^{-2k''_{1zs}d} - e^{-2k''_{1z}d}\right],$$

$$C^s_{32} = \frac{1}{8}\left[\frac{k''_{1zs}-k''_{1z}}{k''_{1zi}-k''_{1zs}}\left(e^{-2(k''_{1zi}-k''_{1zs})d} - e^{2(k''_{1zi}-k''_{1zs})d}\right) + e^{2(k''_{1zi}-k''_{1z})d} + e^{2(k''_{1zs}-k''_{1z})d} - 1 - e^{2(k''_{1zi}-k''_{1zs})d}\right],$$

$$C^s_{33} = \frac{1}{8}\left[\frac{k''_{1zs}-k''_{1z}}{k''_{1zi}-k''_{1zs}}\left(e^{2(k''_{1zi}-k''_{1zs})d} - e^{-2(k''_{1zi}-k''_{1zs})d}\right) + e^{-2(k''_{1zi}-k''_{1z})d} + e^{-2(k''_{1zs}-k''_{1z})d} - e^{-2(k''_{1zi}-k''_{1zs})d} - 1\right],$$

APPENDIX D: FUNCTIONS H_n, V_n, U_n $(n = 1,2,3)$ FOR CALCULATIONS OF $\gamma^{(2)}_{hh}$

We have

$$H_1 = \left|\frac{R_{12}}{D_2} - \frac{S_{12}k^2_{1z}}{F_2 k^2_1}\right|^2 + 2\left[\left|\frac{R_{12}}{D_2}\right|^2 + \left|\frac{S_{12}k^2_{1z}}{F_2 k^2_1}\right|^2\right],$$

$$H_2 = \left|\frac{R_{10}}{D_2} - \frac{S_{10}k_{1z}^2}{F_2k_1^2}\right|^2 + 2\left[\left|\frac{R_{10}}{D_2}\right|^2 + \left|\frac{S_{10}k_{1z}^2}{F_2k_1^2}\right|^2\right]$$

$$H_3 = \left|\frac{R_{10}R_{12}}{D_2} + \frac{S_{10}S_{12}k_{1z}^2}{F_2k_1^2}\right|^2 + 2\left[\left|\frac{R_{10}R_{12}}{D_2}\right|^2 + \left|\frac{S_{10}S_{12}k_{1z}^2}{F_2k_1^2}\right|^2\right],$$

$$H_4 = \left|\frac{1}{D_2} + \frac{k_{1z}^2}{F_2k_1^2}\right|^2 + 2\left[\left|\frac{1}{D_2}\right|^2 + \left|\frac{k_{1z}^2}{F_2k_1^2}\right|^2\right],$$

$$V_1 = \left|\frac{R_{12}k_{1zi}^2}{D_2k_1^2} - \frac{S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{R_{12}k_{1zi}^2}{D_2k_1^2} - \frac{2S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2$$
$$+4\left|\frac{2S_{12}k_{1z}k_{1zi}k_\rho k_{\rho i}}{F_2k_1^4}\right|^2 + 4\,\mathrm{Re}\left[\frac{S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\left(\frac{2S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right)^*\right],$$

$$V_2 = \left|\frac{R_{10}k_{1zi}^2}{D_2k_1^2} - \frac{S_{10}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{R_{10}k_{1zi}^2}{D_2k_1^2} - \frac{2S_{10}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{S_{10}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2$$
$$+4\left|\frac{2S_{10}k_{1z}k_{1zi}k_\rho k_{\rho i}}{F_2k_1^4}\right|^2 + 4\,\mathrm{Re}\left[\frac{S_{10}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\left(\frac{2S_{10}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right)^*\right],$$

$$V_3 = \left|\frac{R_{10}R_{12}k_{1zi}^2}{D_2k_1^2} + \frac{S_{10}S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{R_{10}R_{12}k_{1zi}^2}{D_2k_1^2} + \frac{2S_{10}S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{S_{10}S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2$$
$$+4\left|\frac{2S_{10}S_{12}k_{1z}k_{1zi}k_\rho k_{\rho i}}{F_2k_1^4}\right|^2 + 4\,\mathrm{Re}\left[\frac{S_{10}S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\left(\frac{2S_{10}S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right)^*\right],$$

$$V_4 = \left|\frac{k_{1zi}^2}{D_2k_1^2} + \frac{k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{k_{1zi}^2}{D_2k_1^2} + \frac{2k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2$$
$$+4\left|\frac{2k_{1z}k_{1zi}k_\rho k_{\rho i}}{F_2k_1^4}\right|^2 + 4\,\mathrm{Re}\left[\frac{k_{1z}^2k_{1zi}^2}{F_2k_1^4}\left(\frac{2k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right)^*\right],$$

$$U_1 = \left|\frac{R_{12}k_{1zi}^2}{D_2k_1^2} - \frac{S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{R_{12}k_{1zi}^2}{D_2k_1^2} + \frac{2S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4} - \frac{2S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2,$$

$$U_2 = \left|\frac{R_{10}k_{1zi}^2}{D_2k_1^2} - \frac{S_{10}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{R_{10}k_{1zi}^2}{D_2k_1^2} + \frac{2S_{10}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{S_{10}k_{1z}^2k_{1zi}^2}{F_2k_1^4} - \frac{2S_{10}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2,$$

$$U_3 = \left|\frac{R_{10}R_{12}k_{1zi}^2}{D_2k_1^2} + \frac{S_{10}S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{R_{10}R_{12}k_{1zi}^2}{D_2k_1^2} + \frac{2S_{10}S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{S_{10}S_{12}k_{1z}^2k_{1zi}^2}{F_2k_1^4} - \frac{2S_{10}S_{12}k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2,$$

$$U_4 = \left|\frac{k_{1zi}^2}{D_2k_1^2} + \frac{k_{1z}^2k_{1zi}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{k_{1zi}^2}{D_2k_1^2} + \frac{2k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2 + 2\left|\frac{k_{1z}^2k_{1zi}^2}{F_2k_1^4} - \frac{2k_\rho^2k_{\rho i}^2}{F_2k_1^4}\right|^2.$$

APPENDIX E: FUNCTIONS $D_{hh}^{\gtrless}(\phi)_{ss'pp'}$ FOR CALCULATIONS OF $\gamma_{hh}^{(2)}$

For $z_1 > z_2$,

$$D_{hh}^{>}(\phi)_{1,1,1,1} = (1)_s\cdot[(7)+(11)]^{>}\cdot(13)_i = \Gamma_{hh}^{s}R_{12i}R_{12s}e^{i2k_{1zi}d}e^{i2k_{1zs}d}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi - \frac{k_{1z}^2S_{12}}{k_1^2F_2}\sin^2\phi\right], \tag{E1}$$

$$D_{hh}^{>}(\phi)_{1,1,-1,1} = (2)_s\cdot[(7)+(11)]^{>}\cdot(13)_i = \Gamma_{hh}^{s}R_{12i}e^{i2k_{1zi}d}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi - \frac{k_{1z}^2S_{12}}{k_1^2F_2}\sin^2\phi\right], \tag{E2}$$

$$D_{hh}^{>}(\phi)_{1,1,1,-1} = (1)_s\cdot[(8)+(12)]^{>}\cdot(13)_i = \Gamma_{hh}^{s}R_{12i}R_{12s}e^{i2k_{1zi}d}e^{i2k_{1zs}d}e^{i2k_{1z}d}\left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2S_{10}S_{12}}{k_1^2F_2}\sin^2\phi\right], \tag{E3}$$

$$D_{hh}^{>}(\phi)_{1,-1,1,1} = (1)_s\cdot[(5)+(9)]^{>}\cdot(13)_i = \Gamma_{hh}^{s}R_{12i}R_{12s}e^{i2k_{1zi}d}e^{i2k_{1zs}d}\left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2F_2}\sin^2\phi\right], \tag{E4}$$

$$D_{hh}^{>}(\phi)_{-1,1,1,1} = (1)_s\cdot[(7)+(11)]^{>}\cdot(14)_i = \Gamma_{hh}^{s}R_{12s}e^{i2k_{1zs}d}e^{i2k_{1z}d}\left[\frac{R_{12}}{D_2}\cos^2\phi - \frac{k_{1z}^2S_{12}}{k_1^2F_2}\sin^2\phi\right], \tag{E5}$$

$$D^{>}_{hh}(\phi)_{1,1,-1,-1} = (2)_s\cdot[(8)+(12)]^{>}\cdot(13)_i = \Gamma^{s}_{hh} R_{12i} e^{i2k_{1zi}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi\right], \tag{E6}$$

$$D^{>}_{hh}(\phi)_{-1,-1,1,1} = (1)_s\cdot[(5)+(9)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} R_{12s} e^{i2k_{1zs}d} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi\right], \tag{E7}$$

$$D^{>}_{hh}(\phi)_{-1,1,-1,1} = (2)_s\cdot[(7)+(11)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} e^{i2k_{1z}d} \left[\frac{R_{12}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{12}}{k_1^2 F_2}\sin^2\phi\right], \tag{E8}$$

$$D^{>}_{hh}(\phi)_{-1,1,1,-1} = (1)_s\cdot[(8)+(12)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} R_{12s} e^{i2k_{1zs}d} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi\right], \tag{E9}$$

$$D^{>}_{hh}(\phi)_{1,-1,-1,1} = (2)_s\cdot[(5)+(9)]^{>}\cdot(13)_i = \Gamma^{s}_{hh} R_{12i} e^{i2k_{1zi}d} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi\right], \tag{E10}$$

$$D^{>}_{hh}(\phi)_{1,-1,1,-1} = (1)_s\cdot[(6)+(10)]^{>}\cdot(13)_i = \Gamma^{s}_{hh} R_{12i} R_{1zs} e^{i2k_{1zi}d} e^{i2k_{1zi}d} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi\right], \tag{E11}$$

$$D^{>}_{hh}(\phi)_{-1,-1,-1,-1} = (2)_s\cdot[(6)+(10)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi\right], \tag{E12}$$

$$D^{>}_{hh}(\phi)_{1,-1,-1,-1} = (2)_s\cdot[(6)+(10)]^{>}\cdot(13)_i = \Gamma^{s}_{hh} R_{12i} e^{i2k_{1zi}d} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi\right], \tag{E13}$$

$$D^{>}_{hh}(\phi)_{-1,-1,-1,1} = (2)_s\cdot[(5)+(9)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} \left[\frac{1}{D_2}\cos^2\phi + \frac{k_{1z}^2}{k_1^2 F_2}\sin^2\phi\right], \tag{E14}$$

$$D^{>}_{hh}(\phi)_{-1,1,-1,-1} = (2)_s\cdot[(8)+(12)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} e^{i2k_{1z}d} \left[\frac{R_{10}R_{12}}{D_2}\cos^2\phi + \frac{k_{1z}^2 S_{10}S_{12}}{k_1^2 F_2}\sin^2\phi\right], \tag{E15}$$

$$D^{>}_{hh}(\phi)_{-1,-1,1,-1} = (1)_s\cdot[(6)+(10)]^{>}\cdot(14)_i = \Gamma^{s}_{hh} R_{12s} e^{i2k_{1zs}d} \left[\frac{R_{10}}{D_2}\cos^2\phi - \frac{k_{1z}^2 S_{10}}{k_1^2 F_2}\sin^2\phi\right]. \tag{E16}$$

For $z_1 < z_2$,

(E1′) = (E1),

(E2′) = (E2),

(E3′) = (E4),

(E4′) = (E3),

(E5′) = (E5),

(E6′) = (E7),

(E7′) = (E6),

(E8′) = (E8),

(E9′) = (E10),

(E10′) = (E9),

(E11′) = (E11),

(E12′) = (E12),

(E13′) = (E13),

(E14′) = (E15),

(E15′) = (E14),

(E16′) = (E16),

where

$$\Gamma^{s}_{hh} \equiv \frac{1}{(8\pi^2)^2}\frac{X_{10s}}{D_{2s}}\frac{X_{01i}}{D_{2i}}\frac{1}{k_{1zs}k_{1z}}.$$

[1]L. Tsang and J. A.Kong, IEEE Trans. Antennas Propag. **AP-30**, 292 (1981).

[2]L. Tsang and J. A. Kong, Radio Sci. **16**, 303 (1981).

[3]Y. Q. Jin and J. A. Kong, J. Appl. Phys. **55**, 1364 (1984).

[4]Y. Kuga and A. Ishimaru, J. Opt. Soc. Am. A **1**, 831 (1984).

[5]L. Tsang and A. Ishimaru, J. Opt. Soc. Am. A **1**, 836 (1984).

[6]M. Zuniga and J. A. Kong, J. Appl. Phys. **51**, 2315 (1980).

[7]U. Frisch, "Wave propagation in random media," in *Probabilistic Methods in Applied Mathematics*, edited by B. Reid (Academic, New York, 1968), Vol. 1, pp. 76–198.

[8]Y. Q. Jin and J. A. Kong, "Modified radiative transfer equation in strong fluctuation approach," presented at the National Radio Science Meeting, 25–28 June 1984, p. 52 NRSI.

[9]L. Tsang and J. A. Kong, Radio Sci. **14**, 259 (1979).

Calculating resonances (natural frequencies) and extracting them from transient fields

A. G. Ramm
Mathematics Department, Cardwell Hall, Kansas State University, Manhattan, Kansas 66506

(Received 13 September 1984; accepted for publication 21 December 1984)

Mathematical formulation and analysis of numerical methods for calculating the natural frequencies (resonances) are given. Stability of these methods towards roundoff errors and small perturbations of the obstacles is established. Some formulas for the variations of the natural frequencies due to small perturbations of the surface of the obstacle are given. A simple new method for extraction of resonances from transient fields is given.

I. INTRODUCTION

Let $\mathscr{D}$ be a finite obstacle with a smooth surface Γ, and let Ω be the exterior domain. The obstacle (scatterer) is three dimensional. The smoothness of Γ is of the type that ensures the applicability of Green's formulas. Roughly speaking, cusps-type singular points of the surface are not admissible, but edges (as in a cube) or conical points are admissible. We will discuss for simplicity scalar wave scattering, but the results and arguments are valid for electromagnetic wave scattering. The Green's function for a reflecting obstacle satisfies the equations

$$(-\nabla-k^2)G(x,y,k)=\delta(x-y) \quad \text{in } \Omega,$$
$$k>0, \quad x=(x_1,x_2,x_3), \tag{1.1}$$
$$G=0, \quad \text{if } x\in\Gamma \tag{1.2}$$
$$r\left(\frac{\partial G}{\partial r}-ikG\right)\to 0 \quad \text{as } r=|x|\to\infty. \tag{1.3}$$

Here y is the position vector of the source and ∇^2 is the Laplacian. The function G is uniquely determined by the conditions (1.1)–(1.3) and can be continued analytically on the whole complex plane of k as a meromorphic function of k. Its poles lie in the half-plane $\operatorname{Im} k<0$ and are called resonances, natural frequencies, or complex poles. The meromorphic nature of G as a function of k and the (closely connected with it) behavior of solutions of the time-dependent wave equation as $t\to+\infty$ were studied in the series of papers starting with Ref. 1. In Ref. 2 there is a bibliography of the subject. In Ref. 3 one can find a collection of papers and an extensive bibliography of the singularity and eigenmode expansion methods. In Refs. 4 and 5 there are reviews of the subject for engineers. The connection of the complex poles asymptotic with the behavior of solutions to the time-dependent wave equation is the foundation of the singularity expansion method (SEM). If

$$\nabla^2u=u_{tt} \quad \text{in } \Omega, \quad t>0, \quad u=0 \quad \text{on } \Gamma,$$
$$u(x,0)=0, \quad u_t(x,0)=f(x), \tag{1.4}$$

then the function v, defined as

$$v(x,k)=\int_0^\infty \exp(ikt)u(x,t)dt, \tag{1.5}$$

satisfies the equations

$$(\nabla^2+k^2)v=-f, \quad v=0 \quad \text{on } \Gamma,$$
$$r\left(\frac{\partial v}{\partial r}-ikv\right)\to 0, \quad r\to\infty, \tag{1.6}$$

$$v=\int_\Omega G(x,y,k)f\,dy, \tag{1.7}$$
$$u=(2\pi)^{-1}\int_{-\infty}^{\infty}\exp(-ikt)v(x,k)dk. \tag{1.8}$$

Assume that f is a smooth function which vanishes outside of a bounded domain (compactly supported). In the engineering literature (e.g., in Ref. 4) the complex variable $s=-ik$ is often used. In the physical and mathematical literature k is usually the complex variable. The half-plane $\operatorname{Im} k<0$ (which we use in this paper) corresponds to the half-plane $\operatorname{Re} s<0$ on the s plane. If one knows that[2]

$$v \quad \text{is meromorphic} \quad (\text{and analytic if } \operatorname{Im} k>0), \tag{1.9}$$
$$|v|<C(b)(1+|k|)^{-a}, \quad a>\tfrac{1}{2}, \quad b=\operatorname{Im} k, \quad |\operatorname{Re} k|\to\infty, \tag{1.10}$$
$$|\operatorname{Im} k_j|<\operatorname{Im} k_{j+1}|\to\infty \quad \text{as } j\to\infty, \tag{1.11}$$

then one can move the contour of integration in (1.8) in the k plane down and obtain the SEM expansion[2]

$$u(x,t)=\sum_{j=1}^{N}c_j(x,t)e^{-ik_jt}+O(e^{-|\operatorname{Im} k_{N+1}|t}), \quad t\to+\infty. \tag{1.12}$$

Here k_j are the complex poles of $v(x,k)$, $c_j(x,t)=i\operatorname{Res}_{k=k_j}\{e^{-ikt}v(x,k)\}$ and N is the number of the poles in the strip $0>\operatorname{Im} k\geqslant\operatorname{Im} k_N$. Usually it is assumed by engineers[4] that the poles k_j are simple, in which case $c_j(x,t)=c_j(x)$. If m_j+1 is the multiplicity of the pole k_j, then $c_j(x,t)=O(t^{m_j})$, and one can write (1.12) as

$$u(x,t)=\sum_{j=1}^{N}\sum_{m=0}^{m_j}c_{jm}(x)t^me^{-ik_jt}+O(e^{-|\operatorname{Im} k_{N+1}|t})$$
$$\text{as } t\to\infty. \tag{1.13}$$

Expansions (1.12) and (1.13) were called in Ref. 2 asymptotic SEM expansions. These expansions are proved under the assumption that $\mathscr{D}$ is strictly convex. The expansion

$$u=\sum_{j=1}^{\infty}c_j(x,t)e^{-ik_jt}, \tag{1.14}$$

which one can see in the literature, is not proved and probably is not valid in general. Formula (1.13) is not valid, in general, for nonconvex obstacles. For example, in Ref. 6 it is proved that if the obstacle consists of two strictly convex bodies, then $G(x,y,k)$ has countably many complex poles on the line $\operatorname{Im} k=c_0$. These poles asymptotically are equidistant and the distance between the poles depends on the distance between the bodies and the curvatures and principal

 0022-2488/85/051012-09$02.50

directions of the surfaces Γ_1 and Γ_2 of the bodies $\mathscr{D}_1$ and $\mathscr{D}_2$ at the two closest points $a_1 \in \Gamma_1$ and $a_2 \in \Gamma_2$. (That is, $|a_1 - a_2| = \min_{s_1 \in \Gamma_1, s_2 \in \Gamma_2} |s_1 - s_2|$.) This result shows that SEM expansion (1.13) is not valid for two convex bodies. It suggests also that (1.13) is not valid for a single body with nonconvex boundary which can hold a trapping mode (i.e., a standing wave in the geometrical optics approximation). This, however, is not proved yet.

In principle, one can tell the difference between convex obstacles and nonconvex obstacles, capable of holding a trapping mode, by the behavior of complex poles k_j for large j: for convex obstacles (1.11) holds and $|\mathrm{Im}\, k_j| \to +\infty$, while in the other case there exist infinitely many poles on a line $\mathrm{Im}\, k_j = c_0 < 0$.

The significance of the complex poles is manyfold. We mention only two areas important in applications. First, one can tabulate the complex poles and use them for target identification. Practically, it is expected that different scatterers produce different sets of complex poles. Although this is not proved, there are some supporting arguments (see Ref. 2, pp. 585 and 586). Second, systems theory uses representations of impulse responses as sums of exponentials. The problem is to find these exponentials from transient fields.

It is a long-standing open problem to prove that there exist infinitely many complex not purely imaginary poles of G for any reflecting obstacle. So far it was proved that there exist infinitely many purely imaginary poles (this is a result from Ref. 7, a simple proof one can find in Ref. 2).

The objectives of this paper include (1) formulation of the mathematical methods for numerically calculating the complex poles, (2) analysis of convergence and stability of these methods, and (3) formulation of a simple technique for extracting resonances (natural frequencies) from transient fields.[8]

An extensive bibliography on the third question can be found in Ref. 9. The techniques used the literature and reviewed in Ref. 9 are based mostly on Prony's method. Some other methods were also used.[9,10] Here we present a very simple numerical technique which seems to be new and does not require solving nonlinear or even linear equations. The question most difficult in this problem, that of noisy data, is discussed. The paper is organized as follows: Section II deals with questions (1) and (2). Section III deals with question (3).

Appendix A contains some formulas for perturbations of complex poles under perturbations of the surface of the scatterer. Numeration of formulas is separate in each of the sections. The material in Sec. III appeared in Ref. 8.

II. METHODS OF CALCULATING NATURAL FREQUENCIES

A. Basic equations

From the Green's formula one obtains

$$G(x,y,k) = g(x,y,k) - \int_\Gamma g(x,s')h(s',y)ds', \tag{2.1}$$

$$g = e^{ik|x-y|}/(4\pi|x-y|), \tag{2.2}$$

$$h = \frac{\partial G}{\partial N_s}, \tag{2.3}$$

where N_s is the outer normal to Γ at the point s, and the dependence on k is suppressed in some of the function for brevity. Let $x = s \in \Gamma$ in (2.1). Then

$$\int_\Gamma gh\, ds' = g. \tag{2.4}$$

If k_j is a pole of G then it is a pole of h, so that $h = \psi/(k-k_j)^m$. Multiply (2.4) by $(k-k_j)^m$ and let $k = k_j$ to obtain

$$Qh \equiv \int_\Gamma g(s,s',k_j)\psi\, ds' = 0. \tag{2.5}$$

Therefore the complex poles are the points k_j at which Eq. (2.5) has a nontrivial solution.

Let us differentiate (1) in the direction N_s and then take $x \to s \in \Gamma$ to obtain

$$[I + A(k)]h \equiv h + Ah = 2\frac{\partial g}{\partial N_s}, \tag{2.6}$$

$$Ah = \int_\Gamma 2\frac{\partial g}{\partial N_s} h\, ds'. \tag{2.7}$$

This gives the second way to characterize the complex poles: They are the points at which the equation

$$B\psi \equiv [I + A(k)]\psi = 0 \tag{2.8}$$

has a nontrivial solution.

B. Projection methods for calculating the poles

First, consider Eq. (2.8). Take a complete, in $L^2(\Gamma) = H$, set of linearly independent functions $\{\phi_j\}$. The linear span of the first n functions is a linear subspace H_n, $H_n \subset H_{n+1}$. Since the system $\{\phi_j\}$ is complete in H one concludes that the system of subspaces H_n is limit dense in H, that is, $\mathrm{dist}(\psi,H_n) \to 0$ as $n \to \infty$ for any $\psi \in H$. This property is crucial for the analysis below. Here dist is the distance between the element ψ and the subspace H_n. Let $\psi_n = \sum_{j=1}^n c_j^{(n)}\phi_j$. Consider the projection method for solving (2.8):

$$(B\psi_n,\phi_m) = 0, \quad 1 \leqslant m \leqslant n,$$

$$\sum_{j=1}^n (B\phi_j,\phi_m)c_j = 0, \quad 1 \leqslant m \leqslant n. \tag{2.9}$$

The necessary and sufficient condition for (2.9) to have a nontrivial solution is

$$\det b_n(k) = 0,$$

$$b_n(k) \equiv [b_{jm}(k)]_{j,m=1,\dots,n}, \quad b_{jm} \equiv (B\phi_j,\phi_m). \tag{2.10}$$

The parentheses denote the inner product in $L^2(\Gamma)$, $(u,v) = \int_\Gamma u\bar{v}\, ds$. The elements $b_{jm}(k)$ are entire functions of k since the operator A in (2.7) is an entire analytic operator function of k. Therefore (1) it is not obvious that Eq. (2.10) has zeros [e.g., $\exp(k)$ does not have zeros] and (2) if (2.10) has zeros $k_j^{(n)}$, $j = 1,2,\dots$, then one should prove convergence of this method, that is, one should prove that

$$\lim_{n\to\infty} k_j^{(n)} = k_j, \tag{2.11}$$

where k_j are the complex poles of G, and that all of the complex poles can be obtained in this way. This will be done later.

Consider Eq. (2.5). In the same way as in the previous section one can derive the equation

$$\det Q_n(k) = 0,$$

$$Q_n(k) \equiv [Q_{jm}(k)]_{j,m=1,\dots,n}, \quad Q_{jm} = (Q\phi_j,\phi_m). \quad (2.12)$$

This equation is of the same structure as (2.10), and the same questions (1) and (2) should be investigated for Eq. (2.12). The difference between operators Q and B is that Q is compact while B is of the Fredholm type, so that Eq. (2.5) is of the first kind, while (2.8) is of the second kind. The elements of Q_{jm} are easier to compute than b_{jm}.

C. Variational methods for calculating the poles

Consider the problem

$$|Qf| = \min, \quad |f| = 1, \quad (2.13)$$

where $|f|$ is the $L^2(\Gamma)$ norm, $|f|_p$ is the Sobolev space $W^{2,p}(\Gamma) = H^p$ norm, $|f| = |f|_0$, $|f|_p^2 = \int_\Gamma \{|u|^2 + |Du|^2 + \cdots + |D^p u|^2\} ds$, and D denotes the first-order derivative on Γ. For $p<0$ the space $W^{2,p}$ is defined as a dual to $W^{2,|p|}$.

Take $f_n = \sum_{j=1}^n c_j^{(n)}\phi_j$, substitute in (2.13), and obtain the problem

$$\sum_{j=1}^n q_{mj}^{(n)} c_j^{(n)} = \lambda c_m^{(n)}, \quad 1 < m < n, \quad q_{mj}^{(n)} \equiv (Q\phi_j, Q\phi_m), \quad (2.14)$$

where λ is an eigenvalue of the matrix $q_{mj}^{(n)}$. This matrix is an entire function of k. Its minimal eigenvalue $\lambda_1^{(n)}(k)$ is the minimum of the functional $|Qf|$ under the constraint (13). The points $k_j^{(n)}$, which are zeros of $\lambda_1^n(k)$,

$$\lambda_1^{(n)}(k_j^{(n)}) = 0, \quad (2.15)$$

converge to the complex poles of $G(x,y,k)$ and all of the complex poles can be obtained as limits of $k_j^{(n)}$ as $n \to \infty$,

$$k_j = \lim_{n\to\infty} k_j^{(n)}. \quad (2.16)$$

A similar idea was used in Ref. 11. Convergence of the methods is given in Sec. II B and a study of their stability is given in the next subsection.

D. Convergence and stability of the methods for calculating the poles

The basic ideas and methods of the analysis and proofs are taken from Ref. 12 (see also Refs. 2, 11, and 13). The basic results consist of a proof of convergence and stability of the methods given in Sec. II B towards roundoff errors and perturbations of the data.

(1) We start with the method given in the first part of Sec. II B. Let us assume that there exists a countable discrete set $\mathscr{P}$ of points k_j at which Eq. (2.8) has a nontrivial solution. In Sec. II A we proved that any complex pole of G belongs to $\mathscr{P}$. Let us show that any point $k_0 \in \mathscr{P}$ is a complex pole of G. Let ψ be a nontrivial solution to (8). Define the simple layer potential $v = \int_\Gamma g\psi\, ds'$. From the known formula (see Ref. 12, p. 240) $\partial v/\partial N_i = (A\psi + \psi)/2$ [in which $\partial/\partial N_i$ denotes the limit value of the normal derivative on Γ from the interior and A is given in (2.7)] and from Eq. (2.8) it follows that $\partial v/\partial N_i = 0$. We know that $(\nabla^2 + k_0^2)v = 0$ in $\mathscr{D}$. Since k_0^2 is a complex number and the spectrum of the interior Neumann Laplacian consists of positive numbers only, we conclude that $v = 0$ in $\mathscr{D}$. Therefore $V = 0$ on Γ. If G does not have a pole at $k = k_0$ then the problem $(\nabla^2 + k_0^2)v = 0$ in Ω, $v = 0$ on Γ, $v(x,k_0)$ is the limit value of a function $v(x,k)$ analytic in k in a neighborhood of k_0 and belonging to $L^2(\Omega)$ when $\operatorname{Im} k > 0$, has only the trivial solution. Thus, $v = 0$ in Ω if k_0 is not a pole in G. Therefore $\psi = \partial v/\partial N_i - \partial v/\partial N_e = 0$, where $\partial v/\partial N_e$ is the limit value of the normal derivative on Γ from the exterior domain. This contradicts the assumption that $\psi \not\equiv 0$. Therefore k_0 is a pole of G.

Let us prove now that for sufficiently large n, (1) Eq. (2.10) has solutions, (2) Eq. (2.11) holds, (3) all the complex poles can be obtained as limits (2.11), (4) complex poles k_j are stable towards small perturbations of the data; the notion of small perturbation will be specified.

Equation (2.9) can be written as an operator equation $P_n B P_n \psi = 0$, or

$$(I + P_n A)\psi_n = 0, \quad (2.17)$$

where P_n is the orthoprojection onto H_n, $\psi_n = P_n\psi \in H_n$, I is the identity. Since $A(k)$ is compact in H for any k and $P_n \to I$ as $n \to \infty$, where the arrow denotes strong convergence, one has $\|B - P_n B\| \to 0$ as $n \to \infty$. Therefore the operator $I + P_n A = I + A(k) - P^{(n)}A(k)$ is invertible for sufficiently large n in a neighborhood of any point k_0 at which $I + A(k_0)$ is invertible. Here $P^{(n)} \equiv I - P_n, P^{(n)} \to 0$ as $n \to \infty$. This argument shows that if k_0 is not a complex pole then there are no roots $k_j^{(n)}$ of Eq. (2.10) in a neighborhood of k_0. It remains to be proved that if k_0 is a pole of G, then for sufficiently large n there exists a root $k_j^{(n)}$ of Eq. (2.10) which lies in the circle $C_\delta: |k - k_0| \leqslant \delta$, where $\delta > 0$ is an arbitrary small number. Suppose that for some $\delta > 0$ and all n there are no roots $k_j^{(n)}$ of Eq. (2.10) in the circle C_δ. Then the operator $I + P_n A(k)$ is invertible in C_δ, the operator $(I + P_n A(k))^{-1}$ is analytic in k in C_δ, and therefore $\|(I + P_n A(k))^{-1}\| \leqslant c$, where c is a constant which does not depend on n. On the other hand, $(I + A(k))^{-1} = (I + P_n A(k) + P^{(n)}A(k))^{-1} = (I + P_n A(k))^{-1}(I + P^{(n)}A(k)(I + P_n A(k))^{-1})^{-1}$. Since $\|P^{(n)}A(k)\| \to 0$ and $\|(I + P_n A)^{-1}\| \leqslant c$ one concludes that $(I + A(k))^{-1}$ is a bounded operator in C_δ. This is a contradiction since k_0 is a pole of the operator $(I + A(k))^{-1}$. The contradiction proves that for any $\delta > 0$ and sufficiently large n there is a root $k_j^{(n)}$ of Eq. (2.10) in the circle C_δ.

(2) The above argument settles also the question about stability of the poles towards small perturbations of the data and roundoff errors. Indeed, small perturbations of the data and roundoff errors are equivalent to small perturbations of the matrix $b_{jm}(k)$.

Let us assume that a small perturbation of the matrix $b_{jm}(k)$ is caused by a small perturbation of the operator $B = I + A$. Let us denote $\tilde{B} = I + \tilde{A} = I + A + T$ as the perturbed operator. In this formulation the perturbed matrix $\tilde{b}_{jm}^{(n)}$ is the matrix of the operator $P_n \tilde{B} P_n$. The perturbation T can describe both the perturbation of Γ and the roundoff errors in the computing matrix $b_{jm}^{(n)}$. Our aim is to prove that in any finite domain on the k plane the poles $\tilde{k}_j^{(n)}$ of the perturbed operator $(I + P_n\tilde{B}(k)P_n)^{-1}$ differ from the poles $k_j^{(n)}$ of the unperturbed operator $(I + P_n B(k)P_n)^{-1}$ a little: $|\tilde{k}_j^{(n)} - k_j^{(n)}| \leqslant \epsilon(n, \|T\|)$, $\epsilon \to 0$ if $\|T\| \to 0$ and $n \to \infty$. Since we have already established the convergence property (2.11), it is sufficient to prove that

$$|\tilde{k}_j - k_j| \to 0, \quad \text{if } \|T\| \to 0, \quad T = \widetilde{B} - B. \tag{2.18}$$

Let k_j be a pole of $(I + A(k))^{-1}$ and there are no other poles of this operator in the circle C_δ. One has $(I + A(k) + T)^{-1} = (I + A(k))^{-1}(I + T(I + A(k))^{-1})^{-1}$. Suppose k_j is a pole of multiplicity ν. Then

$$\|(I + A(k))^{-1}\| \leqslant c/|k - k_j|^\nu, \quad k \in C_\delta, \quad c = \text{const}.$$

Thus

$$\|T(I + A(k))^{-1}\| \leqslant c\|T\| \, |k - k_j|^{-m}.$$

If $c\|T\|\delta^m < 1$ then the perturbed pole $\tilde{k}_j$, corresponding to the unperturbed pole k_j, lies inside the circle C_δ, that is, $|\tilde{k}_j - k_j| < \delta$. In other words $|\tilde{k}_j - k_j| = 0(\|T\|^{1/m})$ where m is the multiplicity of the pole k_j and $O(\|T\|^{1/m})$ means $\leqslant \text{const}\|T\|^{1/m}$.

The smallness of the perturbation of the surface is described in terms of the smallness of the norm $\|T\|$. One can give a relationship between the equation of the perturbed surface and the norm of T. This is cumbersome and is done in Appendix A.

(3) Let us study the method based on Eq. (2.12). The results will be the same: (1) Eq. (2.12) has roots $k_j^{(n)}$ for sufficiently large n, (2) Eq. (2.11) holds, (3) all the complex poles can be obtained as limits (2.11) and (4) small perturbations of the data lead to small perturbations of the complex poles uniformly on any bounded domain on the complex k plane.

Analysis of Eqs. (2.5) and (2.12) is more complicated than that of Eqs. (2.8) and (2.10) because (2.5) is an equation of the first kind. The basic tool in our analysis is the factorization formula

$$Q(k) = Q_0(I + V), \quad Q_0 \equiv Q(0), \quad V \equiv Q_0^{-1}(Q(k) - Q_0). \tag{2.19}$$

Here $Q_0 f = \int_\Gamma f\, ds'/4\pi|s - s'|$ is a self-adjoint positive definite operator on $H = L^2(\Gamma)$. This operator is an isomorphism between $H = H^0$ and H^1, while the operator V is compact in any space H^p (see Refs. 2 and 12 for details). Therefore the bilinear form $(Q_0 u, v)$ defines an inner product equivalent to the inner product $(u,v)_{-1/2}$ in $H^{-1/2}$. The matrix

$$Q_{jm} = (Q_0(I + V)\phi_j, \phi_m) = ((I + V)\phi_j, \phi_m)_{-1/2}.$$

Our previous arguments in Sec. II D are fully applicable to this matrix because (1) V is compact in $H^{-1/2}$ and depends analytically on k and (2) if a system $\{\phi_j\}$ is complete in H^0 then it is complete in H^p for any $p < 0$. Compactness of V was already mentioned. To explain the second statement assume that $f_p \in H^p, p < 0$. It is well known that $H^p \subset H^q$ if $p > q$, H^p is dense in H^q (that is, for any $\epsilon > 0$ and any $f \in H^q$ there exists an $f_\epsilon \in H^p$ such that $|f_\epsilon - f|_q < \epsilon$, where $|\cdot|_q$ is the norm in H^q), and $|f|_p \leqslant |f|_q$ if $p < q$. Let $f \in H^p$, $p < 0$, and $\epsilon > 0$ is fixed. Find $f_\epsilon \in H^0$ such that $|f - f_\epsilon|_p < \epsilon/2$. Use completeness of the system $\{\phi_j\}$ in H^0 to find $h_\epsilon = \Sigma_{j=1}^{n(\epsilon)} c_j(\epsilon)\phi_j$ such that $|h_\epsilon - f_\epsilon|_0 < \epsilon/2$. Then

$$|f - h_\epsilon|_p \leqslant |f - f_\epsilon|_p + |f_\epsilon - h_\epsilon|_p$$
$$\leqslant \epsilon/2 + |f_\epsilon - h_\epsilon|_0 < \epsilon, \quad \text{if } p < 0.$$

Therefore the system $\{\phi_j\}$ is complete in $H^{-1/2}$ and the matrix Q_{jm} is a matrix of the operator $I + V(k)$ in $H^{-1/2}$, where $V(k)$ is compact in $H^{-1/2}$ and analytic in k. The rest of the argument is the same as in Secs. II D (1) and II D (2) and the conclusions are formulated in the beginning of this section.

(4) In this section we study the variational method given in Sec. II C. If $\lambda_1^{(n)}(k_j^{(n)}) = 0$, then Eq. (2.14) corresponds to the projection method for the equation $Q^*Qf = 0$. The factorization in Sec. II D (3) is sufficient for the arguments of that section to hold for the operator Q^*Q [the reason is that $Q^*Q = (I + V^*)Q_0^2(I + V) = Q_0^2(I + V_1)$ where V_1 is compact]. Here we used the self-adjointness of Q_0. Compactness of V_1 follows from simple properties of pseudodifferential operators: ord $Q_0 = -1$, ord $V = -3$, $V_1 \equiv Q_0^{-2}V^*Q_0^2 + V + Q_0^{-2}V^*Q_0^2V$, ord $V_1 \leqslant$ ord $V < 0$. Here ord Q is the order of the pseudodifferential operator Q. One can find properties of pseudodifferential operators, e.g., in Ref. 14. Thus, one concludes that the results (1)–(4) in Sec. II D (3) hold for the variational method described in Sec. II C.

III. EXTRACTING NATURAL FREQUENCIES FROM TRANSIENT FIELDS

A. Preliminaries

Consider the problem

$$u_{tt} = \nabla^2 u, \quad t \geqslant 0, \quad x \in \Omega \subset \mathbb{R}^3, \quad u = 0 \quad \text{on } \Gamma,$$
$$u(x,0) = 0, \quad u_t(x,0) = f(x), \tag{3.1}$$

where $\mathscr{D} = R^3 \setminus \Omega$ is a bounded connected domain with a smooth strictly convex boundary, $f \in C_0^\infty$. In Ref. 2 the basic results on the asymptotic behavior of u as $t \to +\infty$ are described. In particular, the following asymptotic SEM (singularity expansion method) formula holds:

$$u = \sum_{j=1}^{N} \sum_{m=0}^{m_j} C_{jm}(x) t^m e^{-ik_j t} + O(e^{-|\text{Im}\, k_{N+1}|t}),$$
$$t \to +\infty, \tag{3.2}$$

where $C_{jm}(x)$ do not depend on t, $k_j = a_j - ib_j$, $b_j > 0$, are complex poles of the resolvent kernel G of the Dirichlet Laplacian in Ω

$$(\nabla^2 + k^2)G(x,y,k) = -\delta(x - y) \text{ in } \Omega, \, G = 0 \text{ on } \Gamma,$$
$$|x|\left(\frac{\partial G}{\partial |x|} - ikG\right) \to 0 \quad \text{as } |x| \to \infty,$$

and $m_j + 1$ is the multiplicity of the pole k_j, $\text{Im}\, k_j < 0$, $|\text{Im}\, k_j| \to \infty$ as $j \to \infty$. The poles k_j are called resonances or natural frequencies. The signal (3.2) is the transient field that can be observed experimentally. Knowledge of the resonances k_j may serve for target identification: the scatterers of various shapes produce various sets of resonances, and this is one of the reasons to be interested in resonances.

The other reason is that in systems theory one often models impulse responses as a sum of the type given in formula (3.2). The important problem of system identification can be formulated as follows: From the observation of the transient field (3.2) find the numbers k_j and m_j. There is an extensive literature on the subject. Many researchers contributed to the field (Prony, Bruns, Dale, Lagrange, Kühnen, and quite a few modern researchers). A large bibliography can be found in Refs. 9 and 15. Only the case $m_j = 0$ (simple poles) was treated in the literature as far as I know.

The purpose of the following subsection is (1) to give a simple numerical procedure for computing the number m_j and k_j, $1<j<N$ for any fixed N from the exact transient data, (2) to discuss this problem for noisy data, and (3) to briefly review the classical methods (Prony, Bruns, Lagrange, and Dale).

B. A simple method for extracting resonances from the transient field

Assume first that the scatterer is a strictly convex reflecting body so that (3.2) holds. By $u(n)$ let us denote the sequence $u(x,nh)$, where $h>0$ is a fixed number. It follows from (3.2) that

$$u(n)=c_{1m_1}h^{m_1}n^{m_1}e^{-ia_1nh}e^{-b_1nh}(1+O(1/n))$$

$$\text{as } n\to\infty, \text{ provided that } c_1<c_2<\cdots. \tag{3.3}$$

From (3.3) one obtains

$$\frac{u(n+1)}{u(n)}=e^{-ia_1h}e^{-b_1h}(1+O(1/n)) \quad \text{as } n\to\infty. \tag{3.4}$$

Thus,

$$b_1=\frac{1}{h}\ln\left|\frac{u(n)}{u(n+1)}\right|+O\left(\frac{1}{n}\right) \quad \text{as } n\to\infty, \tag{3.5}$$

$$b_1+ia_1=\frac{1}{h}\ln\frac{u(n)}{u(n+1)}+O\left(\frac{1}{n}\right) \quad \text{as } n\to\infty. \tag{3.6}$$

Suppose that k_j, m_j, and c_{jm}, $1<m<m_j$, $1<j<N$, are computed. In what follows, a method for computing c_{jm} is given. Let u_N denote the sum in formula (3.2), $u-u_N\equiv w_N$. Then, as above,

$$b_{N+1}=\frac{1}{h}\ln\left|\frac{w_N(n)}{w_N(n+1)}\right|=O\left(\frac{1}{n}\right) \quad \text{as } n\to\infty, \tag{3.7}$$

$$b_{N+1}+ia_{N+1}=\frac{1}{h}\ln\frac{w_N(n)}{w_N(n+1)}$$

$$=O\left(\frac{1}{n}\right) \quad \text{as } n\to\infty. \tag{3.8}$$

If a_1 and b_1 are found, then m_1 can be found by the formula

$$m_1=\frac{\ln\{u(n)e^{ia_1nh+b_1nh}\}}{\ln n}+O\left(\frac{1}{\ln n}\right) \quad \text{as } n\to\infty. \tag{3.9}$$

Similarly,

$$m_{j+1}=\frac{\ln\{w_j(n)e^{ia_{j+1}nh+b_{j+1}nh}\}}{\ln n}+O\left(\frac{1}{\ln n}\right)$$

$$\text{as } n\to\infty, \quad j=1,2,\ldots. \tag{3.10}$$

If a_1, b_1, and m_1 are found then

$$c_{1m_1}=\frac{u(n)e^{ia_1nh+b_1nh}}{(nh)^{m_1}}+O\left(\frac{1}{nh}\right) \quad \text{as } n\to\infty. \tag{3.10'}$$

In the literature[15–22] the case of simple resonances ($m_j=0$) only was discussed. In this case $O(1/n)$ in (3.11) can be substituted by $O(e^{-(b_2-b_1)nh})$ as $n\to\infty$.

In a similar way all the coefficients c_{1m} can be computed. Practically, one takes n large, neglects the remainder in formulas of the type (3.10′), and uses the main term in the right-hand side of (3.10′) as the formula for c_{1m_1}, etc.

If k_1, a_1, b_1, and C_{1m}, $1<m<m_1$, are found then one works with $w_1=u-u_1$ and so on. This is a method for computing the coefficients c_{jm} from the transient field.

An alternative method is the usual least squares method. If m_1, a_1, and b_1 are found then c_{1m}, $1<m<m_1$, can be found from the requirement

$$\sum_{n=1}^{\infty}\left|u(n)-\sum_{m=0}^{m_1}c_{1m}(nh)^me^{(-ia_1-b_1)nh}\right|^2=\min.$$

This leads to a uniquely solvable linear system for the coefficients c_{1m}, $1<m<m_1$. If k_1, a_1, b_1, m_1, and c_{1m}, $1<m<m_1$, are found then one works with $u-u_1=w_1$ and considers w_1 as the transient field. The method previously described is quite simple and does not require solving linear systems.

Formulas (3.5)–(3.11) give a simple method for extracting resonances and their multiplicities from the exact transient field. The much more complicated case of noisy data is discussed below.

In systems theory u, being an impulse response of a system, often does not depend on x.

We will discuss the case of noisy data. Assume that $y(n)=u(n)+\epsilon(n)$ is measured instead of $u(n)$. Here $\epsilon(n)$ is noise. Let us assume that $\epsilon(n)$ is uniformly distributed on the interval $[-\epsilon,\epsilon]$, $\epsilon>0$ is a given number. In practice the level of noise is not known exactly since the noise comes not only from the errors in measurements but also from the unknown background noise in the environment of the scatterer. But without some assumptions about the noise nothing can be derived. One has

$$y(n)=c_{1m_1}h^{m_1}n^{m_1}e^{-ia_1nh-b_1nh}(1+O(1/n))+\epsilon(n)$$

$$=c_{1m_1}h^{m_1}n^{m_1}e^{-ia_1nh-b_1nh}\{1+O(1/n)$$

$$+\epsilon(n)c_{1m_1}^{-1}(hn)^{-m_1}e^{ia_1nh+b_1nh}\}. \tag{3.11}$$

From (11) it follows that, regardless of the method used, the extraction of the complex poles k_j from noisy data is highly unstable and depends on the magnitude of $\alpha_n\equiv O(1/n)+\epsilon(hn)^{-m_1}e^{b_1nh}c_{1m_1}^{-1}$. If there exists n such that $\alpha_n\ll 1$ (say $\alpha_n<0.1$), then the pole $k_1=a_1-ib_1$ can be computed by formulas (3.5) and (3.6) in which $y(n)$ (α_n) should take the place of $u(n)$ ($O(1/n)$). Similar considerations hold for other poles. Since $b_j>0$ the factor e^{b_jnh} is growing as $n\to\infty$. Therefore ϵ should be small in order that α_n be small and k_j could be computed. In this case it is not advisable to take n too large because for large n the second term in α_n becomes large. Since the bound on $O(1/n)$ is not available it is not worthwhile to compute the optimal n, but practically, n should be taken as a value for which $(1/h)\ln|y(n)/y(n+1)|$ is stationary when one computes b_1, and for which $(1/h)\ln|z_j(n)/z_j(n+1)|$ is stationary when one computes b_{j+1}.

If the constant c_{1m_1} is small then the second term in α_n is large unless ϵ is sufficiently small. Therefore it is difficult to compute resonance with small Laurent coefficients (coupling coefficients) in front of the singular terms $(k-k_j)^{-m}$. All these arguments are very simple but they show clearly the nature of the difficulties for which noise is responsible and the limitations of any method of resonances extraction from noisy data.

We assumed that the scatterer was convex. This assumption implies the basic result: the validity of Eq. (3.2). In the outstanding paper Ref. 6 it is proved that for the scatterer consisting of two strictly convex reflecting bodies (3.2) does not hold: there exist countably many poles k_j on some line $\operatorname{Im} k = \text{const}$. Therefore one cannot order the poles by the rule $|\operatorname{Im} k_j| < |\operatorname{Im} k_{j+1}|$ in the case of two disjoint convex reflecting bodies. If the scatterer is just one strictly convex reflecting body then it is known that $|\operatorname{Im} k_j| \to +\infty$ as $j \to \infty$ and (3.2) holds.[2]

C. A brief review of the existing methods for the resonances extraction

The most popular method is Prony's method.[8,16] One assumes that $u = u(t) = \sum_{j=1}^{N} c_j e^{s_j t}, s_j \equiv -ik_j, c_j = \text{const}$. One observes $u(t)$ experimentally and wants to compute s_j and c_j. If the data is exact (there is no noise) then Prony's method consists in the following. Let $f_n \equiv u(nh)$ where $h > 0$ is a fixed number, $e^{s_j h} \equiv z_j$. Then $f_n = \sum_{j=1}^{N} c_j z_j^n, n \geqslant 0$. An obvious linear algebra argument shows that $\det A_{pq}^{(m)} = 0$, $m \geqslant 0$, $0 \leqslant p$, $q \leqslant N$, where $A_{p0}^{(m)} \equiv f_{p+m}$, $A_{pq}^{(m)} = z_q^p$, $0 \leqslant p \leqslant N$, $1 \leqslant q \leqslant N$. Therefore,

$$0 = \sum_{p=0}^{N} f_{p+m} A_p, \quad m \geqslant 0, \tag{*}$$

where A_p are the cofactors corresponding to the elements f_{p+m} of the matrix $A_{pq}^{(m)}$. Notice that A_p do not depend on m. Write $N+1$ equations (*) taking $m = 0,1,...,N$, and find a nontrivial solution $(A_0, A_1,...,A_N)$ to the $N+1$ simultaneous equations (*). Consider the equation

$$\sum_{p=0}^{N} A_p z^p = 0\,. \tag{**}$$

From the structure of the matrix $A_{pq}^{(0)}$ it is clear that Eq. (**) has solutions $z_j = e^{s_j h}$. Thus, $s_j = h^{-1} \ln z_j$. If one does not know the number N (and this is usually the case in practice) then there is a problem of choosing the right N. In Ref. 14, p. 140, there is a method (due to Kühnen) for choosing N. If the data are noisy then one faces the difficulties explained in Sec. II B and reported in the literature.[9] If the data are noisy then the matrix f_{p+m}, $0 \leqslant p$, $m \leqslant N$ is nonsingular and system (*) with $0 \leqslant m \leqslant N$ has only the trivial solution $A_p = 0$, which cannot be used since Eq. (**) in this case gives no information. Therefore in practice one takes A_p, $0 \leqslant p \leqslant N$, the components of an eigenvector corresponding to the minimal eigenvalue of the matrix F^*F. Here F^* is the adjoint matrix, and F is the matrix of the noisy data, $F_{p+m} = f_{p+m} + \epsilon_{p+m}$, where ϵ_{p+m} is noise. If there are several eigenvectors corresponding to the minimal eigenvalue, one has no rule to pick up any particular eigenvector. But this situation is not generic in the sense that a small perturbation of the matrix will split up the multiple eigenvalue into a number of simple ones. However, the simple eigenvalues will be close to each other and it will be difficult to find the minimal eigenvalue numerically. (Recall that an eigenvalue is called simple if there is only one linearly independent eigenvector corresponding to this eigenvalue.) One can find an extensive bibliography and a discussion of Prony's method in Ref. 9. The Bruns' method described in Ref. 15 is essentially the Prony method for real resonances.

Let us outline another method for extracting the resonances. Let $f(t) = \sum_{j=1}^{N} c_j e^{s_j t}$ and N is assumed known. Then $f^{(m)}(t) = \sum_{j=1}^{N} c_j s_j^m e^{s_j t}$. Taking $t = 0$ and $m = 0,...,N,...$, one obtains

$$\begin{vmatrix} f(0) & 1 & \cdots & 1 \\ f'(0) & s_1 & & s_N \\ \vdots & \vdots & & \vdots \\ f^{(N)}(0) & s_1^N & & s_N^N \end{vmatrix} = 0, \quad \begin{vmatrix} f'(0) & 1 & 1 \\ \vdots & \vdots & \vdots \\ f^{(N+1)} & s_1^N & s_N^N \end{vmatrix} = 0, \ldots\,. \tag{3.12}$$

Therefore

$$\sum_{p=0}^{N} f^{(m+p)}(0) A_p = 0, m = 0,1,2,\ldots\,. \tag{3.13}$$

Here A_p is the cofactor corresponding to the element $f^{(p)}(0)$ in the first matrix in (3.12). The argument above is very similar to that given in the previous method and $f^{(m+p)}(0)$ plays the role of f_{m+p} in (*) in that section. Taking $m = 0,1,...,N$ in (3.13) one obtains a homogeneous system of linear equations. If $A_0,...,A_N$ is a solution to this system then the N roots of equation (**) for this method are equal to s_j, $1 \leqslant j \leqslant N$. This gives a method of extracting the resonances s_j from the exact transient data. It is interesting that only the data near $t = 0$ is used in this procedure, while the procedure in the previous method needs the data at large t. On the other hand, the procedure based on Eq. (3.13) is very sensitive to the noise in the data because one needs to differentiate the data.

We assumed above that the number N of resonances was known. If N is not known, then one can find N as the smallest number for which $\det f_{p+m} = 0$, $0 \leqslant p$, $m \leqslant N$.

The simple algorithm in Sec. III B *requires that* $b_1 < b_2 < b_3 \cdots$. In practice the poles $a_j - ib_j$ occur in pairs $\pm a_j - ib_j$ and the measured transient field is a real-valued function. Assuming that the poles are simple, i.e., $m_j = 0$, one has

$$u(t) = \sum_{j=1}^{N} c_j \exp(-b_j t) \cos(a_j t + \phi_j) + O(e^{-b_{N+1} t})$$

as $t \to +\infty$. Therefore for large t one obtains $u(t) = c_1 \exp(-b_1 t) \cos(a_1 t + \phi_1) + O(e^{-b_2 t})$, $t \to +\infty$. If the values $u_n \equiv u(nh)$ are measured, then the values c_1, b_1, a_1, and ϕ_1, $0 \leqslant \phi_1 < 2\pi$, can be numerically obtained from the requirement

$$F(c_1,b_1,a_1,\phi_1) \equiv \frac{1}{m} \sum_{n=n_0}^{n_0+m} |c_1 \exp(-b_1 hn) \times \cos(a_1 hn + \phi_1) - u_n|^2 = \min. \tag{3.14}$$

Here n_0 is a large number such that $\exp(-b_1 hn_0) \gg \exp(-b_2 hn_0)$, $m > 4$ is a fixed number, and the function $F(c_1,b_1,a_1,\phi_1)$ is to be minimized numerically. If this minimization problem is solved one can consider $w_1 \equiv u(t) - u_1(t)$, where $u_1(t) \equiv c_1 \exp(-b_1 t) \cos(a_1 t + \phi_1)$, and apply the same procedure for finding c_2, b_2, a_2, ϕ_2. Each step requires minimization of a function of four variables only. The basic new idea in our method is to use the asymptotic behavior as $t \to +\infty$ of the transient field. One should have in mind that the basic asymptotic SEM expansion (3.2) is proved only if the scatterer is convex[1] (or, more generally, star shaped). It does not hold, for example, when the scatterer consists of two convex obstacles. In this case there exists infinitely many poles k_j, such that $|k_j - ic_0 - \pi d^{-1} j| \leqslant c(1 + |j|)^{-1/2}$ for all large j, $j = \pm j_0, \pm (j_0 + 1), \ldots$. Here d is the distance between the two obstacles and c_0 depends on d, on the principal curvatures, and principal directions of the surfaces Γ_1 and Γ_2 of the two obstacles at the points $s_1 \in \Gamma_1$ and $s_2 \in \Gamma_2$, such that $|s_1 - s_2| = d$. This remarkable result was proven recently by Ikawa.[6]

IV. BIBLIOGRAPHICAL REMARKS

First we mention some of the old papers. Of these only Prony's paper[16] is often mentioned by modern authors. There is a translation of this paper in Ref. 8. Bruns[17] used practically the same idea as Prony. His work is discussed in Ref. 14. There are several authors, astronomers mostly, which were interested in detection of hidden periodicities.[17–22] Although only the case $m_j = 0$, $b_j = 0$ was discussed in these papers, the basic questions (extracting the resonances from the transient field, determining the number N of resonances, etc.) were actually identical with the questions discussed in a very recent review[9] of the state of the art in this field. References[17–22] are not cited by modern western authors in the field. A review of these papers can be found in Ref. 15.

There exists a very extensive modern literature on the extracting of resonances. One can find a large bibliography and a review of the basic results in Ref. 9. We did not discuss here some of the methods mentioned in Ref. 9.

There are many reasons for being interested in the extracting of resonances. We mention only two major theories: singularity expansion method (see Refs. 2, 12, and 23 for the mathematical results) and systems identification (see the bibliography in Ref. 9).

ACKNOWLEDGMENTS

The author thanks Dr. C. Baum and T. Brown for useful discussions. This paper is dedicated to Professor C. L. Dolph for his 66th birthday.

The author thanks the Dikewood Corporation for support.

APPENDIX A: PERTURBATION OF RESONANCES

1. Abstract scheme

First let us present an abstract scheme. Assume that a compact operator function $A(k)$ on a Hilbert space H is analytic in k in a domain on the complex plane k, and $A(k_0)$ has an eigenvalue -1. Then $(I + A(k))^{-1}$ has a pole at $k = k_0$. Suppose that $A(k,\epsilon)$ is a compact operator function such that $A(k,0) = A(k)$, which is analytic in k and ϵ, $|k - k_0| < \delta_0$, $|\epsilon| < \delta_1$. Assume that k_0 is an isolated pole of $I = A(k)$. [This is the case if $I + A(k)$ is invertible for at least one k in the disk $|k - k_0| < \delta_0$.] Then, in a neighborhood of k_0, there exist a finite number m_0 of points $k_j(\epsilon) = k_0 + \Sigma_{m=1}^{\infty} a_{mh} \epsilon^{m/p}$, $1 \leqslant j \leqslant j_0$, such that the operator $(I + A(k,\epsilon))^{-1}$ has poles $k(\epsilon)$. Here j_0 is the multiplicity of the pole k_0, and the meaning of the integer p will be explained in the proof which is based on the idea in Ref. 2, p. 582.

Let $\phi_1,\ldots,\phi_n$ be an orthonormal basis of $N(I + A(k_0))$, where $N(A)$ is the null space of an operator A. Let $\psi_1,\ldots,\psi_n$ be an orthonormal basis of $N(I + A^*(k_0))$, where the star denotes the adjoint operator. Let $Th \equiv \Sigma_{j=1}^{n} (h,\phi_j)\psi_j$. The operator $I + A(k_0) + T$ is invertible in H. Indeed, $(I + A(k_0) + T)h = 0$ implies that $(Th,\psi_j) = 0$, $1 \leqslant j \leqslant n$. This leads to $(h,\phi_j) = 0$, $1 \leqslant j \leqslant n$, i.e., $Th = 0$ and $(I + A(k_0))h = 0$. Thus $h \in N(I + A(k_0))$ and $h \perp N(I + A(k_0))$. Therefore, $h = 0$, and by Fredholm's alternative $(I + A(k_0) + T)^{-1} = \Gamma$ is bounded. Consider $(I + A(k,\epsilon))^{-1} = (I + A(k_0) + T + A(k,\epsilon) - A(k_0) - T)^{-1} = (I + a(k,\epsilon))^{-1}$. Then $\Gamma(k,\epsilon)$, where $\Gamma(k,\epsilon) = (I + A(k_0) + T + A(k,\epsilon) - A(k_0))^{-1}$, is analytic in k and ϵ in a neighborhood Δ of $(k_0,0)$, and $a(k,\epsilon) = -\Gamma(k,\epsilon)T$ is a finite-dimensional operator analytic in k and ϵ in Δ, $ah = -\Sigma_{j=1}^{n} (h,\phi_j)\Gamma(k,\epsilon)\psi_j$. Since $\Gamma(k,\epsilon)$ is an isomorphism of H onto H for $k,\epsilon \in \Delta$, the elements $\psi_j(k,\epsilon) = -\Gamma(k,\epsilon)\psi_j$ are linearly independent and analytic in $k,\epsilon \in \Delta$. Therefore the operator $(I + a(k,\epsilon))^{-1}$ can be constructed explicitly. If $(I + a)h = f$, then $h + \Sigma_{j=1}^{n} h_j \psi_j(k,\epsilon) = f$, $h_j \equiv (h,\phi_j)$. Multiply by ϕ_m to obtain $h_m + \Sigma_{j=1}^{n} c_{mj} h_j = f_m$, where $f_m = (f,\phi_m)$, $c_{mj} = (\psi_j,(k,\epsilon),\phi_m)$, c_{mj} are analytic in $k,\epsilon \in \Delta$. Thus, $h_m = d_m(k,\epsilon)/d(k,\epsilon)$, where d_m and $d = \det(\delta_{mj} + c_{mj})$ are analytic in Δ. One has $(I + a(k,\epsilon))^{-1} f = f - \frac{1}{d}\Sigma_{j=1}^{n} d_j(k,\epsilon)\psi_j(k,\epsilon)$. From this formula it is clear that the poles of $(I + a(k,\epsilon))^{-1}$ can occur only at the zeros of $d(k,\epsilon)$. Thus the equation for the perturbed poles is

$$\begin{aligned} &d(k,\epsilon) = \det[\delta_{jm} - (\Gamma(k,\epsilon)\psi_j,\phi_m)] = 0, \\ &\Gamma(k,\epsilon) = (I + T + A(k,\epsilon))^{-1}. \end{aligned} \tag{A1}$$

For $\epsilon = 0$ the function $d(k) = d(k,0)$ has, by assumption, a zero of multiplicity j_0. By the Weierstrass' preparation theorem (see, e.g., Ref. 2, p. 583) one has $d(k,\epsilon) = [k^{j_0} + \Sigma_{j=1}^{j_0-1} c_j(\epsilon)k^j]g(k,\epsilon)$, where $g(0,0) \neq 0$, $c_j(0) = 0$, and c_j and $g(k,\epsilon)$ are holomorphic functions. Therefore Eq. (A1) is equivalent to

$$k^{j_0} + \sum_{j=1}^{j_0-1} c_j(\epsilon)k^j = 0. \tag{A2}$$

This equation has j_0 roots. These roots can be divided into several groups so that the p roots $(k_1(\epsilon),\ldots,k_p(\epsilon))$ in the νth group can be expanded in a Puiseux series $k_j(\epsilon) = k_0 + \Sigma_{m=1}^{\infty} a_{mj} \epsilon^{m/p_\nu}$, $\Sigma_\nu p_\nu = j_0$. The number of groups and the integers p_ν can be computed by the algorithm known as Newton's diagram method (see, e.g., Ref. 24). Let us summarize our arguments: a method for computing the poles $k_j(\epsilon)$ of the perturbed operator $(I + A(k,\epsilon))^{-1}$ is given. The method is valid under the following assumptions: $A(k,\epsilon)$ is a compact analytic (in k and ϵ) linear operator function on a Hilbert space, the operator $(I + A(k,0))^{-1}$ has an isolated pole at $k = k_0$. To use the method computationally one needs to compute (or to know) the bases in the subspaces $N(I + A(k_0))$ and $N(I + A^*(k_0))$. This is a linear algebra problem.

2. Reduction of a concrete perturbation problem to the abstract one

Suppose that the surface Γ of the obstacle is perturbed. Let $x_j = x_j(u,v)$, $1 \leqslant j \leqslant 3$ be a parametric equation of Γ, and

$z_j = x_j(u,v) + \epsilon y_j(u,v)$ be the equation of the perturbed surface Γ_ϵ, where ϵ is a small parameter. Assume that the functions x_j and y_j, $1 \leqslant j \leqslant 3$, $u,v \in S \equiv \{u,v: 0 \leqslant u,\ v \leqslant 1\}$ are smooth. Consider Eq. (2.8). Assume that k_0 is a pole of the operator $(I + A(k))^{-1}$. Suppose that the bases of the subspaces $N(I + A(k_0))$ and $N(I + A^*(k_0))$ are computed. Consider the problem corresponding to the perturbed surface Γ_ϵ. The operator $A(k,\epsilon)$, associated with this problem, is of the form

$$\int_{\Gamma_\epsilon} 2 \frac{\partial g}{\partial N_s} h\, ds = \int_S A(u,v,u',v',k,\epsilon) h\, du' dv'.$$

Here $A(u,v,u',v',k,\epsilon)$ is the kernel of the integral operator in the variables u,v,u',v'. If $\epsilon = 0$ then $\Gamma_\epsilon = \Gamma$ and we assume that the sets $\{\phi_j\}$ $(\{\psi_j\})$, $1 \leqslant j \leqslant n$ of all linearly independent solutions of the equation $\phi + \int_S A(u,v,u',v',k_0,0)\phi \times du'\, dv' = 0$ $(\psi + \int_S \overline{A(u',v',u,v,k_0,0)}\psi\, du'\, dv' = 0)$ are known. Then the abstract scheme is applicable.

Since small perturbations of the kernel cause small perturbations of the poles, one can approximate the kernel A by a degenerate kernel and consider the corresponding matrix problem.

As an example consider a simple case when the matrix is 2×2. Let

$$\begin{pmatrix} 1 & e^{\pi k} \\ e^{\pi k} & 1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}.$$

Then the inverse matrix is

$$\begin{vmatrix} 1 & -e^{\pi k} \\ -e^{\pi k} & 1 \end{vmatrix} \frac{1}{1 - e^{2\pi k}}.$$

It has simple poles $k_m = im$, $m = 0, \pm 1, \dots$. Consider, for example, pole $k_0 = 0$. The set $\{\phi_j\}$ corresponding to this pole is the set of linearly independent solutions to the equation

$$\begin{cases} c_1 + c_2 = 0 \\ c_1 + c_2 = 0 \end{cases}.$$

Thus, there exists one linearly independent solution

$$\phi_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

In our example the matrix

$$\begin{pmatrix} 1 & e^{\pi k} \\ e^{\pi k} & 1 \end{pmatrix}$$

is self-adjoint for $k = 0$. Thus

$$\psi_1 = \phi_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Consider the perturbed matrix

$$\begin{pmatrix} 1 + \epsilon a_{11} & e^{\pi k} + \epsilon a_{12} \\ e^{\pi k} + \epsilon a_{21} & 1 + \epsilon a_{22} \end{pmatrix}.$$

The poles of this matrix are the roots of the equation $(1 + \epsilon a_{11})(1 + \epsilon a_{22}) - (e^{\pi k} + \epsilon a_{12})(e^{\pi k} + \epsilon a_{21}) = 0$. We are interested in the root $k(\epsilon)$ such that $k(\epsilon) \to 0$ as $\epsilon \to 0$. One has $1 - e^{2\pi k} + \epsilon[a_{11} + a_{22} - e^{\pi k}(a_{12} + a_{21})] + \epsilon^2(a_{11}a_{22} - a_{12}a_{21}) = 0$. Let $e^{2\pi k} = z$, $a_{11} + a_{22} = a$, $a_{12} + a_{21} = b$, $a_{11}a_{22} - a_{12}a_{21} = c$. Then

$$z^2 + z\epsilon b - 1 - \epsilon a - \epsilon^2 c = 0,$$

$$z(\epsilon) = -\epsilon b/2 + \sqrt{(\epsilon^2 b^2/4) + 1 + \epsilon a + \epsilon^2 c}$$
$$\approx 1 + (\epsilon/2)(a - b).$$

The plus sign in front of the radical is chosen because $z = 1$ if $k = 0$. Thus, in this example the perturbed pole $k_0 = 0$ can be computed for small ϵ as $k(\epsilon) = \pi^{-1} \ln z(\epsilon) = (a - b)\epsilon/2\pi$. Depending on the values of a and b the perturbed pole can move in any direction. If $a \neq b$ then $k(\epsilon) = O(\epsilon)$. If $a = b$ then $k(\epsilon) = O(\epsilon^2)$. The bifurcation theory and the Newton diagram method solve the following problem. Given an equation $F(k,\epsilon) = 0$, find its solutions $k(\epsilon)$ such that $k(\epsilon) \to 0$ as $\epsilon \to 0$. It is assumed that $F(0,0) = 0$. If $F_k(0,0) \neq 0$ then the solution is well known and is given by the standard implicit function theorem. If $F_k(0,0) = 0$ then the solution is more complicated. Methods and algorithms for solving this problem can be found in Ref. 24.

APPENDIX B: EXTRACTION OF RESONANCES FROM THE TRANSIENT FIELD

The case of simple poles symmetrical with respect to the imaginary axis is as follows.

Let the poles be simple and occur in pairs $\pm a_j - ib_j$, $0 < b_1 < b_2 < \cdots$. Then the transient field is $u = \Sigma_{j=1}^N u_j + O(\exp(-b_{N+1}t))$, as $t \to +\infty$, $u_j = c_j \exp(-b_j t) \cos(a_j t + \phi_j)$, $c_j > 0, a_j > 0, b_j > 0, 0 \leqslant \phi_j < 2\pi$. We sketch a method for finding c_j, b_j, a_j, and ϕ_j, which is simpler than the one given by formula (3.14).

Step 1: One has $u(t) = u_1 + O(\exp(-b_2 t))$, so $u(t) \approx u_1(t)$ as $t \gg 1$. Thus

$$\frac{1}{mh} \ln \frac{u[(n+m)h]}{u(nh)} \approx -b_1 + \frac{1}{mh} \ln \frac{\cos(a_1 nh + a_1 mh + \phi_1)}{\cos(a_1 nh + \phi_1)},$$
$$h > 0. \tag{B1}$$

Therefore for $m \gg 1$ one can find b_1 from (B1). Some caution is needed: if $\cos(a_1 nh + a_1 mh + \phi_1) \approx 0$ then the ln term is large. But this can happen rarely. One can compute the left side of (B1) for a number of consecutive m and neglect large values of this quantity.

Step 2: If b_1 is found then a_1 can be found by the formula $a_1 \approx (-\ddot{v}/v)^{1/2}$, $t \gg 1$, where $v(t) \equiv u(t)\exp(b_1 t)$, and $\ddot{v} = d^2 v/dt^2$.

Step 3: If a_1 and b_1 are found then $c_1 \approx (\dot{v}^2/a_1^2 + v^2)^{1/2}$, $t \gg 1$.

Step 4: If a_1, b_1, and c_1 are found, then ϕ_1 can be found from the equations

$$c_1^{-1} v(2\pi n/a_1) = \cos\phi_1, \quad -c_1^{-1} v(2\pi n/a_1 + \pi/2a_1)$$
$$= \sin\phi_1, \quad n \gg 1.$$

Since $0 \leqslant \phi_1 < 2\pi$, these equations determine ϕ_1 uniquely. If b_1, a_1, c_1, and ϕ_1 are found one can use $u - u_1$ for finding u_2, etc. The basic idea is the same as in Sec. III. The difficulties are similar to the ones discussed after formula (3.11).

[1]A. G. Ramm, Usp. Mat. Nauk. **19**, 192 (1964).
[2]A. G. Ramm, J. Math. Anal. Appl. **86**, 562 (1982).

[3]A. G. Ramm, Electromagnetics **1**, 385 (1981).
[4]C. Baum, Proc. IEEE **64**, 1598 (1976).
[5]C. Dolph and R. Scott, "Recent developments in the use of complex singularities in electromagnetic theory and elastic wave propagation," in *Electromagnetic scattering*, edited by P. Vslenghi (Academic, New York, 1978).
[6]M. Ikawa, J. Math. Kyoto Univ. **23**, 127, 795 (1983).
[7]P. Lax and R. Phillips, Commun. Pure Appl. Math. **22**, 737 (1969).
[8]A. G. Ramm, IEEE Trans. **AP-33**, 223 (1985) © 1985 IEEE.
[9]J. Auton and M. Van Blaricum, AFWL Math. Notes **79**, (1981).
[10]T. Senior and J. Pond, "Pole extraction in the frequency domain," Interaction Note 411, AFWL, Dec. 1981.
[11]A. G. Ramm, J. Math. Phys. **21**, 2052 (1980); **23**, 1112 (1982).
[12]A. G. Ramm, *Theory and Applications of Some New Classes of Integral Equations* (Springer, New York, 1980).
[13]A. G. Ramm, J. Math. Phys. **23**, 1123, 2408 (1982); **24**, 2619 (1983) (the last paper is joint with G. Kristensson and S. Ström).
[14]H. Kumano-go, *Pseudodifferential Operators* (MIT, Cambridge, MA, 1981).
[15]M. Serebrennikov and A. Pervozvanskii, *Detection of Hidden Periodicities* (Nauka, Moscow, 1965) (in Russian).
[16]R. Prony, J. Ec. Polytech. **1**, 24 (1795).
[17]H. Bruns, Astron. Nachr. **188**, (1911).
[18]J. Dale, Mon. Not. R. Astron. Soc. **74**, 628 (1914).
[19]H. Plumer, Mon. Not. R. Astron. Soc. **76**, 60 (1913).
[20]F. Hopfner, Wien. Sitzungsber. **119**, 351 (1910).
[21]S. Oppenheim, Wien. Sitzungsber. **118**, (1909).
[22]F. Kühnen, Astron. Nachr. **182**, (1909).
[23]A. G. Ramm, IEEE Trans. **AP-28**, 897 (1980).
[24]M. Vainberg and V. Trenogin, *Bifurcation of Solutions of Nonlinear Equations* (Noordhoff, Leiden, 1974).

Dual physical interpretation of the energy tensor: Nonexistence of the Kerr–Newman example

D. W. Ferdon[a] and J. L. Safko
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208

(Received 28 February 1984; accepted for publication 4 January 1985)

Several recent papers have dealt with the possibility of interpreting the Kerr–Newman metric as a viscous fluid as well as its usual interpretation of a rotating, charged black hole. In this paper we show that there is no possible viscous fluid with the Kerr–Newman metric.

I. INTRODUCTION

General relativity can be interpreted as a theory in which the geometry and the "gravitational fields" are identified. The geometry is related to other fields through the Einstein field equation $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -kT_{\mu\nu}$, where $G_{\mu\nu}$ is a function of the metric tensor $g_{\mu\nu}$ only and $T_{\mu\nu}$ is a function of both $g_{\mu\nu}$ and the other fields ϕ^i. We let the index i denote any indices, metrical and field, needed to describe the nongravitational fields. For example, if ϕ^i is the Maxwell field tensor $f_{\mu\nu}$, then $T_{\mu\nu}$ is the electromagnetic stress tensor.

The ϕ^i are not arbitrary, they must also satisfy a dynamical set of equations themselves. For example, if the ϕ^i is an electromagnetic field, the $f_{\mu\nu}$ must satisfy Maxwell's equations.

In general this problem could be considered as arising from a variational principle $S = \int L(g_{mn}, \phi^i) d^4 \chi$. Physical solutions result from an extremum

$$\frac{\delta S}{\delta g_{\mu\nu}} = 0 \rightarrow G_{\mu\nu} = -kT_{\mu\nu},$$

$$\frac{\delta S}{\delta \phi^i} = 0 \rightarrow \text{dynamical equations of the field } \phi^i.$$

Any solution must, in general, satisfy both sets of equations.

There exist some special cases where the second set of equations is a consequence of the Bianchi relations which the $g_{\mu\nu}$ must satisfy. The most common case is the ideal fluid, viscous or nonviscous. In this case the dynamical equations of the fluid are contained in the gravitational field equations.

II. DUAL INTERPRETATION

Tupper[1,2] in 1977 and Raychaudhuri and Saha[3–5] have presented several viscous fluid solutions which are physically acceptable and which produce the same $g_{\mu\nu}$ as some other physically acceptable fluid solutions. Such a possibility will be termed a dual interpretation of the energy tensor. In this paper we show that such a proposed dual interpretation is not possible for the Kerr–Newman metric.

In general the dual interpretation idea can be stated as follows: It is possible to relate the metric parameters to two different sets of physical parameters, so that

$$G_{\mu\nu} = E_{\mu\nu} = M_{\mu\nu}, \tag{1}$$

where $E_{\mu\nu}$ is the electromagnetic stress energy tensor, given by

$$E_{\mu\nu} = F_{\mu\alpha} F^{\alpha}{}_{\nu} - \tfrac{1}{4} g_{\mu\nu} F_{\alpha\rho} F^{\alpha\rho}, \tag{2}$$

and $M_{\mu\nu}$ is the viscous fluid matter tensor given by

$$M_{\mu\nu} = \rho\, u_\mu u_\nu + p^* P_{\mu\nu} - 2\eta\, \sigma_{\mu\nu}, \tag{3}$$

where u_α is the fluid velocity vector, $P_{\alpha\rho} = g_{\alpha\rho} + u_\alpha u_\rho$ is the projection tensor perpendicular to $\bar{u}$, $\sigma_{\alpha\beta} = (u_{\alpha;\mu} P^\mu{}_\beta + u_{\beta;\mu} P^\mu{}_\alpha)/2 - \theta P_{\alpha\beta}/3$ is the shear tensor, trace-free part of expansion tensor $\theta_{\alpha\beta}$, $\widetilde{\Theta} = u^\alpha{}_{;\alpha}$ is the expansion of fluid lines, ρ is the density, $p^* = (p - \zeta\widetilde{\Theta})$ is the dynamic pressure, and η and ζ are coefficients of shear and bulk viscosity, respectively. We must have, from the definitions, the following algebraic identities:

$$\sigma_{\mu\nu} u^\mu = 0, \quad \sigma_\mu{}^\mu = 0, \quad u_\mu u^\mu = -1, \tag{4}$$

and for obvious physical reasons,

$$\eta, \zeta \geqslant 0, \quad E_{\mu\nu} \text{ traceless and } \rho = 3p^*.$$

A decisive criterion is to examine the relation between the parameters of the two fields as seen in the canonical coordinate system of the metric. Equations (1)–(3) are sufficient to "completely determine" the elements of one field expressed in terms of those of the other. If such a dual interpretation is possible, the resulting relationships must be physically reasonable.

This procedure will, however, be extremely complicated to do generally. One can make the identifications in the tetrad frame, as do Raychaudhuri and Saha and Tupper, but the necessity of transforming back to the canonical metric coordinate frame may well create more difficulty than performing all the calculations there to begin with.

One example agreed upon by Tupper and Raychaudhuri and Saha is the Kerr–Newman electrovac solution. They imply there is a physically acceptable viscous fluid which is a dual interpretation of the Kerr–Newman solution. In the next section we show that this solution is not possible.

III. KERR–NEWMAN SOLUTION

The Kerr–Newman metric is normally considered to represent a charged, rotating black hole. The metric form, in Boyer–Lindquist coordinates (t, r, θ, ϕ) with rotation in the ϕ direction, is

$$ds^2 = -\Delta\,(dt - a \sin^2\theta\, d\phi)^2/\tilde{\rho}^2 + \sin^2\theta\,(R^2\, d\phi - a\, dt)^2/\tilde{\rho}^2 + \tilde{\rho}^2\, dr^2/\Delta + \tilde{\rho}^2\, d\theta^2,$$

where $\Delta \equiv r^2 - 2mr + a^2 + Q^2$, $\tilde{\rho}^2 \equiv r^2 + a^2\cos^2\theta$, $R^2 \equiv r^2$

[a] Submitted in partial fulfillment of the degree of Doctor of Philosophy.

 0022-2488/85/051021-03$02.50

$+a^2$, Q is the charge, and $a \equiv s/m$ is the intrinsic angular momentum/unit mass.

The associated electromagnetic field has components

$$\begin{aligned} F_{10} &= (2)^{1/2}Q(r^2 - a^2\cos^2\theta)/\tilde{\rho}^4, \\ F_{13} &= -(2)^{1/2}Qa\sin^2\theta(r^2 - a^2\cos^2\theta)/\tilde{\rho}^4, \\ F_{20} &= -(2)^{1/2}Qra^2\cos\theta\sin\theta/\tilde{\rho}^4, \\ F_{23} &= (2)^{1/2}QR^2ra\cos\theta\sin\theta/\tilde{\rho}^4, \end{aligned} \tag{5}$$

and energy tensor

$$\begin{aligned} E_3^3 &= -E_0^0 = -\tfrac{1}{2}g^{00}(A) + \tfrac{1}{2}g^{33}(B) \\ &= Q^2(R^2 + a^2\sin^2\theta)/(2\tilde{\rho}^6), \\ E_1^1 &= -E_2^2 = \tfrac{1}{2}g^{00}(a_1 - a_2) + \tfrac{1}{2}g^{33}(b_1 - b_2) + g^{03}(c_2 - c_1) \\ &= -Q^2/(2\tilde{\rho}^4), \\ E_0^3 &= g^{03}(A) - g^{33}(C) = -Q^2a/\tilde{\rho}^6, \\ E_3^0 &= g^{03}(B) - g^{00}(C) = Q^2R^2a\sin^2\theta/\tilde{\rho}^6, \end{aligned} \tag{6}$$

where

$$\begin{aligned} A &= (a_1 + a_2) = g^{11}(F_{10})^2 + g^{22}(F_{02})^2, \\ B &= (b_1 + b_2) = g^{11}(F_{31})^2 + g^{22}(F_{23})^2, \\ C &= (c_1 + c_2) = g^{11}F_{10}F_{31} + g^{22}F_{02}F_{23}. \end{aligned}$$

To test the possibility of a dual or alternate interpretation of this metric, namely that it could also describe a viscous fluid, the fluid energy tensor $M_{\mu\rho}$ must be constructed within the Kerr–Newman geometry.

Equation (4) forces the general fluid velocity for this metric to be

$$u^\alpha = (u^0, u^1, 0, u^3).$$

Furthermore, from the symmetries of the metric we will assume that

$$u^\alpha = u^\alpha(r,\theta)$$

only, since all the other associated fields are t and ϕ independent. With the above assumptions the viscous fluid energy tensor components are

$$\begin{aligned} M_{11} &= \rho u_1u_1 + p^*P_{11} + \eta u_1(g_{00,1}u^0u^0 \\ &\quad + 2g_{03,1}u^0u^3 + g_{33,1}u^3u^3) + 2\eta\,\theta P_{11}/3 \\ &\quad - \eta(2u_{1,1} - g_{11,1}u^1)(1 + u_1u^1), \\ M_{22} &= p^*P_{22} + 2\eta\,\theta P_{22}/3 - \eta u^1 g_{22,1}, \\ M_{00} &= \rho u_0u_0 + p^*P_{00} + 2\eta\,\theta P_{00}/3 \\ &\quad - \eta u^1(2u_{01}u_0 + g_{00,1}), \\ M_{33} &= \rho u_3u_3 + p^*P_{33} + 2\eta\,\theta P_{33}/3 \\ &\quad - \eta u^1(2u_{3,1}u_3 + g_{33,1}), \\ M_{03} &= \rho u_0u_3 + p^*P_{03} + 2\eta\,\theta P_{03}/3 \\ &\quad - \eta u^1(u_{0,1}u_3 + u_{3,1}u_0 + g_{03,1}), \\ M_{01} &= (\rho + p^*)u_0u_1 + 2\eta\,\theta u_0u_1/3 - \eta[u_{0,1}(1 + u_1u^1) \\ &\quad + u_{1,1}u_0u^1] + \eta[g_{11,1}u_0u^1u^1 \\ &\quad + g_{00,1}u^0(2 + u_0u^0) + 2g_{03,1}u^3(1 + u_0u^0) \\ &\quad + g_{33,1}u_0u^3u^3]/2, \\ M_{31} &= (\rho + p^*)u_3u_1 + 2\eta\,\theta u_3u_0/3 - \eta[u_{3,1}(1 + u_1u^1) \\ &\quad + u_{1,1}u_3u^1] + \eta[g_{11,1}u_3u^1u^1 + g_{00,1}u_3u^0u^0 \\ &\quad + 2g_{03,1}u^0(1 + u_3u^3) + g_{33,1}u^3(2 + u_3u^3)]/2, \\ 2M_{20} &= -\eta(2u_{0,2} - g_{11,2}u_0u^1u^1) \\ &\quad + \eta[g_{00,2}u^0(2 + u_0u^0) \\ &\quad + 2g_{03,2}u^3(1 + u_0u^0) + g_{33,2}u_0u^3u^3], \\ 2M_{23} &= -\eta(2u_{3,2} - g_{11,2}u_3u^1u^1) + \eta[g_{00,2}u_3u^0u^0 \\ &\quad + 2g_{03,2}u^0(1 + u_3u^3) + g_{33,2}u^3(2 + u_3u^3)], \\ 2M_{21} &= \eta[2u_{1,2} - g_{11,2}u^1(2 + u_1u^1)] \\ &\quad + \eta u_1(g_{00,2}u^0u^0 + 2g_{03,2}u^0u^3 \\ &\quad + g_{33,2}u^3u^3). \end{aligned} \tag{7}$$

Setting $M_\alpha{}^\beta = E_\alpha{}^\beta$ in Eq. (6), we obtain, after some rewriting and use of repeated substitutions (see the Appendix), Eqs. (8) which follow. Related sets are denoted by Roman numerals.

$$\begin{array}{ll} \text{I} & \begin{cases} \rho = -E_1^1, \\ p^1 = -E_1^1 = \eta u^1 g_{22,1}\,g^{22}, \end{cases} \\ \text{II} & \begin{cases} 2\eta u^1(u_{0,1}u_0) = \rho^1u_0u_0 - g_{00}E_1^1 - E_{00} \\ \quad + \eta u^1(g_{22,1}\,g^{22}g_{00,1}), \\ 2\eta u^1[E_1,E\,]u_0u_0 = \eta u^1[E^2g_{00,1} - g_{33,1} \\ \quad + g_{22,1}\,g^{22}(g_{33} - E^2g_{00})] + E_1^1(E^2g_{00} - g_{33}) \\ \quad + E_0^0(E^2g_{00} + g_{33} - 2Eg_{03}), \end{cases} \\ \text{III} & 0 = \eta u^1u_0u_0[E^2g_{00,1} + g_{33,1} - 2Eg_{03,1} \\ & \quad + g_{22,1}\,g^{22}(2Eg_{03} - E^2g_{00} - g_{33})], \\ \text{IV} & \begin{cases} u_3 = Eu_0, E = E_3^0/(E_0^0 + E_1^1) = (E_1^1 - E_0^0)/E_0^3, \\ 0 = \eta u_0[g_{00,1}\,E(g^{00} + Eg^{03}) - g_{33,1}(g^{03} \\ \quad + Eg^{33}) + g_{03,1}(E^2g^{33} - g^{00})], \end{cases} \\ \text{V} & \begin{cases} u^1{}_{,2} = u^1(u_{1,2}u^1 + u_{0,2}u^0 + u_{3,2}u^3), \\ u^0{}_{,2} = u^0(u_{1,2}u^1 + u_{0,2}u^0 + u_{3,2}u^3) \quad \text{iff } \eta \neq 0, \\ u^3{}_{,2} = u^3(u_{1,2}u^1 + u_{0,2}u^0 + u_{3,2}u^3), \end{cases} \end{array} \tag{8}$$

where $p^1 = (p^* + 2\eta\,\tilde{\theta}/3)$, $\rho^1 = \rho + p^1$.

The above equations should be regarded as conditions necessary for a dual interpretation to be possible, since their solutions, establishing a correspondence between the viscous fluid and electromagnetic field elements, will determine whether or not such a dual interpretation is physically satisfactory. At this point, therefore, substitution of the explicit forms for $g_{\mu\beta}$ and $E_\mu{}^\nu$ must be made, to introduce the actual physical parameters.

Equations (8) V give us first that

$$u^1 = f(r)u^0, \tag{9a}$$

$$u^3 = h(r)u^0. \tag{9b}$$

From Eq. (8) IV,

$$u_3 = -a\sin^2\theta u_0. \tag{10}$$

Combining (9b) and (10) yields

$$h(r) = a/R^2.$$

Equation (8) III is satisfied as an identity, but the second equation of Eq. (8) II gives

$$4\eta u^1 r a^2 \sin^4\theta/\tilde{\rho}^2 = 0 \tag{11a}$$

and the second equation (8) IV

$$2\eta u_0 r a \sin^2\theta/\tilde{\rho}^2 = 0. \tag{11b}$$

Therefore, $\eta\, u_0 a = 0$, which makes $\eta = 0$ or $a = 0$.

Thus, a dual interpretation for the Kerr–Newman metric is not possible. The preceding argument fails if $a = 0$ so no criticism has been directed at the dual interpretation of the Reissner–Nordstrom metric.

ACKNOWLEDGMENT

We thank B. O. J. Tupper for his helpful revisions of an earlier draft of this paper.

APPENDIX: SOLUTION OF $M^{\alpha\beta} = E^{\alpha\beta}$

We provide a sketch of the solution of the equation $M^{\alpha\beta} = E^{\alpha\beta}$.

To determine the implications of a dual interpretation of the energy tensor, the relationship between the hypothetical viscous fluid quantities and the well-understood electromagnetic quantities must be found. That is, expressions must be obtained for u_0, u_1, u_3, ρ, p^*, η, and s which contain only electromagnetic quantities.

Examining Eqs. (7), note that the covariant components contain metric derivatives from the affine connection as well as derivatives of the fluid velocities and products of the velocities with the other fluid quantities. This is intuitively meaningful, as, for example, $\rho u_\alpha\, u_\beta =$ kinetic energy, but for the purpose of "solving" for the individual fluid quantities such products must be simplified or eliminated. To do this it is useful to combine the metric derivatives with the velocity derivatives to gain simpler expressions.

For example, since the fluid is normalized, i.e., $-1 = u_\alpha\, u^\alpha$, then

$$(u_1 u^1)_{,1} = -(u_0 u^0 + u_3 u^3)_{,1},$$

which gives

$$\begin{aligned} u^1(u_{1,1} - g_{11,1}u^1/2) &= +(g_{00,1}u^0u^0 + 2g_{03,1}u^0u^3 + g_{33,1}u^3u^3)/2 \\ &\quad - (u_{0,1}u^0 + u_{3,1}u^3). \end{aligned}$$

Rewriting (9) using the above identity, and redefining $p^1 = p^* + 2\eta\,\tilde{\theta}/3, \rho^1 = \rho + p^1$ gives the equivalence—with metric and fluid identities:

$$\mathrm{I}\ \left\{\begin{aligned} E_{11} &= \rho^1 u_1 u_1 + p^1 g_{11} - \eta(2u_{1,1} - g_{11,1}u^1) \\ &\quad + 2\eta u_1(u_{0,1}u^0 + u_{3,1}u^3), \\ E_{22} &= p^1 g_{22} - \eta u^1\, g_{22,1}, \end{aligned}\right.$$

$$\mathrm{II}\ \left\{\begin{aligned} E_{00} &= \rho^1 u_0 u_0 + p^1 g_{00} - \eta u^1(2u_{0,1}u_0 + g_{00,1}), \\ E_{33} &= \rho^1 u_3 u_3 + p^1 g_{33} - \eta u^1(2u_{3,1}u^3 + g_{33,1}), \end{aligned}\right.$$

$$\mathrm{III}\quad \begin{aligned} E_{03} &= \rho^1 u_0 u_3 + p^1 g_{03} - \eta u^1(u_{0,1}u_3 + u_{3,1}u_0 \\ &\quad + g_{03,1}), \end{aligned}$$

$$\mathrm{IV}\ \left\{\begin{aligned} 0 &= \rho^1 u_0 u_1 + \eta(g_{00,1}u^0 + g_{03,1}u^3) \\ &\quad + \eta u_{0,1}(u_0u^0 + u_3u^3) \\ &\quad + \eta u_0(u_{0,1}u^0 + u_{3,1}u^3), \\ 0 &= \rho^1 u_3 u_1 + \eta(g_{03,1}u^0 + g_{33,1}u^3) \\ &\quad + \eta u_{3,1}(u_0u^0 + u_3u^3) \\ &\quad + \eta u_3(u_{0,1}u^0 + u_{3,1}u^3), \end{aligned}\right.$$

$$\mathrm{V}\ \left\{\begin{aligned} 0 &= u_{0,2} - (g_{00,2}u^0 + g_{03,2}u^3) \\ &\quad - u_0(u_{1,2}u^1 + u_{0,2}u^0 + u_{3,2}u^3), \\ 0 &= u_{3,2} - (g_{03,2}u^0 + g_{33,2}u^3) \\ &\quad - u_3(u_{1,2}u^1 + u_{0,2}u^0 + u_{3,2}u^3), \\ 0 &= u_{1,2} - g_{11,2}u^1 - u_1(u_{1,2}u^1 + u_{0,2}u^0 + u_{3,2}u^3). \end{aligned}\right. \tag{A1}$$

In this form, the following can be seen.

Group I will reduce to a fairly simple form for (ρ^1) and (p^1) once the velocity derivative terms are determined.

Groups II and V are expressions for the velocity derivatives which presumably can be integrated for the velocities, provided the viscosity coefficients are known.

Groups III and IV can be regarded as eventual conditions on the viscosity coefficients.

To begin, note the following details.

(i) Groups I, III, and IV contain derivatives of both u_0 and u_3. Substitution from Eq. (A1) II will eliminate these terms.

(ii) The metric terms of Eq. (A1) IV can then be combined with the previous identity to eliminate the $u_{1,1}$ term from (A1) I leaving simple expressions for (ρ^1) and (p^1).

(iii) These expressions from (A1) I can then eliminate ρ^1 and p^1 from Eqs. (A1) II–IV.

(iv) Group V contains only derivatives with respect to θ of the velocity and metric. These can be rearranged using metric identies to produce derivative expressions for the velocity only, since

$$u_\beta = g_{\alpha\beta}\, u^\alpha, \quad u_{\beta,i} = g_{\alpha\beta,i}u^\alpha + g_{\alpha\beta}u^\alpha{}_{,i}.$$

(v) The second equation of Eq. (A1) IV can be replaced by subtracting from it [$E \times$ first equation of (A1) IV] and the results simplified. Collecting the results of (i)–(v) above becomes Eq. (8).

[1]B. O. J. Tupper, in *Abstracts of Presented Papers, 8th International Conference on General Relativity and Gravitation*, Canada, 7–12 August 1977 (University of Waterloo, Waterloo, Ontario, 1977).

[2]B. O. J. Tupper, J. Math. Phys. **22**, 2666 (1981).

[3]A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. **22**, 2237 (1981).

[4]A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. **23**, 2554 (1982).

[5]S. K. Saha, Pramana **20**, 245 (1983).

Symmetries of cosmological Cauchy horizons with exceptional orbits

James Isenberg[a)]
Department of Mathematics, University of Oregon, Eugene, Oregon 97403
Vincent Moncrief
Department of Physics and Department of Mathematics, Yale University, P. O. Box 6666, New Haven, Connecticut 06511

(Received 22 August 1984; accepted for publication 6 December 1984)

We show here that if an analytic space-time satisfying the vacuum (or electrovacuum) Einstein equations contains a compact null hypersurface with closed generators, then the space-time must have a nontrivial Killing vector field. This result is an extension of an earlier theorem, which required that the generators of the hypersurface must not only be closed, but in addition must satisfy a local product bundle (LPB) condition. This LPB condition (which is known to be violated in certain of the Kerr–Taub–NUT space-time models) is equivalent to the requirement that the generators must all be ordinary fibers of a Seifert fibration. We prove here that the LPB condition can be dropped. Thus we have stronger support for our conjecture that causality violations (as evidenced by Cauchy horizons) in cosmological solutions of the Einstein equations are essentially an artifact of symmetry, and are therefore nongeneric.

I. INTRODUCTION

In previous work,[1] we showed that if an analytic space-time satisfying the vacuum (or electrovacuum) Einstein equations contains a compact Cauchy horizon N, and if the generators of this horizon have closed orbits which satisfy a certain local product bundle (LPB) condition (described below), then the space-time must contain a Killing vector field. In that work (which we shall refer to as "Paper I") we noted that this result provides some support for the strong cosmic censorship conjecture,[2] since it indicates that space-times with Taub–NUT-like extensions into nonglobally hyperbolic regions are very nongeneric. Here, we make this support stronger by showing that our result is true even if we drop the LPB condition.

The key to this strengthened version of our theorem is the use of results from the theory of Seifert fibrations.[3,4] These results guarantee that if the orbits are all closed, then the LPB condition must in fact be satisfied everywhere on N except in neighborhoods of certain "exceptional orbits"; further, one can always unwrap a region of the space-time in such a way that the LPB condition is satisfied everywhere in the region. Our theorem in Paper I may then be applied to the unwrapped space-time regions, and we obtain a Killing field on the original space-time by projecting back.

It is not difficult to construct space-times which have Cauchy horizons containing exceptional (closed) orbits. The Kerr–Taub–NUT space-time models, for example, contain them if one makes an appropriate choice of the parameters which characterize these metrics.[5] We note that there are other choices of the parameters in the Kerr–Taub–NUT family for which some of the horizon generators have nonclosed orbits. These examples are of interest, since we hope to be able to extend our results further by dropping the closedness condition along with the LPB condition.

II. SEIFERT HORIZONS

The theorem which we will prove in this paper is as follows.

Theorem: Let $({}^4V, g)$ be an analytic space-time which (i) satisfies the vacuum or electrovacuum Einstein equations, and (ii) contains a compact, orientable null hypersurface[6] N with closed (i.e., S^1) generators. Then in some neighborhood of N in ${}^4V$, there is an analytic Killing vector field Y. This vector field has closed integral curves which, on N, are null and coincide with the generators of N. If a Maxwell field F is present then F (along with the metric g) is invariant with respect to Y.

The difference between this theorem and that which we proved in Paper I is the dropping of the local product bundle condition. Before we state exactly what this condition is, it is useful to recall some results from the theory of circle foliations of three-manifolds: We first note a theorem of Epstein[4] which states that if M is a compact orientable three-manifold (without boundary) and if λ is a smooth foliation of M by circles, then λ must be a Seifert fibration. The definition of a Seifert fibration[3] of a manifold M is based upon certain special foliations of the solid torus $S^1 \times D^2$. These special foliations are labeled by pairs of coprime integers μ and ν. One obtains the (μ,ν)-fibered solid torus (which we shall denote by $T_{[\mu,\nu]}$) by starting with the simple cylinder $[0,1]\times D^2$, and then rotating one end of this cylinder through an angle $2\pi(\nu/\mu)$ before identifying the two ends to obtain the solid torus (see Fig. 1). Using these $T_{[\mu,\nu]}$'s, one defines a Seifert fibration of M as a circle foliation which is "locally like a $T_{[\mu,\nu]}$." That is, for each of the circles γ, there exists a pair (W_γ, ρ_γ), where W_γ is an open set in M which consists of γ together with nearby complete circles of the foliation (a "fiber neighborhood of γ"), and where ρ_γ is a fiber-preserving diffeomorphism from W_γ to some fibered solid torus $T_{[\mu,\nu]}$ (see Fig. 2).

There are two important classes of these fibered solid tori: The *ordinary* ones have $\mu = 1$ so the rotation angle is an integer multiple of 2π, while for the *exceptional* ones this condition fails. This distinction is important because, for the ordinary ones, all of the circles of the foliation close after one circuit of the torus, while this is not true for the exceptional ones. For an exceptional $T_{[\mu,\nu]}$ the central fiber closes after one circuit, but all of the rest of the fibers require μ complete circuits before closing (see Fig. 3).

While the exceptional fibered solid tori might appear to

[a)]Portions of this research were done while visiting the Department of Mathematics, Rice University, Houston, Texas 77251.

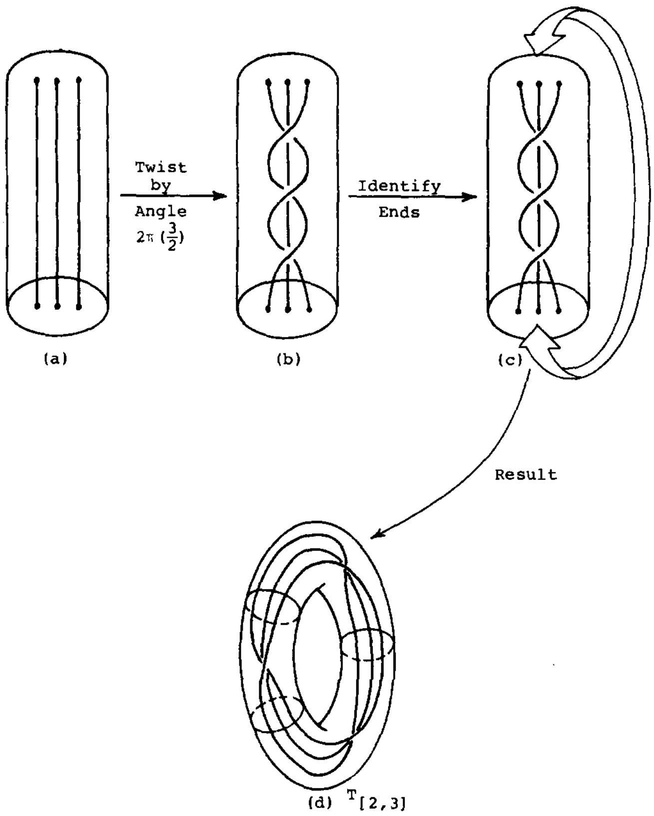


FIG. 1. Here we "construct" an example of a (μ,ν)-fibered solid torus—specifically, $T_{[2,3]}$. We start with $[0,1]\times D^2$, a solid cylinder. As shown in (a), we presume the cylinder to be foliated by segments parallel to the axis. We then twist the cylinder by holding one end fixed and rotating the other by the angle $2\pi(\frac{3}{2}) = 3\pi$; see (b). All the segments of the foliation, except the central one, now spiral about. Finally we identify the two ends of the cylinder—see (c)—and obtain the solid torus (d). Note that in this example, $T_{[2,3]}$, the central fiber's ends are joined to each other, while the ends of the noncentral fibers are joined from one fiber to another.

be the generic ones, Seifert[3] shows that for any Seifert fibration of any compact, connected, Hausdorff three-manifold M, there is only a *finite* subset of all of the orbits γ of the foliation for which ρ_γ maps W_γ to an exceptional $T_{[\mu,\nu]}$. These are called the *exceptional orbits* of the foliation. For all of the rest of the orbits, W_γ is diffeomorphic to an ordinary $T_{[\mu,\nu]}$.

All of this is relevant to our results on symmetries in space-times with compact null hypersurfaces because, in

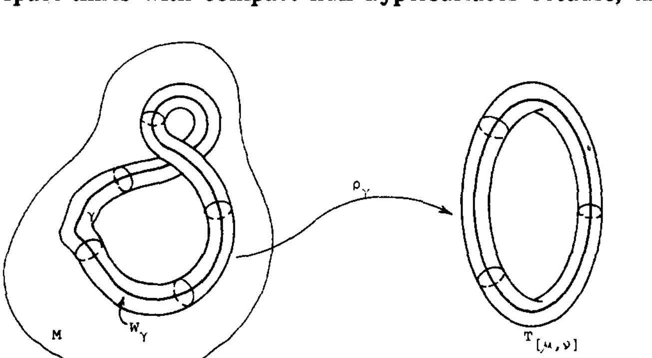


FIG. 2. M is a three-dimensional manifold (compact without boundary) which has a Seifert fibration. If γ is one of the fibers, then it follows from the definition of Seifert fibration that there is a neighborhood W_γ of γ, which is diffeomorphic to a standard fibered torus $T_{[\mu,\nu]}$ as shown. Here ρ_γ labels the diffeomorophism.

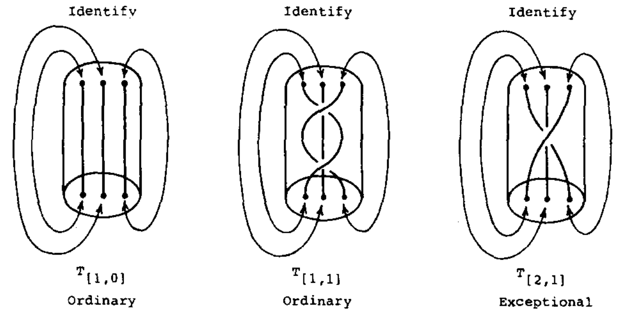


FIG. 3. Here we have three fibered tori, $T_{[1,0]}$, $T_{[1,1]}$, and $T_{[2,1]}$, drawn in their "preidentification" form for clarity. Two of these, $T_{[1,0]}$ and $T_{[1,1]}$, are called *ordinary* because after a single circuit of the torus, all fibers close (i.e., when ends are identified, *all* fibers have their two ends joined). The other, $T_{[2,1]}$, is called *exceptional* because it does not have this feature. Indeed, for $T_{[2,1]}$, one must make two circuits of the torus before completing an orbit of any but the central fiber.

proving the main theorem of Paper I, we relied upon coordinate systems which are well defined only if all neighboring generators of N close simultaneously. More precisely, that proof goes through only if the null generators of N constitute a Seifert fibration for which *all generators are ordinary*. This is the local product bundle condition which we required in Paper I.

To prove the existence of the Killing field Y without using the LPB condition (i.e., to prove the theorem of this paper), we first note that since we assume that the null generators of N have closed orbits the Epstein theorem guarantees that these generators are the fibers of a Seifert fibration of N. It then follows that, about each generator γ, one can choose a fibered neighborhood W_γ of the sort defined above. We can use these to build a patch covering for N; and since N is compact, we may choose a finite number of such patches W_γ to cover N. Among these patches, some may be exceptional—we denote them by W_{γ_e}. The rest will be ordinary—we denote those by W_{γ_o}.

Now the W_{γ_o} patches are exactly of the form of the "elementary regions" used in Paper I. Hence, using the construction outlined there (in Sec. III A) we can show that in a space-time neighborhood U_{γ_o} of each of the W_{γ_o} patches, there is a local Killing field Y of the desired form.

For the patches W_{γ_e} containing an exceptional orbit, the demonstration that a local Killing field exists requires a bit more work. We proceed as follows: We first use a field of null geodesics, transverse to the null hypersurface W_{γ_e}, to extend the fibration of W_{γ_e} so that it fills a space-time neighborhood of W_{γ_e}. We call this neighborhood U_{γ_e}, and note that topologically $U_{\gamma_e} = W_{\gamma_e}\times I$, where I is an open interval. Next, we construct the space-time region $\widetilde{U}_{\gamma_e}$ which is just the μ-fold cover of U_{γ_e} (where μ, we recall, is one of the Seifert indices). To make this concrete, we define a map

$$\phi : \widetilde{U}_{\gamma_e} \to U_{\gamma_e}, \tag{1}$$

which is μ-to-1, fiber preserving, and locally a diffeomorphism. The metric $\tilde{g}$ on $\widetilde{U}_{\gamma_e}$ is given by

$$\tilde{g} = \phi^* g, \tag{2}$$

so ϕ is also an isometry. (See Fig. 4.)

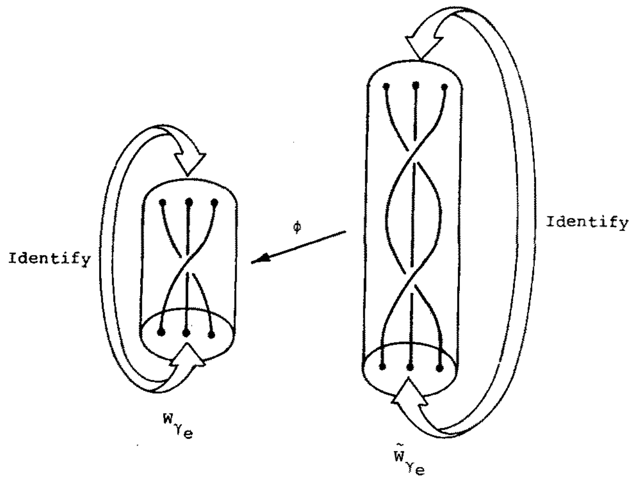


FIG. 4. In this figure, W_{γ_e} is an exceptional fibered torus of type $T_{[2,1]}$. However, the double cover $\tilde{W}_{\gamma_e}$ of W_{γ_e} is an ordinary fibered torus (of type $T_{[1,1]}$) as shown.

One easily verifies that, while the Seifert fibration in U_{γ_e} is exceptional, on $\tilde{U}_{\gamma_e}$ it is ordinary. (In this sense $\tilde{U}_{\gamma_e}$ is the "unwrapped" version of U_{γ_e}.) Hence, one can use the methods of Paper I to show that $\tilde{U}_{\gamma_e}$ contains a space-time Killing field $\tilde{Y}$ (see Ref. 7).

Of course, in general a vector field defined on a covering manifold does not project down to a well-defined vector field on the original manifold. In the present case, however, we can show that $\tilde{Y}$ does project down. Our argument consists of three parts: We first consider $\tilde{Y}$ restricted to the μ-fold cover $\tilde{\gamma}_e$ of the exceptional fiber γ_e, and show analytically that $\tilde{Y}|_{\tilde{\gamma}_e}$ is μ-fold periodic and projects down to γ_e. We then look at the ordinary fibers γ_o in a neighborhood of γ_e in N, show that the Killing fields defined about them agree with $\tilde{Y}$ in $\tilde{W}_{\gamma_e} - \tilde{\gamma}_e$, and thereby verify that $\tilde{Y}$ projects down properly in $W_{\gamma_e} \subset N$. Finally, we verify that the projection works in the space-time neighborhood U_{γ_e} of W_{γ_e}.

The argument that $\tilde{Y}|_{\tilde{\gamma}_e}$ is μ-fold periodic is most easily carried out in terms of the coordinates used in Sec. III A of Paper I (as adapted for $\tilde{U}_{\gamma_e}$). We work entirely in the null hypersurface region $\tilde{W}_{\gamma_e}$, on which we have the coordinates (x^3, x^a), for $a = 1$ or 2. The coordinates x^a are constant along the fibers, and for convenience we choose $x^a = 0$ on $\tilde{\gamma}_e$. The vector field $\partial/\partial x^3$ thus lies tangent to the fibers.

In Paper I, we found an analytic expression for the local Killing vector field in terms of a function ψ (called "$\mathring{\phi}_{,t}$" in Paper I) which depends upon the geometry of the space-time. Explicitly we get

$$\tilde{Y}|_{\tilde{W}_{\gamma_e}} = u^{-1}(x^3, x^a)\frac{\partial}{\partial x^3}, \tag{3}$$

where in the degenerate ($k = 0$) (see Ref. 8) case we have

$$u(x^3, x^a) = \frac{2\pi p(x^3, x^a)}{\int_0^{2\pi} ds\, p(s, x^a)}, \tag{4}$$

while in the nondegenerate ($k \neq 0$) case we have

$$u(x^3, x^a) = 2p(x^3, x^a)/kD(x^3, x^a), \tag{5}$$

with

$$p(x^3, x^a) := \exp\Big[-\int_0^{x^3} ds\, \frac{\psi}{2}(s, x^a)\Big] \tag{6}$$

and

$$D(x^3, x^a) := \frac{1}{1 - e^{-\pi k}}\int_0^{2\pi} ds\, p(s, x^a) - \int_0^{x^3} ds\, p(s, x^a). \tag{7}$$

Now the coordinate x^3 ranges from 0 to 2π in $\tilde{U}_{\gamma_e}$. Hence in Paper I it was necessary to verify that $u(x^3 + 2\pi, x^a) = u(x^3, x^a)$ so that $\tilde{Y}$ is well defined in $\tilde{U}_{\gamma_e}$. Here, we must show further that, at least for $x^a = 0$, we get $u(x^3 + 2\pi/\mu, 0) = u(x^3, 0)$, so that $\tilde{Y}$ is μ-fold periodic on $\tilde{\gamma}_e$.

What makes this work is the μ-fold periodicity of ψ on $\tilde{\gamma}_e$,

$$\psi(x^3 + 2\pi/\mu, 0) = \psi(x^3, 0), \tag{8}$$

which is a consequence of the construction of $\tilde{W}_{\gamma_e}$ as a μ-fold cover of W_{γ_e} with $\phi|_{\tilde{\gamma}_e}: \tilde{\gamma}_e \to \gamma_e$ being a μ-to-1 map. Using this periodicity, together with the definition $2\pi k = \int_0^{2\pi} ds\, \psi(s, x^a)$, we calculate from Eq. (6) the following result:

$$\begin{aligned} p\Big(x^3 + \frac{2\pi}{\mu}, 0\Big) &= \exp\Big[-\int_0^{2(\pi/\mu) + x^3} ds\, \frac{\psi}{2}(s, 0)\Big] \\ &= \exp\Big[-\int_0^{2\pi/\mu} ds\, \frac{\psi}{2}(s, 0) \\ &\quad - \int_{2\pi/\mu}^{2\pi/\mu + x^3} ds\, \frac{\psi}{2}(s, 0)\Big] \\ &= \exp(-\pi k/\mu)\, p(x^3, 0). \end{aligned} \tag{9}$$

A similar calculation shows that

$$D(x^3 + 2\pi/\mu, 0) = \exp(-\pi k/\mu) D(x^3, 0). \tag{10}$$

These two results show that in both the degenerate ($k = 0$) and nondegenerate ($k \neq 0$) cases, we have

$$u(x^3 + 2\pi/\mu, 0) = u(x^3, 0). \tag{11}$$

Therefore, since the points on $\tilde{\gamma}_e$ that are projected down to the same point on γ_e by ϕ are exactly those for which the x^3 coordinates differ by multiples of $2\pi/\mu$, we find that $\phi_* \tilde{Y}|_{\tilde{\gamma}_e}$ is a well-defined vector field on γ_e.

One might try to apply the same argument to $\tilde{Y}$ away from $\tilde{\gamma}_e$, since expressions (3)–(7) are true for any value of x^a (not just $x^a = 0$). However the fibers neighboring $\tilde{\gamma}_e$ in $\tilde{W}_{\gamma_e}$ are *not* μ-fold covers of fibers of the foliation of N. Hence $\psi(x^3 + 2\pi/\mu, x^a) \neq \psi(x^3, x^a)$ for $x^a \neq 0$, and the above argument fails away from $\tilde{\gamma}_e$.

The key to the second part of our argument that $\tilde{Y}$ projects down is that, rather than these neighboring fibers in $\tilde{W}_{\gamma_e}$ being μ-fold covers of closed fibers in N, they are *single* covers of those closed fibers. This follows from the fibration topology of the Seifert exceptional solid tori and the construction of the covering projection map ϕ (see Fig. 4). Hence if we consider an arbitrary ordinary fiber $\gamma_{\hat{o}}$ in W_{γ_e}, then we can always find a fiber neighborhood $W_{\gamma_{\hat{o}}}$ of $\gamma_{\hat{o}}$ such that $\phi|_{W_{\gamma_{\hat{o}}}}$ is a diffeomorphism on $W_{\gamma_{\hat{o}}}$.

Now using the standard argument of Sec. III A in Paper I, we construct a Killing field $Y_{\hat{o}}$ in a space-time neighborhood of $\phi^{-1}(W_{\gamma_{\hat{o}}}) \subset \tilde{W}_{\gamma_e}$. This field projects down (since

$\phi|_{\phi^{-1}(W_{\gamma_\delta})}$ is a diffeomorphism). Let us compare Y_δ with $\widetilde{Y}$, which is the local Killing field in $\widetilde{U}_{\gamma_e}$ and hence in $\widetilde{W}_{\gamma_e}$. The arguments of Sec. III B of Paper I may be directly applied to show that in fact $\widetilde{Y}$ and Y_δ must be the same vector field on $\phi^{-1}(W_{\gamma_\delta}) \subset \widetilde{W}_{\gamma_e}$. Hence in $\phi^{-1}(W_{\gamma_\delta})$, $\widetilde{Y}(=Y_\delta)$ projects down.

Every point in $\widetilde{W}_{\gamma_e}$ lies either on $\tilde{\gamma}_e$ or on some fiber γ_δ. In both cases the local Killing vector field $\widetilde{Y}$ projects down. Hence $\phi|_{\widetilde{W}_{\gamma_e}}$ maps $\widetilde{Y}$ to a well-defined vector field on W_{γ_e}.

To complete our argument that there is a well-defined Killing vector field on U_{γ_e}, we note that in each contractible open set $S \subset U_{\gamma_e}$ for which $S \cap W_{\gamma_e}$ is not empty, we have a set of μ possibly inequivalent local Killing fields, all of which agree on $S \cap W_{\gamma_e}$. But since $S \cap W_{\gamma_e}$ is a codimension-one hypersurface in S, the uniqueness theorems for Killing vector fields require that these all agree throughout S. Hence $Y = \phi_*(\widetilde{Y})$ is a well-defined Killing vector field everywhere in S, and throughout U_{γ_e} as well.

We now have a Killing vector field defined on each of the patches U_{γ_e} and U_{γ_0} which cover a space-time neighborhood of the compact null hypersurface N. Since the Killing vector fields agree on patch overlaps (see the arguments of Sec. III B in Paper I), we have a well-defined Killing vector field Y in a space-time neighborhood of N. This completes the proof of our theorem.

III. CONCLUSION

It is a direct consequence of our new theorem that if an analytic space-time satisfying the vacuum (or electrovacuum) Einstein equations contains a compact Cauchy horizon with closed generators, then the space-time contains a symmetry. The result is true whether or not some of the horizon generators are exceptional. Thus we see that in a wider class of space-times than we considered in Paper I, the existence of a complete Cauchy horizon is a very unstable (i.e., nongeneric) condition. This supports the conjecture of strong cosmic censorship, which claims that in the class of *all* space-times, the existence of any Cauchy horizon is an unstable property.

Clearly we would have even stronger support for strong cosmic censorship if we could drop the condition that the generator orbits be closed. This possibility is the focus of our present research.

ACKNOWLEDGMENTS

We are thankful to M. Anderson, R. Bryant, T. Duchamp, W. Dunbar, and J. Wolfson for helpful discussions.

One of us (JI) was partially supported by National Science Foundation (NSF) Grant No. MCS 83-0-3998, and the other (VM) was partially supported by NSF Grant No. PHY 82-0-4995.

[1]V. Moncrief and J. Isenberg, Commun. Math. Phys. **89**, 387 (1983).

[2]See, for example, the articles by R. Penrose and by R. Geroch and G. Horowitz, in *General Relativity: An Einstein Centenary Survey*, edited by S. Hawking and W. Israel (Cambridge U.P., London, 1979).

[3]H. Seifert, Acta. Math. **60**, 147 (1933).

[4]D. Epstein, Ann. Math. **95**, 66 (1972).

[5]J. G. Miller, J. Math. Phys. **14**, 486 (1973).

[6]Recall that, as in paper I, our results refer to "compact null hypersurfaces" rather than to "compact Cauchy horizons." However, since every compact Cauchy horizon is in fact a compact null hypersurface, our results pertain to compact Cauchy horizons as a special case.

[7]Note that the existence proof does not guarantee that $\widetilde{Y}$ will fill $\widetilde{U}_{\gamma_e}$. It *does*, however, guarantee that $\widetilde{Y}$ fills an open subset of $\widetilde{U}_{\gamma_e}$, and that this open subset contains the inverse image of W_{γ_e} (under the map ϕ). Let us, for the sake of notational simplicity, use the notation $\widetilde{U}_{\gamma_e}$ to refer to this open subset.

[8]As in paper I, "k" is defined as $2\pi k = \int_0^{2\pi} ds\, \psi(s,x^a)$. This quantity is constant on $\widetilde{W}_{\gamma_e}$ (independent of x^a). If $k \neq 0$, N is the boundary of a globally hyperbolic space-time region (and hence N is a Cauchy horizon); while if $k = 0$, N lies in a space-time region which may not be globally hyperbolic.

Relativistic superfluid in a rigid solid. Exact solutions

G. Papini and M. Weiss
Department of Physics and Astronomy, University of Regina, Saskatchewan S4S 0A2, Canada

(Received 16 October 1984; accepted for publication 28 December 1984)

Exact solutions of Einstein's equations are obtained for a superfluid flowing in a rigid solid.

I. INTRODUCTION

The purpose of this work is to present some exact solutions of Einstein's equations for a neutral superfluid free to flow in a rigid solid. The system, reminiscent of a neutron superfluid in the crust and interior of a neutron star,[1] is studied here in a conformally flat metric and special coordinates.[2] Though the latter choice makes the solutions found impractical for astrophysical applications, the relativistic "material" so constructed is nevertheless interesting as it fully complies with Einstein's equations and offers some insight into the nature of the problem. Relativistic materials are also of interest in atomic gauge problems[3] of conformally invariant theories of the Weyl–Dirac type.[4]

Specifically, it is assumed that the superfluid is neutral and does not interact with the solid, which in turn satisfies Born's rigidity condition[5]

$$\widetilde{U}_{\alpha;\beta} + \widetilde{U}_{\beta;\alpha} - \widetilde{U}_\alpha \widetilde{U}_{\beta;\sigma} \widetilde{U}^\sigma - \widetilde{U}_\beta \widetilde{U}_{\alpha;\sigma} \widetilde{U}^\sigma = 0, \tag{1}$$

where $\widetilde{U}_\alpha$ denotes the four-velocity of the solid and

$$\widetilde{U}_\alpha \widetilde{U}^\alpha = 1. \tag{2}$$

Einstein's equations then become

$$\widetilde{R}_{\mu\nu} - \frac{1}{2}\tilde{g}_{\mu\nu}\widetilde{R} + \Lambda \tilde{g}_{\mu\nu} = -8\pi\{\tilde{\mu}\tilde{n}\tilde{u}_\mu \tilde{u}_\nu + \tilde{\rho}^s \widetilde{U}_\mu \widetilde{U}_\nu + \widetilde{P}_{\mu\nu}\}, \tag{3}$$

where $\tilde{\mu} = \partial\tilde{\rho}^f/\partial\tilde{n}$, $\tilde{n}$, and $\tilde{u}_\mu$ are chemical potential, particle number density, and four-velocity of the fluid, respectively, while $\tilde{\rho}^s$ and $\widetilde{P}_{\mu\nu}$ are the energy density and the pressure tensor of the solid. In (3) the pressure of the fluid has been neglected relative to that of the solid and the solid energy-momentum tensor is the one prescribed by Carter and Quintana[6] under the condition that there be no energy transport in the solid rest frame. This also requires

$$\widetilde{P}_{\mu\nu}\widetilde{U}^\nu = 0. \tag{4}$$

It is assumed that under the conformal transformation

$$\tilde{g}_{\mu\nu}(x) = \beta^2(x)\eta_{\mu\nu}, \tag{5}$$

$\widetilde{P}_{\mu\nu}$ transforms according to

$$\widetilde{P}_{\mu\nu} = \beta^{-2}P_{\mu\nu},$$

as may be inferred from the isotropic form of $\widetilde{P}_{\mu\nu}$,

$$\widetilde{P}_{\mu\nu} = -\widetilde{P}\tilde{g}_{\mu\nu} + \widetilde{P}\widetilde{U}_\mu \widetilde{U}_\nu, \tag{6}$$

and the corresponding transformation for the pressure $\widetilde{P}$.

The superfluid obeys the equations[7]

$$(\tilde{\mu}\tilde{u}_\alpha)_{;\sigma} - (\tilde{\mu}\tilde{u}_\sigma)_{;\alpha} = 0, \tag{7}$$

$$(\tilde{n}\tilde{u}^\alpha)_{;\alpha} = 0, \tag{8}$$

$$\tilde{\mu}\tilde{n} = \tilde{\rho}^f + \tilde{p}, \tag{9}$$

$$\tilde{u}_\alpha \tilde{u}^\alpha = 1. \tag{10}$$

In terms of the variables[2]

$$s = (x_\mu x^\mu)^{1/2}, \quad v = x^0/s$$

and of the associated vectors

$$s_\mu = \frac{\partial s}{\partial x^\mu}, \quad v_\mu = \frac{\partial v}{\partial x^\mu},$$

with

$$s_\mu s^\mu = 1, \quad v_\mu v^\mu = (1 - v^2)/s^2, \quad s_\mu v^\mu = 0,$$

the four-velocity U_μ in the rest frame of the solid is

$$U_\mu = \delta^0_\mu = vs_\mu + sv_\mu. \tag{11}$$

The dependence of all other field quantities is also restricted to s and v. Thus the most general form of $P_{\mu\nu}$ is

$$P_{\mu\nu} = h(s, v)s_\mu s_\nu + l(s, v)v_\mu v_\nu + k(s, v)(s_\mu v_\nu + v_\mu s_\nu) + w(s, v)\eta_{\mu\nu}. \tag{12}$$

Two of the unknown functions h, l, k, and w can be determined from Eq. (4). They are

$$\begin{aligned} h(s, v) &= -[w + k(1 - v^2)/sv], \\ l(s, v) &= -[(ws^2 + kvs)/(1 - v^2)]. \end{aligned} \tag{13}$$

In the rest frame of the solid Eq. (1) reduces to

$$\beta' v + \dot{\beta}[(1 - v^2)/s] = 0$$

and has the solution

$$\beta = s^\lambda (1 - v^2)^{\lambda/2}, \tag{14}$$

where λ is an arbitrary constant. Similarly, one writes for the fluid

$$u_\mu = f(s,v)s_\mu + g(s,v)v_\mu, \tag{15}$$

where $f(s,v)$ and $g(s,v)$ are as yet unknown functions. Equations (7) and (8) then become

$$\mu' g + \mu g' - \dot{\mu} f - \mu \dot{f} = 0, \tag{16}$$

$$n' f + \dot{n} g \frac{(1 - v^2)}{s^2} + n\left[f' + \frac{3f}{s} + \dot{g}\frac{(1 - v^2)}{s^2} - \frac{3gv}{s^2}\right] = 0, \tag{17}$$

where $' = \partial/\partial s$ and $\cdot = \partial/\partial v$.

Using (11), (13), (15), and (5) Eqs. (3) yield

$$\frac{4\beta'^2}{\beta^2} - \frac{2\beta''}{\beta} + \frac{2\beta'}{\beta s} - \frac{2\dot{\beta}v}{\beta s^2} = \frac{8\pi}{\beta^2}\left\{\mu n f^2 + \rho^s v^2 - \frac{k(1 - v^2)}{sv} - w\right\}, \tag{18}$$

$$\frac{4\dot{\beta}\beta'}{\beta^2} - \frac{2\dot{\beta}'}{\beta} + \frac{2\dot{\beta}}{\beta s} = \frac{8\pi}{\beta^2}\{\mu n f g + \rho^s v s + k\}, \qquad (19)$$

$$\frac{4\dot{\beta}^2}{\beta^2} - \frac{2\ddot{\beta}}{\beta} = \frac{8\pi}{\beta^2}\left\{\mu n g^2 + \rho^s s^2 - \frac{kvs + ws^2}{1-v^2}\right\}, \qquad (20)$$

$$\frac{4\beta'}{\beta s} - \frac{4\beta v}{\beta s^2} - \frac{\beta'^2}{\beta^2} - \frac{\dot{\beta}^2(1-v^2)}{\beta^2 s^2} + \frac{2\beta''}{\beta} + \frac{2\ddot{\beta}(1-v^2)}{\beta s^2} - \Lambda\beta^2 = \frac{8\pi w}{\beta^2}, \qquad (21)$$

where β is now given by (16) while (10) becomes

$$f^2 + g^2[(1-v^2)/s^2] = 1. \qquad (22)$$

II. THE SOLUTIONS

From (14), (18), and (22) one gets

$$f^4 - f^2 = v^4 - v^2. \qquad (23)$$

Equation (23) has the solutions

$$\text{(a) } f = \pm v,\ g = \pm s,\ \text{or } u_\mu = \delta^0_\mu, \qquad (24)$$

which correspond to the rest frame for both solid and superfluid, and

$$\text{(b) } f = \pm(1-v^2)^{1/2},\ g = \pm vs/(1-v^2)^{1/2}. \qquad (25)$$

For case (a) one obtains

$$\begin{aligned} 8\pi k &= s^{2\lambda-1}v(1-v^2)^{\lambda-1}(-3\lambda^2 - 4\lambda) - \Lambda s^{4\lambda+1}v(1-v^2)^{2\lambda}, \\ 8\pi w &= \lambda^2 s^{2\lambda-2}(1-v^2)^{\lambda-1} + \Lambda s^{4\lambda}(1-v^2)^{2\lambda}, \\ 8\pi\rho^s &= \lambda(\lambda+2)s^{2\lambda-2}(1-v^2)^{\lambda-1} + \Lambda s^{4\lambda}(1-v^2)^{2\lambda} - 8\pi\mu n, \end{aligned} \qquad (26)$$

while Eqs. (16) and (17) yield, respectively,

$$\mu = A s^\sigma v^\sigma \qquad (27)$$

and

$$n = \beta s^\delta(1-v^2)^{\delta/2}, \qquad (28)$$

where A, B, δ, σ are constants.

For case (b) one finds

$$\begin{aligned} 8\pi k &= \lambda(-3\lambda - 4)s^{2\lambda-1}v(1-v^2)^{\lambda-1} - \Lambda s^{4\lambda+1}v(1-v^2)^{2\lambda} + 8\pi\mu n\, vs, \\ 8\pi w &= \lambda^2 s^{2\lambda-2}(1-v^2)^{\lambda-1} + \Lambda s^{4\lambda}(1-v^2)^{2\lambda}, \end{aligned} \qquad (29)$$

$$8\pi\rho^s = \lambda(\lambda+2)s^{2\lambda-2}(1-v^2)^{\lambda-1} + \Lambda s^{4\lambda}(1-v^2)^{2\lambda},$$

$$\mu = A s^\sigma(1-v^2)^{-\sigma/2-1}, \qquad (30)$$

$$n = C s^{-\delta}v^{-4+\delta}(1-v^2)^{-1}, \qquad (31)$$

where C is a constant.

The conditions for the solid to be isotropic can be derived by comparing (6) with (12) and (13):

$$P = -w, \qquad (32)$$

$$k = -wsv. \qquad (33)$$

In case (a) the solid becomes isotropic for $\lambda = 0, -2$. For case (b), Eq. (33) is satisfied if $\lambda = -1, \delta = 4, \sigma = 0$, and $AC = (4\pi)^{-1}$. It is then easy to calculate P from (32) for both solutions. Anisotropy is recovered in both cases for any other value of λ.

Finally, it is worth noticing that while solid rigidity and motion determine β to within an arbitrary constant, thus affecting the value of n; the properties and motion of the superfluid in turn affect the structure of the solid functions w and k through μ, f, and g, even though solid and fluid are formally noninteracting.

ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

[1]G. Baym and C. Pethick, Ann. Rev. Nucl. Sci. **25**, 27 (1975).

[2]G. E. Tauber, J. Math. Phys. **8**, 118 (1967); Gen. Relativ. Gravit. **3**, 17 (1972).

[3]G. Papini, Phys. Lett. A **75**, 173 (1980); A **91**, 105 (1982); Nuovo Cimento B **70**, 113 (1982).

[4]P. A. M. Dirac, Proc. R. Soc. London Ser. A **333**, 403 (1973); V. Canuto, P. J. Adams, S-H. Hsieh, and E. Tsiang, Phys. Rev. D **16**, 1643 (1977); D. Gregorash and G. Papini, Phys. Lett. A **82**, 67 (1981); Nuovo Cimento B **63**, 487 (1981) **64**, 55 (1981).

[5]M. Born, Ann. Phys. **30**, 1 (1909); A. Trautman, in *General Relativity*, edited by S. Deser and K. W. Ford (Prentice-Hall, Englewood Cliffs, NJ, 1965).

[6]B. Carter and H. Quintana, Proc. R. Soc. London Ser. A **331**, 57 (1972).

[7]F. Rothen, Helv. Phys. Acta **41**, 591 (1968); W. Israel, Phys. Lett. A **86**, 79 (1981); G. Papini and M. Weiss, Phys. Lett. A **89**, 329 (1982).

The relationship between monopole harmonics and spin-weighted spherical harmonics

Tevian Dray
Instituut voor Theoretische Fysica, Princetonplein 5, Postbus 80.006, 3508 TA Utrecht, The Netherlands[a] *and School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540*

(Received 1 October 1984; accepted for publication 28 December 1984)

We compare two independent generalizations of the usual spherical harmonics, namely monopole harmonics and spin-weighted spherical harmonics, and make precise the sense in which they can be considered to be the same. By analogy with the spin-gauge language, raising and lowering operators for the monopole index of the monopole harmonics can immediately be written down.

I. INTRODUCTION

Once again physicists in two completely different areas have independently developed the same mathematics. Wu and Yang[1] introduced[2] monopole harmonics as particular solutions of the Schrödinger equation for an electron in the field of a Dirac magnetic monopole. Newman and Penrose[3] introduced[2] spin-weighted spherical harmonics as a means to describe certain quantities exhibiting a particular "spin-gauge" behavior which occur naturally in the asymptotic expansion of the gravitational field in null directions.

In what follows we compare these two generalizations of the usual spherical harmonics and show that, for a particular choice of spin gauge, the spin-weighted spherical harmonics reduce to the monopole harmonics. As a simple application of this result, we note that the fundamental operators in the spin-gauge language raise or lower the spin weight by 1. Thus, writing these operators in the appropriate gauge immediately yields operators which raise or lower the monopole index of the monopole harmonics by 1. Going in the other direction, we adapt the angular momentum operators of the Schrödinger picture to the spin-gauge language and derive the corresponding operators there.

In Sec. II we first review monopole harmonics and in Sec. III we do the same for spin-weighted spherical harmonics. We compare the two in Sec. IV and then discuss our results in Sec. V.

II. MONOPOLE HARMONICS

The term "monopole harmonics" was first used by Wu and Yang[1] to describe solutions of the Schrödinger equation for an electron in the field of a magnetic monopole. However, the functions used in this description are almost as old as the relevant Schrödinger equation itself, which dates back to the original paper on monopoles by Dirac.[4]

These functions were first discussed by Tamm[5] and Fierz[6] and then by numerous other authors.[7]

The fundamental difference in the approach of Wu and Yang[1] is that elements of their Hilbert space are not functions at all, but rather sections of a particular fiber bundle. This eliminates the string singularity of the original description of the Dirac monopole. Although the presentation below follows Wu and Yang[1] we will deliberately deemphasize the underlying fiber bundle structure.

Define the regions R_a and R_b on the sphere by

$$R_a = \{0 \leqslant \theta < \pi\}, \quad R_b = \{0 < \theta \leqslant \pi\}. \tag{1}$$

The relevant Schrödinger equation is

$$[-(1/r^2)\partial_r(r^2\partial_r) + (1/r^2)[L^2 - q^2] + V - E]\psi = 0, \tag{2}$$

where $V(r)$ is the potential, E is the energy eigenvalue, L^2 is the total angular momentum operator, and $q = eg$ (see Ref. 8).

One makes the ansatz

$$\psi(r,\theta,\varphi) = R(r)Y_{qlm}(\theta,\varphi), \tag{3}$$

where the Y_{qlm} are characterized by their angular momentum eigenvalues

$$L^2 Y_{qlm} = l(l+1)Y_{qlm}, \quad L_z Y_{qlm} = mY_{qlm}. \tag{4a}$$

We also have

$$L_\pm Y_{qlm} = [(l \mp m)(l + 1 \mp m)]^{1/2} Y_{qlm\pm 1}. \tag{4b}$$

The fiber bundle structure can be interpreted as follows: The angular momentum operators take different forms in regions R^a and R^b, leading to different functions Y^a_{qlm} and Y^b_{qlm} which *together* make up a monopole harmonic Y_{qlm}. In this paper, however, we will only be concerned with the *functions* Y^a_{qlm} and Y^b_{qlm}.

The angular momentum operators are

$$L^a_z = -i\partial_\varphi - q, \tag{5a}$$

$$L^a_\pm = e^{\pm i\varphi}\left(\pm\partial_\theta + \frac{i\cos\theta}{\sin\theta}\partial_\varphi - \frac{q(1-\cos\theta)}{\sin\theta}\right), \tag{5b}$$

$$\begin{aligned}(L^2)^a &= -\Delta + \frac{2iq}{\sin^2\theta}(1-\cos\theta)\partial_\varphi + \frac{2q^2}{\sin^2\theta}(1-\cos\theta)\\ &= -\Delta + \frac{2q}{\sin^2\theta}(\cos\theta - 1)L^a_z,\end{aligned} \tag{5c}$$

$$L^b_z = -i\partial_\varphi + q, \tag{5d}$$

$$L^b_\pm = e^{\pm i\varphi}\left(\pm\partial_\theta + \frac{i\cos\theta}{\sin\theta}\partial_\varphi - \frac{q(1+\cos\theta)}{\sin\theta}\right), \tag{5e}$$

$$\begin{aligned}(L^2)^b &= -\Delta - \frac{2iq}{\sin^2\theta}(1+\cos\theta)\partial_\varphi + \frac{2q^2}{\sin^2\theta}(1+\cos\theta)\\ &= -\Delta + \frac{2q}{\sin^2\theta}(\cos\theta + 1)L^b_z,\end{aligned} \tag{5f}$$

[a] Permanent address.

 0022-2488/85/051030-04$02.50

where

$$\Delta = \partial_\theta^2 + \frac{\cos\theta}{\sin\theta}\partial_\theta + \frac{1}{\sin^2\theta}\partial_\varphi^2$$

is the Laplace operator on the two-sphere.

With appropriate normalization the Y_{qlm} satisfy

$$\int_s Y_{qlm}\overline{Y}_{ql'm'}\,dS = \delta_{ll'}\delta_{mm'}, \tag{6a}$$

where the integral is over the full two-sphere; we note that the integrand is the same in the regions R^a and R^b. We also have

$$Y_{0lm} = Y_{lm}, \tag{6b}$$

$$Y_{qlm} = 0, \quad \text{for } l<|q|, \tag{6c}$$

where the Y_{lm} denote the usual spherical harmonics. Finally, we note that $\{Y_{qlm}\}$ for given q is complete in the following sense: Given any section $f=(f^a,f^b)$, where f^a and f^b are functions on R^a and R^b, respectively, satisfying $f^a = e^{2qi\varphi}f^b$, then f can be expanded as a linear combination of the Y_{qlm}.

III. SPIN-WEIGHTED SPHERICAL HARMONICS

Newman and Penrose[3] introduced spin-weighted spherical harmonics based on ideas in Janis and Newman[9] in order to describe the asymptotic behavior of the gravitational field of isolated systems at large null distances from the source. Although they did this for a particular choice of spin gauge (the "standard" spin gauge) the concept can be immediately generalized to an arbitrary spin gauge. Except for this minor difference our presentation follows Newman and Penrose.[3]

Consider a two-sphere with the usual metric

$$g_{ab}\,dx^a\,dx^b = d\theta^2 + \sin^2\theta\,d\varphi^2. \tag{7}$$

Instead of the usual orthonormal basis $[\partial_\theta,(1/\sin\theta)\partial_\varphi]$, we introduce a complex null basis $(m^a,\overline{m}^a)$ via

$$g_{ab}m^am^b = 0, \quad g_{ab}m^a\overline{m}^b = 2, \tag{8}$$

where the bar denotes complex conjugation. The general m^a can thus be written

$$m^a(\partial_x)_a = e^{i\gamma}[\partial_\theta + (i/\sin\theta)\partial_\varphi]. \tag{9}$$

The choice of the function $\gamma(\theta,\varphi)$ will be called the choice of a spin gauge. We are thus led to consider transformations of the form

$$m^a \mapsto e^{i\Lambda}m^a. \tag{10}$$

A quantity Q whose behavior under this gauge transformation is

$$Q \mapsto e^{is\Lambda}Q \tag{11}$$

is said to have spin weight s [$\mathrm{sw}(Q)=s$]. The simplest example of this is

$$\mathrm{sw}(m^a) = +1, \quad \mathrm{sw}(\overline{m}^a) = -1. \tag{12}$$

Note that not all quantities have a well-defined spin weight. An example of this is

$$2\alpha = -\tfrac{1}{2}m^a\overline{m}^b\nabla_b\overline{m}_a, \tag{13}$$

where ∇_a denotes covariant differentiation on the two-sphere, which transforms under (10) as

$$2\alpha \mapsto e^{-i\Lambda}(2\alpha + i\overline{m}^a\partial_a\Lambda). \tag{14}$$

We can, however, combine m^a and $\overline{\alpha}$ into operators which raise or lower the spin weight. For $\mathrm{sw}(Q)=s$ define[10]

$$ðQ = m^a\partial_aQ + 2\overline{\alpha}sQ, \quad \overline{ð} = \overline{m}^a\partial_aQ - 2\alpha sQ, \tag{15}$$

where ð is the Icelandic letter "edth"; note that $\overline{ðQ}$ is the complex conjugate of $ðQ$ since $\mathrm{sw}(\overline{Q}) = -\mathrm{sw}(Q)$. The fundamental property of these operators is

$$\mathrm{sw}(Q)=s \Rightarrow \begin{array}{l}\mathrm{sw}(ðQ)=s+1,\\ \mathrm{sw}(\overline{ð}Q)=s-1,\end{array} \tag{16}$$

i.e., $\mathrm{sw}(ð)=1$, $\mathrm{sw}(\overline{ð})=-1$. We also have

$$[ð,\overline{ð}]Q = -2sQ. \tag{17}$$

The standard gauge is given by choosing $\gamma=0$ in (9), thus

$$\begin{aligned} ð_0 &= \partial_\theta + (i/\sin\theta)\partial_\varphi - s(\cos\theta/\sin\theta),\\ \overline{ð}_0 &= \partial_\theta - (i/\sin\theta)\partial_\varphi + s(\cos\theta/\sin\theta). \end{aligned} \tag{18}$$

In an arbitrary gauge we have

$$\begin{aligned} ð &= e^{i\gamma}\left[\partial_\theta + \frac{i}{\sin\theta}\partial_\varphi + s\left(-\frac{\cos\theta}{\sin\theta} - i\gamma_{,\theta} + \frac{1}{\sin\theta}\gamma_{,\varphi}\right)\right],\\ \overline{ð} &= e^{-i\gamma}\left[\partial_\theta - \frac{i}{\sin\theta}\partial_\varphi - s\left(-\frac{\cos\theta}{\sin\theta} + i\gamma_{,\theta} + \frac{1}{\sin\theta}\gamma_{,\varphi}\right)\right]. \end{aligned} \tag{19}$$

We can now obtain the spin-weighted spherical harmonics (for integer spin) ${}_sY_{lm}$ by raising and lowering the spin weight of the usual spherical harmonics $Y_{lm}(\theta,\varphi)$ [$\mathrm{sw}(Y_{lm})=0$][11]

$${}_sY_{lm} = \begin{cases} \left[\dfrac{(l-s)!}{(l+s)!}\right]^{1/2} ð^sY_{lm}, & 0\leqslant s\leqslant l,\\ \left[\dfrac{(l+s)!}{(l-s)!}\right]^{1/2} (-1)^s\overline{ð}^{-s}Y_{lm}, & -l\leqslant s\leqslant 0,\\ 0, & l<|s|. \end{cases} \tag{20}$$

We summarize the properties of the ${}_sY_{lm}$

$$\mathrm{sw}({}_sY_{lm}) = s, \tag{21a}$$

$$ð({}_sY_{lm}) = +[(l-s)(l+s+1)]^{1/2}\,{}_{s+1}Y_{lm}, \tag{21b}$$

$$\overline{ð}({}_sY_{lm}) = -[(l+s)(l-s+1)]^{1/2}\,{}_{s-1}Y_{lm}, \tag{21c}$$

$${}_0Y_{lm} = Y_{lm}, \tag{21d}$$

$$\int_S {}_sY_{lm}\,{}_s\overline{Y}_{l'm'}\,dS = \delta_{ll'}\delta_{mm'}. \tag{21e}$$

We can ask if there are generalizations of the usual angular momentum operators, i. e., operators $L_z, L_\pm, L^2$ satisfying [cf. (4)]

$$\begin{aligned} L_z\,{}_sY_{lm} &= m\,{}_sY_{lm},\\ L_\pm\,{}_sY_{lm} &= [(l\mp m)(l+1\pm m)]^{1/2}\,{}_sY_{lm\pm1},\\ L^2\,{}_sY_{lm} &= l(l+1)\,{}_sY_{lm}. \end{aligned} \tag{22}$$

Since these imply that

$$[L,ð] = 0 = [L,\overline{ð}], \tag{23}$$

where L represents any of the angular momentum operators, one can easily solve for these operators. The result is [12]

$$L_z = -i\,\partial_\varphi - s\gamma_{,\varphi},$$

$$L_\pm = e^{\pm i\varphi}\Big[\pm\partial_\theta + \frac{i\cos\theta}{\sin\theta}\partial_\varphi + s\Big(-\frac{1}{\sin\theta} \mp i\gamma_{,\theta} + \frac{\cos\theta}{\sin\theta}\gamma_{,\varphi}\Big)\Big],$$

$$L^2 = -\Delta' + \frac{2s\cos\theta}{\sin^2\theta}L_z + \frac{s^2}{\sin^2\theta}$$
$$= -\Delta + is(\Delta\gamma) - 2s\gamma_{,\theta}(L_y\cos_\varphi - L_x\sin_\varphi) + \frac{2s}{\sin^2\theta}(\cos\theta - \gamma_{,\varphi})L_z - s^2\Big(\gamma_{,\theta}^2 + \frac{\gamma_{,\varphi}^2 - 1}{\sin^2\theta}\Big),$$

where

$$\Delta' = \Delta - is(\Delta\gamma) - 2is(\gamma_{,\theta}\partial_\theta + \gamma_{,\varphi}\partial_\varphi/\sin^2\theta) - s^2(\gamma_{,\theta}^2 + \gamma_{,\varphi}^2/\sin^2\theta),$$

and

$$L_\pm = L_x \pm iL_y. \tag{24}$$

Here, Δ' is just the operator obtained from Δ by the substitutions

$$\partial_\varphi \mapsto \partial_\varphi - is\gamma_{,\varphi}, \quad \partial_\theta \mapsto \partial_\theta - is\gamma_{,\theta}.$$

Note that in the standard gauge, denoted "0" the $L^0_\pm$ are just the angular momentum operators $\hat{J}_\pm$ given in Landau and Lifschitz[13] for the symmetric top (with k there identified with $-s$ here). The similarity between the symmetric top operators and the Y_{qlm} has already been pointed out, e. g., in Ref. 6.

IV. COMPARISON OF MONOPOLE AND SPIN-WEIGHTED SPHERICAL HARMONICS

Comparing (24) with (5) we see that if we introduce the gauges A, defined by $\gamma = +\varphi$, and B, defined by $\gamma = -\varphi$ [in (9)], and if we make the identification $q = s$, then

$$L^a \equiv L^A, \qquad L^b \equiv L^B, \tag{25}$$

where L again represents any of the angular momentum operators. But since the Y_{qlm} are fully determined *up to a constant phase factor for each* q by specifying q, the behavior of the angular momentum operators [Eq. (4)], and the normalization condition (6a), and since the ${}_sY_{lm}$ have the same behavior with respect to angular momentum [Eq. (22)] and the same normalization [Eq. (21e)], we see that Y_{qlm} and ${}_qY_{lm}$ differ at most by a constant (q-dependent) phase factor. With our ${}_sY_{lm}$ as defined in (20) we have

$$Y^a_{qlm} = {}_qY^A_{lm}, \quad Y^b_{qlm} = {}_qY^B_{lm}. \tag{26}$$

This is our main result.

Note that we can now immediately give raising and lowering operators for the monopole index of the monopole harmonics; these are just $\eth$ and $\bar{\eth}$ in the appropriate gauge:

$$\eth^A = e^{+i\varphi}\Big(\partial_\theta + \frac{i}{\sin\theta}\partial_\varphi + q\frac{(1-\cos\theta)}{\sin\theta}\Big),$$
$$\eth^B = e^{-i\varphi}\Big(\partial_\theta + \frac{i}{\sin\theta}\partial_\varphi - q\frac{(1+\cos\theta)}{\sin\theta}\Big),$$
$$\bar{\eth}^A = e^{-i\varphi}\Big(\partial_\theta - \frac{i}{\sin\theta}\partial_\varphi - q\frac{(1-\cos\theta)}{\sin\theta}\Big), \tag{27}$$
$$\bar{\eth}^B = e^{+i\varphi}\Big(\partial_\theta - \frac{i}{\sin\theta}\partial_\varphi + q\frac{(1+\cos\theta)}{\sin\theta}\Big).$$

[To obtain the correct normalization merely divide these by the constant on the right side of (21b) or (21c) with $s = q$.]

V. DISCUSSION

Our result (26) should not be surprising. The monopole harmonics are analytic, whereas the operator $\eth_0$ has a direction-dependent limit at $\theta = 0$ and $\theta = \pi$. Going to the gauge A or B is necessary in order to turn $\eth$ into an analytic operator on the region R^a or R^b![14]

Futhermore, since the ${}_sY_{lm}$ of course have spin weight s, our result can be interpreted as follows: Remove the explicit q dependence (i. e., $e^{\pm iq\varphi}$) from the ${}_qY^{a,b}_{lm}$. The result is precisely the spin-weighted spherical harmonics ${}_qY^0_{lm}$ *in standard gauge.*

We have only explicitly treated the spin-weighted spherical harmonics for integer spin. However, the argument used in Sec. III to introduce the angular momentum operators L can be inverted: we could equally well *define* the spin-weighted spherical harmonics as eigenfunctions of L. It is then obvious that the results of Sec. IV are also valid for half-integer spin.

Note added: In fact, if we let ${}_sY^\gamma_{lm}$ denote the spin-weighted spherical harmonics in *spin gauge* γ [Eq. (9)] then[11]

$${}_sY^\gamma_{lm} \equiv e^{is\gamma}\,{}_sY^0_{lm}(\theta,\varphi) \equiv \frac{(-1)^s(2l+1)^{1/2}}{(4\pi)^{1/2}}D^l_{-sm}(\varphi,\theta,\gamma), \tag{28}$$

where the D^l_{-sm} are the Wigner D functions as given by Goldberg *et al.*[15] Thus, choosing a gauge γ in the sense of this paper corresponds to fixing a Euler angle $(-\gamma)$ in the argument of the Wigner D functions. As pointed out by the referee, the spin-weighted spherical harmonics in standard gauge ${}_sY^0_{lm}$ and the monopole harmonics $Y^{a,b}_{qlm}$ merely correspond to different choices of this Euler angle.

ACKNOWLEDGMENTS

I would like to thank Professor Chen Ning Yang for providing Ref. 7, Peter Batenburg for bringing the symmetric top (Ref. 13) to my attention and for pointing out some mistakes in the original manuscript, Malcolm Perry and Annti Niemi for discussions on the Wigner D functions, and the referee for suggesting Ref. 15.

This work was supported in part by the Stichting voor Fundamenteel Onderzoek der Materie.

[1] T. T. Wu and C. N. Yang, Nucl. Phys. B **107**, 365 (1976).
[2] In both cases the original ideas can be traced to earlier work. This is discussed in Secs. II and III below.
[3] E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[4] P. A. M. Dirac, Proc. R. Soc. London Ser. A **133**, 60 (1931).
[5] Ig. Tamm, Z. Phys. **71**, 141 (1931).
[6] M. Fierz, Helv. Phys. Acta **17**, 27 (1944).
[7] P. P. Banderet, Helv. Phys. Acta **19**, 503 (1946); Harish-Chandra, Phys. Rev. **74**, 883 (1948); A. S. Goldhaber, Phys. Rev. **140**, B1407 (1965).

[8]We have set $\hbar = c = 1$; e is the electron charge and g the strength of the monopole.

[9]A. I. Janis and E. T. Newman, J. Math. Phys. **6**, 902 (1965).

[10]These differ by a factor of (-1) from the operators defined in Ref. 3.

[11]These ${}_sY_{lm}$ differ from those in Ref. 3 by a factor $(-1)^s$. The reason for this choice will become clear in Sec. IV.

[12]These operators reduce to the standard ones for $s = 0$. Note that strictly speaking "s" must be interpreted as an operator.

[13]L. D. Landau and E. M. Lifschitz, *Quantum Mechanics* (Pergamon, New York, 1977), 3rd ed., p. 417; see also F. Reich and H. Rademacher, Z. Phys. **39**, 444 (1926).

[14]This idea is implicit in Ref. 3.

[15]J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudershan, J. Math. Phys. **8**, 2155 (1967).

Separability of the Dirac equation in a class of perfect fluid space-times with local rotational symmetry

B. R. Iyer and C. V. Vishveshwara
Raman Research Institute, Bangalore-560080, India

(Received 4 June 1984; accepted for publication 12 October 1984)

Chandrasekhar's technique for separation of the Dirac equation in the Kerr background is applied to perfect fluid space-times with local rotational symmetry. These space-times fall into three distinct types. It is found that in case (1) the Dirac equation separates if the space-time is at least "locally static" while in case (3) it separates if the space-time is at least "locally diagonal," in contrast to the massless case where Dhurandhar, Vishveshwara, and Cohen showed that the Hertz potential is separable in all cases. In case (2), however, the Dirac equation is separable in all those cases where the Hertz potential for neutrinos is separable.

I. INTRODUCTION

In a series of papers Dhurandhar, Vishveshwara, and Cohen[1] have been systematically studying massless perturbations of varying spins in perfect fluid space-times with local rotational symmetry. These space-times first discussed by Ellis[2] and Ellis and Stewart[3] form a subclass of the generalized Goldberg–Sachs class. This class includes a wide spectrum of interesting space-times like Friedmann, Godel, Kantowski–Sachs universes, Taub–NUT, anisotropic cosmological models, etc. The above investigation used the Hertz–Debye formalism developed in detail for curved space-times by Cohen and Kegeles.[4]

The above treatment for neutrinos is in need of modification if the neutrinos turn out to be massive.[5] As a preliminary to the study of massive spin half-perturbations, it is interesting to ask whether in the above background space-times wherein the Hertz potential for every massless spin field—in particular the neutrino—is separable,the massive Dirac equation is also separable. This motivates us to look into the question of whether the Chandrasekhar separation for the Dirac equation in Kerr background[6] can be extended to the above class of space-times. In the next section, we write down the general form of the background metric for perfect fluid space-times with local rotational symmetry and collect together geometrical details of relevance. In Sec. III the Dirac equation is written down in the Newman–Penrose[7] spinor form. In Sec. IV, we show that Chandrasekhar's method separates the Dirac equation in a certain subclass of space-times. In the last section we discuss this subclass and obtain decoupled equations for the angular and radial parts. Finally we also briefly discuss the behavior of the angular and radial functions.

II. PERFECT FLUID SPACE-TIMES (PFST) WITH LOCAL ROTATIONAL SYMMETRY (LRS)

Perfect fluid space-times with local rotational symmetry are described by the line element[2,3]

$$\begin{aligned} ds^2 = (dx^0)^2/F^2 - X^2(dx^1)^2 - Y^2[(dx^2)^2 + t^2(dx^3)^2] \\ - (y/F^2)(2dx^0 - y\, dx^3)dx^3 \\ + hX^2(2dx^1 - h\, dx^3)dx^3, \end{aligned} \tag{1}$$

where F, X, Y are functions of x^0 and x^1 and t, y, h are functions of x^2 only. The space-times fall into three different classes as follows:

$$\text{(i)}\quad X=1,\quad Y=Y(x^1),\quad F=F(x^1),\quad h=0. \tag{2a}$$

$$\text{(ii)}\quad h=y=0. \tag{2b}$$

$$\text{(iii)}\quad F=1,\quad X=X(x^0),\quad Y=Y(x^0),\quad y=0. \tag{2c}$$

It is to be noted that t can take one of the following functional forms:

$$\begin{aligned} &\text{(a) } t=\sin(x^2),\quad \text{(b) } t=\sinh(x^2),\\ &\text{(c) } t=x^2,\quad \text{(d) } t=\text{const}, \end{aligned} \tag{3}$$

while h and y are obtained from t by the relation

$$h_{,2} = ct,\quad y_{,2} = c^1 t, \tag{4}$$

where c and c^1 are constants.

A convenient null tetrad for these space-times has been given by Wainwright[8] as

$$k^a = \frac{1}{\sqrt{2}}\Big(F,\frac{1}{X},0,0\Big), \tag{5a}$$

$$n^a = \frac{1}{\sqrt{2}}\Big(F,\frac{-1}{X},0,0\Big), \tag{5b}$$

$$m^a = \frac{1}{\sqrt{2}}\Big(\frac{-iy}{Yt},\frac{-ih}{Yt},\frac{-1}{Y},\frac{-i}{Yt}\Big), \tag{5c}$$

$$\overline{m}^a = \frac{1}{\sqrt{2}}\Big(\frac{iy}{Yt},\frac{ih}{Yt},\frac{-1}{Y},\frac{i}{Yt}\Big). \tag{5d}$$

As usual we have for the only nonvanishing innerproducts of the above null vectors

$$k_a n^a = 1,\quad m_a \overline{m}^a = -1. \tag{6}$$

The associated directional derivatives are given by

$$D \equiv k^a\partial_a = \frac{1}{\sqrt{2}}\Big(F\partial_0 + \frac{1}{X}\partial_1\Big), \tag{6a}$$

$$\Delta \equiv n^a\partial_a = \frac{1}{\sqrt{2}}\Big(F\partial_0 - \frac{1}{X}\partial_1\Big), \tag{6b}$$

$$\delta \equiv m^a\partial_a = \frac{-1}{\sqrt{2}Y}\Big(\frac{iy}{t}\partial_0 + \frac{ih}{t}\partial_1 + \partial_2 + \frac{i}{t}\partial_3\Big), \tag{6c}$$

$$\delta^* \equiv \overline{m}^a\partial_a = \frac{1}{\sqrt{2}Y}\Big(\frac{iy}{t}\partial_0 + \frac{ih}{t}\partial_1 - \partial_2 + \frac{i}{t}\partial_3\Big). \tag{6d}$$

After a lengthy computation the spin coefficients are computed straightforwardly and are given by

 0022-2488/85/051034-06$02.50

$$\alpha = -\beta^* = \frac{1}{2\sqrt{2}Yt}\left[t_{,2} - i\frac{(yY_{,0} + hY_{,1})}{Y}\right],$$

$$\gamma = -\frac{1}{2\sqrt{2}}\left[\frac{F}{X}\left(X_{,0} + \frac{F_{,1}}{F^2}\right) + \frac{i}{2Y^2t}\left(Xh_{,2} + \frac{y_{,2}}{F}\right)\right],$$

$$\epsilon = \frac{1}{2\sqrt{2}}\left[\frac{F}{X}\left(X_{,0} - \frac{F_{,1}}{F^2}\right) + \frac{i}{2Y^2t}\left(Xh_{,2} - \frac{y_{,2}}{F}\right)\right],$$

$$\mu = -\frac{1}{\sqrt{2}Y}\left[-\left(FY_{,0} - \frac{Y_{,1}}{X}\right) + \frac{i}{2Yt}\left(Xh_{,2} + \frac{y_{,2}}{F}\right)\right], \quad (7)$$

$$\rho = -\frac{1}{\sqrt{2}Y}\left[\left(FY_{,0} + \frac{Y_{,1}}{X}\right) - \frac{i}{2Yt}\left(Xh_{,2} - \frac{y_{,2}}{F}\right)\right],$$

$$\nu = \kappa = -\frac{i}{2\sqrt{2}Yt}\left[h\left(\frac{F_{,1}}{F} + \frac{X_{,1}}{X}\right) + y\left(\frac{F_{,0}}{F} + \frac{X_{,0}}{X}\right)\right],$$

$$\pi = \tau = -\frac{i}{2\sqrt{2}Yt}\left[h\left(\frac{F_{,1}}{F} - \frac{X_{,1}}{X}\right) + y\left(\frac{F_{,0}}{F} - \frac{X_{,0}}{X}\right)\right],$$

$$\lambda = \sigma = 0.$$

III. THE DIRAC EQUATION IN PFST WITH LRS

Following Chandrasekhar[6] the Dirac equation in curved space-time is written as a set of four coupled first-order differential equations

$$(D + \epsilon - \rho)F_1 + (\delta^* + \pi - \alpha)F_2 = i\tilde{\mu}_e G_1, \quad (8a)$$

$$(\Delta + \mu - \gamma)F_2 + (\delta + \beta - \tau)F_1 = i\tilde{\mu}_e G_2, \quad (8b)$$

$$(D + \epsilon^* - \rho^*)G_2 - (\delta + \pi^* - \alpha^*)G_1 = i\tilde{\mu}_e F_2, \quad (8c)$$

$$(\Delta + \mu^* - \gamma^*)G_1 - (\delta^* + \beta^* - \tau^*)G_2 = i\tilde{\mu}_e F_1, \quad (8d)$$

where

$$F_1 = P^0, \quad F_2 = P^1, \quad G_1 = \bar{Q}^{1'}, \quad G_2 = -\bar{Q}^{0'}. \quad (9)$$

The four-component Dirac wave function ψ is given by $\psi = (P^A, \bar{Q}_{A'})^T$ and mass of the particle is $\mu_e = \sqrt{2}\tilde{\mu}_e$. As explained earlier the PFST with LRS fall into three distinct classes and we shall write down the Dirac equation in each of the three cases.

A. Case (1)

Employing the particular values $X = 1$, $Y = Y(x^1)$, $F = F(x^1)$, and $h = 0$ in Eqs. (6) and (7) and substituting in Eqs. (8), we obtain the Dirac equation for this case:

$$\left(F\partial_0 + \partial_1 - \frac{F_{,1}}{2F} + \frac{Y_{,1}}{Y} + \frac{iy_{,2}}{4Y^2tF}\right)F_1 + \frac{1}{Y}\left(\frac{iy}{t}\partial_0 - \partial_2 + \frac{i}{t}\partial_3 - \frac{t_{,2}}{2t}\right)F_2 = i\mu_e G_1, \quad (10a)$$

$$\left(F\partial_0 - \partial_1 + \frac{F_{,1}}{2F} - \frac{Y_{,1}}{Y} - \frac{iy_{,2}}{4Y^2tF}\right)F_2 - \frac{1}{Y}\left(\frac{iy}{t}\partial_0 + \partial_2 + \frac{i}{t}\partial_3 + \frac{t_{,2}}{2t}\right)F_1 = i\mu_e G_2, \quad (10b)$$

$$\left(F\partial_0 + \partial_1 - \frac{F_{,1}}{2F} + \frac{Y_{,1}}{Y} - \frac{iy_{,2}}{4Y^2tF}\right)G_2 + \frac{1}{Y}\left(\frac{iy}{t}\partial_0 + \partial_2 + \frac{i}{t}\partial_3 + \frac{t_{,2}}{2t}\right)G_1 = i\mu_e F_2, \quad (10c)$$

$$\left(F\partial_0 - \partial_1 + \frac{F_{,1}}{2F} - \frac{Y_{,1}}{Y} + \frac{iy_{,2}}{4Y^2tF}\right)G_1 - \frac{1}{Y}\left(\frac{iy}{t}\partial_0 - \partial_2 + \frac{i}{t}\partial_3 - \frac{t_{,2}}{2t}\right)G_2 = i\mu_e F_1. \quad (10d)$$

The above equations do not depend on x^0 and x^3 and hence the x^0, x^3 dependence is of the form

$$\exp i(wx^0 + mx^3). \quad (11)$$

Writing

$$F_i = \exp i(wx^0 + mx^3)F_i(x^1,x^2), \quad i = 1,2,$$
$$G_i = \exp i(wx^0 + mx^3)G_i(x^1,x^2), \quad i = 1,2, \quad (12)$$

and noting that $y_{,2}/t = c^1$, a constant, Eq. (10) can be rewritten as

$$(\mathscr{D}_1 + ic^1/4YF)F_1 - \mathscr{L}_1F_2 = i\mu_e YG_1, \quad (13a)$$

$$(\mathscr{D}_1^\dagger + ic^1/4YF)F_2 + \mathscr{L}_1^\dagger F_1 = -i\mu_e YG_2, \quad (13b)$$

$$(\mathscr{D}_1 - ic^1/4YF)G_2 + \mathscr{L}_1^\dagger G_1 = i\mu_e YF_2, \quad (13c)$$

$$(\mathscr{D}_1^\dagger - ic^1/4YF)G_1 - \mathscr{L}_1G_2 = -i\mu_e YF_1, \quad (13d)$$

where

$$\mathscr{D}_1 \equiv Y(\partial_1 + iwF + Y_{,1}/Y - F_{,1}/2F), \quad (14a)$$

$$\mathscr{D}_1^\dagger \equiv Y(\partial_1 - iwF + Y_{,1}/Y - F_{,1}/2F), \quad (14b)$$

$$\mathscr{L}_1 \equiv \partial_2 + (wy + m)/t + t_{,2}/2t, \quad (15a)$$

$$\mathscr{L}_1^\dagger \equiv \partial_2 - (wy + m)/t + t_{,2}/2t. \quad (15b)$$

To separate the x^1 and x^2 dependence in Eqs. (13) one introduces

$$F_1 = R_-(x^1)S_-(x^2), \quad F_2 = R_+(x^1)S_+(x^2),$$
$$G_1 = R_+(x^1)S_-(x^2), \quad G_2 = R_-(x^1)S_+(x^2), \quad (16)$$

and rearranges terms to obtain

$$(\mathscr{D}_1 + ic^1/4YF)R_- = (\lambda_1 + i\mu_e Y)R_+, \quad (17a)$$

$$\mathscr{L}_1 S_+ = \lambda_1 S_-, \quad (17b)$$

$$(\mathscr{D}_1^\dagger + ic^1/4YF)R_+ = -(\lambda_2 + i\mu_e Y)R_-, \quad (18a)$$

$$\mathscr{L}_1^\dagger S_- = \lambda_2 S_+, \quad (18b)$$

$$(\mathscr{D}_1 - ic^1/4YF)R_- = (-\lambda_3 + i\mu_e Y)R_+, \quad (19a)$$

$$\mathscr{L}_1^\dagger S_- = \lambda_3 S_+, \quad (19b)$$

$$(\mathscr{D}_1^\dagger - ic^1/4YF)R_+ = (\lambda_4 - i\mu_e Y)R_-, \quad (20a)$$

$$\mathscr{L}_1 S_+ = \lambda_4 S_-, \quad (20b)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are four constants of separation. Consistency of Eqs. (17b) and (20b) imply $\lambda_1 = \lambda_4$ while that of (18b) and (19b) give $\lambda_2 = \lambda_3$. Substituting this in the equation for $R_\pm$ we get

$$(\mathscr{D}_1 + ic^1/4YF)R_- = (\lambda_1 + i\mu_e Y)R_+, \quad (21a)$$

$$(\mathscr{D}_1^\dagger + ic^1/4YF)R_+ = -(\lambda_2 + i\mu_e Y)R_-, \quad (21b)$$

$$(\mathscr{D}_1 - ic^1/4YF)R_- = (-\lambda_2 + i\mu_e Y)R_+, \quad (21c)$$

$$(\mathscr{D}_1^\dagger - ic^1/4YF)R_+ = (\lambda_1 - i\mu_e Y)R_-. \quad (21d)$$

Consistency of the above equations is possible only if $\lambda_1 = -\lambda_2$ and $c^1 = 0$. Thus in space-times of type I, the Dirac equation is separable in the subclass given by $c^1 = 0$. The above analysis however, does not preclude separability by some other method when $c^1 \neq 0$. For $c^1 = 0$, the Dirac equation becomes

$$\mathscr{D}_1 R_- = (\lambda + i\mu_e Y)R_+, \tag{22a}$$

$$\mathscr{D}_1^\dagger R_+ = (\lambda - i\mu_e Y)R_-, \tag{22b}$$

$$\mathscr{L}_1 S_+ = \lambda S_-, \tag{23a}$$

$$\mathscr{L}_1^\dagger S_- = -\lambda S_+. \tag{23b}$$

B. Case (3)

Before going to case (2) we treat case (3) since the treatment is very similar to that of case (1). In this case, $F = 1$, $X = X(x^0)$, $Y = Y(x^0)$, and $y = 0$. Employing the directional derivatives and spin coefficients appropriate to this instance the Dirac equation may be written down as before. The equations, however, now are translationally invariant with respect to x^1 and x^3 and further $h_{,2}/t = c$, a constant. Writing

$$F_i = \exp i(k_1x^1 + k_3x^3)F_i(x^0,x^2), \quad i = 1,2,$$
$$G_i = \exp i(k_1x^1 + k_3x^3)G_i(x^0,x^2), \quad i = 1,2, \tag{24}$$

the equations may be written in the form

$$(\mathscr{D}_3 - icX/4Y)F_1 - \mathscr{L}_3F_2 = i\mu_e YG_1, \tag{25a}$$

$$(\mathscr{D}_3^\dagger - icX/4Y)F_2 - \mathscr{L}_3^\dagger F_1 = i\mu_e YG_2, \tag{25b}$$

$$(\mathscr{D}_3 + icX/4Y)G_2 + \mathscr{L}_3^\dagger G_1 = i\mu_e YF_2, \tag{25c}$$

$$(\mathscr{D}_3^\dagger + icX/4Y)G_1 + \mathscr{L}_3 G_2 = i\mu_e YF_1, \tag{25d}$$

where

$$\mathscr{D}_3 \equiv Y\left(\partial_0 + \frac{ik_1}{X} + \frac{X_{,0}}{2X} + \frac{Y_{,0}}{Y}\right), \tag{26a}$$

$$\mathscr{D}_3^\dagger \equiv Y\left(\partial_0 - \frac{ik_1}{X} + \frac{X_{,0}}{2X} + \frac{Y_{,0}}{Y}\right), \tag{26b}$$

$$\mathscr{L}_3 \equiv \partial_2 + \frac{hk_1 + k_3}{t} + \frac{t_{,2}}{2t}, \tag{27a}$$

$$\mathscr{L}_3^\dagger \equiv \partial_2 - \frac{hk_1 + k_3}{t} + \frac{t_{,2}}{2t}. \tag{27b}$$

Writing

$$F_1 = T_-(x^0)S_-(x^2), \quad F_2 = T_+(x^0)S_+(x^2),$$
$$G_1 = T_+(x^0)S_-(x^2), \quad G_2 = T_-(x^0)S_+(x^2) \tag{28}$$

and repeating consistency arguments similar to the previous case, one obtains the result that the Dirac equation is separable if we have $c = 0$. As before, this does not rule out separability by different procedures when $c \neq 0$. For $c = 0$, the separated equations are of the form

$$\mathscr{D}_3 T_- = (\lambda + i\mu_e Y)T_+, \tag{29a}$$

$$\mathscr{D}_3^\dagger T_+ = (-\lambda + i\mu_e Y)T_-, \tag{29b}$$

$$\mathscr{L}_3 S_+ = \lambda S_-, \tag{30a}$$

$$\mathscr{L}_3^\dagger S_- = -\lambda S_+. \tag{30b}$$

C. Case (2)

Finally we look into case (2) which is characterized by $h = y = 0$. In this case consequently the Dirac equation may be written down using Eqs. (6), (7), and (8). Further noting that in general only ∂_3 is a Killing vector the x^3 dependence of F and G is given by $\exp ik_3x^3$. Consequently, the equation may be simplified to the form

$$\mathscr{D}_2F_1 - \mathscr{L}_2F_2 = i\mu_e YG_1, \tag{31a}$$

$$\mathscr{D}_2^\dagger F_2 - \mathscr{L}_2^\dagger F_1 = i\mu_e YG_2, \tag{31b}$$

$$\mathscr{D}_2G_2 + \mathscr{L}_2^\dagger G_1 = i\mu_e YF_2, \tag{31c}$$

$$\mathscr{D}_2^\dagger G_1 + \mathscr{L}_2G_2 = i\mu_e YF_1, \tag{31d}$$

where $\mathscr{L}_2$ and $\mathscr{L}_2^\dagger$ are the "angular" operators depending on x^2 only while $\mathscr{D}_2$ and $\mathscr{D}_2^\dagger$ are in general "radial-temporal" operators depending on both x^0 and x^1

$$\mathscr{D}_2 \equiv Y\left[F\partial_0 + \frac{1}{X}\partial_1 + \frac{F}{2X}\left(X_{,0} - \frac{F_{,1}}{F^2}\right) + \frac{1}{Y}\left(FY_{,0} + \frac{Y_{,1}}{X}\right)\right], \tag{32a}$$

$$\mathscr{D}_2^\dagger \equiv Y\left[F\partial_0 - \frac{1}{X}\partial_1 + \frac{F}{2X}\left(X_{,0} + \frac{F_{,1}}{F^2}\right) + \frac{1}{Y}\left(FY_{,0} - \frac{Y_{,1}}{X}\right)\right], \tag{32b}$$

$$\mathscr{L}_2 \equiv \partial_2 + k_3/t + t_{,2}/2t, \tag{33a}$$

$$\mathscr{L}_2^\dagger \equiv \partial_2 - k_3/t + t_{,2}/2t. \tag{33b}$$

The "angular" dependence may be extracted out by introducing

$$F_1 = Z_-(x^0,x^1)S_-(x^2), \quad F_2 = Z_+(x^0,x^1)S_+(x^2),$$
$$G_1 = Z_+(x^0,x^1)S_-(x^2), \quad G_2 = Z_-(x^0,x^1)S_+(x^2). \tag{34}$$

Consistency arguments along the lines of case (1) now show that the angular part separates in all cases. We then obtain for the following system of equations for the angular part and the radial-temporal part:

$$\mathscr{D}_2 Z_- = (\lambda + i\mu_e Y)Z_+, \tag{35a}$$

$$\mathscr{D}_2^\dagger Z_+ = (-\lambda + i\mu_e Y)Z_-, \tag{35b}$$

$$\mathscr{L}_2 S_+ = \lambda S_-, \tag{36a}$$

$$\mathscr{L}_2^\dagger S_- = -\lambda S_+. \tag{36b}$$

Equations (35) are more complicated than in the earlier cases since the time and spatial dependence are still coupled. Though in general X, Y, and F are functions of x^0 and x^1 a useful restriction obtains if one assumes

$$X = X(x^1), \quad Y = Y(x^1), \quad \text{and} \quad F = F(x^0). \tag{37}$$

In this case, Eqs. (35) become

$$F\partial_0 Z_- = \frac{\lambda + i\mu_e Y}{Y}Z_+ - \frac{1}{X}\left(\partial_1 + \frac{Y_{,1}}{Y}\right)Z_-, \tag{38a}$$

$$F\partial_0 Z_+ = -\left(\frac{\lambda - i\mu_e Y}{Y}\right)Z_- + \frac{1}{X}\left(\partial_1 + \frac{Y_{,1}}{Y}\right)Z_+. \tag{38b}$$

The above equations can be separated by writing

$$Z_- = T(x^0)R_-(x^1), \quad Z_+ = T(x^0)R_+(x^1), \tag{39}$$

whence one obtains

$$F\frac{\partial_0 T}{T}=\frac{1}{R_-}\left[\frac{\lambda+i\mu_e Y}{Y}R_+-\frac{1}{X}\left(\partial_1+\frac{Y_{,1}}{Y}\right)R_-\right]$$
$$=\frac{1}{R_+}\left[-\frac{\lambda-i\mu_e Y}{Y}R_-+\frac{1}{X}\left(\partial_1+\frac{Y_{,1}}{Y}\right)R_+\right] \tag{40}$$
$$= \text{constant } w \text{ say.}$$

Integrating the equation for T one obtains

$$T=\exp\left[w\int\frac{dx^0}{F}\right] \tag{41}$$

while the x^1 dependence is given by the coupled equations

$$((1/X)\partial_1+w)YR_-=(\lambda+i\mu_e Y)R_+, \tag{42a}$$

$$((1/X)\partial_1-w)YR_+=(\lambda-i\mu_e Y)R_-. \tag{42b}$$

Thus in this restricted case the Dirac equation is completely separable.

Finally we consider another particular instance of case (2) which also yields a completely separable system. In this case

$$X=X(x^1),\quad Y=Y(x^0),\quad F=F(x^0). \tag{43}$$

Proceeding as before in this case, we are led to the equations

$$Z_-=T_-(x^0)R(x^1),\quad Z_+=T_+(x^0)R(x^1), \tag{44}$$

where

$$R(x^1)=\exp\left[k_1\int X\,dx^1\right]. \tag{45}$$

The x^0 dependence in this instance is given by the coupled system

$$Y(F\partial_0+F(Y_{,0}/Y)+k_1)T_-=(\lambda+i\mu_e Y)T_+, \tag{46a}$$

$$Y(F\partial_0+F(Y_{,0}/Y)-k_1)T_+=(-\lambda+i\mu_e Y)T_-. \tag{46b}$$

Thus in the case (2) the Dirac equation is completely separable in the two subclasses specified by Eqs. (37) and (43).

IV. THE PARTICULAR SPACE-TIMES AND DECOUPLED EQUATIONS

In this section we examine the subclasses of space-times in which the Dirac equation separates. We shall mention some characteristics of such space-times based essentially on the acceleration, rotation, expansion, and shear of the fluid world-lines. As demonstrated in the previous section the Dirac equation separates in case (1) if $c^1=0$. This implies, using Eq. (4), that y is a constant. Thus the space-times where the Dirac equation is separable are of the form

$$ds^2=(1/F^2)(dx^0-y\,dx^3)^2-(dx^1)^2-Y^2((dx^2)^2+t^2(dx^3)^2). \tag{47}$$

Since y is a constant, introducing

$$\bar{x}^0=x^0-yx^3 \tag{48}$$

puts the above metric in the form

$$ds^2=(1/F^2)(d\bar{x}^0)^2-(dx^1)^2-Y^2((dx^2)^2+t^2(dx^3)^2). \tag{49}$$

If x^3 is a cyclic coordinate (the usual spherical coordinate ϕ) the above transformation is an allowed transformation only locally. In this case $c^1=0$ represents space-times which can be called "locally static." We have thus proved that for metrics of type I the Dirac equation is separable if the metric is at least locally static. A straightforward calculation shows that the fluid world-lines have nonvanishing acceleration with the other parameters, viz., rotation, expansion, and shear, being zero.

We next proceed to obtain the decoupled equation for the angular and radial parts. Operating on Eq. (23b) by $\mathscr{L}_1$ and employing Eq. (23a) yields the equation satisfied by S_-:

$$(\mathscr{L}_1\mathscr{L}_1^\dagger+\lambda^2)S_-=0. \tag{50a}$$

Employing Eqs. (15) $\mathscr{L}_1\mathscr{L}_1^\dagger$ can be explicitly obtained. Thus

$$\mathscr{L}_1\mathscr{L}_1^\dagger=\partial_2^2+\frac{t_{,2}}{t}\left(\partial_2+\frac{wy+m}{t}-\frac{t_{,2}}{4t}\right)$$
$$+\frac{t_{,22}}{2t}-\left(\frac{wy+m}{t}\right)^2 \tag{50b}$$

so that Eqs. (50) yield the decoupled equation satisfied by S_-. Similarly, $S_+(w,m,\lambda;x^2)$ satisfies the same equation as $S_-(-w,-m,\lambda;x^2)$.

To obtain the decoupled eqution for the radial part we operate on Eq. (22a) by $\mathscr{D}_1^\dagger$ and using Eq. (22b) we obtain

$$\left[\mathscr{D}_1^\dagger\mathscr{D}_1-\frac{i\mu_e Y_{,1}Y}{\lambda+i\mu_e Y}\mathscr{D}_1-(\lambda^2+\mu_e^2Y^2)\right]R_-(w,\lambda;x^1)=0. \tag{51}$$

Using Eqs. (14) $\mathscr{D}_1^\dagger\mathscr{D}_1$ may be explicitly computed. Substituting this in Eq. (51) and after some simplifications we obtain the decoupled equation satisfied by $R_-(w,\lambda;x^1)$:

$$Y^2\left[\partial_1^2+\left(2\frac{Y_{,1}}{Y}-\frac{F_{,1}}{F}+Q_1^-\right)+iwF\left(Q_1^-+\frac{F_{,1}}{F}\right)\right.$$
$$+\frac{Y_{,11}}{Y}-\frac{F_{,11}}{F}+\frac{3}{4}\frac{F_{,1}^2}{F^2}-\frac{F_{,1}}{2F}\left(2\frac{Y_{,1}}{Y}+Q_1^-\right)$$
$$\left.+\frac{Y_{,1}}{Y}Q_1^-+w^2F^2-\frac{\lambda^2+\mu_e^2Y^2}{Y^2}\right]R_-(w,\lambda;x^1)=0, \tag{52a}$$

where

$$Q_1^-\equiv Q_1^-(\lambda,x^1)$$
$$=\frac{Y_{,1}}{Y}\left(1-\frac{i\mu_e Y}{\lambda+i\mu_e Y}\right). \tag{52b}$$

$R_+(w,\lambda;x^1)$ satisfies the same equation as $R_-(-w,-\lambda;x^1)$.

Case (3): In this case the particular space-time for which the Dirac equation separates are given by $c=0$ so that Eq. (4) implies h is a constant. Restricting to this subclass the space-times are characterized by the line element of the form

$$ds^2=(dx^0)^2-X^2(dx^1-h\,dx^3)^2$$
$$-Y^2(dx^2)^2-t^2Y^2(dx^3)^2. \tag{53}$$

Introducing, as before,

$$\bar{x}^1=x^1-hx^3 \tag{54}$$

reduces the metric to a diagonal form. If x^3 is a cyclic coordinate this is possible only locally and we may call the metric "locally diagonal." In this case fluid lines are geodetic and nonrotating, but have nonzero expansion and shear.

Applying the procedure of the previous section to Eqs. (30) and (29) and using Eqs. (21) and (26) we obtain the decoupled angular radial equations. Thus,

$$\Big[\partial_2^2 + \frac{t_{,2}}{t}\Big(\partial_2 + \frac{hk_1 + k_3}{t} - \frac{t_{,2}}{4t}\Big) + \frac{t_{,22}}{2t} - \Big(\frac{hk_1+k_3}{t}\Big)^2 + \lambda^2\Big] S_- = 0; \tag{55}$$

$$Y^2\Big[\partial_0^2 + \Big(\frac{2Y_{,0}}{Y} + \frac{X_{,0}}{X} + Q_3^-\Big)\partial_0 + \frac{ik_1}{X}\Big(Q_3^- - \frac{X_{,0}}{X}\Big) + \frac{X_{,00}}{2X} + \frac{Y_{,00}}{Y} - \frac{X_{,0}^2}{4X^2} + \frac{X_{,0}}{2X}\Big(\frac{2Y_{,0}}{Y} + Q_3^-\Big) + \frac{Y_{,0}}{Y}Q_3^- + \frac{k_1^2}{X^2} + \frac{\lambda^2 + \mu_e^2 Y^2}{Y^2}\Big] T_-(k_1,\lambda;x^0) = 0; \tag{56a}$$

$$Q_3^-(\lambda) = \frac{Y_{,0}}{Y}\Big(1 - \frac{i\mu_e Y}{\lambda + i\mu_e Y}\Big); \tag{56b}$$

$S_+(k_1,k_3,\lambda;x^2)$ satisfies the same equation as $S_-(-k_1,-k_3,\lambda;x^2)$ while $T_+(k_1,\lambda;x^0)$ satisfies the same equation as $T_-(-k_1,-\lambda;x^0)$.

Case (2): In this case the angular part is separable in all the cases. The decoupled equation for S_- by following a procedure indicated previously is given by

$$\Big[\partial_2^2 + \frac{t_{,2}}{t}\Big(\partial_2 - \frac{t_{,2}}{4t} + \frac{k_3}{t}\Big) + \frac{t_{,22}}{2t} - \frac{k_3^2}{t^2} + \lambda^2\Big] S_-(k_3,\lambda;x^2) = 0. \tag{57}$$

$S_+(k_3,\lambda;x^2)$ satisfies the same equation as $S_-(-k_3,\lambda;x^2)$.

The radial (temporal) part as shown in the previous section is separable in two particular cases. In the first case given by Eq. (37) the metric is

$$ds^2 = \frac{(dx^0)^2}{F^2(x_0)} - X^2(x^1)(dx^1)^2 - Y^2(x^1) \times [(dx^2)^2 + t^2(x^2)(dx^3)^2]. \tag{58}$$

This is a static metric with the fluid lines having all the parameters zero.

In this subclass $R_-(w,\lambda;x^1)$ satisfies

$$Y\Big[\frac{1}{X^2}\partial_1^2 + \frac{1}{X^2}\Big(\frac{2Y_{,1}}{Y} - \frac{X_{,1}}{X} + Q_3^-\Big)\partial_1 + \frac{Y_{,11}}{X^2 Y} - \frac{Y_{,1}}{X^2 Y}\Big(\frac{X_{,1}}{X} - Q_3^-\Big) + \frac{wQ_3^-}{X} - w^2 - \frac{\lambda^2 + \mu_e^2 Y^2}{Y^2}\Big] R_-(w,\lambda;x^1) = 0, \tag{59a}$$

$$Q_3^- = \lambda Y_{,1}/Y(\lambda + i\mu_e Y). \tag{59b}$$

$R_+(w,\lambda;x^1)$ satisfies the same equation as $R_-(-w,-\lambda;x^1)$.

The second subclass [Eq. (4)] corresponds to the line element

$$ds^2 = \frac{(dx^0)^2}{F^2(x^0)} - X^2(x^1)(dx^1)^2 - Y^2(x^0)((dx^2)^2 + t^2(x^2)(dx^3)^2). \tag{60}$$

Here the space-time is nonexpanding in the x^1 direction. In contrast to the first subclass, the fluid world-lines, though geodetic and nonrotating, have nonvanishing expansion and shear. The decoupled temporal equation in this instance is given by

$$Y\Big[F^2\partial_0^2 + F^2\Big(\frac{2Y_{,0}}{Y} + \frac{F_{,0}}{F} + Q_4^- F\Big)\partial_0 + \frac{Y_{,00}F^2}{Y} + \frac{F^2 Y_{,0}}{Y}\Big(\frac{F_{,0}}{F} + Q_4^-\Big) + k_1 F Q_4^- - k_1^2 + \frac{\lambda^2 + \mu_e^2 Y^2}{Y^2}\Big] T_-(k_1,\lambda;x^0) = 0, \tag{61a}$$

$$Q_4^- = \lambda Y_{,0}/Y(\lambda + i\mu_e Y). \tag{61b}$$

$T_+(k_1,\lambda;x^0)$ satisfies the same equation as $T_-(-k_1,-\lambda;x^0)$.

V. DISCUSSION

In the previous sections we have obtained the subclass of perfect fluid space-times with local rotational symmetry wherein the Dirac equation is separable. For space-times belonging to case (1), the Dirac equation is separable if the background is at least "locally static" while in case (3), it separates if the space-time is at least "locally diagonal." In case (2) the massive Dirac equation is separable in those cases where the Hertz potential for the massless spin-$\frac{1}{2}$ equation is separable.[1] We may mention in passing that though cases (1) and (3) described, respectively, by Eqs. (49) and (53), resemble case (2) they are distinct from it. This is because $F = F(x^1)$ in case (1) while $F = F(x^0)$ in case (2) and similarly $X = X(x^0)$ in case (3) while $X = X(x^1)$ in case (2). However, whether the Dirac equation separates out in other cases for different, judicious choices of tetrads and variables remains an open question.

We have also obtained the second-order decoupled equations satisfied by the angular and radial (temporal) parts of the wave function. To discuss further the angular and radial equations, we note that the form of the various equations in different cases is of the same nature so that we need to discuss in detail only a prototype for each. For instance, comparing the radial and angular parts in cases (1) and (3), i.e., Eqs. (50) and (55), Eqs. (52) and (56), respectively, we find that the equation for case (3) may be obtained for those of case (1) by the identifications

$$x^0 \leftrightarrow x^1,\quad w \leftrightarrow k_1,\quad m \leftrightarrow k_3,\quad y \leftrightarrow h,\qquad F \leftrightarrow 1/X,\quad (\lambda^2 + \mu_e^2 Y^2) \leftrightarrow -(\lambda^2 + \mu_e^2 Y^2). \tag{62}$$

Further, the angular equation in case (2) also is of the same form as for case (1) since y is a constant. They may be obtained by the identification $wy + m \leftrightarrow k_3$. One caution, however. In case x^3 is a cyclic coordinate, boundary conditions—like single valuedness of ψ—would imply a different spectrum for m as compared to k_3. In the two subclasses of case (2) where the Dirac equation is separable the radial (temporal) equations have a similar structure. From Eqs. (59) and (61) it can be seen that the equations go into one another with the identifications

$$x^0 \leftrightarrow x^1,\quad F \leftrightarrow 1/X,\quad (\lambda^2 + \mu_e^2 Y^2) \leftrightarrow (\lambda^2 + \mu_e^2 Y^2). \tag{63}$$

Further studies of the explicit solutions of the angular and radial equations are in progress and will be published elsewhere.

ACKNOWLEDGMENTS

We thank Joseph Samuel for useful discussions.

[1]See, e.g., S. V. Dhurandhar, C. V. Vishveshwara, and J. M. Cohen, Phys. Rev. D **21**, 2794 (1980); **26**, 2598 (1982).
[2]G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).
[3]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[4]J. M. Cohen and L. S. Kegeles, Phys. Rev. D **10**, 1070 (1974); L. S. Kegeles and J. M. Cohen, *ibid.* **19**, 1641 (1979).
[5]V. A. Lyubimov, E. G. Novikov, V. Z. Nozik, E. F. Tretyakov, and V. S. Kosik, Phys. Lett. B **94**, 266 (1980); see also D. W. Sciama, in *Astrophysical Cosmology*, edited by H. A. Bruck, G. V. Cuyne, and M. S. Longair (Pontificia Academia Scientiarum, Vatican City, 1982).
[6]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[7]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[8]J. Wainwright, Commun. Math. Phys. **17**, 42 (1970).

Solutions of the Einstein–Cartan–Dirac equations with vanishing energy-momentum tensor

A. Dimakis
Institut für Theoretische Physik, Universität Göttingen, D-3400 Göttingen, Federal Republic of Germany

F. Müller-Hoissen
Max-Planck-Institut für Physik und Astrophysik, Werner-Heisenberg-Institut für Physik, D-8000 München 40, Federal Republic of Germany

(Received 17 July 1984; accepted for publication 14 September 1984)

It is shown that a frame can always be chosen with respect to which a Dirac spinor ψ has a particular simple form. This result is used to investigate the existence of solutions of the Einstein–Cartan–Dirac equations with vanishing energy-momentum tensor, but nonvanishing current ("ghost solutions"). We concentrate on solutions with a massive Dirac field, subject to the restriction $(\bar\psi\psi)^2 + (\bar\psi\gamma_5\psi)^2 > 0$. In this case, the field equations can be reduced to a set of (covariant) equations on the group SL(2, $\mathbb{R}$). A family of new exact solutions is presented.

I. INTRODUCTION

It is known that certain anomalies appear if the Dirac field is minimally coupled to a space-time geometry. The local geometry enters the energy-momentum tensor of this field in such a way that, for example, the sign of the energy-density may become observer dependent and even change if the field propagates through space.[1] An extreme case occurs if the whole (minimally coupled) energy-momentum tensor vanishes although the field gives rise to a nonvanishing current. Exact solutions demonstrating this possibility are known as "ghost solutions." Most examples given so far represent massless neutrinos in general relativity.[2,3] The existence of massive ghost neutrino solutions has been proven by Griffiths.[4] However, there seems to be no explicit example up to now.

Griffiths[5] also obtained the general (massless) ghost neutrino solution of the Einstein–Cartan (EC) theory.[6] The corresponding metric is one of the plane wave type.

The present work is devoted to ghost solutions of the EC theory with Dirac fields which do not satisfy the chirality condition. More precisely, we will consider a Dirac spinor field ψ with

$$(\bar\psi\psi)^2 + (\bar\psi\gamma_5\psi)^2 > 0. \tag{1.1}$$

In Appendix A it is shown that a g-orthonormal coframe θ^i exists, with respect to which[7] ψ takes the form[8]

$$\psi = \sqrt{\rho}\begin{pmatrix} \cos\beta/2 \\ 0 \\ -\sin\beta/2 \\ 0 \end{pmatrix}, \tag{1.2}$$

with two real functions ρ and β. Using this result, it is proven that the EC theory (without a cosmological constant) does not admit massive ghosts. Allowing the presence of a cosmological constant, it is possible to reduce our problem to that of solving a linear first-order partial differential equation, supplemented by some nonlinear constraints.

Among the exact solutions presented in this work, a solution is recovered which was obtained earlier.[9] This special solution has the following properties.

(1) The mass squared of the Dirac field is proportional to the cosmological constant.

(2) The metric is conformally flat (cf. Appendix C).

(3) In the limit of vanishing mass, a teleparallel solution of the EC theory is obtained.[10] The latter is then also an exact solution of the Poincaré gauge theory,[11] with a Lagrangian of the form $R + R^2$, where R^2 represents an arbitrary quadratic curvature term. Furthermore, this solution provides an example for a relation between simply transitive Lie groups and a class of teleparallel connections (cf. Appendix D).[12]

Only the first property turns out to be a general feature.

The methods used in this work are quite different from those applied in earlier investigations of ghost solutions. Apart from our results in Appendix A, which include the simple form (1.2) of the spinor field, we profit from writing the Dirac equation as a Pfaff system.[13] This is of particular help for the evaluation of the integrability conditions.

Section II presents the field equations and their specialization to the frame in which (1.2) holds.

In Section III, the field equations are analyzed in the case of massive ghosts. An explicit family of solutions is then presented in Section IV. In order to obtain these solutions, it is necessary to solve the Maurer–Cartan equation for the Lie group SL(2, $\mathbb{R}$) [respectively, SO(2,1)]. Appendix B is added for this purpose.

Section V reveals an action of SL(2, $\mathbb{R}$) on the set of solutions. Finally, Section VI contains some concluding remarks.

II. THE FIELD EQUATIONS IN A SPECIAL FRAME

The Dirac equation in a space-time M with metric g_{ij} and a metric-compatible linear connection reads[14,15]

$$\gamma^i(\nabla_i - \tfrac{1}{2}Q_i)\psi + m\psi = 0. \tag{2.1}$$

Here, as well as in the following, all fields (and indices) refer to a g-orthonormal coframe field θ^i. Then $Q_i = Q^k{}_{ki}$ is the vectorial part of the torsion tensor, which is given by

$$\Theta^k \equiv d\theta^k + \omega^k{}_j \wedge \theta^j = \tfrac{1}{2}Q^k{}_{ij}\theta^i \wedge \theta^j, \tag{2.2}$$

involving the connection one-form $\omega^i{}_j = \omega^i{}_{jk}\theta^k$. For the covariant derivative part in (2.1) we obtain

$$\begin{aligned}\gamma^i\nabla_i\psi &= \gamma^i(\partial_i + \tfrac{1}{8}\omega_{kli}[\gamma^k,\gamma^l])\psi \\ &= \gamma^i[\partial_i + \tfrac{1}{2}({}_{\mathscr{V}}\omega_i + {}_{\mathscr{A}}\omega_i\gamma^5)]\psi,\end{aligned} \tag{2.3}$$

where we have introduced a "vectorial" and an "axial" part of the connection

$${}_{\mathscr{V}}\omega_i = \omega^j{}_{ij}, \quad {}_{\mathscr{A}}\omega_i = \tfrac{1}{2}\epsilon_{klmi}\omega^{klm}. \tag{2.4}$$

Restricting the Dirac field by (1.1) we may choose θ^i such that ψ takes the form (1.2). The Dirac equation can then be written as a Pfaff system:

$$d(\ln\rho) + {}_{\mathscr{V}}\omega - Q = 2m\sin\beta\,\theta^3, \tag{2.5}$$

$$d\beta + {}_{\mathscr{A}}\omega = 2m\cos\beta\,\theta^3, \tag{2.6}$$

with the one-forms

$${}_{\mathscr{V}}\omega = {}_{\mathscr{V}}\omega_i\theta^i, \quad {}_{\mathscr{A}}\omega = {}_{\mathscr{A}}\omega_i\theta^i, \quad Q = Q_i\theta^i. \tag{2.7}$$

The canonical energy-momentum three-form[16] of the Dirac field is (using the Hodge $*$-operator)

$$t_i = T_{ij}*\theta^j, \tag{2.8}$$

with

$$T_{ij} = \tfrac{i}{2}(\nabla_i\bar\psi\gamma_j\psi - \bar\psi\gamma_j\nabla_i\psi). \tag{2.9}$$

In our special frame this becomes

$$t_i = \tfrac{1}{2}\rho(\partial_i\beta*\theta^3 - \tfrac{1}{2}\omega_{kli}\theta^k\wedge\theta^l\wedge\theta^3). \tag{2.10}$$

In particular, we find

$$T^i{}_i = \tfrac{1}{2}\rho(\partial_3\beta + {}_{\mathscr{A}}\omega_3) = m\rho\cos\beta. \tag{2.11}$$

The field equations of the Einstein–Cartan theory with the Dirac field as source are[17]

$$(\kappa/2)\epsilon_{ijkl}\bar\psi\gamma^5\gamma^l\psi = Q_{ijk}, \tag{2.12}$$

$$R_{ji} + (\Lambda - \tfrac{1}{2}R)\eta_{ij} = \kappa T_{ij}, \tag{2.13}$$

where $R_{ij} = R^{\ k}_{ikj}$ and

$$\Omega^i{}_j \equiv d\omega^i{}_j + \omega^i{}_k\wedge\omega^k{}_j = \tfrac{1}{2}R^i{}_{jkl}\theta^k\wedge\theta^l. \tag{2.14}$$

As a consequence of (2.12) we have

$$Q = 0. \tag{2.15}$$

In the special frame, Eq. (2.12) reads

$$Q_{ijk} = (\kappa/2)\rho\epsilon_{ijk3}, \tag{2.16}$$

respectively,

$$\Theta^i = (\kappa/2)\rho*(\theta^i\wedge\theta^3). \tag{2.17}$$

III. MASSIVE GHOSTS

In this section we investigate the system of equations (2.5), (2.6), (2.13), and (2.16), supplemented by the "ghost condition" of a vanishing (canonical) energy-momentum tensor,

$$T_{ij} = 0. \tag{3.1}$$

Furthermore, we restrict our considerations to the case $m\neq 0$. Then (2.11) implies

$$\cos\beta = 0, \tag{3.2}$$

i.e., $\beta = \pi/2$ or $\beta = 3\pi/2$. Now (2.6) and (2.10) yield

$$\omega_{kli}\theta^k\wedge\theta^l\wedge\theta^3 = 0, \quad {}_{\mathscr{A}}\omega = 0, \tag{3.3}$$

and the solution of these two equations is given by

$$\omega_{ij} = 2\delta^A{}_{[i}\delta^3{}_{j]}\zeta_A, \tag{3.4}$$

with $\zeta_A = \zeta_{Ai}\theta^i$, $\zeta_{AB} = \zeta_{BA}$ $(A,B = 0,1,2)$. It follows that (2.5) can be written as

$$d(\ln\rho) = (\pm 2m - \zeta)\theta^3 + \chi, \tag{3.5}$$

with

$$\zeta = \eta^{AB}\zeta_{AB}, \quad \chi = \zeta_{A3}\theta^A. \tag{3.6}$$

The upper (lower) sign in (3.5) corresponds to $\beta = \pi/2$ $(\beta = 3\pi/2)$.

From (2.2), (2.17), and (3.4) we deduce that

$$d\theta^A = (\kappa/4)\rho\epsilon^A{}_{BC}\theta^B\wedge\theta^C - \zeta^A\wedge\theta^3, \tag{3.7}$$

$$d\theta^3 = -\chi\wedge\theta^3, \tag{3.8}$$

with $\epsilon_{ABC}\equiv\epsilon_{ABC3}$. The integrability condition of (3.7) enforces

$$\chi = 0, \tag{3.9}$$

by the use of (3.5), and leads to

$$0 = \pm m\kappa\rho*\theta^A - (\kappa/2)\rho*\zeta^A - d\zeta^A\wedge\theta^3. \tag{3.10}$$

Taking the exterior product of the last equation with θ_A and using the symmetry of ζ_{AB}, we get

$$\zeta = \pm 3m. \tag{3.11}$$

Equations (3.8) and (3.9) show that $\theta^3$is closed. Hence, the Poincaré lemma implies

$$\theta^3 = dz, \tag{3.12}$$

with a function z. We are now able to integrate (3.5):

$$\rho = \rho_0 e^{\mp mz} \quad (\rho_0 > 0). \tag{3.13}$$

The spinor ψ is completely determined by (3.2) and (3.13):

$$\psi = \sqrt{\rho_0/2}\,e^{\mp mz/2}\begin{pmatrix}\pm 1\\ 0\\ -1\\ 0\end{pmatrix}. \tag{3.14}$$

Furthermore, the connection components are reduced to

$$\omega_{A3} = -\omega_{3A} = \zeta_{AB}\theta^B, \quad \omega_{AB} = 0, \tag{3.15}$$

with symmetric ζ_{AB} $(A,B = 0,1,2)$.

So far, the main EC equation, i.e. (2.13), has not been used. Taking account of (3.1) this equation reads

$$R_{ij} = \Lambda\eta_{ij}. \tag{3.16}$$

The curvature of the connection (3.15) is given by

$$\Omega_{AB} = -\zeta_A\wedge\zeta_B, \quad \Omega_{A3} = d\zeta_A, \tag{3.17}$$

and the components of the Ricci tensor are

$$R_{AB} = -\partial_3\zeta_{AB} - \zeta\zeta_{AB},$$

$$R_{A3} = R_{3A} = 0, \quad R_{33} = -\zeta^{AB}\zeta_{AB}, \tag{3.18}$$

where $\partial_3 \lrcorner\, \theta^i = \delta^i{}_3$. Now (3.16) takes the form

$$\partial_3\zeta_{AB} \pm 3m\zeta_{AB} = -\Lambda\eta_{AB}, \tag{3.19}$$

$$\zeta^{AB}\zeta_{AB} = -\Lambda. \tag{3.20}$$

Contracting (3.19) with η^{AB} and paying attention to (3.11) we obtain a relation between the cosmological constant and the mass of the Dirac field:

$$\Lambda = -3m^2. \tag{3.21}$$

This proves the following theorem.

Theorem: Massive ghost solutions of the EC–Dirac equations, with the restriction (1.1), exist only for the negative cosmological constant given by (3.21). □

In the following, we show that a considerable further reduction of the remaining equations (3.7), (3.11), (3.19), and (3.20) can be achieved.

The equation $z = z_0 = \text{const}$ determines a local hypersurface. We denote the corresponding imbedding by μ. Taking the pullback of (3.7) we find

$$d\hat{\theta}^A = (\kappa/4)\rho(z_0)\epsilon^A{}_{BC}\hat{\theta}^B \wedge \hat{\theta}^C, \tag{3.22}$$

with

$$\hat{\theta}^A \equiv \mu^*\theta^A. \tag{3.23}$$

By a rescaling of the one-forms $\hat{\theta}^A$, i.e.,

$$q^A \equiv (\kappa/2)\rho(z_0)\hat{\theta}^A, \tag{3.24}$$

Eq. (3.22) attains the form of the Maurer–Cartan equation for the Lie group SO (2,1), respectively, SL (2, $\mathbb{R}$),

$$dq^A = \tfrac{1}{2}\epsilon^A{}_{BC}q^B \wedge q^C. \tag{3.25}$$

Appendix B provides a way of solving this equation. Now it follows that

$$\theta^A = [2/\kappa\rho(z)]q^A - \rho(z)^2 f^A\, dz, \tag{3.26}$$

where f^A are functions which may depend on all space-time coordinates. The introduction of the factor ρ^2 in (3.26) simplifies some of the following expressions. Acting with the exterior derivative d on (3.26), using (3.13) and (3.25), and comparing the result with (3.7), leads to

$$\zeta_{AB} = \pm m\eta_{AB} + (\kappa/2)\rho^3 X_{(A}f_{B)}, \tag{3.27}$$

$$X_{[A}\, f_{B]} = -\epsilon_{ABC}f^C. \tag{3.28}$$

The vector fields X_A are dual to the one-forms q^A, i.e.,

$$X_A \lrcorner\, q^B = \delta^B{}_A. \tag{3.29}$$

Insertion of (3.27) in (3.11), (3.19), and (3.20) transforms these equations into

$$X^A f_A = 0, \tag{3.30}$$

$$X_{(A}\, f_{B)} X^A f^B = 0, \tag{3.31}$$

$$\left(\frac{\partial}{\partial z} + \frac{\kappa}{2}\rho^3 f^C X_C\right) X_{(A}\, f_{B)} = 0. \tag{3.32}$$

Let $\tilde{d}$ denote the exterior derivative in the three-dimensional space $\{z = z_0\}$, and $\tilde{*}$ the Hodge operator with respect to the (Cartan–Killing) metric

$$\tilde{g} = \eta_{AB}q^A \otimes q^B. \tag{3.33}$$

Introducing the one-form

$$\mathscr{F} = f_A q^A, \tag{3.34}$$

and using the Maurer–Cartan equation (3.25), Eq. (3.28) becomes

$$\tilde{d}\mathscr{F} + \tilde{*}\mathscr{F} = 0. \tag{3.35}$$

The integrability condition tells us that f is coclosed, i.e.,

$$\tilde{\delta}\mathscr{F} = 0, \tag{3.36}$$

where the coderivative acting on a p form is given by

$$\tilde{\delta} = (-1)^p \tilde{*}\tilde{d}\tilde{*}. \tag{3.37}$$

With the help of (3.25), Eq. (3.36) is seen to be equivalent to (3.30). Furthermore, as a consequence of (3.35) we find the eigenvalue equations

$$\begin{aligned} \tilde{\Delta}\mathscr{F} &= \mathscr{F}, \\ \tilde{\Delta} f_B &= X^A X_A f_B = 2 f_B, \end{aligned} \tag{3.38}$$

for the Laplace–Beltrami operator of the metric g, i.e.,

$$\tilde{\Delta} = \tilde{\delta}\tilde{d} + \tilde{d}\tilde{\delta}. \tag{3.39}$$

Equations (3.38) suggest the application of harmonic analysis on SL (2,$\mathbb{R}$). We will, however, not proceed on this route.

The problem, which we started with, is now reduced to that of solving the linear partial differential equation (3.35). Every solution of the set of equations (3.31), (3.32), and (3.35) determines a solution of the original set of equations. The coframe field θ^i can be obtained from (3.12) and (3.26), and the connection is given by (3.15) and (3.27). The metric then takes the following form:

$$\begin{aligned} g = (2/\kappa\rho)^2\tilde{g} &- (2/\kappa)\rho\,(\mathscr{F} \otimes dz + dz \otimes \mathscr{F}) \\ &+ (1 + \rho^4 f_A f^A) dz \otimes dz. \end{aligned} \tag{3.40}$$

A natural choice for the underlying manifold is $M = \mathrm{SL}(2,\mathbb{R}) \times \mathbb{R}$. However, in order to avoid closed timelike curves, we have to pass over to the universal covering space of M. As a pseudo-Riemannian space, SL(2,$\mathbb{R}$) can be identified[18] with the three-dimensional anti-de Sitter space, from which these problems are well-known.[19]

Referring to the orthonormal coframe (3.26), the Levi–Civita connection of the metric (3.40) is given by

$$\begin{aligned} \omega(g)^A{}_B &= (\kappa/4)\,\rho\epsilon^A{}_{BC}\theta^C, \\ \omega(g)^A{}_3 &= -\omega(g)_3{}^A = \zeta^A = \pm m\theta^A + (\kappa/2)\,\rho^3 X^A{}_B\theta^B, \end{aligned} \tag{3.41}$$

with

$$X_{AB} \equiv X_{(A}\, f_{B)}. \tag{3.42}$$

The corresponding components of the Riemann tensor are

$$\begin{aligned} R(g)^{AB}{}_{CD} &= -2(m^2 + (\kappa^2/16)\,\rho^2)\delta^A{}_{[C}\,\delta^B{}_{D]} \\ &\quad - (\kappa^2/2)\,\rho^6 X^A{}_{[C}X^B{}_{D]} \\ &\quad \mp \kappa m\rho^3(X^A{}_{[C}\,\delta^B{}_{D]} - X^B{}_{[C}\,\delta^A{}_{D]}), \\ R(g)_{ABD3} &= (-\kappa^2/8)\,\rho^4\epsilon_{ABC}X^C{}_D, \\ R(g)_{3ACD} &= (\kappa^2/4)\,\rho^4\epsilon_{AB[C}X^B{}_{D]}, \\ R(g)_{3A3B} &= -m^2\eta_{AB} + \frac{\kappa}{2}\rho^3\left(\frac{\partial}{\partial z}X_{AB} \pm mX_{AB}\right) \\ &\quad - (\kappa^2/4)\,\rho^6 X_{AC}X^C{}_B. \end{aligned} \tag{3.43}$$

For the calculation of the expressions (3.43) we assumed that the field equations [i.e., (3.28), (3.31), and (3.32)] are satisfied.

IV. EXACT SOLUTIONS

In terms of the coordinates used in the expression (B19) for the Maurer–Cartan one-forms q^A, the one-form

$$\begin{aligned} \mathscr{F} = e^{-y}h(x)dx = e^{-y}h(x)\,[\tfrac{1}{2}(x^2+1)q^0 \\ - \tfrac{1}{2}(x^2-1)q^1 + xq^2], \end{aligned} \tag{4.1}$$

with an arbitrary function $h(x)$ solves the system of equations (3.31), (3.32), and (3.35). According to the prescription given in the preceding section, it is now straightforward to obtain the metric and the connection associated with $\mathscr{F}$. Using (C1) and $f_A f^A = 0$, Eq. (3.40) reads

$$\begin{aligned} g = (4/\kappa\rho)^2 e^{-y}(-dt\,dx + \tfrac{1}{4}e^y\,dy^2) \\ - (4/\kappa)\rho e^{-y}h(x)dx\,dz + dz^2. \end{aligned} \tag{4.2}$$

The nonvanishing components of the connection one-form [referring to the coframe (3.26)] are

$$\omega^A{}_3 = \pm m\theta^A + (\kappa/2)\rho^3 X^A{}_B \theta^B, \tag{4.3}$$

with the nilpotent matrix

$$(X^A{}_B) = \frac{1}{4} e^{-y} \frac{dh}{dx} \times \begin{pmatrix} -(1+x^2)^2 & -(1-x^4) & -2x(1+x^2) \\ 1-x^4 & (1-x^2)^2 & 2x(1-x^2) \\ 2x(1+x^2) & 2x(1-x^2) & 4x^2 \end{pmatrix}. \tag{4.4}$$

Using (3.43), we can show that the conformal (Weyl) tensor of the metric g vanishes if h is constant (so that X_{AB} vanishes). In this special case, the geometry becomes equivalent to that determined by

$$\mathscr{F} = 0 \tag{4.5}$$

(respectively, $h = 0$), which leads to

$$\theta^A = (2/\kappa\rho)q^A, \quad \theta^3 = dz, \tag{4.6}$$

$$\omega^A{}_3 = -\omega_3{}^A = \pm m\theta^A, \quad \omega_{AB} = 0. \tag{4.7}$$

Choosing the representation (B12) of the Maurer–Cartan one-forms q^A, we recover the metric of Ref. 9:

$$g = (4/\kappa\rho_0)^2 e^{\pm 2mz}(-dt^2 + e^{-2x}dt\,dy + dx^2) + dz^2. \tag{4.8}$$

This metric admits a six-parameter group of motions with isotropy group (locally) isomorphic to SO(2,1).[9] In Appendix C a conformally flat form of the metric is derived.

All massive ghost solutions have a regular limit as the mass tends to zero. The result is a ghost solution of the EC theory without cosmological constant, i.e., $\Lambda = 0$. In the particular case of the solution determined by (4.5) the connection vanishes in the limit $m \to 0$, i.e.,

$$\omega^i{}_j = 0, \tag{4.9}$$

with reference to the frame (4.6), where ρ has now to be replaced by ρ_0. The expression (2.17) for the torsion two-form becomes (up to a rescaling) the Maurer–Cartan equation for the group SL(2,$\mathbb{R}$) $\otimes$ $\mathbb{R}$:

$$d\theta^A = (\kappa/4)\rho_0 \epsilon^A{}_{BC}\theta^B \wedge \theta^C, \quad d\theta^3 = 0. \tag{4.10}$$

This illustrates a relation betweeen a class of teleparallel connections and simply transitive Lie groups, as explained in Appendix D.

V. THE ACTION OF SL(2,$\mathbb{R}$) ON THE SET OF SOLUTIONS

Let X_A be vector fields which constitute a basis for the Lie algebra of SL(2,$\mathbb{R}$) and satisfy the canonical commutation relations

$$[X_A, X_B] = -\epsilon^C{}_{AB}X_C. \tag{5.1}$$

For a linear combination $Y = c^A X_A$ with constants c^A, let ϕ_s denote the one-parameter group of right translations generated by Y. According to the following lemma, the group SL(2,$\mathbb{R}$) [respectively, SO (2,1)] acts on the set of massive ghost solutions of the EC–Dirac equations.

Lemma: If a one-form $\mathscr{F}$ is a solution of (3.31), (3.32), and (3.35), then the pullback $\mathscr{F}' = \phi_s^* \mathscr{F}$ is again a solution.

The tetrad field, the metric, and the connection determined by $\mathscr{F}$ and $\mathscr{F}'$ are related as follows:

$$\theta'^i = (L^{-1})^i{}_j \phi_s^* \theta^j, \tag{5.2}$$

$$g' = \phi_s^* g, \tag{5.3}$$

$$\omega' = L^{-1}\phi_s^* \omega L, \tag{5.4}$$

where

$$L = \left(\begin{array}{c|c} \Lambda^A{}_B & 0 \\ \hline 0 & 1 \end{array}\right), \tag{5.5}$$

with a constant SO (2,1) matrix $(\Lambda^A{}_B)$.

Proof: Let q^A be the Maurer–Cartan one-forms dual to X_A. From

$$\mathscr{L}_{X_B} q^A = \tilde{d}(X_B \lrcorner\, q^A) + X_B \lrcorner\, \tilde{d}q^A = \epsilon^A{}_{BC}q^C, \tag{5.6}$$

it follows that

$$\phi_s^* q^A = \Lambda^A{}_B q^B, \tag{5.7}$$

with a constant SO (2,1) matrix $(\Lambda^A{}_B)$. The maps ϕ_s are isometries of the metric $\tilde{g}$, hence they commute with the $\tilde{*}$-operator and we have

$$\tilde{d}\mathscr{F}' + \tilde{*}\mathscr{F}' = \phi_s^*(\tilde{d}\mathscr{F} + \tilde{*}\mathscr{F}). \tag{5.8}$$

With the help of

$$\mathscr{F}' = \phi_s^*(f_A q^A) = (f_A \circ \phi_s)\Lambda^A{}_B q^B = f'_A q^A, \tag{5.9}$$

and

$$\phi_{s*} X_A = \Lambda^B{}_A X_B, \tag{5.10}$$

we find, furthermore,

$$X_A f'_B = \Lambda^D{}_B [(\phi_{s*} X_A) f_D] \circ \phi_s = \Lambda^C{}_A \Lambda^D{}_B \phi_s^* X_C f_D, \tag{5.11}$$

$$f'^C X_C X_{(A} f'_{B)} = \Lambda^C{}_A \Lambda^D{}_B \phi_s^* [f^E X_E X_{(C} f_{D)}]. \tag{5.12}$$

Now the first part of the lemma follows. Using (5.9) and (5.11) we obtain the expressions

$$\begin{aligned} \phi_s^* \theta^A &= \Lambda^A{}_B \theta'^B, \\ \zeta'_{AB} &= \Lambda^C{}_A \Lambda^D{}_B \phi_s^* \zeta_{CD}, \\ \omega'_{A3} &= \Lambda^B{}_A \phi_s^* \omega_{B3}, \end{aligned} \tag{5.13}$$

which immediately imply (5.2)–(5.4). □

The second part of the lemma shows that the effect of an SL(2,$\mathbb{R}$) transformation on the geometry is simply given by a constant Lorentz (gauge) transformation, together with a diffeomorphism. The spinor ψ is, however, not affected by these transformations. But we can move the SL(2,$\mathbb{R}$) freedom from the geometry to the spinor. The complete set of solutions is then obtained in the following way. Choose a single one-form $\mathscr{F}_{(\alpha)}$ from each orbit $\mathscr{O}_\alpha$ of the SL(2,$\mathbb{R}$) action. To every $\mathscr{F}_{(\alpha)}$ then belongs an "orbit of spinors," constructed by acting on (3.14) with arbitrary constant SL(2,$\mathbb{R}$) transformations

$$\psi' = S(L)\psi. \tag{5.14}$$

Here, S denotes the corresponding representation of the Lorentz group, and L is of the form (5.5).

We mention that the transformations ϕ_s are isometries, only if $X_{(A} f_{B)} = 0$.

In order to obtain the action of SL (2,$\mathbb{R}$) explicitly, it is convenient to consider the vector fields

$$Y_0 = \frac{\partial}{\partial x}, \quad Y_1 = x\frac{\partial}{\partial x} + \frac{\partial}{\partial y},$$

$$Y_2 = x^2\frac{\partial}{\partial x} + 2x\frac{\partial}{\partial y} + e^y\frac{\partial}{\partial t}, \tag{5.15}$$

which are dual to the one-forms β^A appearing in Appendix B. They generate the following transformations:

$$Y_0 \to \phi_a: \begin{cases} x \to x + a, \\ y \to y, \\ t \to t; \end{cases}$$

$$Y_1 \to \phi_b: \begin{cases} x \to xe^b, \\ y \to y + b, \\ t \to t; \end{cases} \tag{5.16}$$

$$Y_2 \to \phi_c: \begin{cases} x \to x(1-cx)^{-1}, \\ y \to y - 2\ln(1-cx), \\ t \to t + ce^y(1-cx)^{-1}. \end{cases}$$

The effect of these transformations on the vector-valued one-form $\boldsymbol{\beta} = (\beta^0, \beta^1, \beta^2)^t$ is given by

$$\phi_a^* \boldsymbol{\beta} = \begin{pmatrix} 1 & -a & a^2 \\ 0 & 1 & -2a \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{\beta},$$

$$\phi_b^* \boldsymbol{\beta} = \begin{pmatrix} e^b & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-b} \end{pmatrix} \boldsymbol{\beta},$$

$$\phi_c^* \boldsymbol{\beta} = \begin{pmatrix} 1 & 0 & 0 \\ 2c & 1 & 0 \\ c^2 & c & 1 \end{pmatrix} \boldsymbol{\beta}. \tag{5.17}$$

The corresponding transformations of the one-forms q^A used in the expression of our solution (4.1) can be obtained by the use of

$$\mathbf{q} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \boldsymbol{\beta} \tag{5.18}$$

[cf. (B19)]. It is, however, simpler to work with the one-forms β^A.

Our solution (4.1) is already a complete orbit of the SL(2,**R**) action, since the effect of the transformations (5.16) can be absorbed in the function $h(x)$.

VI. CONCLUSION

With the restrictions[20] (1.1) and $m \neq 0$ for the Dirac spinor, we reduced the problem of finding ghost solutions of the Einstein–Cartan–Dirac equations to that of solving a certain system of linear partial differential equations. The nonlinear nature of the gravitational equations is, however, still maintained in the subsidiary conditions.

The solutions were found to possess a preferred spatial direction. Hence, the suitably reduced field equations, referred to an orthonormal frame, will show an SO (2,1) covariance. In the case under consideration, the role of the group SO (2,1), which is locally isomorphic to SL (2,**R**), turned out to be much more involved. In particular, we discovered an invariance of the field equations with respect to SL (2,**R**) transformations of the spinor field, keeping the geometry fixed.

A family of exact solutions, depending on an arbitrary function of one coordinate, was obtained explicitly. Certainly, it does not exhaust the set of all solutions. A possible way of finding more general solutions was mentioned in Sec. III.

We must admit that the physical relevance of solutions with fermions treated as classical fields is not certain. The Dirac field in our solutions is not normalizable and hence does not admit a one-particle interpretation. In order to associate it with a current of fermions, we have to take care of the Pauli exclusion principle. This problem has been discussed by Ray[21] for the particular case of a classical neutrino solution in a plane wave space-time. He concludes that, in this case, the classical solution has indeed physical significance. For our solutions, further discussion is needed.

In order to anticipate the fermionic nature of the quantized Dirac field, it has been suggested to treat spinor fields as (anticommuting) Grassmann variables.[22] We expect that some of the techniques used in the present work can be applied in this case also.

Besides that, it would be advantageous to include an electromagnetic field, so that the massive Dirac fields could represent electrons.

We finally mention that the notion "ghost" is not really adequate for the class of solutions investigated in this work. Although the energy-momentum tensor vanishes, the geometry is still influenced by the spin tensor. The latter is different from zero if the condition (1.1) holds.

In the case $\bar{\psi}\psi = 0 = \bar{\psi}\gamma_5\psi$ there are, however, also nontrivial solutions, for which both the energy-momentum and spin tensor vanish. These "true ghosts" will be the subject of a separate work.

APPENDIX A: THE DIRAC SPINOR IN A SPECIAL FRAME

With respect to orthonormal frames, the metric is given by $\eta = \mathrm{diag}(-1,1,1,1)$. For the γ matrices we use the representation

$$\gamma_0 = i\begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}, \quad \gamma_\alpha = \begin{pmatrix} 0 & \sigma_\alpha \\ \sigma_\alpha & 0 \end{pmatrix}$$

$$(\alpha = 1,2,3)$$

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{A1}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} 0 & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}.$$

The following identities are useful:

$$[\gamma_i\gamma_j]\gamma_5 = \tfrac{1}{2}\epsilon_{ijkl}[\gamma^k,\gamma^l], \tag{A2}$$

$$\tfrac{1}{2}\gamma_i[\gamma_j,\gamma_k] = 2\eta_{i[j}\gamma_{k]} - \epsilon_{ijkl}\gamma^l\gamma_5. \tag{A3}$$

Here, ϵ_{ijkl} is totally antisymmetric with $\epsilon_{0123} = 1$, and (anti-) symmetrization is understood with factors, e.g., $\gamma_{[i}\gamma_{j]} = \frac{1}{2}(\gamma_i\gamma_j - \gamma_j\gamma_i)$.

For an arbitrary element A of the *real* Dirac Clifford algebra[23] $\mathscr{C}$, let

$$\overline{A} \equiv i\gamma^0 A^\dagger i\gamma^0, \tag{A4}$$

where $A^\dagger$ denotes the Hermitian conjugate. This defines an anti-involution of $\mathscr{C}$ with $\overline{AB} = \overline{B}\,\overline{A}$ $(\forall A,B\in\mathscr{C})$.

An involutive automorphism of $\mathscr{C}$ is induced by

$$\hat{\mathbf{1}} \equiv \mathbf{1}, \quad \hat{\gamma}^i \equiv -\gamma^i. \tag{A5}$$

It satisfies $\widehat{AB} = \hat{A}\hat{B}$.

Lemma 1: Let $A\in\mathscr{C}$. Then $\hat{A} = \overline{A} = -A$ implies $A = a_i\gamma^i$, with $a_i\in\mathbb{R}$.

Proof: Here, A can be written as a sum of products of γ's:

$$A = A_0 + A_1 + A_2 + A_3 + A_4, \tag{A6}$$

where an index indicates the number of γ's occurring in the corresponding term. Now

$$-A = \hat{A} = A_0 - A_1 + A_2 - A_3 + A_4 \tag{A7}$$

leads to $A = A_1 + A_3$, and $\overline{A} = -A$ eliminates A_3. □

Lemma 2: Let $S\in\mathscr{C}$ with $\hat{S} = S$ and $\overline{S}S = \mathbf{1}$. Then

$$S\gamma^i\overline{S} = L^i{}_j\gamma^j, \tag{A8}$$

with a Lorentz matrix L.

Proof: Let $A^i \equiv S\gamma^i\overline{S}$. Then

$$\overline{A}^i = S\overline{\gamma}^i\overline{S} = -A^i, \tag{A9}$$

and with the help of $\hat{S} = S$, we find

$$\hat{A}^i = \hat{S}\hat{\gamma}^i\hat{\overline{S}} = -A^i. \tag{A10}$$

Now Lemma 1 shows that

$$A^i = L^i{}_j\gamma^j, \tag{A11}$$

with real $L^i{}_j$. Using the defining relation of the Clifford algebra, we have

$$\begin{aligned} 0 &= S(\gamma^i\gamma^j + \gamma^j\gamma^i - 2\eta^{ij}\mathbf{1})\overline{S} \\ &= S(\gamma^i\overline{S}S\gamma^j + \gamma^j\overline{S}S\gamma^i - 2\eta^{ij}\mathbf{1})\overline{S} \\ &= 2(L^i{}_kL^j{}_l\eta^{kl} - \eta^{ij}). \end{aligned} \tag{A12}$$

Hence, L is a Lorentz matrix. □

For a Dirac spinor ψ, the adjoint is given by

$$\overline{\psi} \equiv \psi^\dagger i\gamma^0. \tag{A13}$$

Theorem: Let (M,g) be a four-dimensional (noncompact) space-time admitting a spinor structure,[24] and ψ a Dirac spinor field on M.

(1) If $(\overline{\psi}\psi)^2 + (\overline{\psi}\gamma_5\psi)^2 > 0$, an SL (2,C)frame exists with respect to which the components of ψ are

$$\psi' = \sqrt{\rho}\begin{pmatrix} \cos\beta/2 \\ 0 \\ -\sin\beta/2 \\ 0 \end{pmatrix}, \tag{A14}$$

with real functions β and $\rho > 0$.[8]

(2)If $\overline{\psi}\psi = \overline{\psi}\gamma_5\psi = 0$, a frame can be chosen such that

$$\psi' = \mu\begin{pmatrix} 1 \\ 0 \\ if^3 \\ -f^2 + if^1 \end{pmatrix}, \tag{A15}$$

with real functions μ, f^α $(\alpha = 1,2,3)$, restricted by $f^\alpha f_\alpha = 1$.

Proof: The assumption of a spinor structure implies the existence of a global "SL (2,C) frame" on M, i.e., a global cross section of the SL (2,C) principal fiber bundle covering the $SO_+(3,1)$ bundle of (restricted) g-orthonormal frames.[24] Let $\psi = (\phi_1,\phi_2,\chi_1,\chi_2)^t$ be the corresponding components of the Dirac field. We define

$$A = \begin{pmatrix} \phi_1 & -\phi_2^* \\ \phi_2 & \phi_1^* \end{pmatrix}, \quad B = \begin{pmatrix} \chi_1 & -\chi_2^* \\ \chi_2 & \chi_1^* \end{pmatrix},$$

$$\Psi = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}, \tag{A16}$$

where an asterisk means complex conjugation. The matrix Ψ has the properties

$$\begin{aligned} &\overline{\Psi}\Psi = (\overline{\psi}\psi)\mathbf{1} + (\overline{\psi}\gamma_5\psi)\gamma^5 \equiv \rho e^{\beta\gamma^5}, \\ &\det\Psi = \rho^2, \\ &\psi = \Psi(1\,0\,0\,0)^t. \end{aligned} \tag{A17}$$

(1) It follows that

$$S \equiv \Psi(1/\sqrt{\rho})e^{-(1/2)\beta\gamma^5} \tag{A18}$$

satisfies the equations

$$\overline{S}S = \mathbf{1}, \quad \det S = 1. \tag{A19}$$

Furthermore, Ψ and thus also S can be written as a real linear combination of $\mathbf{1}$, $[\gamma^i,\gamma^j]$, and γ^5, so that $\hat{S} = S$. Now we infer from Lemma 2 that S represents an SL (2,C) transformation. The components of the spinor in the new gauge are

$$\psi' = S^{-1}\psi = \overline{S}\psi = \overline{S}\Psi\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \sqrt{\rho}\,e^{(1/2)\beta\gamma^5}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{A20}$$

(2) In this case, we have $\rho = 0$ and, assuming $\psi \neq 0$,

$$\mu^2 \equiv \phi^\dagger\phi = \chi^\dagger\chi > 0. \tag{A21}$$

Then $U \equiv (1/\mu)A$ has values in SU (2), and

$$S \equiv \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix} \tag{A22}$$

represents an SL (2,C) transformation, with $\overline{S} = S^\dagger$. We find

$$S^\dagger\Psi = \mu\begin{pmatrix} \mathbf{1} & -V \\ V & \mathbf{1} \end{pmatrix}, \tag{A23}$$

where

$$V \equiv (1/\mu)U^\dagger B \tag{A24}$$

is an anti-Hermitian SU(2) matrix as a consequence of $\overline{\Psi}\Psi = 0$. It follows that

$$V = if^\alpha\sigma_\alpha, \quad \delta_{\alpha\beta}f^\alpha f^\beta = 1, \tag{A25}$$

with real functions f^α $(\alpha,\beta = 1,2,3)$. Now

$$\psi' \equiv S^{-1}\psi = S^\dagger\Psi\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{A26}$$

leads to the form of the spinor as stated above. □

Massless neutrino fields satisfy the assumption of part (2) of the theorem. The chirality condition

$$\psi = i\gamma_5\psi \tag{A27}$$

reduces ψ' to

$$\psi' = \mu \begin{pmatrix} 1 \\ 0 \\ i \\ 0 \end{pmatrix}. \tag{A28}$$

APPENDIX B: SOLUTION OF THE SL (2,ℝ) MAURER–CARTAN EQUATION

Let $g(t)$ be a one-parameter subgroup of a Lie group G. Then L_g and R_g denote the left and right action on G, respectively, by an element $g\in G$. Let Y be the vector field on G, which generates the transformation $R_{g(t)}$. From

$$L_{g'} \circ R_{g(t)}(g) = R_{g(t)}(g'.g) \quad (\forall g,g' \in G), \tag{B1}$$

it follows that

$$L_{g',*}(Y|_g) = Y|_{g'g}, \tag{B2}$$

i.e., Y is left invariant and thus an element of the Lie algebra of G.[25]

In the case $G = \mathrm{SL}(2,\mathbb{R})$ we use the Iwasawa decomposition

$$g(t,x,y) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \times \begin{pmatrix} e^x & 0 \\ 0 & e^{-x} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}. \tag{B3}$$

In the coordinates t,x,y, the right translation by $g(0,x',0)$ and $g(0,0,y')$ is easily obtained:

$$\begin{aligned} &R_{g(0,x',0)}: t \to t, \quad x \to x + x', \quad y \to y e^{2x'}, \\ &R_{g(0,0,y')}: t \to t, \quad x \to x, \quad y \to y + y'. \end{aligned} \tag{B4}$$

The corresponding generators are therefore

$$Y_1 = \frac{\partial}{\partial x} + 2y \frac{\partial}{\partial y}, \quad Y_2 = \frac{\partial}{\partial y}. \tag{B5}$$

For the right action of $g(t',0,0)$ the new coordinates $\tilde{t},\tilde{x},\tilde{y}$, determined by

$$g(t,x,y) \cdot g(t',0,0) = g(\tilde{t},\tilde{x},\tilde{y}), \tag{B6}$$

are more complicated. Explicitly, the last equation reads

$$\begin{aligned} e^{-\tilde{x}} \sin \tilde{t} &= (e^x \cos t + y e^{-x} \sin t) \sin t' \\ &\quad + e^{-x} \sin t \cos t', \\ e^{-\tilde{x}} \cos \tilde{t} &= (-e^x \sin t + y e^{-x} \cos t) \sin t' \\ &\quad + e^{-x} \cos t \cos t', \\ e^{\tilde{x}} \cos \tilde{t} + \tilde{y} e^{-\tilde{x}} \sin \tilde{t} &= (e^x \cos t + y e^{-x} \sin t) \cos t' \\ &\quad - e^{-x} \sin t \sin t'. \end{aligned} \tag{B7}$$

Differentiation of these equations with respect to t', and evaluation of the result at $t' = 0$ gives the components of the generator

$$Y_0 = e^{2x} \frac{\partial}{\partial t} - y \frac{\partial}{\partial x} + (e^{4x} - 1 - y^2) \frac{\partial}{\partial y}. \tag{B8}$$

Passing over to the linear combinations

$$X_0 = \tfrac{1}{2} Y_0, \quad X_1 = \tfrac{1}{2} Y_1, \quad X_2 = \tfrac{1}{2} Y_0 + Y_2, \tag{B9}$$

we obtain the canonical form of the commutation relations:

$$[X_B, X_C] = -\epsilon^A{}_{BC} X_A. \tag{B10}$$

The left-invariant Maurer–Cartan one-forms q^A are now determined by

$$X_A \lrcorner\, q^B = \delta^B{}_A. \tag{B11}$$

This leads to

$$\begin{aligned} q^0 &= (y^2 e^{-2x} + 2 \cosh 2x) dt - dy + 2y\, dx, \\ q^1 &= 2y e^{-2x} dt + 2\, dx, \\ q^2 &= -(y^2 e^{-2x} + 2 \sinh 2x) dt + dy - 2y\, dx, \end{aligned} \tag{B12}$$

which provides a solution of (3.25).

Another way of solving (3.25) will be outlined in the following. The vector fields

$$Y_0 = \frac{\partial}{\partial v}, \quad Y_1 = v \frac{\partial}{\partial v}, \quad Y_2 = v^2 \frac{\partial}{\partial v}, \tag{B13}$$

on $\mathbb{R}^1$, provide a realization of the Lie algebra of SL (2,ℝ). The commutation relations are

$$\begin{aligned} &[Y_0, Y_1] = Y_0, \quad [Y_0, Y_2] = 2Y_1, \\ &[Y_1, Y_2] = Y_2. \end{aligned} \tag{B14}$$

Let β^A be a solution of the Maurer–Cartan equation on SL (2,ℝ) with the structure constants given by (B14), i.e.,

$$\begin{aligned} &d\beta^0 = -\beta^0 \wedge \beta^1, \quad d\beta^1 = -2\beta^0 \wedge \beta^2, \\ &d\beta^2 = -\beta^1 \wedge \beta^2. \end{aligned} \tag{B15}$$

The (finite) transformations $\bar{v} = f(v,a), a \in \mathrm{SL}(2,\mathbb{R})$ are then determined by

$$d_G \bar{v} = \beta^0 + \bar{v}\beta^1 + \bar{v}\beta^2, \tag{B16}$$

where d_G is the exterior derivative with respect to the group coordinates. The first and second extension (in the sense of Eisenhart[12]) are given by

$$\begin{aligned} &d_G \bar{v}_1 = \bar{v}_1 \beta^1 + 2\bar{v}\, \bar{v}_1 \beta^2, \\ &d_G \bar{v}_2 = \bar{v}_2 \beta^1 + 2(\bar{v}_1{}^2 + \bar{v}\, \bar{v}_2) \beta^2. \end{aligned} \tag{B17}$$

The equations (B16) and (B17) can be solved for β^A. In terms of the new coordinates given by

$$\bar{v} = x, \quad \bar{v}_1 = e^y, \quad \bar{v}_2 = 2e^y t, \tag{B18}$$

we finally obtain the following solution of (3.25):

$$\begin{aligned} q^0 &\equiv \beta^0 + \beta^2 = (x^2 + 1) e^{-y} dt + dx - x\, dy, \\ q^1 &\equiv \beta^0 - \beta^2 = (x^2 - 1) e^{-y} dt + dx - x\, dy, \\ q^2 &\equiv \beta^1 = -2x e^{-y} dt + dy. \end{aligned} \tag{B19}$$

Any two solutions of a Maurer–Cartan equation are related by a diffeomorphism.[26] This applies to (B12) and (B19). In general, it is a difficult task to find this diffeomorphism. Since a suitable choice for the expression of the one-forms q^A can simplify a given problem considerably, it is worthwhile to have different methods for solving the Maurer–Cartan equation available.

APPENDIX C: CONFORMALLY FLAT FORM OF THE METRIC (4.8)

Using the representation (B19) of the Maurer–Cartan forms on SL(2,ℝ), the Cartan–Killing metric (3.33) attains the form

$$\tilde{g} = -4e^{-y}\,dt\,dx + dy^2. \tag{C1}$$

Expressed in the new coordinates

$$\tau = \tfrac{1}{2}(t+x), \quad \xi = \tfrac{1}{2}(t-x), \quad r = e^{y/2}, \tag{C2}$$

the metric becomes conformally flat,

$$\tilde{g} = (4/r^2)(-d\tau^2 + d\xi^2 + dr^2). \tag{C3}$$

Assuming $m \neq 0$, and choosing the new coordinate

$$\phi = (\kappa/4m)\rho_0 e^{\mp mz} \tag{C4}$$

instead of z, a conformally flat form of the metric determined by (4.6) is obtained:

$$g = (mr\phi)^{-2}(-d\tau^2 + d\xi^2 + dr^2 + r^2\,d\phi^2)$$
$$(\tau,\xi \in \mathbb{R};\ r > 0,\ \phi > 0). \tag{C5}$$

Interpreting r, ϕ, ξ as cylindrical coordinates on $\mathbb{R}^3$, we may think of g as being defined on the universal covering space $\widetilde{M}$ of $\mathbf{R}\times(\mathbf{R}^3\setminus\{r=0\})$. From the point of view of an intrinsic observer, the regions $\{r=0\}$ and $\{\phi=0\}$ are infinitely far away. Geodesics (with affine parameter) satisfy the equations

$$(\dot{r}/r)^2 = (m\phi)^2[\epsilon + (c_1{}^2 - c_2{}^2)(mr\phi)^2] - \dot{\phi}^2,$$
$$\dot{\phi}^2 = c_3 e^{4\phi} + (\epsilon/8)m^2(1+4\phi), \quad \text{if } \dot{\phi} \neq 0, \tag{C6}$$
$$\phi = \phi_0, \quad \text{if } \epsilon = 0 \text{ and } \dot{\phi} = 0,$$

where c_1, c_2, c_3 are constants, and $\epsilon \in \{-1, 0, +1\}$ decides whether the curve is timelike, null, or spacelike. Using these equations, it can be shown that each point in $\widetilde{M}$ has an infinite affine parameter distance from $\{r=0\}$ and $\{\phi=0\}$.

In order to cover also the case $m=0$, the definition of the new coordinate ϕ has to be replaced by $\phi = \kappa\rho_0 z/4$, and the conformal factor in (C5) has to be exchanged by $(4/\kappa\rho_0 r)^2$. The metric is then geodesically complete on $\mathbf{R}^4\setminus\{r=0\}$, which is homeomorphic to $\mathbf{R}^3\times S^1$. In this space, "one can walk around infinity."

APPENDIX D: RELATION BETWEEN TELEPARALLEL CONNECTIONS AND SIMPLY TRANSITIVE LIE GROUPS

On an n-dimensional manifold M, we consider a teleparallel linear connection with covariantly constant torsion, i.e.,

$$\Omega^i{}_j = 0, \tag{D1}$$

$$\nabla_l Q^k{}_{ij} = 0. \tag{D2}$$

The first Bianchi identity

$$D\Theta^i = \Omega^i{}_j \wedge \theta^j, \tag{D3}$$

where D denotes the covariant exterior derivative,[16,25] then reduces to

$$Q^i{}_{j[m}Q^j{}_{kl]} = 0. \tag{D4}$$

Equation (D1) implies the local existence of a "teleparallel" coframe θ^i, with respect to which the connection vanishes, i.e.,

$$\omega^i{}_j = 0, \tag{D5}$$

so that (D2) requires the components of the torsion tensor to be constant:

$$Q^i{}_{kl} = C^i{}_{kl}. \tag{D6}$$

Now (D4) turns out to be the Jacobi "identity"

$$C^j{}_{[kl}C^i{}_{m]j} = 0. \tag{D7}$$

According to Lie's third fundamental theorem,[12] the constants $C^i{}_{kl}$ are the structure constants of a Lie group G. The structure equation (2.2) for the torsion two-form becomes

$$d\theta^i = \tfrac{1}{2}C^i{}_{kl}\theta^k \wedge \theta^l, \tag{D8}$$

and the vector fields X_i, dual to θ^i, satisfy

$$[X_i, X_j] = -C^k{}_{ij}X_k, \tag{D9}$$

showing that the X_i, $i = 1,\ldots,n$, generate a simply transitive action of G on M (if we require θ^i to exist globally on M).

Starting with a simply transitive group action on a manifold M, we can construct a left-invariant frame X_i on M and define a teleparallel connection via $\omega^i{}_j = 0$ with respect to it. The torsion components in this frame then coincide with the structure constants of the group.

This establishes a correspondence between teleparallel connections (in the restricted sense that a global teleparallel frame exists) with covariantly constant torsion, and simply transitive Lie groups.[27]

[1] J. B. Griffiths, Int. J. Theor. Phys. **5**, 141 (1972); Gen. Relativ. Gravit. **8**, 365 (1977); Nukleonika **25**, 1415 (1980).

[2] J. B. Griffiths, Commun. Math. Phys. **28**, 295 (1972); C. D. Collinson and P. B. Morris, J. Phys. A **6**, 915 (1973); J. Audretsch, Phys. Lett. A **56**, 15 (1976); K. R. Pechenick and J. M. Cohen, Phys. Rev. D **19**, 1635 (1979); K. D. Krori, T. Chaudhury, and R. Bhattacharjee, Phys. Rev. D **25**, 1492 (1982).

[3] T. M. Davis and J. R. Ray, Phys. Rev. D **9**, 331 (1974); J. Math. Phys. **16**, 75, 80 (1975); **17**, 1049 (1976).

[4] J. B. Griffiths, J. Phys. A **12**, 2429 (1979).

[5] J. B. Griffiths, Phys. Lett. A **75**, 441 (1980); Gen. Relativ. Gravit. **13**, 227 (1981). See also P. S. Letelier, Phys. Lett. A **54**, 351 (1975); **79**, 283 (1980).

[6] F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).

[7] In order to be more rigorous we would have to introduce a spinor structure (see Ref. 24). The spinor ψ then refers to an SL (2,C) gauge. This is, however, a mere technical complication which we will ignore for the sake of a simpler notation. We simply refer ψ to a g-orthonormal (co-)frame.

[8] A similar result has been obtained in a different framework by D. Hestenes, J. Math. Phys. **8**, 798 (1967); **14**, 893 (1973).

[9] A. Dimakis and F. Müller-Hoissen, Phys. Lett. A **92**, 431 (1982). See also A. Dimakis, Ph. D. dissertation, University of Göttingen, 1983.

[10] A. Dimakis and F. Müller–Hoissen, in *10th International Conference on General Relativity and Gravitation, Contributed Papers*, edited by B. Bertotti, F. de Felice, and A. Pascolini (Consiglio Nazionale delle Ricerche, Rome, 1983), p. 505.

[11] See, e.g., F. W. Hehl, in *Cosmology and Gravitation*, edited by P. G. Bergmann and V. de Sabbata (Plenum, New York, 1980), p. 5.

[12] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).

[13] Y. Choquet-Bruhat, C. de Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, Amsterdam, 1977).

[14] B. K. Datta, Nuovo Cimento B **6**, 1, 16 (1971).

[15] Indices $i,j,\ldots$ run from 0 to 3. Capital Latin indices run from 0 to 2. See also Appendix A for conventions used in this work.

[16] A. Trautman, Sym. Math. **12**, 139 (1973).

[17] $\kappa = 8\pi G$, G is Newton's gravitational constant. We allow the presence of a cosmological constant Λ.

[18] K. Nomizu, Osaka J. Math. **16**, 143 (1979).

[19] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., London, 1973).

[20] The case $\bar{\psi}\psi = \bar{\psi}\gamma_5\psi = 0$ will be treated elsewhere (see also Appendix A).

As a subcase, this includes massless neutrinos.
[21]J. R. Ray, Lett. Nuovo Cimento **21**, 453 (1978).
[22]T. Dereli and R. W. Tucker, Phys. Lett. A **82**, 229 (1981).
[23]$\mathscr{C}$ consists of real linear combinations of products of the γ's.
[24]R. Geroch, J. Math. Phys. **9**, 1739 (1968).
[25]See, e.g., S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963).
[26]S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, NJ, 1964), p. 220 Theorem 2.4.
[27]This result is due to Eisenhart, see Ref. 12, pp. 192–197. Compare also É. Cartan, Actual. Sci. Ind., **44** (1932). Appendix D is a translation into more modern language.

Equilibrium states of inhomogeneous mean-field quantum spin systems with random sites

A. Blobel and J. Messer
Sektion Physik, Theoretische Physik, Universität München, Theresienstrasse 37, D-8000 München 2, Federal Republic of Germany

(Received 1 November 1983; accepted for publication 7 September 1984)

For inhomogeneous mean-field quantum spin systems with random sites, we prove that the weak*-limit points of the sequence of n-particle Gibbs states, on a suitable C^*-algebra, are superpositions of inhomogeneous product states, which are given by solutions of a generalized Curie–Weiss equation. Under certain conditions there is only one weak*-limit point. As applications, the BCS model and a spin system with cosine interaction are studied.

I. INTRODUCTION

We consider a quantum mechanical system, given by the Hamiltonian

$$H_n = \sum_{i=1}^{n} \mathbf{B}(x_i)\boldsymbol{\sigma}_i + (2n)^{-1} \sum_{\substack{i,j=1\\ i\neq j}}^{n} J(x_i,x_j)\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j. \tag{1.1}$$

Here, $\mathbf{B}$ represents an external magnetic field, J is a space-dependent two-body spin interaction, and $\boldsymbol{\sigma}_i$ is the vector of Pauli matrices. Let $x_i \in \Lambda \subset \mathbf{R}^\nu$ be the locations of the sites at which the particles are attached. The Hamiltonian (1.1) describes a system of atoms, for which the positional degrees of freedom are frozen in, or "quenched." The internal degrees of freedom, the spins, are thermally distributed; the locations of the particles are random with uniform distribution.

We study the correlation functions averaged in a quenched way. These objects are obtained by first forming the quantum mechanical expectation value with respect to the thermal distribution of the spin degrees of freedom at a fixed configuration of sites, and afterwards averaging over the possible locations with a uniform distribution. In this way, e.g., the average spin density is obtained as

$$\frac{1}{|\Lambda|^n}\int dx_1 \ldots dx_n \, \frac{1}{n}\sum_{i=1}^{n} \delta(x-x_j)\langle \boldsymbol{\sigma}_j\rangle (x_1,\ldots,x_n),$$

with the quantum mechanical average

$$\langle \boldsymbol{\sigma}_j \rangle = \operatorname{tr} \boldsymbol{\sigma}_j e^{-\beta H_n}/\operatorname{tr} e^{-\beta H_n},$$

at a fixed configuration of sites $(x_1,\ldots,x_n)$. Similarly, one obtains the higher-order correlation functions.

We compute these equilibrium correlation functions in a limit $n\to\infty$, but with fixed volume $\Lambda \subset \mathbf{R}^\nu$. In this limit the distance between two neighboring sites becomes small compared to the range of the interaction. Due to the mean-field character of the Hamiltonian, we obtain then a generalized Curie–Weiss equation for the average spin density from a variational problem. Although the distribution of the random sites is uniform, the average magnetization density is position dependent. This inhomogeneity is caused by the spatial dependence of the interaction. If the interaction is constant, our results are identical with those of a mean-field quantum lattice system.

With respect to the scaling of the Hamiltonian and the inhomogeneity of the resulting average magnetization density in the limit, our quantum spin system with random locations bears some resemblance to Thomas–Fermi theory.[1–4] In particular, we expect a similar outcome for a ν-dimensional quantum lattice system in a fixed box, with interaction scaled by the inverse number of spins, and a lattice constant shrinking as the $(1/\nu)$-th power of the number of spins. This is the quantum analog of a classical problem solved by Eisele and Ellis.[5,6] The analog to the Curie–Weiss equation, for a continuous system of fermions, is the Thomas–Fermi equation;[1–4] for a continuous system of classical particles, the Lane–Emden equation.[7] For bosons, a set of self-consistency equations has been derived recently for the Jellium model.[8] The connection with Thomas–Fermi theory motivated our investigation.

In this paper we are able to treat more general mean-field Hamiltonians than (1.1) with finite spin and a more general notion of equilibrium correlation functions. We describe our system in the framework of algebraic statistical mechanics. As an example of an observable, the spin density at location $a\in\Lambda$, for n particles, is given by

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i f_a(x_i),$$

where $\boldsymbol{\sigma}_i$ is the vector of Pauli matrices and f_a is a continuous function on Λ, supported in a small neighborhood of $a\in\Lambda$. According to this example, the C^*-algebra of observables is chosen as the infinite tensor product of identical copies of the space of continuous functions from Λ to B, where B denotes the algebra of bounded operators in a finite-dimensional Hilbert space of spin degrees of freedom. We now ask for the Gibbs equilibrium states as positive, normalized linear functionals on the C^*-algebra in the limit, where the number of spins tends toward infinity, and the box is kept fixed.

Our results are the following: The sequence of n-particle Gibbs states on our algebra has accumulation points in the weak*-topology. It turns out that each of these limit points is a superposition of inhomogeneous product states, which globally minimize the free energy functional. They satisfy a matrix self-consistency equation, which is a general-

 0022-2488/85/051049-08$02.50

ized Curie–Weiss equation. For high enough temperature, in general or under sufficient symmetry conditions, there is a unique limiting Gibbs state. If the interaction J is independent of the positions, our results reduce to those for homogeneous mean-field systems.[9]

As an application, we study a spin system with cosine interaction and external field, where we determine the phase diagram which shows three different phase regions and a continuum of pure equilibrium states. Our model still allows for the following abstraction: the particles may be viewed as states, which are empty or filled with real particles. For example, in the BCS model the density of Cooper pair states around the Fermi surface can be considered as uniform. This leads us to the gap equation of van Hemmen.[10] Another possible application (which is not elaborated here) is a model of gravitating fermions. There, one might view the phase space cells as uniformly distributed. This yields the Thomas–Fermi equation.

The general theory is developed in Secs. II–IV, whereas we apply our methods to the BCS and cosine model in Sec. V. In order not to reproduce already known facts and techniques, we have stressed brevity.

II. BASIC NOTIONS

Let h be a complex Hilbert space with dimension $d<\infty$. Then h represents the internal degrees of freedom of a particle. Define

$$h_n := \underbrace{h\otimes\cdots\otimes h}_{n},$$

and denote by B_n the set of bounded, linear operators on h_n equipped with the usual operator norm $\|\cdot\|_{B_n}$. B_n is a C^*-algebra with identity 1_{B_n}.

Given an orthonormal basis (ONB) $\{\varphi_i | i=1,\dots,d\}$ of h, the set

$$\{\varphi_{i_1}\otimes\cdots\otimes\varphi_{i_n} | i_1,\dots,i_n = 1,\dots,d\}$$

forms an ONB of h_n. With respect to this basis, B_n may be regarded as the set of complex $d^n\times d^n$ matrices. We will often make use of this interpretation. Now let $\Lambda\subset\mathbb{R}^\nu$; $\nu\in\mathbb{N}$, be compact. Λ stands for the "box" containing the particles. Define

$$A_n := \{f{:}\Lambda^n\to B_n | f \text{ continuous}\},$$

with the norm

$$\|\cdot\|_n := \sup\{\|\cdot(x)\|_{B_n} | x\in\Lambda^n\},$$

and

$$L^1_n := \{f{:}\Lambda^n\to B_n | f \text{ integrable with respect to the Lebesgue measure } \lambda \text{ on } \Lambda^n\},$$

with the norm

$$\|\cdot\|^1_n := |\Lambda|^{-n}\int_{\Lambda^n} dx\|\cdot(x)\|_{B_n}\quad (|\Lambda| := \lambda(\Lambda)).$$

The attributes "continuous" and "integrable" refer to the components of f, viewed as a matrix-valued function.

A_n and L^1_n are separable Banach spaces, and A_n is dense in L^1_n. This follows from the corresponding well-known properties of the sets $C(\Lambda^n,\mathbb{C})$ and $L^1(\Lambda^n,\mathbb{C})$ of continuous, resp. λ-integrable, complex-valued functions on Λ^n. Moreover, A_n is a C^*-algebra with identity $1_n(x)=1_{B_n}$; $x\in\Lambda^n$ (the product of two elements is defined pointwise).

Now for $m<n$ consider the injection $i_{mn}{:}L^1_m\to L^1_n$, defined by

$$i_{mn}(f)=f\otimes 1_{n-m};\quad f\in L^1_m,$$

wherein $\otimes$ is defined pointwise as a tensor product between operators. One easily verifies that

$$\|i_{mn}(f)\|^1_n = \|f\|^1_m;\quad f\in L^1_m,$$

and

$$\|i_{mn}(B)\|_n = \|B\|_m;\quad B\in A_m.$$

Therefore, identifying A_m and L^1_m with their images under i_{mn}, we may look upon them as subsets of A_n resp. L^1_n and form

$$A=\bigcup_{n\geqslant 1} A_n,\quad L^1=\bigcup_{n\geqslant 1} L^1_n,$$

with norms $\|\cdot\|$, $\|\cdot\|^1$, and smallest completions $\overline{A}$, $\overline{L}^1$. The completions $\overline{A}$ and $\overline{L}^1$ are separable Banach spaces, and $\overline{A}$ is dense in $\overline{L}^1$. Since $\overline{A}$ is the C^*-inductive limit of $\{A_n | n\in\mathbb{N}\}$ it forms a separable C^*-algebra with identity 1. Therefore the corresponding state space E, equipped with the weak*-topology is compact and metrizable. A state $\omega\in E$ is called *symmetric*, if

$$\omega(T^n_\pi(f))=\omega(f);\quad n\in\mathbb{N},\ f\in A_n,\ \pi\in P_n,$$

where P_n denotes the set of permutations of $\{1,\dots,n\}$, and $T^n_\pi{:}L^1_n\to L^1_n$ is defined by

$$T^n_\pi(f)(x_1,\dots,x_n) = P^n_\pi f(x_{\pi^{-1}(1)},\dots,x_{\pi^{-1}(n)})P^n_\pi;\quad f\in L^1_n,$$

with

$$P^n_\pi(\varphi_{i_1}\otimes\dots\otimes\varphi_{i_n}) = \varphi_{i_{\pi(1)}}\otimes\dots\otimes\varphi_{i_{\pi(n)}};\ \varphi_{i_1},\dots,\varphi_{i_n}\in\text{ONB}.$$

The set S of symmetric states is a convex, closed subset of E. A state $\omega\in E$ is said to be a *product state*, if

$$\omega(f_1\otimes f_2)=\omega(f_1)\omega(f_2);\quad f_1,f_2\in\overline{A}.$$

Let P denote the set of product states.

Another class of states is given by

$$R := \{\omega\in E \mid |\omega(f)|\leqslant\|f\|^1; f\in\overline{A}\}.$$

Obviously R is convex and closed. As $\overline{A}$ is dense in $\overline{L}^1$, each $\omega\in R$ has a unique extension $\overline{\omega}$ onto $\overline{L}^1$. Note the following important feature of R. First define

$$R_n := \{\rho\in L^1_n | \rho(x)\geqslant 0,\ \mathrm{tr}_n(\rho(x)) = |\Lambda|^{-n}; x\in\Lambda^n\}.$$

Now each $\omega\in R$ determines a unique hierarchy $\{\rho^\omega_n\in R_n | n\in\mathbb{N}\}$, such that

$$\omega(f)=\int_{\Lambda^n} dx\,\mathrm{tr}_n(\rho^\omega_n(x)f(x));\quad n\in\mathbb{N},\ f\in A_n. \tag{2.1}$$

For consistency $\{\rho^\omega_n | n\in\mathbb{N}\}$ must satisfy the relations

$$\rho^\omega_m = \mathrm{Tr}_{n,\{m+1,\dots,n\}}(\rho^\omega_n);\quad m<n, \tag{2.2}$$

where $\mathrm{Tr}_{n,\{m+1,\ldots,n\}}(\cdot)$ abbreviates

$$\int_{\Lambda^{n-m}} dx_{m+1} \ldots dx_n \, \mathrm{tr}_{n,\{m+1,\ldots,n\}}(\cdot(x_1,\ldots,x_n)),$$

$\mathrm{tr}_{n,\{m+1,\ldots,n\}}$ denoting the partial trace over the Hilbert spaces $m+1,\ldots,n$. The proof is based on the Riesz representation theorem $L^1(\Lambda^n,\mathbb{R})^* = L^\infty(\Lambda^n,\mathbb{R})$. Conversely, each consistent hierarchy constitutes by (2.1) a unique state $\omega \in R$. Accordingly, the states $\omega \in R$ are said to be *spatially homogeneous.*

Within the scope of this paper, by the notion "random system," we mean a set of identical, stationary particles, distributed uniformly over Λ. Hence $R \cap S$ is called the set of random system states, the symmetry taking into account the identity of the particles.

We now introduce the Hamiltonians $\{H_n \mid n \in \mathbb{N}\}$ and the corresponding Gibbs sequence $\{\omega_{\mathrm{G}n} \mid n \in \mathbb{N}\}$. It will turn out that the weak*-limit points are random system states.

Choose the external potential $U \in L^1_1$ and the two-body interaction $V \in L^1_2$, both self-adjoint and bounded with constants C_U and C_V, respectively. Furthermore, let V be symmetric, i.e.,

$$T^2_\pi(V) = V; \quad \pi \in P_2.$$

For $n \geqslant 2$, $i,j \in \{1,\ldots,n\}$, $i \neq j$, the elements

$$U^n_i = T^n_\pi(U \otimes 1_{n-1}); \quad \pi \in P_n, \quad \pi(1) = i,$$

and

$$V^n_{ij} = T^n_\pi(V \otimes 1_{n-2}); \quad \pi \in P_n, \quad \pi(1) = i, \quad \pi(2) = j$$

are well-defined. They set up the n-particle Hamiltonian

$$H_n = \sum_{i=1}^n U^n_i + (2n)^{-1} \sum_{\substack{i,j=1 \\ i \neq j}}^n V^n_{ij} \in L^1_n.$$

With

$$\rho^\beta_{\mathrm{G}n} := |\Lambda|^{-n} \exp(-\beta H_n)/\mathrm{tr}_n(\exp(-\beta H_n)),$$

define the consistent hierarchy $\{\rho^{\beta k}_{\mathrm{G}n} \in R_k \mid k \in \mathbb{N}\}$ by

$$\rho^{\beta k}_{\mathrm{G}n} = \begin{cases} \mathrm{Tr}_{n,\{k+1,\ldots,n\}}(\rho^\beta_{\mathrm{G}n}), & \text{if } k < n, \\ \rho^\beta_{\mathrm{G}n}, & \text{if } k = n, \\ \rho^\beta_{\mathrm{G}n} \otimes (d\,|\Lambda|)^{n-k} 1_{k-n}, & \text{if } k > n; \quad k \in \mathbb{N}. \end{cases}$$

Therefore, a state $\omega^\beta_{\mathrm{G}n} \in R$ is determined by

$$\omega^\beta_{\mathrm{G}n}(f) = \mathrm{Tr}_k(\rho^{\beta k}_{\mathrm{G}n} f) := \int_{\Lambda^k} dx_1 \ldots dx_k \times \mathrm{tr}_k(\rho^{\beta k}_{\mathrm{G}n}(x_1,\ldots,x_k) f(x_1,\ldots,x_k)); \quad k \in \mathbb{N}, \quad f \in A_k.$$

Then $\{\omega^\beta_{\mathrm{G}n} \mid n \in \mathbb{N}\} \subset R$ is called *Gibbs sequence* at the inverse temperature β. Now let ω_0 be an accumulation point of $\{\omega^\beta_{\mathrm{G}n} \mid n \in \mathbb{N}\}$. Since R is closed, ω_0 is spatially homogeneous. But the symmetry of V entails the symmetry of the H_n's, i.e., $T^n_\pi(H_n) = H_n; n \in \mathbb{N}, \pi \in P_n$, which implies $\omega_0 \in S$. Hence every accumulation point of the Gibbs sequence is a random system state.

Our next aim is the construction of a free energy density on $S \cap R$, with global minima at the accumulation points of the Gibbs sequence. To this end, define the n-particle energy E_n, entropy S_n, and free energy F^β_n on R by

$$E_n(\omega) = \overline{\omega}(H_n),$$

$$S_n(\omega) = -\mathrm{Tr}_n(\rho^\omega_n \ln \rho^\omega_n); \quad \omega \in R,$$

$$F^\beta_n = E_n - \beta^{-1} S_n.$$

The corresponding densities are

$$e_n = E_n/n, \quad s_n = S_n/n, \quad f^\beta_n = F^\beta_n/n.$$

Straightforward generalizations of analogous results in Ref. 11 yield the following:

(a) the S_n's are upper semicontinuous;

(b) $S_{m+n} \leqslant S_m + S_n$; $m,n \in \mathbb{N}$ (subadditivity); (2.3)

(c) the family $\{s_n \mid n \in \mathbb{N}\}$ converges pointwise on $S \cap R$, and $\lim_{n\to\infty} s_n(\omega) = \inf\{s_n(\omega) \mid n \in \mathbb{N}\}; \omega \in S \cap R$;

(d) the entropy density $s: R \cap S \to \mathbb{R}$, given by $s(\omega) = \lim_{n\to\infty} s_{n|R\cap S}(\omega)$ is affine and upper semicontinuous. Therefore the free energy density $f^\beta = e - \beta^{-1} s = \lim_{n\to\infty} f^\beta_{n|R\cap S}$ is affine and lower semicontinuous.

III. DECOMPOSITION OF THE ACCUMULATION POINTS OF THE GIBBS SEQUENCE

Let ω_0 be an accumulation point of $\{\omega^\beta_{\mathrm{G}n} \mid n \in \mathbb{N}\}$.

Step 1: There exists a unique regular Borel probability measure μ_{ω_0} on S, such that

$$\mu_{\omega_0}(P) = 1, \tag{3.1}$$

and

$$\omega_0 = \int_S \mu_{\omega_0}(d\omega)\omega. \tag{3.2}$$

The proof is analogous to the one given in Ref. 12.

Step 2: The support of μ_{ω_0} is reduced by Theorem 1.

Theorem 1: $\mu_{\omega_0}(P \cap R) = 1$.

Proof: Given $n,k \in \mathbb{N}$, $g \in A_n$. Define $\hat{g}: S \to \mathbb{R}^2$ by $\hat{g}(\omega) = (\mathrm{Re}\,\omega(g), \mathrm{Im}\,\omega(g)); \omega \in S$, and $\hat{g}_k: S \to \mathbb{R}$ by $\hat{g}_k(\omega) = \omega(g \otimes \cdots \otimes g \otimes g^* \otimes \cdots \otimes g^*)$; $\omega \in S$. (3.1) and (3.2) imply

$$\hat{g}_k(\omega) = \int_P \mu_{\omega_0}(d\omega)\hat{g}_k(\omega) = \int_P \mu_{\omega_0}(d\omega)\hat{g}(\omega)^{2k} = \int_{\mathbb{R}^2} \hat{g}(\mu_{\omega_0})(dx)|x|^{2k}; \; k \in \mathbb{N}.$$

But $\omega_0 \in R$; therefore $\hat{g}_k(\omega_0) \leqslant (\|g\|^1)^{2k}$, hence

$$\int_{\mathbb{R}^2} \hat{g}(\mu_{\omega_0})(dx)|x|^{2k} \leqslant (\|\Lambda\|^1)^{2k}; \; k \in \mathbb{N},$$

from which it follows

$$\hat{g}(\mu_{\omega_0})(\{x \in \mathbb{R}^2 \mid |x| > \|g\|^1\}) = 0,$$

or

$$\mu_{\omega_0}(\{\omega \in S \mid |\omega(g)| > \|\Lambda\|^1\}) = 0; \; g \in A. \tag{3.3}$$

As $\overline{A}$ is separable, there exists a dense sequence $D = \{D_n \mid n \in \mathbb{N}\} \subset A$. Define $N := \bigcup_{n \geqslant 1} \{\omega \in S \mid |\omega(D_n)|$

$>\|D_n\|^1\}$. From (3.3) we conclude

$$\mu_{\omega_0}(N)=0. \tag{3.4}$$

Now, given $\omega\in S-N$, $g\in\overline{A}$, and $\{D_{k(n)}|n\in\mathbb{N}\}\subset D$, such that $D_{k(n)}\overset{n\to\infty}{\longrightarrow}g$,

$$\big|\|D_{k(n)}\|^1-\|g\|^1\big|\leqslant\|D_{k(n)}-g\|^1 \leqslant\|D_{k(n)}-g\|\overset{n\to\infty}{\longrightarrow}0,$$

hence

$$\lim_{n\to\infty}\|D_{k(n)}\|^1=\|g\|^1,$$

and consequently

$$|\omega(g)|=\Big|\lim_{n\to\infty}\omega(D_{k(n)})\Big| \leqslant\lim_{n\to\infty}\|D_{k(n)}\|^1=\|g\|^1.$$

Therefore $\omega\in R$, and as ω was arbitrary, $S-N=S\cap R$, yielding together with (3.4) $\mu_{\omega_0}(P\cap R)=1$. Q.E.D.

Step 3: We now show that $M:=\{\omega\in P\cap R\,|f^\beta(\omega)\leqslant f^\beta(\omega');\ \omega'\in P\cap R\}$ comprises the support of μ_{ω_0}.

Lemma 2: f^β has a global minimum at ω_0.

Proof: (cf. Ref. 7). Suppose for simplicity $\omega^\beta_{Gn}\overset{n\to\infty}{\longrightarrow}\omega_0$. We start from the relations [cf. (2.3)]

$$S_{k+l}(\omega^\beta_{Gn})\leqslant S_k(\omega^\beta_{Gn})+S_l(\omega^\beta_{Gn});\ k+l\leqslant n, \tag{3.5}$$

which are derived as, in Ref. 11, from the symmetry of the H_n's. Decomposing $n=km+r$; $k,r\in\mathbb{N}_0$, $r<m$, (3.5) yields

$$S_n(\omega^\beta_{Gn})\leqslant kS_m(\omega^\beta_{Gn})+S_r(\omega^\beta_{Gn}),$$

whence

$$s_n(\omega^\beta_{Gn})\leqslant s_m(\omega^\beta_{Gn})+r/n(s_r(\omega^\beta_{Gn})-s_m(\omega^\beta_{Gn})).$$

Therefore, because of the boundedness and upper semicontinuity of the s_n's

$$\lim_{n\to\infty}\sup s_n(\omega^\beta_{Gn})\leqslant\lim_{n\to\infty}\sup s_m(\omega^\beta_{Gn}) =s_m(\omega_0);\ m\in\mathbb{N},$$

giving

$$\lim_{n\to\infty}\sup s_n(\omega^\beta_{Gn})\leqslant s(\omega_0). \tag{3.6}$$

As for the energy densities, one easily verifies that $\lim_{n\to\infty}e_n(\omega^\beta_{Gn})=e(\omega_0)$, implying, together with (3.6)

$$\lim_{n\to\infty}\inf f^\beta_n(\omega^\beta_{Gn})\geqslant f^\beta(\omega_0). \tag{3.7}$$

On the other hand, by definition of the Gibbs sequence, $f^\beta_n(\omega^\beta_{Gn})\leqslant f^\beta_n(\omega)$; $\omega\in R\cap S$, whence

$$\lim_{n\to\infty}\sup f^\beta_n(\omega^\beta_{Gn})\leqslant f^\beta(\omega);\ \omega\in R\cap S. \tag{3.8}$$

Inserting $\omega=\omega_0$ in (3.8) one gets, with (3.7) $f^\beta(\omega_0)=\lim_{n\to\infty}f^\beta_n(\omega^\beta_{Gn})$, which in turn yields, together with (3.8) $f^\beta(\omega_0)\leqslant f^\beta(\omega);\omega\in R\cap S$. Q.E.D.

Theorem 3: There exists a unique regular probability measure μ_{ω_0} on $S\cap R$, such that $\mu_{\omega_0}(M)=1$, and $\omega_0=\int_{S\cap R}\mu_{\omega_0}(d\omega)\omega$.

Proof: (3.2) and Theorem 1 involve

$$0=\int_{P\cap R}\mu_{\omega_0}(d\omega)(f^\beta(\omega)-f^\beta(\omega_0)).$$

But since $f^\beta(\omega)-f^\beta(\omega_0)>0$ for $\omega\in P\cap R-M$ (Lemma 2), we have $\mu_{\omega_0}(P\cap R-M)=0$, whence $\mu_{\omega_0}(M)=1$. Q.E.D.

IV. THE CURIE–WEISS EQUATION

There is a one-to-one correspondence between the elements of R_1 and $P\cap R$. For, suppose $\omega\in P\cap R$, then $\rho^\omega_1\in R_1$. Conversely, if $\rho\in R_1$, fix $\omega\in P\cap R$ through the hierarchy $\rho^\omega_n:=\underbrace{\rho\otimes\cdots\otimes\rho}_{n}$; $n\in\mathbb{N}$. Accordingly, $P\cap R$ and R_1 will often be identified.

In this section we show that each $\rho\in M$ obeys the generalized Curie–Weiss equation

$$\rho=\exp(-\beta H_\rho)/(|\Lambda|\operatorname{tr}(\exp(-\beta H_\rho))),$$

with

$$H_\rho(x):=U(x)+\int_\Lambda dy\,\operatorname{tr}_{2,\{2\}}(V(x,y)1_{B_1}\otimes\rho(y)).$$

Standard manipulations yield

$$f^\beta(\rho)=\operatorname{Tr}_1(\rho U)+\tfrac12\operatorname{Tr}_2(\rho\otimes\rho V) +(1/\beta)\operatorname{Tr}_1(\rho\ln\rho);\quad \rho\in R_1.$$

Thus we are posed with the problem of minimizing the nonlinear functional $\operatorname{Tr}(\cdot U)+\frac12\operatorname{Tr}_2(\cdot\otimes\cdot V)+(1/\beta)\operatorname{Tr}_1(\cdot\ln\cdot)$ on R_1.

Lemma 4: Let $\rho\in M$ and $\lambda_i(x)$; $i=1,\dots,d$ be the eigenvalues of $\rho(x)$; $x\in\Lambda$. Then there exists an $\epsilon_0>0$, such that $\lambda_i(x)\geqslant\epsilon_0$; $i=1,\dots,d$, λ-a.e. on Λ.

Proof: For simplicity we restrict ourselves to the case $|\Lambda|=1$. With $\Delta\subset\Lambda$ measurable and $\epsilon>0$, define $\rho^{\Delta\epsilon}\in R_1$ by

$$\rho^{\Delta\epsilon}:=\begin{cases}\rho(x), & \text{if } x\notin\Delta,\\ (\rho(x)+\epsilon 1_{B_1})/(1+\epsilon d), & \text{if } x\in\Delta.\end{cases}$$

By the definition of M,

$$\begin{aligned}0&\leqslant f^\beta(\rho^{\Delta\epsilon})-f^\beta(\rho)\\ &=\int_\Delta dx\,\operatorname{tr}[(\rho(x)-\rho^{\Delta\epsilon}(x))U(x)]+\int_\Delta dx\int_{\Lambda-\Delta}dy\,\operatorname{tr}_2\\ &\quad\times[V(x,y)(\rho^{\Delta\epsilon}(x)-\rho(x))\otimes\rho(y)]\\ &\quad+\int_\Delta dx\int_\Delta dy\,\operatorname{tr}_2[V(x,y)(\rho^{\Delta\epsilon}(x)-\rho(x))\\ &\quad\otimes(\rho^{\Delta\epsilon}(y)+\rho(y))/2]\\ &\quad+\frac1\beta\int_\Delta dx\,\operatorname{tr}[\rho^{\Delta\epsilon}(x)\ln\rho^{\Delta\epsilon}(x)-\rho(x)\ln\rho(x)].\end{aligned} \tag{4.1}$$

The mean value theorem supplies upper bounds for the integrands in (4.1):

$$\text{(a)}\ \frac{[\operatorname{tr}(\rho^{\Delta\epsilon}(x)U(x))-\operatorname{tr}(\rho(x)U(x))]}{\epsilon}=\operatorname{tr}\Big[\frac{\partial}{\partial\eta}\rho^{\Delta\eta}(x)_{|\eta=\tilde\epsilon(x)}U(x)\Big]\leqslant\frac{\|U(x)\|_{B_1}\operatorname{tr}[|1_{B_1}-d\rho^{\Delta\tilde\epsilon(x)}(x)|]}{(1+\tilde\epsilon(x)d)}\leqslant C_U2d;$$

$x\in\Delta$, $0<\tilde\epsilon(x)<\epsilon$.

Similarly

$$\text{(b)}\left.\begin{array}{l}(1/\epsilon)\,\mathrm{tr}_2[V(x,y)(\rho^{\Delta\epsilon}(x)-\rho(x))\otimes\rho(y)]\\(1/\epsilon)\,\mathrm{tr}_2[V(x,y)(\rho^{\Delta\epsilon}(x)-\rho(x))\otimes(\rho^{\Delta\epsilon}(y)+\rho(y))/2)]\end{array}\right\}<C_V2d;\ x\in\Delta.$$

$$\text{(c)}\ (1/\epsilon)\,\mathrm{tr}[\rho^{\Delta\epsilon}(x)\ln\rho^{\Delta\epsilon}(x)-\rho(x)\ln\rho(x)]$$
$$=\mathrm{tr}\left[\frac{\partial}{\partial\eta}\rho^{\Delta\eta}(x)_{|\eta=\tilde\epsilon(x)}(\ln\rho^{\Delta\tilde\epsilon(x)}(x)+1)\right]=\mathrm{tr}[(1_{B_1}-d\rho^{\Delta\tilde\epsilon(x)}(x))(1+\ln\rho^{\Delta\tilde\epsilon(x)}(x))](1+\tilde\epsilon(x)d)^{-1}$$
$$\leqslant[2d-|\ln[(\lambda_{\min}(x)+\tilde\epsilon(x))/(1+\tilde\epsilon(x)d)]|+d^2/e](1+\tilde\epsilon(x)d)^{-1}$$
$$\leqslant k-|\ln(\lambda_{\min}(x)+\epsilon)|(1+\epsilon d)^{-1};\quad x\in\Delta,\ 0<\tilde\epsilon(x)<\epsilon,\ \lambda_{\min}(x):=\min\{\lambda_i(x)|i=1,\dots,d\},\quad k:=2d+d^2/e.$$

Combining (a), (b), and (c) yields, with (4.1)

$$0\leqslant[f^\beta(\rho^{\Delta\epsilon})-f^\beta(\rho)]/\epsilon\leqslant|\Delta|C_U2d+|\Delta|C_V2d+(1/\beta)|\Delta|k$$
$$-\frac{1}{\beta}(1+\epsilon d)^{-1}\int_\Delta dx\,|\ln(\lambda_{\min}(x)+\epsilon)|=|\Delta|k_1-\frac{1}{\beta}(1+\epsilon d)^{-1}\int_\Delta dx\,|\ln(\lambda_{\min}(x)+\epsilon)|;$$
$$k_1:=2d(C_U+C_V)+(1/\beta)k. \tag{4.2}$$

Now choose $1>\eta>0$, with $|\ln\eta|>k_1\beta(1+d)$, $\epsilon,\epsilon_0=\eta/2$, and $\Delta=\{x\in\Lambda\,|\lambda_{\min}(x)<\epsilon\}$. Inserting this into (4.2) gives

$$0\leqslant|\Lambda|k_1-(1/\beta)(1+d)^{-1}|\Delta|\,|\ln\eta|=|\Delta|(k_1-(1/\beta)(1+d)^{-1}|\ln\eta|)\leqslant0,$$

therefore $|\Delta|=0$. Q.E.D.

Theorem 5: Each $\rho\in M$ obeys the generalized Curie–Weiss equation

$$\rho=\exp(-\beta H_\rho)/(|\Lambda|\mathrm{tr}(\exp(-\beta H_\rho))), \tag{4.3}$$

where

$$H_\rho(x)=U(x)+\int_\Lambda dy\,\mathrm{tr}_{2,\{2\}}(V(x,y)1_{B_1}\otimes\rho(y)).$$

Proof: Again we assume $|\Lambda|=1$. Set $\rho\in M$ and $\epsilon_0>0$, such that $\rho>\epsilon_0 1_1$ (Lemma 4). With $\Delta\subset\Lambda$ measurable, $g\in B_1$, $g=g^*$, $\mathrm{tr}(g)=0$, and $-\epsilon_0/\|g\|_{B_1}\leqslant\epsilon\leqslant\epsilon_0/\|g\|_{B_1}$ define $\rho^{\Delta g\epsilon}\in R_1$ by

$$\rho^{\Delta g\epsilon}(x):=\begin{cases}\rho(x), & \text{if } x\notin\Delta,\\ \rho(x)+\epsilon g, & \text{if } x\in\Delta,\quad x\in\Lambda.\end{cases}$$

We have

$$\frac{[f^\beta(\rho^{\Delta g\epsilon})-f^\beta(\rho)]}{\epsilon}=\int_\Delta dx\,\mathrm{tr}(gU(x))+\int_\Delta dx\int_{\Lambda-\Delta}dy\,\mathrm{tr}_2(V(x,y)g\otimes\rho(y))$$
$$+\int_\Delta dx\int_\Delta dy\,\mathrm{tr}_2\left[V(x,y)g\otimes\left(\rho(y)+\left(\frac{\epsilon}{2}\right)g\right)\right]+\frac{1}{\beta}\int_\Delta dx\left(\frac{1}{\epsilon}\right)\mathrm{tr}[\rho^{\Delta g\epsilon}(x)\ln\rho^{\Delta g\epsilon}(x)$$
$$-\rho(x)\ln\rho(x)];\quad 0<|\epsilon|\leqslant\epsilon_0/\|g\|_{B_1}. \tag{4.4}$$

But

$$\lim_{\substack{\epsilon\to0\\0<|\epsilon|<\epsilon_0/\|g\|_{B_1}}}\left(\frac{1}{\epsilon}\right)\mathrm{tr}[\rho^{\Delta g\epsilon}(x)\ln\rho^{\Delta g\epsilon}(x)-\rho(x)\ln\rho(x)]=\mathrm{tr}\left[\frac{\partial}{\partial\eta}\rho^{\Delta g\eta}(x)_{|\eta=0}(\ln\rho^{\Delta g\eta=0}(x)+1)\right]=\mathrm{tr}(g\ln\rho(x));\ x\in\Delta. \tag{4.5}$$

From

$$(1/\beta)|\mathrm{tr}[\rho^{\Delta g\epsilon}(x)\ln\rho^{\Delta g\epsilon}(x)-\rho(x)\ln\rho(x)]|$$
$$=\left|\mathrm{tr}\left[\frac{\partial}{\partial\eta}\rho^{\Delta g\eta}(x)_{|\eta=\tilde\epsilon(x)}(\ln\rho^{\Delta g\tilde\epsilon(x)}(x)+1)\right]\right|$$
$$=|\mathrm{tr}[g\ln(\rho(x)+\tilde\epsilon(x)g)]|\leqslant\mathrm{tr}[|g|]|\ln\epsilon_0/2|;$$
$$0<\tilde\epsilon(x)<\epsilon\leqslant\epsilon_0/2\|g\|_{B_1},$$

it follows that the functions

$$(1/\epsilon)\,\mathrm{tr}[\rho^{\Delta g\epsilon}(\cdot)\ln\rho^{\Delta g\epsilon}(\cdot)-\rho(\cdot)\ln\rho(\cdot)]$$

on Λ are majorized by $\mathrm{tr}[|g|]\,|\ln\epsilon_0/2|$ for $0<\epsilon<\epsilon_0/2\|g\|_{B_1}$.

Thus, using (4.4), (4.5), Lebesgue's convergence theorem, and the minimality of $f^\beta(\rho)$, one gets

$$0\leqslant\lim_{0<\epsilon\to0}[f^\beta(\rho^{\Delta g\epsilon})-f^\beta(\rho)]/\epsilon$$
$$=\int_\Delta dx\,\mathrm{tr}(gU(x))+\int_\Delta dx\int_\Lambda dy\,\mathrm{tr}_2(V(x,y)g\otimes\rho(y))$$
$$+\frac{1}{\beta}\int_\Delta dx\,\mathrm{tr}(g\ln\rho(x))$$
$$=\lim_{0<\epsilon\to0}[f^\beta(\rho^{\Delta g-\epsilon})-f^\beta(\rho)]/-\epsilon\leqslant0,$$

which involves

$$\mathrm{tr}(gD(\Delta))=0;\ \Delta\subset\Lambda\text{ measurable},\ g\in B_1,\ g=g^*,\ \mathrm{tr}(g)=0, \tag{4.6}$$

where

$$D(\Delta):=\int_\Delta dx\, D(x),$$

and

$$D(x):=U(x)+\int_\Lambda dy\, \mathrm{tr}_{2,\{2\}}(V(x,y)1_{B_1}\otimes\rho(y))$$

$$+\frac{1}{\beta}\ln\rho(x). \tag{4.7}$$

Inserting $g=D(\Delta)-\mathrm{tr}(D(\Delta))1_{B_1}$ in (4.6) yields

$$\mathrm{tr}[(D(\Delta)-\mathrm{tr}(D(\Delta))1_{B_1})^2]=0,$$

and hence $D(\Delta)=\mathrm{tr}(D(\Delta))1_{B_1}$; $\Delta\subset\Lambda$ measurable. Writing D in the form[13]

$$D(x)=\lim_{m\to\infty}|K_m(x)|^{-1}D(K_m(x));\quad \lambda\text{-a.e. on }\Lambda,$$

$K_m(x)$ being the open sphere with radius m^{-1} and center x, one arrives at

$$D(x)=C(x)1_{B_1};\quad \lambda\text{-a.e. on }\Lambda, \tag{4.8}$$

with a real function $C(x)$. Finally, insert (4.7) into (4.8) and solve for $\rho(x)$, having regard to the normalization condition $\mathrm{tr}\,\rho(x)=|\Lambda|^{-1}; x\in\Lambda$. This gives

$$\exp(\beta C(x))=[\mathrm{tr}\exp(-H_\rho(x))]^{-1};\quad \lambda\text{-a.e. on }\Lambda,$$

from which (4.3) follows. Q.E.D.

Theorem 6: (4.3) has a unique solution $\rho\in R_1$ for $\beta<(2C_V)^{-1}$.

Proof: (cf. Ref. 9). Define the real Banach space

$$L^1_{1s}:=\{g\in L^1_1\,|\,g=g^*\},$$

with the norm

$$\|\cdot\|^1_{\mathrm{tr}}:=\int_\Lambda dx\,\mathrm{tr}(|\cdot(x)|),$$

and $H,A^\beta,T^\beta:L^1_{1s}\to L^1_{1s}$, through

$$H(g)(x):=U(x)+\int_\Lambda dy\,\mathrm{tr}_{2,\{2\}}(V(x,y)1_{B_1}\otimes g(y));\ x\in\Lambda,$$

$$A^\beta(g):=-\beta H(g)-\ln\mathrm{tr}\exp(-\beta H(g)),$$

and

$$T^\beta(g):=|\Lambda|^{-1}\exp(A^\beta(g));\ g\in L^1_{1s}.$$

Then

$$\begin{aligned}
&|\Lambda|\,\mathrm{tr}[|T^\beta(g_1)(x)-T^\beta(g_2)(x)|]\\
&\quad\leqslant\|A^\beta(g_1)(x)-A^\beta(g_2)(x)\|_{B_1}\\
&\quad\leqslant\beta\|H(g_1)(x)-H(g_2)(x)\|_{B_1}+\left|\ln\frac{\mathrm{tr}\exp(-\beta H(g_1)(x))}{\mathrm{tr}\exp(-\beta H(g_2)(x))}\right|\\
&\quad\leqslant 2\beta\|H(g_1)(x)-H(g_2)(x)\|_{B_1}\\
&\quad\leqslant 2\beta C_V\|g_1-g_2\|^1_{\mathrm{tr}};\ x\in\Lambda,
\end{aligned}$$

where the first inequality follows from an application of Hadamard's three line principle[14] to

$$\begin{aligned}
&\exp(A^\beta(g_1)(x))-\exp(A^\beta(g_2)(x))\\
&\quad=\int_0^1 dt\,\exp(tA^\beta(g_1)(x))[A^\beta(g_1)(x)-A^\beta(g_2)(x)]\\
&\qquad\times\exp((1-t)A^\beta(g_2)(x)).
\end{aligned}$$

Hence

$$\|T^\beta(g_1)-T^\beta(g_2)\|^1_{\mathrm{tr}}\leqslant 2\beta C_V\|g_1-g_2\|^1_{\mathrm{tr}},$$

and the result follows from the contraction mapping principle. Q.E.D.

Corollary 7: Let $\beta<(2C_V)^{-1}$ and $\rho\in R_1$, the unique solution of (4.3). Then the Gibbs sequence converges, and

$$\lim_{n\to\infty}\omega^\beta_{\mathrm{G}n}=\rho.$$

Proof: Since M contains only the single element ρ, all accumulation points of the Gibbs sequence are equal to ρ, which proves our assertion. Q.E.D.

V. APPLICATIONS

A. The gap equation in the BCS theory

Using the BCS n-particle Hamiltonian in the quasispin formulation

$$H_n=-\sum_{i=1}^n\epsilon\sigma_i^z-\frac{c}{n}\sum_{i,j=1}^n\sigma_i^+\sigma_j^-;\quad n\in\mathbb{N}$$

($\boldsymbol{\sigma}$ is the vector of Pauli matrices, c a coupling constant), it has been shown in Ref. 15 that the corresponding Gibbs sequence converges at all temperatures towards a suitable superposition of product states. We wish to derive a similar result starting from the more general ansatz

$$\begin{aligned}H_n(k_1,\ldots,k_n)=\sum_{i=1}^n U(k_i)\sigma_i^z-(2n)^{-1}\\
\times\sum_{\substack{i,j=1\\ i\neq j}}^n V(k_i,k_j)B_{ij},\end{aligned} \tag{5.1}$$

where $B=\sigma^x\otimes\sigma^x+\sigma^y\otimes\sigma^y$. The external potential U and the positive, symmetric, bounded two-body interaction V extend over

$$\Lambda=\{k\in\mathbb{R}^3|\epsilon_{\mathrm{F}}-\epsilon_{\mathrm{D}}\leqslant|k|\leqslant\epsilon_{\mathrm{F}}+\epsilon_{\mathrm{D}}\},$$

for some parameters ϵ_{F} (Fermi energy) and ϵ_{D} (Debye energy). The Curie–Weiss equation corresponding to (5.1) is

$$\rho=\exp(-\beta H_\rho)/(|\Lambda|\,\mathrm{tr}(\exp(-\beta H_\rho))), \tag{5.2}$$

where

$$\begin{aligned}H_\rho(k)=U(k)\sigma^z-\int_\Lambda dk'\,V(k,k')\\
\times[\mathrm{tr}(\rho(k')\sigma^x)\sigma^x+\mathrm{tr}(\rho(k')\sigma^y)\sigma^y].\end{aligned}$$

With the help of the identity

$$\exp(\mathbf{x}\cdot\boldsymbol{\sigma})=\cosh(|\mathbf{x}|)+\sinh(|\mathbf{x}|)(\mathbf{x}/|\mathbf{x}|)\cdot\boldsymbol{\sigma};\quad \mathbf{x}\in\mathbb{R}^3, \tag{5.3}$$

(5.2) may be manipulated to give

$$a^i(k)=\beta|\Lambda|^{-1}\int_\Lambda dk'\,\frac{V(k,k')(\tanh a(k'))}{a(k')}a^i(k');\ i=1,2,$$

$$a^z(k)=-\beta U(k), \tag{5.2a}$$

where

$$\mathbf{a}(k):=\begin{bmatrix}\beta\int_\Lambda dk'\,V(k,k')\mathrm{tr}(\rho(k')\sigma^x)\\ \beta\int_\Lambda dk'\,V(k,k')\mathrm{tr}(\rho(k')\sigma^y)\\ -\beta U(k)\end{bmatrix},$$

and $a(k) := |\mathbf{a}(k)|$. Solutions of these nonlinear integral equations are the following.

(i) $a^x = a^y = 0$. The corresponding product state is called $\overset{0}{\omega}$.

(ii) Suppose $\alpha \neq 0$ satisfies

$$\alpha(k) = \beta |\Lambda|^{-1} \int_\Lambda dk' V(k,k') \times \frac{\tanh[\alpha(k')^2 + a^z(k')^2]^{1/2}}{[\alpha(k')^2 + a^z(k')^2]^{1/2}} \alpha(k'). \tag{5.4}$$

Then all pairs $(a^x, a^y) = (\alpha \cos\varphi, \alpha \sin\varphi)$; $\varphi \in [0,2\pi)$, solve (5.2a). Denote the corresponding product state by ω_φ. An explicit evaluation of the free energy density f (the index β will be dropped henceforth) yields the following lemma.

Lemma 8: Suppose $\alpha \neq 0$ is a solution of (5.4). Then (i) $f(\omega_\varphi)$ is independent of φ, and (ii) $f(\omega_\varphi) < f(\overset{0}{\omega})$.

Introducing the group $G = \{T^\varphi | \varphi \in [0,2\pi)\}$ of *-automorphisms on $\overline{A}$ by

$$T^\varphi(g)(k) = \exp(-i\varphi S_n^z) g(k) \exp(i\varphi S_n^z);$$

$$n \in \mathbb{N},\ g \in A_n,\ k \in \Lambda^n,$$

where $S_i^z := \Sigma_{i=1}^n \frac{1}{2}\sigma_i^z$, one derives from $[H_n, S_n^z] = 0; n \in \mathbb{N}$, that each accumulation point ω_0 of the Gibbs sequence is G invariant, i.e.,

$$\omega_0(T^\varphi(g)) = \omega_0(g);\ \varphi \in [0,2\pi),\ g \in \overline{A}. \tag{5.5}$$

Besides, applying the identity

$$\exp(i(\varphi_2/2)\sigma^z)(\cos\varphi_1 \sigma^x + \sin\varphi_1 \sigma^y)\exp(-i(\varphi_2/2)\sigma^z) = \cos(\varphi_1 - \varphi_2)\sigma^x + \sin(\varphi_1 - \varphi_2)\sigma^y,$$

one concludes

$$\omega_{\varphi_1}(T^{\varphi_2}(g)) = \omega_{\varphi_1 - \varphi_2}(g);\ \varphi_1, \varphi_2 \in [0,2\pi),\ g \in \overline{A}. \tag{5.6}$$

Now Theorem 3 and Lemma 8 imply, together with (5.5) and (5.6)

$$\int_{[0,2\pi)} \mu_{\omega_0}(d\varphi_1)\omega_{\varphi_1}(g) = \omega_0(g) = \int_{[0,2\pi)} \mu_{\omega_0}(d\varphi_1)\omega_{\varphi_1 + \varphi_2}(g) = \int_{[0,2\pi)} \tau^{\varphi_2}(\mu_{\omega_0})(d\varphi_1)\omega_{\varphi_1}(g);$$

$$\varphi_2 \in [0,2\pi),\ g \in \overline{A},$$

where τ^{φ_2} is given by $\tau^{\varphi_2}(\varphi) = \varphi + \varphi_2$ (modulo 2π). Therefore, since μ_{ω_0} is unique, $\mu_{\omega_0} = \tau^\varphi(\mu_{\omega_0})$; $\varphi \in [0,2\pi)$. Thus, μ_{ω_0} is a Haar measure on the compact topological group ($[0,2\pi)$, $+ =$ addition modulo 2π), and by the uniqueness theorem for Haar measures, $\mu_{\omega_0}(d\varphi) = d\varphi/2\pi$, for all accumulation points ω_0. This enables us to state Theorem 9.

Theorem 9: There is a unique limiting Gibbs state ω_G.

For high enough temperatures (see Theorem 6) we have $\omega_G = \overset{0}{\omega}$. However, if there exists a solution $\alpha \neq 0$ of (5.4), then $\omega_G = (1/2\pi)\int_0^{2\pi} d\varphi\, \omega_\varphi$, and evaluating the energy density e explicitly, one finds $e(\omega_G) < e(\overset{0}{\omega})$. Therefore (5.4) is called gap equation.[10]

B. Spins with a cosine interaction on a circle in a homogeneous magnetic field

As it will come out, the limit of the Gibbs sequence exists for all temperatures and magnetic field strengths and takes three different forms according to the three phases of the system. The "box," which contains the particles, is the compact topological group $\Lambda = ([0,2\pi), + =$ addition modulo 2π) representing the points on the unit circle. The n-particle Hamiltonian is given by

$$H_n(x_1,...,x_n) = -\sum_{i=1}^n \mathbf{B}\cdot\boldsymbol{\sigma} - (2n)^{-1} \times \sum_{\substack{i,j=1 \\ i \neq j}}^n \cos(x_i - x_j)\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j;\ x_1,...,x_n \in \Lambda. \tag{5.7}$$

The corresponding Curie–Weiss equation reads

$$\rho = (1/2\pi)[\exp(-\beta H_\rho)/\mathrm{tr}(\exp(-\beta H_\rho))], \tag{5.8}$$

wherein

$$H_\rho(x) = -\left[\mathbf{B} + \int_0^{2\pi} dy \cos(x-y)\mathrm{tr}(\rho(y)\boldsymbol{\sigma})\right]\boldsymbol{\sigma};\ x \in \Lambda.$$

Using (5.3) together with the notations

$$\mathbf{m}(x) = \mathrm{tr}(\rho(x)\boldsymbol{\sigma}),\ \boldsymbol{\mu} = \int_0^{2\pi} dy \cos y\, \mathbf{m}(y),$$

$$\mathbf{v} = \int_0^{2\pi} dy \sin y\, \mathbf{m}(y),\quad \mathbf{a}(x) = \mathbf{B} + \cos x\boldsymbol{\mu} + \sin x\mathbf{v},$$

$$a(x) = |\mathbf{a}(x)|,$$

and $K(x) = (1/2\pi)[\tanh(\beta a(x))/a(x)]$, (5.8) may be written in the form

$$\begin{aligned}\boldsymbol{\mu} &= \int_0^{2\pi} dy \cos(y)^2 K(y)\boldsymbol{\mu} + \int_0^{2\pi} dy \cos(y)\sin(y)K(y)\mathbf{v} + \int_0^{2\pi} dy \cos(y)K(y)\mathbf{B},\\ \mathbf{v} &= \int_0^{2\pi} dy \sin(y)^2 K(y)\mathbf{v} + \int_0^{2\pi} dy \cos(y)\sin(y)K(y)\boldsymbol{\mu} + \int_0^{2\pi} dy \sin(y)K(y)\mathbf{B}.\end{aligned} \tag{5.9}$$

We distinguish three classes of solutions of (5.9): (A) $(\boldsymbol{\mu},\mathbf{v}) = (0,0)$. Let $\overset{0}{\omega}$ denote the corresponding product state; (B) if $\mathbf{B} \neq 0$, and $(\mu\hat{\mathbf{B}},0)$ solves (5.9), then so does $(\mu\cos\varphi\hat{\mathbf{B}}, \mu\sin\varphi\hat{\mathbf{B}})$ for all $\varphi \in [0,2\pi)$ ($\boldsymbol{\mu}$ and $\mathbf{B}$ are always parallel, $\mu = |\boldsymbol{\mu}|$, $B = |\mathbf{B}|$, $\hat{\mathbf{B}} = \mathbf{B}/B$); and (C) if $\mathbf{B} = 0$, and $(\boldsymbol{\mu},0)$ solves (5.9), then all pairs $(\mu\cos\varphi\hat{\mathbf{n}}\, \mu\sin\varphi\hat{\mathbf{n}})$; $\varphi \in [0,2\pi)$, $\hat{\mathbf{n}} \in \mathbf{R}^3$, $|\hat{\mathbf{n}}| = 1$, satisfy (5.9). Denote by $\omega(\varphi,\boldsymbol{\mu})$ the product state belonging to $(\cos\varphi\boldsymbol{\mu}, \sin\varphi\boldsymbol{\mu})$. The free-energy density of these solutions amounts to

$$f(\mu) = \frac{1}{2}\mu^2 - (2\pi\beta)^{-1}\int_0^{2\pi} dy \times \ln[2\cosh[\beta(B + \cos(y)\mu)]] - \frac{\ln(2\pi)}{\beta}.$$

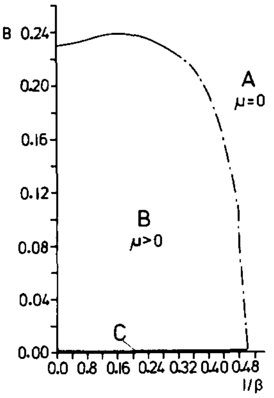


FIG. 1. Phase diagram for a spin system with cosine interaction and external magnetic field.

Notice that f is neither φ- nor $\hat{\mathbf{n}}$-dependent (cf. Lemma 8), reflecting the radial symmetry of the system and the isotropy for $\mathbf{B}=0$. Therefore a reasoning analogous to the one established in the first example of application leads to Theorem 10.

Theorem 10: (cf. Theorem 9). The limit ω_G of the Gibbs sequence exists for all temperatures and all values of **B**. According to which class the f minimizing solutions of (5.9) belong to, it is given by

$$\omega_G = \begin{cases} \overset{0}{\omega} & \text{(phase (A))}, \\ \left(\dfrac{1}{2\pi}\right)\displaystyle\int_0^{2\pi} d\varphi\, \omega(\varphi,\mu\hat{\mathbf{B}}) & \text{(phase (B))}, \\ \displaystyle\int_{S^2} \frac{d\Omega}{4\pi}\int_0^{2\pi}\frac{d\varphi}{2\pi}\,\omega(\varphi,\mu\hat{\mathbf{n}}(\Omega)) & \text{(phase (C))}, \end{cases}$$

where $\hat{\mathbf{n}}(\Omega)$ points in Ω direction, and $S^2 := \{\mathbf{x}\in\mathbf{R}^3 \mid |\mathbf{x}| = 1\}$. Figure 1 shows the phase diagram. On crossing the solid line, the order parameter μ jumps discontinuously, whereas μ is continuous at the dashed line, which consists of critical points.

ACKNOWLEDGMENT

We are grateful to Professor H. Spohn for giving us the idea to the proof of Theorem 1.

[1] W. Thirring, *Lehrbuch der Mathematischen Physik* (Springer, Vienna, 1980), Vol. 4.
[2] P. Hertel, H. Narnhofer, and W. Thirring, Commun. Math. Phys. **28**, 159 (1972).
[3] H. Narnhofer and W. Thirring, Ann. Phys. (N.Y.) **134**, 128 (1981).
[4] J. Messer, "Temperature Dependent Thomas–Fermi Theory," in *Lecture Notes in Physics*, Vol. 147 (Springer, Berlin, 1981).
[5] Th. Eisele and R. S. Ellis, Z. Wahrscheinlichkeitstheorie verw. Geb. **63**, 297 (1983).
[6] Th. Eisele and R. S. Ellis, J. Phys. A: Math. Gen. **16**, 195 (1983).
[7] J. Messer and H. Spohn, J. Stat. Phys. **29**, 561 (1982).
[8] B. Baumgartner, H. Narnhofer, and W. Thirring, Ann. Phys. (N.Y.) **150**, 373 (1983).
[9] M. Fannes, H. Spohn, and A. Verbeure, J. Math. Phys. **21**, 355 (1980).
[10] L. van Hemmen, Fortschr. Phys. **26**, 397 (1978).
[11] D. W. Robinson, "The Thermodynamic Pressure in Quantum Statistical Mechanics," in *Lecture Notes in Physics*, Vol. 9 (Springer, Berlin, 1971).
[12] R. Hudson and G. Moody, Z. Wahrscheinlichkeitstheorie verw. Geb. **33**, 343 (1976).
[13] A. C. Zaanen, *Integration* (North-Holland, Amsterdam, 1967) (exercise 37.6).
[14] M. Reed and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[15] W. Thirring, Commun. Math. Phys. **7**, 181 (1968).

Confinement of quarks and gluons in gauge quantum field theory

A. H. Völkel
Institut für Mathematik III, Freie Universität Berlin, Berlin, West Germany

(Received 15 October 1984; accepted for publication 14 December 1984)

A SU(3)-gauge quantum field theory with a quark triplet, an antiquark triplet, and a self-conjugate gluon octet as basic fields is investigated. In virtue of a nontrivial coupling between the representation of the translation group and the SU(3)-color charge of the basic fields it is proved: (i) The basic quark, antiquark, and gluon fields are confined. (ii) Every state vector of the physical Hilbert space is a SU(3)-color singlet state. (iii) Poincaré invariance holds in the physical Hilbert space. These results rest on a nonstandard transformation law of the basic fields with respect to translations. Its meaning within the frame of classical Lagrangian field theory for chromodynamics is investigated.

I. INTRODUCTION

In recent years the experimental and theoretical evidence that non-Abelian gauge quantum field theories (GQFT's) are (apart from quantum electrodynamics) the only relevant ones for elementary particle physics has become overwhelming.[1] In spite of the spectacular results (both experimental and theoretical), which have been gained in this field of modern research, the fundamental problem of quark and gluon confinement in quantum chromodynamics is still unsolved. The reason may lie in the fact that confinement cannot be understood on the level of perturbation theory without some drastic change in the basic concepts on which it relies. Whatever the final basis may be, today everyone agrees that quantum chromodynamics has to exhibit at least two features: The colored basic fields must be confined and all physical states are color neutral (i.e., color singlets).

Five years ago Strocchi[2] published some arguments indicating that within GQFT's of the Wightman–Gårding type[2–6] charge confinement may be linked to the unboundedness of the translation operators. In a recent systematic investigation[7] (quoted from now on as I) of unbounded representations of symmetry groups in GQFT's the present author proved that a class of nonisometric representations of the translation group enforces the confinement of some or all basic fields and leaves certain mixed products unconfined. This at once seems to open the path for a nonperturbative description of quark and gluon confinement in quantum chromodynamics on the same level of mathematical rigor as is customary for Wightman quantum field theories.

In this note we present a GQFT of the Wightman–Gårding type with confinement, in which the gauge group is the color group SU(3) and the set of basic field operators is formed by a quark triplet, an antiquark triplet, and a self-conjugate gluon octet. Assuming that these basic field operators transform under translations with a (nonstandard) representation of the translation group which is essentially determined by the color charge $Q = I^3 + \frac{1}{2}Y$ ($I^3 \sim$ third component of the "color isopin" and $Y \sim$ "color hypercharge") carried by the fields, the following are proved.

(i) All (basic) quark, antiquark, and gluon fields are confined.

(ii) Every state from the physical Hilbert space is a SU(3)-color singlet.

(iii) Poincaŕe symmetry, which is broken in the underlying unphysical Hilbert space, is restored via the existence of a continuous unitary representation of the Poincaŕe group in the physical Hilbert space.

The paper is organized as follows: Section II contains a collection of some standard notations, conventions and formulas which will be intensively used later on. In Sec. III we formulate the general assumptions of a GQFT of colored quarks and gluons and derive some restrictions for the representations of the translation group. In Sec. IV we prove the three consequences (i)–(iii) stated above. Some final remarks are made in Sec. V. In the Appendix we investigate the meaning of the nonstandard substitution rules of the basic fields with respect to the translation group in the classical Lagrangian field theory of chromodynamics. It is shown that the Lagrangian and thus the corresponding field equations are form invariant under these substitution rules. Hence the nonstandard basic fields represent just new solutions of the field equations.

Since the flavor degrees of freedom are irrelevant for the following considerations, we omit them for the sake of notational simplicity. Their inclusion can easily be achieved by taking into account a further internal symmetry group SU(n), $n = ?$, and correspondingly adding some more indices at the basic field operators.

II. NOTATIONS AND CONVENTIONS

In the following $S(\mathbb{R}_4,\mathbb{C})$ ($\mathbb{R}$ real and $\mathbb{C}$ complex numbers) denotes a countably normed space of test functions $f: \mathbb{R}_4 \to \mathbb{C}$, namely either the Schwartz space[8] of strongly decreasing C^∞ functions or a Jaffe space[9] of strictly localizable C^∞ functions. By $(\Lambda(\alpha)^\mu{}_\nu)$ and $(S^{1/2}(\alpha)_{\mu\nu})$ $(\mu,\nu = 0,...,3)$ with α an element from the universal covering group SL(2,C) of the Lorentz group we denote the usual 4×4 matrices according to which a Minkowski four-vector, respectively, a Dirac spinor, transforms under Lorentz transformations. Here, $\Lambda(\alpha)$ is real and $S^{1/2}(\alpha)$ satisfies the relations

$$S^{1/2}(\alpha)^{-1}\gamma^\mu S^{1/2}(\alpha) = \Lambda(\alpha)^\mu{}_\nu \gamma^\nu, \qquad \gamma^0 S^{1/2}(\alpha)^* \gamma^0 = S^{1/2}(\alpha)^{-1}, \tag{2.1}$$

γ^μ $(\mu = 0,...,3)$ are the Dirac matrices. We assume γ^0 to be Hermitian and its transpose $(\gamma^0)^t$ equals γ^0. For the sake of

 0022-2488/85/051057-08$02.50

notational compactness we glue the two matrices $\Lambda(\alpha)$ and $S^{1/2}(\alpha)$ together into a new symbol:

$$S^{(p,q)}(\alpha)^{\mu}{}_{\nu} := \begin{cases} S^{1/2}(\alpha^{-1})_{\nu\mu}, & \text{for } (p,q)=(1,0), \\ \Lambda(\alpha)^{\mu}{}_{\nu}, & \text{for } (p,q)=(1,1), \\ S^{1/2}(\alpha)_{\mu\nu}, & \text{for } (p,q)=(0,1). \end{cases} \quad (2.2)$$

For the Minowski metric we take $(+\ -\ -\ -)$.

By a representation R of a group G on a vector space E we mean a homomorphism $R: G \rightarrow \text{Aut } E$, $g \rightarrow R(g)$ of G into the automorphism group of E. For the special case of the color group SU(3) we adopt the conventions and apart from minor changes also the notations of de Swart.[10] Thus, if $\mathbb{N}^0$ is the set of non-negative integers, for every pair $(p,q) \in \mathbb{N}^0 \times \mathbb{N}^0$, $R^{(p,q)}$ denotes an unitary irreducible (matrix) representation of SU(3) on $\mathbb{C}_N$ with

$$N = N(p,q) = (1+p)(1+q)(1+1/2[p+q]). \quad (2.3)$$

The basis for such a representation in $\mathbb{C}_N$ is uniquely (up to an overall phase factor) fixed by the two assumptions: First its elements are eigenvectors of the two Casimir operators, of the squared of the "color isopin" operator with eigenvalues $I(I+1)$, $I = 0, \frac{1}{2}, 1, \ldots$, of the third component of the "color isopin" with eigenvalues $I^3 = -I, -I+1, \ldots, I$, and last but not least of the "color hypercharge" operator with eigenvalues Y. Second, the Condon–Shortley phase conventions[10] hold. Hence the indices a,b in the representation matrices $R^{(p,q)}(u)^b{}_a$ are triples (I, I^3, Y) which run over the index set $\Gamma(p,q)$ defined by the Okubo rule[11]

$$\begin{aligned} \Gamma(p,q) := \{(I,I^3,Y) \mid\ & I = 1/2(\lambda - \rho); I_3 = -I, \ldots, I; \\ & Y = \lambda + \rho - 1/3(2p+4q); \\ & \lambda, \rho \in \mathbb{N}^0; 0 \leqslant \rho \leqslant q \leqslant p+q\}. \end{aligned} \quad (2.4)$$

An immediate consequence of this definition reads

$$(I,I^3,Y) \in \Gamma(p,q) \Leftrightarrow (I, -I^3, -Y) \in \Gamma(q,p). \quad (2.5)$$

In accordance with this relation the negative $-a$ of an index triple $a = (I,I^3,Y)$ is by definition the index triple $-a := (I, -I^3, -Y)$. In the following considerations a fundamental role is played by the color charge quantum number:

$$Q(a) = Q(I,I^3,Y) := I^3 + 1/2Y, \quad a \in \Gamma(p,q), \quad (2.6)$$

which due to the definition of $-a$ has the property

$$Q(-a) = -Q(a). \quad (2.7)$$

Finally the definitions and conventions above imply the following important connection between a representation $R^{(q,p)}$ and the complex conjugate representation $\overline{R^{(p,q)}}$:

$$R^{(q,p)}(u)^b{}_a = (-1)^{Q(b)-Q(a)}\, \overline{R^{(p,q)}}(u)^{-b}{}_{-a}, \quad (2.8)$$
$$a,b \in \Gamma(q,p), \quad u \in \text{SU}(3).$$

III. GQFT OF COLORED QUARKS AND GLUONS

In this section we present a list of the basic assumptions for a GQFT of colored quarks and gluons and derive some restrictions for the nonunitary representations of the translation group.

Assumption I: Hilbert spaces and the metric operator: Let $\mathscr{H}$ denote a Hilbert space with elements $\Psi, \Phi, \ldots$, scalar product (Ψ,Φ), and norm $\|\Psi\|_{\mathscr{H}} = (\Psi,\Psi)^{1/2}$. There exists a linear, bounded, Hermitian, and nonsingular operator η in $\mathscr{H}$ which generates a nontrivial and nonpositive semidefinite sesquilinear form $\langle \cdot,\cdot \rangle := (\cdot, \eta \cdot)$ on $\mathscr{H}$. Furthermore there exists a nontrivial and maximal linear subspace $H \subset \mathscr{H}$ such that $\langle \Psi,\Psi \rangle \geqslant 0$ for all $\Psi \in H$. If H_0 denotes the linear subspace of all $\Psi \in H$ with $\langle \Psi,\Psi \rangle = 0$, then the completion of the factor space H/H_0 (with elements $[\Psi] := \Psi + H_0$) in the natural scalar $([\Psi],[\Phi])_H := \langle \Psi,\Phi \rangle$ is called the Hilbert space of physical states $\mathscr{H}_{\text{ph}} = \overline{H/H_0}$.

Comments: The metric operator η is nonsingular means by definition that η^{-1} exists and is bounded on its domain $D(\eta^{-1}) = \eta\mathscr{H}$. Together with the Hermiticity of η this implies for every dense linear subspace $S \subseteq \mathscr{H}$ that ηS is again dense in $\mathscr{H}$ and $S \subseteq \eta\mathscr{H}$. Hence η is bijective and $(\eta^{-1})^* = (\eta^*)^{-1}$ ($*$ ~ adjoint in $\mathscr{H}$).

Assumption II: Symmetries and the vacuum: On a dense linear subspace $D_s \subseteq \mathscr{H}$ there exist a representation T of the vector group of $\mathbb{R}_4$ and a representation $W = L \times U$ of the direct product $G = \text{SL}(2,\mathbb{C}) \times \text{SU}(3)$ of the universal covering of the Lorentz group and the internal color group with the following properties.

(a) (η-isometry): For all $y \in \mathbb{R}_4$, $(\alpha,u) \in G$, and $\Psi,\Phi \in D_S$:

$$\begin{aligned} &\langle T(y)\Psi, T(y)\Phi \rangle = \langle \Psi,\Phi \rangle, \\ &\langle W(\alpha,u)\Psi, W(\alpha,u)\Phi \rangle = \langle \Psi,\Phi \rangle. \end{aligned} \quad (3.1)$$

(b) (Vacuum): There exists a unique state $\Psi_0 \in D_S \cap H$ with $\langle \Psi_0,\Psi_0 \rangle = 1$ (called the vacuum) which is invariant under T and W:

$$\forall y \in \mathbb{R}_4, \quad T(y)\Psi_0 = \Psi_0; \quad \forall (\alpha,u) \in G, \quad W(\alpha,u)\Psi_0 = \Psi_0. \quad (3.2)$$

(c) $H \cap D_S$ is invariant under T and W.

Assumption III: Field operators and completeness: There exists a dense linear subspace $D \subseteq D_S$ with $\eta D = D$ and for every $f \in S(\mathbb{R}_4,\mathbb{C})$, $(p,q) \in \{(1,0),(1,1)\}$, $a \in \Gamma(p,q)$, and $\mu \in \{0,1,2,3\}$ a linear operator $\phi^{(p,q)}_{\mu,a}(f)$ with domain $D(\phi^{(p,q)}_{\mu,a}(f))$ which together with its adjoint operator $\phi^{(p,q)*}_{\mu,a}(f) := \phi^{(p,q)}_{\mu,a}(\bar{f})^*$ ($\bar{f}$ complex conjugate of f) satisfies the following conditions.

(a) $D \subseteq D(\phi^{(p,q)}_{\mu,a}(f)) \cap D(\phi^{(p,q)*}_{\mu,a}(f))$,

$\phi^{(p,q)}_{\mu,a}(f)D \subseteq D, \quad \phi^{(p,q)*}_{\mu,a}(f)D \subseteq D.$

(b) For every $\Phi \in \mathscr{H}$ and $\Psi \in D$ the mappings $f \rightarrow (\Phi, \phi^{(p,q)}_{\mu,a}(f)\Psi)$ and $f \rightarrow (\Phi, \phi^{(p,q)*}_{\mu,a}(f)\Psi)$ are linear continuous functionals on $S(\mathbb{R}_4,\mathbb{C})$.

(c) (Completeness): The vacuum Ψ_0 is cyclic with respect to the polynomial *algebra $\mathscr{F}(\phi)$ over $\mathbb{C}$ generated from the set $\{\text{id}_{\mathscr{H}}, \phi^{(p,q)}_{\mu,a}(f), \phi^{(p,q)*}_{\mu,a}(h) \mid f,h \in S(\mathbb{R}_4,\mathbb{C}); (p,q) \in \{(1,0),(1,1)\}; a \in \Gamma(p,q); \mu = 0,\ldots,3\}$ and $\mathscr{F}(\phi)\Psi_0 \cap H$ is dense in H.

(d) (η-stability): $\eta\mathscr{F}(\phi)\Psi_0 = \mathscr{F}(\phi)\Psi_0$.

Since all field operators together with their adjoint ones possess a dense domain, they are all closable. For the sake of notational simplicity we assume they are closed, i.e.,

$$\phi^{(p,q)**}_{\mu,a}(f) = \phi^{(p,q)}_{\mu,a}(f). \quad (3.3)$$

The completeness assumption means that the linear subspace $D_\Pi = \mathscr{F}(\phi)\Psi_0$ is dense in $\mathscr{H}$. As a consequence $\mathscr{H}$ is separable, since $S(\mathbb{R}_4,\mathbb{C})$ is separable.

The fundamental assumption for all the following considerations is the next one.

Assumption IV: Covariance (Substitution rules): The field operators are covariant with respect to Lorentz transformations, color transformations, and translations. This means that on D the following substitution rules hold for all $(p,q)\in\{(1,0),(1,1)\}$, $a\in\Gamma(p,q)$, $\mu\in\{0,\dots,3\}$, and $(\alpha,u)\in G$ (resp. $y\in\mathbb{R}_4$).

(a) $W(\alpha,u)\phi^{(p,q)}_{\mu,a}(f)W(\alpha,u)^{-1}\Psi$

$$= \sum_{\nu=0}^{3}\sum_{b\in\Gamma(p,q)} S^{(p,q)}(\alpha)^{\nu}{}_{\mu}\, R^{(p,q)}(u)^{b}{}_{a}\,\phi^{(p,q)}_{\nu,b}(f_\Lambda)\Psi$$

$W(\alpha,u)\,\phi^{(p,q)*}_{\mu,a}(f)W(\alpha,u)^{-1}\Psi$

$$= \sum_{\nu=0}^{3}\sum_{b\in\Gamma(p,q)} (M_\eta(p,q)\,\overline{S^{(p,q)}}(\alpha)M_\eta(p,q)^{-1})^{\nu}{}_{\mu} \times \overline{R^{(p,q)}}(u)^{b}{}_{a}\,\phi^{(p,q)*}_{\nu,b}(f_\Lambda)\Psi,$$

with $M_\eta(p,q)$ a nonsingular matrix and $f_\Lambda(x) := f(\Lambda(\alpha^{-1})x)$.

(b) $T(y)\;\phi^{(p,q)}_{\mu,a}(f)T(y)^{-1}\Psi$

$$= \exp[Q(a)(y\cdot m(N))]\;\phi^{(p,q)}_{\mu,a}(f_y)\Psi$$

$T(y)\,\phi^{(p,q)*}_{\mu,a}(f)T(y)^{-1}\Psi$

$$= \exp[Q(a)(y\cdot m(N)^*]\;\phi^{(p,q)*}_{\mu,a}(f_y)\Psi,$$

where $m(N)$ and $m(N)^*$ are complex four vectors with nonvanishing real part, which only depend on the dimension N of $R^{(p,q)}$, and $f_y(x) := f(x-y)$.

Before we plunge into the discussion of the covariance assumption we only mention the two remaining assumptions of a GQFT: Assumption V, spectrum condition, and Assumption VI, locality (Einstein causality). We do not spell them out in detail, because we are not going to use them explicitly. The interested reader may consult the literature, especially Ref. 5.

In the following we call the field operators $\phi^{(1,0)}_{\mu,a}$ *quark fields* and the gauge fields $\phi^{(1,1)}_{\mu,a}$ *gluon fields.* In classical gauge theories the gauge fields are self-conjugate. According to the conventions of Sec. II this implies for our gluon fields the condition

$$\phi^{(1,1)A*}_{\mu,a}(f) = (-1)^{Q(a)}\sum_{\nu=0}^{3}\phi^{(1,1)}_{\nu,-a}(f)(M_\eta(1,1)^{-1})^{\nu}{}_{\mu}. \tag{3.4}$$

The matrix $M_\eta(p,q)^{-1}$ on the right-hand side is enforced by the difference in the substitution rules for the two fields under Lorentz transformations (see below).

Next we have to make a couple of comments on the substitution rules of Assumption IV and especially on those of the adjoint (with respect to the scalar product in $\mathscr{H}$!) field operators.

First, it obviously follows from the substitution rules that the linear subspace $D_\Pi \subseteq D$ is invariant under W and T:

$$\begin{aligned}&\forall(\alpha,u)\in G,\quad W(\alpha,u)D_\Pi = D_\Pi;\\ &\forall y\in\mathbb{R}_4,\quad T(y)D_\Pi = D_\Pi.\end{aligned} \tag{3.5}$$

For pure color transformations (ϵ,u) with ϵ the neutral element of SL(2,C) the adjoint field operators $\phi^{(p,q)*}_{\mu,a}(f)$ transform with the complex conjugate $\overline{R^{(p,q)}}(u)$ of the matrices with which the fields $\phi^{(p,q)}_{\mu,a}(f)$ themselves transform. Then Theorem 4.2 of I implies the following.

Statement I: If the vacuum Ψ_0 is an eigenstate of η, then the representation $V(u) = L(\epsilon)\times U(u)$ of the color group SU(3) has a unique unitary extension $\overline{V}$ on $\mathscr{H}$ and hence $[\overline{V}(u),\eta] = 0$.

The same theorem leads to the introduction of the matrices $M_\eta(p,q)$ in the substitution rules for the adjoint field operators under Lorentz transformations. For if as in the case of a normal Wightman theory $M_\eta(p,q)$ is the unit matrix, then again due to Theorem 4.2 of I also the representation of the Lorentz group $W(\alpha,e) = L(\alpha)\times U(e)$ [e the neutral element of SU(3)], would possess a unitary extension onto $\mathscr{H}$ in case Ψ_0 is an eigenstate of η. But this, in general, is not true as we know from the Gupta–Bleuler formulation of the electromagnetic potential.[3] However due to the η isometry of W the η-adjoint field operators $\phi^{(p,q)+}_{\mu,a}(f) := \eta^{-1}\,\phi^{(p,q)*}_{\mu,a}(f)\eta$ transform as usual with the complex conjugate representation..

The substitution rules Assumption IV (b) for the translation group differ from the conventional ones in the occurrence of the exponential factors which will turn out to be the sources for the confinement mechanism. Our motivation for the choice of these unconventional representations is based on the widespread belief[1,2] that confinement is closely connected to the unboundedness of the translation operators $T(y)$ in $\mathscr{H}$. However it follows immediately from Theorem 4.2 in I that, at least in theories in which the vacuum is an eigenstate of the metric operator, the usual representation (without the exponential factors) are necessarily unitary in the unphysical Hilbert space $\mathscr{H}$. Hence if there is any truth in the connection between confinement and unboundedness of the translation operators, then one has to admit unconventional representations of the translation group.

In normal Wightman theories ($\eta = \mathrm{id}_{\mathscr{H}}$, $\mathscr{H} = \mathscr{H}_{\rm ph}$) all unconventional representations are excluded via Poincaré symmetry in the physical Hilbert space. It will be shown in Sec. IV below that in the present case ($\mathscr{H} \neq \mathscr{H}_{\rm ph}$) also the unconventional representations of Assumption IV generate in a natural way a unitary representation of the Poincaré group in the physical Hilbert space $\mathscr{H}_{\rm ph}$ in spite of the fact that Poincaré symmetry obviously does not hold in the unphysical Hilbert space $\mathscr{H}$. Even more with the further restricted choice for the four vectors $m(N)$ and $m(N)^*$, which will be made below, the classical Yang–Mills Lagrangian for chromodynamics is form invariant under the substitution rules [Assumption IV (b)] since the exponential factors compensate each other. Hence in gauge quantum field theories there is no physical reason to ban the unconventional representations from consideration.

Since the exponential factors in Assumption IV (b) enforce the confinement of states their exponents have to depend in some way on the color degrees of freedom. Thus the occurrence of the color charge $Q(I,I^3,Y) = I^3 + \frac{1}{2}Y$ seems more or less natural and is finally justified by the success in the derivation of the results.

In the remainder of the present section we derive some important restrictions for the complex vectors $m(N)$ and $m(N)^*$. Since the gluon fields are self-conjugate it follows at once from Eq. (3.4) and the substitution rules [Assumption IV (b)]:

$$\forall y \in \mathbb{R}_4, \quad \forall a \in \Gamma(1,1),$$
$$\exp(Q(a)[(y\cdot m(8)) + (y\cdot m(8)^*)]) = 1. \tag{3.6}$$

Since four of the eight color charges of the gluon octet are nonzero, we obtain from Eq. (3.6) the two conditions

$$\operatorname{Re} m(8) = -\operatorname{Re} m(8)^*, \tag{3.7}$$
$$\forall y \in \mathbb{R}_4, \quad (y \operatorname{Im}[m(8) + m(8)^*]) = O(\operatorname{mod} \pi),$$

which are equivalent to

$$m(8) = -m(8)^*. \tag{3.8}$$

The same connection between $m(N)$ and $m(N)^*$ for both the gluon and the quark fields holds in theories in which Ψ_0 is an eigenstate of the metric operator η.

Theorem I: If the vacuum Ψ_0 is an eigenstate of the metric operator and if the field oprators $\phi^{(p,q)}_{\mu,a}$ are nontrivial, then we have (i) $m(N)^* = -m(N)$ $(N = N(p,q))$; and (ii) every translation operator $T(y)$ with $y \in \mathbb{R}_4 \setminus \{0\}$ is *nonisometric* on D.

Proof: The η isometry of T implies, for all $y \in \mathbb{R}_4$ and $\Psi \in D$,

$$T(y)^* \eta \Psi = \eta T(-y)\Psi. \tag{3.9}$$

Therefore Ψ_0 is invariant under $T(y)^*$ since Ψ_0 is an eigenstate of η and invariant under $T(y)$. But then we obtain, from Lemma 4.1 of I and the substitution rules [Assumption IV (b)],

$$T(y)^* \phi^{(p,q)}_{\mu,a}(f)\Psi_0 \tag{3.10}$$
$$= \exp[-Q(a)(y\cdot \overline{m(N)^*})]\, \phi^{(p,q)}_{\mu,a}(f_{-y})\Psi_0.$$

From this equation it follows, via straightforward calculations, that

$$\|T(y)\phi^{(p,q)}_{\mu,a}(f)\Psi_0\|^2_{\mathscr{H}}$$
$$= \exp[Q(a)(y\cdot m(N))]$$
$$\times(\phi^{(p,q)}_{\mu,a}(f)\Psi_0, T(y)^*\phi^{(p,q)}_{\mu,a}(f_y)\Psi_0)$$
$$= \exp[Q(a)(y\cdot[m(N) - \overline{m(N)^*}])]$$
$$\times\|\phi^{(p,q)}_{\mu,a}(f)\Psi_0\|^2_{\mathscr{H}}. \tag{3.11}$$

On the other hand we get, via a direct application of the substitution rules,

$$\|T(y)\phi^{(p,q)}_{\mu,a}(f)\Psi_0\|^2_{\mathscr{H}} \tag{3.12}$$
$$= \exp[Q(a)(y\cdot[m(N) + \overline{m(N)}])]$$
$$\times\|\phi^{(p,q)}_{\mu,a}(f_y)\Psi_0\|^2_{\mathscr{H}}.$$

From Theorem 7.1 and its proof in I we know that the norm on the right-hand side of Eq. (3.12) is bounded by a polynomial in y. Therefore, since for the quark triplet all three color charge quantum numbers $Q(a)$ are nonzero and for the gluon octet four of them are nonzero, Eqs. (3.11) and (3.12) imply that either for all $f \in S(\mathbb{R}_4, \mathbb{C})$ and all $a \in \Gamma(p,q)$, with $Q(a) \neq 0$,

$$\phi^{(p,q)}_{\mu,a}(f)\Psi_0 = \mathscr{O} \tag{3.13}$$

or the two conditions

$$\operatorname{Re} m(N) = -\operatorname{Re} m(N)^*, \tag{3.14}$$

$$\forall y \in \mathbb{R}_4, \quad (y \operatorname{Im}[m(N) + m(N)^*]) = O(\operatorname{mod} \pi)$$

hold. This proves part (i), since due to the Reeh–Schlieder theorem[2], Eq. (3.13) implies $\phi^{(p,q)}_{\mu,a}$ to be a trivial field. In this last step we have used locality. Part (ii) follows directly from (i), since due to Theorem 4.2 in I, $T(y)$ is isometric on D if and only if $m(N) = \overline{m(N)^*}$. □

For the remaining part of this note we assume in accordance with part (i) of Theorem I that $m(N) = -m(N)^*$. However Ψ_0 does not need to be an eigenstate of η. Finally the form invariance of the classical Yang–Mills Lagrangian for chromodynamics under the substitution rules [Assumption IV (b)] with $m(N) = -m(N)^*$ requires in addition $m(3) = m(8)$ (see Appendix). Hence there remains one single four-vector m in the substitution rules for all quark and gluon fields, if we add to our list the further reasonable assumption

$$m = m(3) = m(8) = -m(3)^* = -m(8)^*. \tag{3.15}$$

Since SU(3)–Clebsch–Gordan calculations will be an essential tool in the following considerations, we have to introduce in place of the adjoint quark fields $\phi^{(1,0)*}_{\mu,a}(f)$, $a \in \Gamma(1,0)$, new fields which transform under color transformations according to a representation satisfying the Condon–Shortley phase conventions,[10] namely the representation $R^{(0,1)}$. This is easily achieved by the definition

$$\phi^{(0,1)}_{\mu,-a}(f) := (-1)^{Q(a)+1/3}$$
$$\times \sum_{\nu=0}^{3} \phi^{(1,0)*}_{\nu,a}(f)(M_\eta(1,0)\gamma^0)^\nu{}_\mu, \tag{3.16}$$

with $a \in \Gamma(1,0)$. Notice that due to the relation (2.5), $a \in \Gamma(1,0)$ implies $-a \in \Gamma(0,1)$ and vice versa. The new field operators $\phi^{(0,1)}_{\mu,a}(f)$ with $a \in \Gamma(0,1)$ are called *antiquark fields*.

A straightforward calculation, which takes into account Assumption IV (a) and Eqs. (2.2) and (2.8), shows that the antiquark fields transform under Lorentz and color transformations with the direct product matrix $S^{(0,1)}(\alpha) \times R^{(0,1)}(u)$. Hence the substitution rules of Assumption IV (a) read in compact form, for all $p,q = 0,1$, $a \in \Gamma(p,q)$, and $\Psi \in D$,

$$W(\alpha,u)\phi^{(p,q)}_{\mu,a}(f)W(\alpha,u)^{-1}\Psi$$
$$= \sum_{\nu=0}^{3} \sum_{b \in \Gamma(p,q)} S^{(p,q)}(\alpha)^\nu{}_\mu R^{(p,q)}(u)^b{}_a \phi^{(p,q)}_{\nu,b}(f_\Lambda)\Psi. \tag{3.17}$$

Similarly in view of Eqs. (3.15) and (3.16) we obtain from Assumption IV (b), for all $p,q = 0,1$, $a \in \Gamma(p,q)$, and $\Psi \in D$,

$$T(y)\phi^{(p,q)}_{\mu,a}(f)T(y)^{-1}\Psi \tag{3.18}$$
$$= \exp[Q(a)(y\cdot m)]\, \phi^{(p,q)}_{\mu,a}(f_y)\Psi, \quad \operatorname{Re} m \neq 0.$$

IV. CONFINEMENT AND POINCARÉ INVARIANCE

An important technical input for the derivation of the confinement properties is the following lemma which is proved along the same lines as in the case of the Hilbert space scalar product.

Lemma I: Let $\{\Psi^{(p,q)}_a | a \in \Gamma(p,q)\}$, $(p,q) \in \mathbb{N}^0 \times \mathbb{N}^0$ be a set of vectors from D_S transforming with an irreducible matrix representation of the color group SU(3)

$$V(u)\,\Psi_a^{(p,q)} = \sum_{b\in\Gamma(p,q)} R^{(p,q)}(u)^b{}_a\,\Psi_b^{(p,q)}, \tag{4.1}$$

then the following condition is true:

$$\forall a,b\in\Gamma(p,q),\quad \langle \Psi_a^{(p,q)},\Psi_a^{(p,q)}\rangle = \langle \Psi_b^{(p,q)},\Psi_b^{(p,q)}\rangle. \tag{4.2}$$

Proof: If ρ denotes the invariant SU(3) measure, then the orthogonality relation[10]

$$\int d\rho(u)\,\overline{R^{(p,q)}(u)^b{}_a}\,R^{(p',q')}(u)^c{}_d = (1/N)\delta_{pp'}\delta_{qq'}\delta_{bc}\delta_{ad} \tag{4.3}$$

together with the η isometry of V imply, for all $a\in\Gamma(p,q)$,

$$\langle \Psi_a^{(p,q)},\Psi_a^{(p,q)}\rangle = \frac{1}{N}\sum_{b\in\Gamma(p,q)}\langle \Psi_b^{(p,q)},\Psi_b^{(p,q)}\rangle. \tag{4.4}$$

Every solution of this homogeneous system of N equations is determined by the condition (4.2). □

Now all the confinement properties are direct consequences of Theorem 8.1 of I, which, for the present case, states that for all $f^i\in S(\mathbb{R}_4,\mathbb{C})$, $p_i,q_i=0,1$, $a_i\in\Gamma(p_i,q_i)$, $\mu_i=0,\dots,3$, and $n\in\mathbb{N}$, we have ($\mathscr{H}_0$ is the set of states with zero η norm)

$$\prod_{i=1}^{n}\phi_{\mu_i,a_i}^{(p_i,q_i)}(f_i)\Psi_0\in\mathscr{H}_0\supset H_0, \tag{4.5}$$

unless the charges carried by the fields satisfy the condition

$$\sum_{i=1}^{n} Q(a_i)=0. \tag{4.6}$$

An immediate consequence is that all three quark fields as well as all three antiquark fields and (due to Lemma I) also all eight gluon fields are confined (in the sense of the next theorem).

Theorem II: Every state created from the vacuum by application of a quark, antiquark, or gluon field has zero η norm; i.e.,

$$\phi_{\mu,a}^{(p,q)}(f)\Psi_0\in\mathscr{H}_0.$$

On the other hand, the products of either three different quark fields $\phi_{\dots,r}^{(1,0)}(h)\phi_{\dots,g}^{(1,0)}(f)\phi_{\dots,b}^{(1,0)}(k)$ or three different antiquark fields $\phi_{\dots,\bar r}^{(0,1)}(h)\phi_{\dots,\bar g}^{(0,1)}(f)\phi_{\dots,\bar b}^{(0,1)}(k)$ as well as the quark–antiquark products [for $a\in\Gamma(1,0)$] $\phi_{\dots,-a}^{(0,1)}(f)\phi_{\dots,a}^{(1,0)}(h)$, respectively, the special gluon products [for $a\in\Gamma(1,1)$] $\phi_{\dots,-a}^{(1,1)}(f)\phi_{\dots,a}^{(1,1)}(h)$ all satisfy the condition (4.6). Hence at a first glance all states created from the vacuum Ψ_0 by these products or, more generally, by any products of at least two quark, antiquark, or gluon fields, for which the sum of the corresponding color charges vanishes, seem to escape the fate of being confined in $\mathscr{H}_0$. However, it turns out that this is not true for all of them. We are going to show that all states from the dense linear subspace $D_\Pi\cap H$ of H are from H_0 unless they are equivalent mod$(D_\Pi\cap H_0)$ to a *SU(3)-singlet* state. Two states $\Psi,\Phi\in\mathscr{H}$ are called equivalent mod$(D_\Pi\cap H_0)$ if and only if $\Psi-\Phi\in(D_\Pi\cap H_0)$. According to Theorem 3.3 of I or Lemma 1 and Theorem 2 of Mintchev and d'Emilio[12] the linear space of equivalence classes

$$\mathscr{D}_{\Pi\mathrm{ph}} := \{[\Psi]_\Pi = \Psi + (D_\Pi\cap H_0)\,|\,\Psi\in D_\Pi\cap H\} \tag{4.7}$$

forms with the natural scalar product $([\Psi]_\Pi,[\Phi]_\Pi)_\Pi := \langle\Psi,\Phi\rangle$ a pre-Hilbert space and its completion $\overline{\mathscr{D}_{\Pi\mathrm{ph}}}$ is unitary equivalent to the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}$. Moreover due to the η isometry and the relations (3.5) $D_\Pi\cap H_0$ and $D_\Pi\cap H$ are invariant with respect to color transformations. Hence all elements of $\mathscr{D}_{\Pi\mathrm{ph}}$ are color singlets. Therefore, provided every state of $D_\Pi\cap H$ is equivalent mod$(D_\Pi\cap H_0)$ to a SU(3) singlet or to the zero vector, we arrive at our main result.

Theorem III: Every state from the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}$ is a SU(3) singlet state.

The proof will be cut into two pieces. In the first step we consider any SU(3)-tensor operator of the form

$$\phi_{\mu_1,\dots,\mu_n;a}^{(p,q)\Delta}(f^1\times\cdots\times f^n|(p_1,q_1),\dots,(p_n,q_n))\Psi := \sum_{a_1\in\Gamma(p_1,q_1)}\cdots\sum_{a_n\in\Gamma(p_n,q_n)} K\left(\begin{matrix}(p_1,q_1),\dots,(p_n,q_n)\\ a_1\ ,\dots,\ a_n\end{matrix}\middle|\begin{matrix}(p,q)\\ a\end{matrix}\middle|\Delta\right)\prod_{i=1}^{n}\phi_{\mu_i,a_i}^{(p_i,q_i)}(f^i)\Psi,$$
$$\Delta\in E_n(p,q), \tag{4.8}$$

which transforms with an irreducible matrix representation $R^{(p,q)}(u)$ under color transformations:

$$V(u)\,\phi_{\dots;a}^{(p,q)\Delta}(\cdots|\cdots)V(u)^{-1}\Psi = \sum_{b\in\Gamma(p,q)} R^{(p,q)}(u)^b{}_a\,\phi_{\dots;b}^{(p,q)}(\cdots|\cdots)\Psi. \tag{4.9}$$

The (multi) index Δ describes the different irreducible tensor operators with fixed (p,q) which can be constructed from one and the same given product of field operators (due to different coupling modes and different symmetry types, etc.). The corresponding index set $E_n(p,q)$, of course, depends on the pairs (p_i,q_i), $i=1,\dots,n$ occurring in the given product. This dependence is not spelled out explicitly but should be kept in mind. For instance, in the case $n=2$, $E_2(p,q)$ contains always one element except for the product of two gluon octets and $(p,q)=(1,1)$. In this case there exist two different irreducible octet operators, a symmetric and an antisymmetric one.[10]

With increasing n the set $E_n(p,q)$ becomes rapidly more and more complex due to the different coupling modes and the occurrence of further mixed symmetry types. The coefficients $K(\cdots|\cdots|\cdots)$ are composed of finite sums of products of SU(3)–Clebsch–Gordan coefficients. The only important point for the following is that

$$K\left(\begin{matrix}(p_1,q_1)\ ,\dots,\ (p_n,q_n)\\ (I_1,I_1^3,Y_1),\dots,(I_n,I_n^3,Y_n)\end{matrix}\middle|\begin{matrix}(p,q)\\ (I,I^3,Y)\end{matrix}\middle|\Delta\right)=0, \tag{4.10}$$
$$\text{for } I^3\neq\sum_{i=1}^{n} I_i^3 \ \vee\ Y\neq\sum_{i=1}^{n} Y_i,$$

since the "color hypercharge" Y and the third component of the "color isopin" I^3 always couple additivity. A first important consequence reads as follows.

Lemma II: Every irreducible SU(3) tensor operator $\phi_{\dots,a}^{(p,q)\Delta}(\cdots|\cdots)$ with $(p,q)\neq(0,0)$ is confined:

$$\forall a\in\Gamma(p,q),\quad \forall\Delta\in E_n(p,q),\quad \phi^{(p,q)\Delta}_{\ldots,a}(\cdots|\cdots)\Psi_0\in\mathscr{H}_0. \tag{4.11}$$

Proof: In view of the condition (4.10) and the definition (4.5) the color charge $Q(a)$ of the tensor operator $\phi^{(p,q)\Delta}_{\ldots,a}(\cdots|\cdots)$ is just the sum of the color charges of its constituents: $Q(a)=\Sigma_{i=1}^n Q(a_i)$. But then the substitution rules (3.18) for the basic fields imply, for all $y\in\mathbb{R}_4$, $a\in\Gamma(p,q)$, and $\Delta\in E_n(p,q)$,

$$T(y)\phi^{(p,q)\Delta}_{\ldots,a}(f^1\times\cdots\times f^n|\cdots)T(y)^{-1}\Psi$$
$$=\exp[Q(a)(y\cdot m)]\,\phi^{(p,q)\Delta}_{\ldots,a}(f^1_a\times\cdots\times f^n_y|\cdots)\Psi. \tag{4.12}$$

For any irreducible SU(3)-tensor operator which is not a SU(3)-singlet operator there exists at least one $a=(I,I^3,Y)\in\Gamma(p,q)$ with $Q(a)\neq 0$. Namely the component with the largest value of I^3 carries the quantum numbers[10] $I=I^3=1/2(p+q)$ and $Y=1/3(p-q)$. The charge is $1/3(2p+q)$. Then the relation (4.11) follows immediately from Eq. (4.12), Theorem 8.1 of I and Lemma I in the same way as for the (anti-)quark and gluon fields themselves. □

For the second step in the proof of Theorem III we have just to invert the Clebsch–Gordan calculations which were applied in the construction of the irreducible SU(3)-tensor operators above.

Lemma III: Every state $\Psi\in D_\Pi\cap H$ is from H_0 unless it is equivalent mod$(D_\Pi\cap H_0)$ to a SU(3)-singlet state.

Proof: It suffices to prove Lemma III for a state of the form $\Psi_n=\Pi_{i=1}^n\phi^{(p_i,q_i)}_{\mu_i,a_i}(f^i)\Psi_0$, since D_Π is the linear hull of such states and the vacuum. Via straightforward Clebsch–Gordan calculations Ψ_n can be expanded into a (finite) linear combination of the irreducible SU(3)-tensor operators introduced above. But then Lemma III follows immediately from Lemma II. □

It remains to investigate the Poincaré invariance which, as was mentioned in Sec. III, is violated in the big unphysical Hilbert space $\mathscr{H}$. The reason is the following. In spite of the fact that for every pair (α,y) from the semidirect product $P^\uparrow_+ = \mathrm{SL}(2,\mathbb{C})\ominus\mathbb{R}_4$ the mapping

$$\mathscr{P}(\alpha,y):D_\Pi\to D_\Pi,\quad \Psi\to\mathscr{P}(\alpha,y)\Psi:=T(y)W(\alpha,e)\Psi \tag{4.13}$$

is an automorphism of D_Π, the corresponding mapping

$$\mathscr{P}:P^\uparrow_+\to\mathrm{Aut}\,D_\Pi,(\alpha,y)\to\mathscr{P}(\alpha,y)=T(y)W(\alpha,e) \tag{4.14}$$

of the Poincaré group into the automorphism group of D_Π cannot be a homomorphism. Due to the occurrence of the exponential factors in the substitution rules of $T(y)$ the composition law of $P^\uparrow_+$ does not hold for the images $\mathscr{P}(\alpha,y)$, $(\alpha,y)\in P^\uparrow_+$. However, the recipe to restore the Poincaré invariance in the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}$ (and this is all we need from the physical point of view) is obvious. Consider the linear subspace of D_Π defined by

$$D_C:=\mathrm{L.H.}\Big\{\Psi_0,\prod_{i=1}^n\phi^{(p_i,q_i)}_{\mu_i,a_i}(f^i)\Psi_0\big|f^i\in S(\mathbb{R}_4,\mathbb{C});$$
$$\mu_i=0,\ldots,3;\ q_i,p_i=0,1;$$
$$a_i\in\Gamma(p_i,q_i);\ i=1,\ldots,n;$$
$$n\in\mathbb{N};\ \sum_{j=1}^n Q(a_j)=0\Big\}. \tag{4.15}$$

Here L. H. means the linear hull; and D_C is dense in $\mathscr{H}\setminus\mathscr{H}_0$ and invariant under $\mathscr{P}(\alpha,y)$ for all $(\alpha,y)\in P^\uparrow_+$. On this subspace the exponential factors disappear from $\mathscr{P}(\alpha,y)$:

$$\mathscr{P}(\alpha,y)\prod_{i=1}^n\phi^{(p_i,q_i)}_{\mu_i,a_i}(f^i)\Psi_0$$
$$=\sum_{\nu_1=0}^3\cdots\sum_{\nu_n=0}^3\prod_{i=1}^n S^{(p_i,q_i)}(\alpha)^{\nu_i}{}_{\mu_i}$$
$$\times\phi^{(p_i,q_i)}_{\nu_i,a_i}(f^i_{(\Lambda,y)})\Psi_0, \tag{4.16}$$

with $f_{(\Lambda,y)}(x):=f(\Lambda(\alpha^{-1})(x-y))$. Hence the restriction $\mathscr{P}_C$ of $\mathscr{P}$ to D_C

$$\mathscr{P}_C:P^\uparrow_+\to\mathrm{Aut}\,D_C,\quad(\alpha,y)\to\mathscr{P}_C(\alpha,y):=\mathscr{P}(\alpha,y)\restriction D_C \tag{4.17}$$

is an η-isometric representation of the Poincaré group $P^\uparrow_+$ on D_C. Moreover it has been shown in Theorem 8.2 of I that the representation $\mathscr{P}_C$ lifts to a continuous unitary representation of $P^\uparrow_+$ on the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}$.

V. FINAL REMARKS

In spite of the rather encouraging results obtained in the sections before, there remains a missing link to a satisfactory gauge quantum field theoretical frame for quantum chromodynamics with confinement. According to Lemma II every state with nonzero (positive *and* negative) η norm is equivalent to a color-singlet state. But only a subset of states with positive η norm are physical states in the sense that they generate the Hilbert space $\mathscr{H}_{\mathrm{ph}}$. On the other hand due to a theorem of Strocchi[5] the set of states with negative η norm is nonempty. Hence a condition is missing which tells us that the color-singlet states built from quark and antiquark fields lie in the subspace H and that the remaining color singlets do not, that is, a well-defined subset of them has negative η norm. Since this problem touches more detailed dynamical features of quantum chromodynamics we have to leave it open for further investigation.

Also the question of "asymptotic freedom" and related problems must be left to further investigations.

The confinement mechanism presented in this note seems to work along completely different lines than that in the local and covariant operator solution of the Schwinger model[13–15] (QED_2) in an indefinite inner product space. A closer inspection, however, exhibits some similarities. For the confinement of the *basic* charged fermion fields (i.e., the fields occurring in the Lagrangian) relies on the fact that charged states created by their powers from the vacuum of the indefinite inner product space are lying *outside* the positive semidefinite subspace from which the physical Hilbert space is constructed via factorization. For instance, the two-point $\langle\cdot,\cdot\rangle$-matrix element of the basic fermion field is either negative or in exceptional cases (depending on the regularization parameter) zero.[16] The positive semidefinite subspace generating the physical Hilbert space is obtained by the application of monomials in the so-called bleached field[14] to the vacuum state. The bleached field is constructed from the basic fields in a subtle way such that it does not carry electric charge.

Hence the confinement mechanism of the present note and of the Schwinger model have two features in common:

(i) The positive semidefinite subspace generating the physical Hilbert space is obtained from the vacuum via application of monomials in *nonbasic* fields which do not carry (color) charge. (ii) The *basic* fields are confined in the sense that all charged states created by them from the vacuum lie either in the complement or the zero part of the positive semidefinite subspace above. This is the deeper reason why they cannot show up as physical states. Both mechanisms mainly differ in what one may call the bleaching process; this means the manner in which the uncharged fields creating the physical states are selected (resp. constructed) from the basic ones. Moreover, in contrast to the present case the charged states of the Schwinger model can have nonzero $\langle\cdot,\cdot\rangle$ norm since the representation of the translation group is the conventional one. It is not yet known if this model also admits unconventional representations.

Up to now also all perturbative treatments of quantum chromodynamics[1] were performed within the conventional representation and hence it is not astonishing if there occurs some evidence of colored states with nonvanishing $\langle\cdot,\cdot\rangle$ norm. The general belief seems to be that these states by some subtle mechanism do not occur in the physical Hilbert space. However, as long as such a mechanism is not established one has to realize the existence of other representations with a (too?) simple confinement mechanism. According to the appendix in classical Lagrangian field theory of chromodynamics the basic fields with nonstandard substitution rules under translations are possible new solutions of the field equations. Moreover, at the moment no general physical principle is known which justifies the preference of any one of these representations above the other. Finally, one should remember that the classical paper of Yang and Mills[17] did not seem to fit into the conventional wisdom of Lagrangian quantum field theory when it was published thirty years ago. It took more than fifteen years to establish that it fits very well.

ACKNOWLEDGMENTS

The author is indebted to J-J. Loeffel, P. Stichel, and G. Wanders for critical discussions and to J-J. Loeffel also for pointing out an error in the original preprint.

APPENDIX: CLASSICAL LCD

In this appendix we consider the classical Lagrangian field theory of chromodynamics. Especially we investigate the consequences of the nonstandard substitution rules of the fields under translations. We show the following.

(1) The Lagrangian and hence the field equations as well as the Poisson brackets are invariant under the translations corresponding to the nonstandard representations (3.18).

(2) The nonisometry of T in the unphysical Hilbert space $\mathscr{H}$ or equivalently the noncommutativity of T and η corresponds on the classical level to the fact that the complex conjugation of the fields does not commute with the translations.

In order to avoid notational confusion we describe the classical fields corresponding to $\phi^{\cdots}_{\cdots}(f)$ by the symbol $\varphi^{\cdots}_{\cdots}(x)$, $x \in \mathbb{R}_4$. If $\bar{\varphi}$ denotes the complex conjugate of φ we have

$$\phi^{(p,q)}_{\mu,a}(f) \leftrightsquigarrow \varphi^{(p,q)}_{\mu,a}(x), \tag{A1}$$

$$\phi^{(p,q)*}_{\nu,a}(f) M_\eta(p,q)^\nu{}_\mu \leftrightsquigarrow \bar{\varphi}^{(p,q)}_{\mu,a}(x) := \overline{\varphi^{(p,q)}_{\mu,a}(x)}.$$

Self-conjugateness of the gluon fields means

$$\overline{\varphi^{(1,1)}_{\mu,a}(x)} = (-1)^{Q(a)} \varphi^{(1,1)}_{\mu,-a}(x). \tag{A2}$$

Dirac indices will be often dropped by the usual conventions. If $\varphi^{(1,0)}_a$ is a column, then by Eq. (3.16) $\varphi^{(0,1)}_a$ is a row! The Dirac adjoint of $\varphi^{(1,0)}_a$ is connected to $\varphi^{(0,1)}_a$ by

$$\tilde{\varphi}^{(1,0)}_a(x) := \overline{\varphi^{(1,0)}_a(x)}^t \gamma^0 = (-1)^{-Q(a)-1/3} \varphi^{(0,1)}_{-a}(x). \tag{A3}$$

Corresponding to Eqs. (3.17) and (3.18), the substitution rules under SL(2,$\mathbb{C}$)$\times$SU(3), respectively, the translation group, reads

$$(W(\alpha,u)\varphi)^{(p,q)}_{\mu,a}(x) = S^{(p,q)}(\alpha)^\nu{}_\mu R^{(p,q)}(u)^b{}_a \varphi^{(p,q)}_{\nu,b}(\Lambda(\alpha)x), \tag{A4}$$

$$(T(y)\varphi)^{(p,q)}_{\mu,a}(x) = \exp[Q(a)(y\cdot m)]\, \varphi^{(p,q)}_{\mu,a}(x+y). \tag{A5}$$

A first important consequence of the Eqs. (A2)–(A4) is that for all (color-) charged fields the translations do not commute with complex conjugation:

$$(\overline{T(y)\varphi})^{(p,q)}_a(x) = \exp[2Q(a)(y \operatorname{Re} m)](T(y)\bar{\varphi})^{(p,q)}_a(x). \tag{A6}$$

Especially the Dirac adjoint of $\varphi^{(1,0)}_a$ transforms according to

$$(T(y)\tilde{\varphi})^{(1,0)}_a(x) = \exp[-Q(a)(y\cdot m)]\, \tilde{\varphi}^{(1,0)}_a(x+y). \tag{A7}$$

The Lagrangian $L(x)$ of chromodynamics is completely determined by the quark Lagrangian [summation over $a \in T(1,0)$]

$$L_q(x) = \tilde{\varphi}^{(1,0)}_a(x)((i/2)\gamma^\mu \overleftrightarrow{\partial}_\mu - k)\, \varphi^{(1,0)}_a(x), \tag{A8}$$

the form invariance under local SU(3)-gauge transformations

$$(G_w\varphi)^{(1,0)}_a(x) = \varphi^{(1,0)}_a(x) - (i/2)(\lambda_\alpha)^b{}_a \varphi^{(1,0)}_b(x)\omega^\alpha(x) + \cdots, \tag{A9}$$

and the principle of minimal coupling of gauge fields. It reads[1,18]

$$L(x) = -\tfrac{1}{4} F^{\mu\nu}_\alpha(x) F^\alpha_{\mu\nu}(x) + \tilde{\varphi}^{(1,0)}_a(x)\big[(i/2)\gamma^\mu\{\delta^b_a \overleftrightarrow{\partial}_\mu + ig(\lambda_\alpha)^b{}_a A^\alpha_\mu(x)\} - \delta^b_a k\big]\, \varphi^{(1,0)}_a(x), \tag{A10}$$

$$F^\alpha_{\mu\nu}(x) := \partial_\mu A^\alpha_\nu(x) - \partial_\nu A^\alpha_\mu(x) - g f_\beta{}^\alpha{}_\gamma A^\beta_\mu(x) A^\gamma_\nu(x). \tag{A11}$$

Here λ_α $(\alpha = 1,\ldots,8)$ denote the traceless Hermitian (3,3) matrices of the fundamental representation of the Lie algebra su(3) and $f_\alpha{}^\beta{}_\gamma$ its real totally antisymmetric structure constants. They satisfy the relation[19]

$$[\lambda_\alpha, \lambda_\beta]_- = 2if_\alpha{}^\gamma{}_\beta \lambda_\gamma \quad \text{and} \quad \operatorname{tr}(\lambda_\alpha\lambda_\beta) = 2\delta_{\alpha\beta}. \tag{A12}$$

The $A^\alpha(x)$ $(\alpha = 1,\ldots,8)$ are real vector fields transforming under local SU(3)-gauge transformations according to

$$(G_w A)^\alpha_\mu(x) = A^\alpha_\mu(x) + f_\gamma{}^\alpha{}_\beta A^\beta_\mu(x)\omega^\gamma(x) + (1/g)\partial_\mu \omega^\alpha(x) + \cdots. \tag{A13}$$

The connection between the eight real gauge fields A^α and our previous complex and self-conjugate gluon fields $\varphi^{(1,1)}_{(I,I^3,Y)}$ is given by[19]

$$\varphi^{(1,1)}_{(1,1,0)} = -(1/\sqrt{2})(A^1 + iA^2), \quad \varphi^{(1,1)}_{(1,0,0)} = A^3,$$
$$\varphi^{(1,1)}_{(1/2,1/2,1)} = -(1/\sqrt{2})(A^4 + iA^5), \tag{A14}$$
$$\varphi^{(1,1)}_{(1/2,-1/2,1)} = -(1/\sqrt{2})(A^6 + iA^7), \quad \varphi^{(1,1)}_{(0,0,0)} = A^8.$$

In analogy with these equations we introduce instead of the Hermitian matrices λ_α the self-conjugate ones τ^a $[a \in T(1,1)]$:

$$\tau^{(1,1,0)} := -(1/\sqrt{2})(\lambda_1 - i\lambda_2), \quad \tau^{(1,0,0)} := \lambda_3,$$
$$\tau^{(1/2,1/2,1)} := -(1/\sqrt{2})(\lambda_4 - i\lambda_5), \quad \tau^{(0,0,0)} := \lambda_8,$$
$$\tau^{(1/2,-1/2,1)} := -(1/\sqrt{2})(\lambda_6 - i\lambda_7), \quad \tau^{-a} := (-1)^{Q(a)}\bar{\tau}^a. \tag{A15}$$

Equations (A12) imply for the τ^a (see Ref. 19)

$$[\tau^a,\tau^b]_- = -2\sqrt{3}\begin{pmatrix} 8 & 8 & 8_A \\ a & b & c \end{pmatrix}\tau^c, \quad \mathrm{tr}(\tau^a\bar{\tau}^b) = 2\delta^{ab}. \tag{A16}$$

By means of these equations it easily follows from (A10) [resp. (A11)] that

$$L(x) = -\tfrac{1}{4}(-1)^{Q(a)}G_{\mu\nu,-a}(x)G^{\mu\nu}{}_{,a}(x) + \tilde{\varphi}^{(1,0)}_a(x)\left[(i/2)\gamma^\mu\{\delta^b_a\overleftrightarrow{\partial}_\mu + ig(\tau^c)^b{}_a\,\varphi^{(1,1)}_c(x)\} - \delta^b_a\,k\right]\varphi^{(1,0)}_b(x), \tag{A17}$$

with

$$G_{\mu\nu,a}(x) := \partial_\mu\,\varphi^{(1,1)}_{\nu,a}(x) - \partial_\nu\,\varphi^{(1,1)}_{\mu,a}(x) + ig\sqrt{3}\begin{pmatrix} 8 & 8 & 8_A \\ b & c & a \end{pmatrix}\varphi^{(1,1)}_{\mu,b}(x)\,\varphi^{(1,1)}_{\nu,c}(x). \tag{A18}$$

Since the Clebsch–Gordan coefficients on the right-hand side vanish unless $Q(a) = Q(b) + Q(c)$, the gluon tensor transforms under translations according to

$$(T(y)G)_{\mu\nu,a}(x) = \exp[Q(a)(y\cdot m)]\,G_{\mu\nu,a}(x+y). \tag{A19}$$

Furthermore, integrating the equation

$$(\tau^c)^b{}_a = R^{(1,0)}(u)^b{}_f R^{(1,1)}(u)^c{}_e\,\overline{R^{(1,0)}(u)^a{}_d}(\tau^e)^f{}_d \tag{A20}$$

with the Haar measure over the whole group SU(3), we obtain[10]

$$(\tau^c)^b{}_a = \frac{1}{3}\begin{pmatrix} 3 & 8 & 3 \\ b & c & a \end{pmatrix}\begin{pmatrix} 3 & 8 & 3 \\ f & e & d \end{pmatrix}(\tau^e)^f{}_d, \tag{A21}$$

and therefore it follows that

$$(\tau^c)^b{}_a \neq 0 \Longrightarrow Q(a) = Q(b) + Q(c). \tag{A22}$$

Now the form invariance of the Lagrangian under the nonstandard substitution rules (A5), (A7) [resp. (A19)] is obvious, i.e., $L(x+y) = (T(y)L)(x)$.

In order to investigate the translational invariance of the Poisson brackets we add to the Lagrangian $L(x)$ the gauge fixing term[20]

$$L_{\mathrm{GF}}(x) = -(-1)^{Q(a)}\left[B^{(1,1)}_{-a}(x)\overleftarrow{\partial}^\mu\,\varphi^{(1,1)}_{\mu,a}(x) - (r/2)\,B^{(1,1)}_{-a}(x)\,B^{(1,1)}_a(x)\right]. \tag{A23}$$

Here $B^{(1,1)}_a$ is a Lorentz scalar, self-conjugate SU(3) octet field which transforms under translations according to Eq. (A5). Considering $\varphi^{(1,0)}_a, \tilde{\varphi}^{(1,0)}_a, \varphi^{(1,1)}_{k,a}$ $(k = 1,2,3)$ and instead of $\varphi^{(1,1)}_{0,a}$ the field $B^{(1,1)}_a$ as independent canonical coordinates then the corresponding canonical momenta read

$$\Pi^{(1,0)}_a(x) = (i/2)\,\tilde{\varphi}^{(1,0)}_a(x)\gamma^0, \quad \tilde{\Pi}^{(1,0)}_a(x) = -(i/2)\gamma^0\,\varphi^{(1,0)}_a(x),$$
$$\Pi^{(1,1)}_{k,a}(x) = -(-1)^{Q(a)}G^{0k}{}_{,a}(x) \quad (k = 1,2,3), \tag{A24}$$
$$\Pi^{(1,1)}_{0,a}(x) := \Pi^B_a(x) = (-1)^{Q(a)}\,\varphi^{(1,1)}_{0,-a}(x).$$

Now obviously the substitution rules (A5), (A7) [resp. (A19)] imply the invariance of the Poisson brackets under translations, since the exponential factors compensate each other in the nonvanishing ones. Of course, it is trivial to include the gauge fixing field $B^{(1,1)}_a$ into the set of basic fields in the main part of this paper without changing the results.

[1] P. Becher, M. Böhm, H. and Joos, *Eichtheorien der starken und elektroschwachen Wechselwirkung* (Teubner, Stuttgart, 1981).
[2] F. Strocchi, Phys. Rev. D **17**, 2010 (1978).
[3] A. S. Wightman and L. Gårding, Ark. Fys. **28**, 129 (1964).
[4] F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[5] F. Strocchi, Commun. Math. Phys. **56**, 57 (1977).
[6] G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré **AXXXIII**, 251 (1980).
[7] A. H. Völkel, "Unbounded Representations of Symmetry Groups in Gauge Quantum Field Theory. I. Confinement and Differentation," FU-Berlin preprint Math. 150, April 1983.
[8] L. Schwartz, *Théories des Distributions* (Hermann, Paris, 1967).
[9] A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
[10] J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1962).
[11] S. Okubo, Progr. Theor. Phys. **27**, 949 (1962).
[12] M. Mintchev and E. d'Emilio, J. Math. Phys. **22**, 1267 (1981).
[13] J. H. Lowenstein and J. A. Swieca, Ann. Phys. **68**, 172 (1971).
[14] A. K. Raina and G. Wanders: Ann. Phys. **132**, 404 (1981).
[15] L. V. Belvedere, J. A. Swieca, K. D. Rothe, and B. Schroer, Nucl. Phys. B **153**, 112 (1979).
[16] G. Wanders (private communication).
[17] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[18] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[19] J. J. de Swart, "Symmetries of Strong Interactions," in *1966 CERN School of Physics* (Noordwijk-aan-Zee, The Netherlands, 1966).
[20] T. Kugo and I. Ojima, Suppl. Progr. Theor. Phys. **66**, 1 (1979).

Proof of the decoupling theorem of field theory in Minkowski space. II. Theories involving particles with vanishingly small masses

Edward B. Manoukian
Department of National Defence, Royal Military College of Canada, Kingston, Ontario K7L 2W3, Canada

(Received 22 February 1984; accepted for publication 21 December 1984)

The decoupling theorem of quantum field theory is studied in Minkowski space for theories which on experimental grounds may contain particles with vanishingly small masses. Rules are set up to prove the distributional vanishing property of the renormalized amplitudes when any subset of the underlying masses is scaled to infinity, and any subset of the remaining masses is scaled to zero. By careful estimates, the analysis in Minkowski space may be reduced to a similar one in Euclidean space. All subtractions of renormalization are carried out at the origin of momentum space with the degree of divergence of a subtraction coinciding with the dimensionality of the corresponding subdiagram.

I. INTRODUCTION

A rigorous proof of the decoupling theorem of field theory in Minkowski space has been recently given[1] for theories with strictly massive particles. The vanishing of the renormalized Feynman amplitudes, in the sense of distributions, was established.[1] The purpose of the present paper is to generalize this theorem to the more difficult problem for theories that may involve particles with vanishingly small masses. Simple rules are set up to prove in Minkowski space, the vanishing of the renormalized amplitudes, in the sense of distributions, when any subset of the masses become large and some of the remaining masses are scaled to zero. The study is very general in that we allow those masses becoming large to grow, in general, at different rates, and those masses becoming small to vanish, in general, at different rates. A key element in our proof is that the analysis in Minowski space may be reduced to a similar one in Euclidean space. The estimates developed in Ref. 1, however, had to be refined as they involved, in an obvious manner, the existence of a smallest nonvanishing mass, and hence are not useful in the present situation. All the subtractions of renormalization are carried out in momentum space about the origin, with the degree of divergence $d(g)$ associated with a subtraction coinciding with the dimensionality of the corresponding subdiagram g.

II. PROOF OF THE THEOREM

We are interested in studying the limit

$$\lim_{\substack{\xi_1,\ldots,\xi_k\to\infty\\ \lambda_1,\ldots,\lambda_s\to 0}} \left(\lim_{\epsilon\to+0} T^{(\epsilon)}_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}(f)\right), \tag{1}$$

with

$$T^{(\epsilon)}_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}(f) = \int_{\mathbf{R}^{4m}} dP f(P) \int_{\mathbf{R}^{4m}} dK\, R_\epsilon(P,K,\lambda_1\mu^1,\ldots,\lambda_1\cdots\lambda_s\mu^s, \xi_1\cdots\xi_k\mu^{s+1},\ldots,\xi_k\mu^{s+k},\mu^{s+k+1},\ldots,\mu^\rho), \tag{2}$$

where R_ϵ is the renormalized (subtracted out) Feynman integrand associated with a *proper and connected graph G*; $\mu=\{\mu^1,\ldots,\mu^s,\mu^{s+1},\ldots,\mu^{s+k},\mu^{s+k+1},\ldots,\mu^\rho\}$ denotes the set of masses, $\mu^i>0$, and $P=\{p_1^0,\ldots,p_m^3\}$, $K=\{k_1^0,\ldots,k_n^3\}$ denote, respectively, the set of external and internal momenta. Here $F(P)\in\mathscr{S}(\mathbf{R}^{4m})$ is a Schwartz function. For $\epsilon>0$, $\lambda_1\neq0,\ldots,\lambda_s\neq0$, $1<\xi_1<\infty,\ldots,1<\xi_k<\infty$ the expression in (2) in absolutely convergent. We have R_ϵ in the familiar form

$$R_\epsilon(P,K,\mu)=A(P,K,\mu,\epsilon)\prod_{l=1}^{L} D_l^{-1},\quad \epsilon>0, \tag{3}$$

$$D_l=Q_l^2+\mu_l^2-i\epsilon[\mathbf{Q}_l^2+\mu_l^2],\quad \mu_l>0, \tag{4}$$

$$Q_l=\sum_{i=1}^{4m} a_{li}\,p_i+\sum_{i=1}^{4n} b_{li}\,k_i\equiv p^l+k^l, \tag{5}$$

where $\mu_l\in\mu$, and A is a polynomial in its argument and may, in general, be a polynomial in $(\mu^{s+1})^{-1},\ldots,(\mu^\rho)^{-1}$ as well. A propagator carrying a momentum Q_l will be written in the form

$$D^+(Q_l,\mu_l)=(\mu_l)^{-\delta_l}\frac{\widetilde{P}(Q_l,\mu_l)}{[Q_l^2+\mu_l^2-i\epsilon(\mathbf{Q}_l^2+\mu_l^2)]} \equiv(\mu_l)^{-\delta_l}\widetilde{D}^+(Q_l,\mu_l), \tag{6}$$

where for $\mu_l\in[\mu^1,\ldots,\mu^s]$, $\delta_l\equiv0$, and for $\mu_l\in[\mu^{s+1},\ldots,\mu^\rho]$, δ_l is some non-negative integer. The latter is well known for massive higher spin fields, and where $\widetilde{P}(Q_l,\mu_l)$ is some polynomial in μ_l. For $\mu_l\in[\mu^1,\ldots,\mu^s]$, $\widetilde{D}(Q_l,0)$ denotes the zero-mass propagator. Also we require that

$$\operatorname*{degr}_{\mu_l} D^+(Q_l,\mu_l)\leqslant-1, \tag{7}$$

$$\operatorname*{degr}_{Q_l,\mu_l} D^+(Q_l,\mu_l)<\operatorname*{degr}_{Q_l} D^+(Q_l,\mu_l). \tag{8}$$

We prove the following theorem.

Theorem: Suppose that there are no proper, connected, and divergent $[d(g)>0]$ subdiagrams $g\subset G$ such that all masses in g are from the set $\{\mu^1,\mu^2,\ldots,\mu^s\}$. Let T_i be the set of *all* subdiagrams $G'\nsubseteq G$ such that all the masses in G/G' are from the set $\{\mu^i,\mu^{i+1},\ldots,\mu^s\}$. If $\mathrm{Max}_{G'\in T_i}(d(G'))<d[G)$ for *all* $i=1,\ldots,s$, then

 0022-2488/85/051065-05$02.50

$$\lim_{\substack{\xi_1,\ldots,\xi_k\to\infty\\ \lambda_1,\ldots,\lambda_s\to 0}} \left(\lim_{\epsilon\to+0} T^{(\epsilon)}_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}(f)\right) = 0\,, \tag{9}$$

for all $f(P)\in \mathscr{S}(\mathbf{R}^{4m})$, and where the limits $\xi_1,\ldots,\xi_k\to\infty$, $\lambda_1,\ldots,\lambda_s\to 0$ are taken independently. Here we recall that G denotes a proper and connected graph. The symbol $\subset$ may include equality, and the symbol $\subsetneq$ excludes equality.

In the proof we will make use of the following two lemmas: Lemma 1, Lemma 2, proved, respectively, in Refs. 2 and 3. The proof of Lemma 1 is also discussed in the Appendix.

Lemma 1:

$$\lim_{\epsilon\to+0} T^{(\epsilon)}_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\xi_s}(f) \equiv \widehat{T}_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}(f) \tag{10}$$

exists, and

$$|\widehat{T}_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}(f)| \leqslant (\lambda_1\cdots\lambda_s)^{d(G)+4m-2N}\int_{\mathbf{R}^{4m}} dP\left[\frac{1}{\lambda_1^2\cdots\lambda_s^2}+\sum_{i=1}^{m} p_{Ei}^2\right]^{-N}$$
$$\times\int_{\mathbf{R}^{4n}} dK\left|R_E\left(P,K,\frac{\mu^1}{\lambda_2\cdots\lambda_s},\ldots,\frac{\mu^{s-1}}{\lambda_s},\mu^s,\xi_1\cdots\xi_k\,\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{s+1},\ldots,\right.\right.$$
$$\left.\left.\times\xi_k\,\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{s+k},\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{s+k+1},\ldots,\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{\rho}\right)\right|\,, \tag{11}$$

where

$$R_E(P,K,\mu) = A(P,K,\mu,0)\prod_{l=1}^{L}\left[\mathbf{Q}_l^2+Q_l^{02}+\mu_l^2\right]^{-1}$$
$$(p_E^2\equiv \mathbf{p}^2+p^{02}) \tag{12}$$

is a Euclidean version of R_ϵ in (3), and N is a positive integer that may be chosen arbitrarily *large*.

To state Lemma 2, we need the following. Consider the elements in P, K, μ as the components[3,4] of a vector $\mathbf{P}'$ in $\mathbf{R}^{4m+4n+\rho}$. Suppose $\mathbf{P}'$ is of the form

$$\mathbf{P}' = \sum_{i=1}^{t}\mathbf{L}_i'\eta_i\eta_{i+1}\cdots\eta_t+\mathbf{C}',\quad t\in[1,\ldots,4m+4n+\rho]\,, \tag{13}$$

where $\mathbf{L}_i',\ldots,\mathbf{L}_t'$ are t independent vectors in $\mathbf{R}^{4m+4n+\rho}$; C' is a vector confined to a finite region in $\mathbf{R}^{4m+4n+\rho}$, and $\eta_1>1,\ldots,\eta_t>1$ are some parameters. In reference to the graph G, we write $Q_l\equiv p^l+k^l$ [see Eq. (5)], for the momentum Q_l carried by a line l in G. Suppose that for some l, k^l depends on the parameter η_r, for r fixed in $[1,\ldots,t]$. Then we may write R_ϵ in (3) in a familiar form[3]:

$$R_\epsilon = \sum_{\mathscr{N}}\prod_{g\in\mathscr{N}}(\delta_g^{\mathscr{N}}-T_g)I_G\,, \tag{14}$$

where the sum is over all $\mathscr{N}$ sets of proper, connected subdiagrams such that (i) $G\in\mathscr{N}$. (ii) If $g_1, g_2\in\mathscr{N}$, then either $g_1\cap g_2=\phi$, or $g_1\subsetneq g_2$, or $g_2\subsetneq g_1$. (iii) Let $g\in\mathscr{N}$. If $g_1,\ldots,g_m$ denote the maximal elements in $\mathscr{N}$ contained in g: $g_i\subsetneq g$, $i=1,\ldots,m$, then set $\bar{g}=g/(g_1\cup g_2\cup\cdots\cup g_m)$ by shrinking $g_1,\ldots,g_m$ in g points. Then all the $(k^l)^g$ (which are linear combinations of the integration variables in K) of $\bar{g}$ are either all dependent on η_r or are all independent of η_r. Also (a) $\delta_G^{\mathscr{N}}=1$. (b) If $g\subsetneq G$ in $\mathscr{N}$, and all the $(k^l)^g$ of $\bar{g}$ are independent of η_r, then $\delta_g^{\mathscr{N}}=0$. (c) If g is one of the maximal elements in $\mathscr{N}$ contained in a subdiagram $g'\in\mathscr{N}$, such that all $(k^l)^g$ of $\bar{g}'$ are independent of η_r, and all the $(k^l)^g$ of $\bar{g}$ are dependent on η_r then $\delta_g^{\mathscr{N}}=1$. (d) If $g\in\mathscr{N}$ ($g\subsetneq G$) is one of the maximal elements contained in $g'\in\mathscr{N}$ such that all the $(k^l)^{g'}$ of $\bar{g}'$ are dependent on η, then $\delta_G^{\mathscr{N}}=0$. If $d(g)<0$, then $T_g\equiv 0$. We write

$$\mathscr{N}=\mathscr{H}_1\cup\mathscr{H}_2\,, \tag{15}$$

where $g\in\mathscr{H}_1$ if all $(k^l)^g$ of $\bar{g}$ are dependent on η_r, and $g\in\mathscr{H}_2$ if all the $(k^l)^g$ of $\bar{g}$ are independent of η_r. We also write

$$\mathscr{H}_1=\mathscr{F}_1\cup\mathscr{F}_2\,, \tag{16}$$

where for $g\in\mathscr{H}_1$, with $g\subsetneq G$, $g\in\mathscr{F}_1$ if $\delta_g^{\mathscr{N}}=0$, $g\in\mathscr{F}_2$ if $\delta_g^{\mathscr{N}}=1$. Equation (14) may be then also rewritten as[3]

$$R_\epsilon = \sum_{\mathscr{N}} F_G(\mathscr{N})\,, \tag{17}$$

where we have recursively

$$F_g(\mathscr{N}) = (\delta_g^{\mathscr{N}}-T_g)I_{\bar{g}}\prod_i F_{g_i}(\mathscr{N})\,, \tag{18}$$

and $\{g_i\}_i$ denotes the set of maximal elements in $\mathscr{N}$ contained in g: $g_i\subsetneq g$. The maximal elements in $\mathscr{N}$ contained in G will be denoted by $G_1,\ldots,G_m$: $G_i\subsetneq G$.

Lemma 2: Suppose $g\in \mathrm{N}-\{G\}$. For $g\in\mathscr{F}_1\cup\mathscr{H}_2$, if (i) there is a subdiagram $g'\subsetneq g$ with $g'\in\mathscr{F}_2$, and/or (ii) there is a (divergent) subdiagram $g''\subset g$, with $g''\in\mathscr{H}_2$, such that at least one of the masses in g'' depends on η_r, then

$$\operatorname*{degr}_{\eta_r} F_g(\mathscr{N}) < d(g)-\sigma(g)\,, \tag{19}$$

where

$$\sigma(g) = 4\sum_{\substack{g'\in\mathscr{H}_1\\ g'\subset g}} L(\bar{g}')\,, \tag{20}$$

and $L(g)$ denotes the number of independent loops in g. Also if $G\in\mathscr{F}_2$ and one or both of the conditions (i), (ii) above are true with g in them formally replaced by G, then [for $d(G)\geqslant 0$]

$$\operatorname*{degr}_{\eta_r}(-T_G)I_{\bar{G}}\prod_{i=1}^{m}F_{G_i}(\mathscr{N}) < d(G)-\sigma(G)\,. \tag{21}$$

Also if we scale all the $(k^l)^{g'}$ of all $\bar{g}'$, $g'\subset g$, $g'\in\mathscr{H}_1$, and the masses in g, that are dependent on the parameter η_r, by a parameter λ, then

$$\operatorname*{degr}_{\lambda}\mathrm{F}_g(\mathscr{N}) \leqslant -1-\sigma(g),\quad g\in\mathscr{H}_2 \tag{22a}$$

[if no λ dependence occurs in (22a) replace $\operatorname{degr} F_g(\mathcal{N})$ simply by zero],

$$\operatorname*{degr}_{\lambda} F_g(\mathcal{N}) \leqslant -1-\sigma(g), \quad g\in\mathcal{F}_2 . \tag{22b}$$

We now apply the above lemma to the graph G itself.

A. Dimensional analysis in reference to a parameter $1/\lambda_i$, $i\in[1,\ldots,s]$

In the light of the conditions stated in the theorem, any proper and connected and divergent subdiagram $g\subset G$ must necessarily have masses from the set $\{\mu^1,\ldots,\mu^{i-1},\mu^{s+1},\ldots,\mu^{\rho}\}$. The latter masses will necessarily depend on the parameter $1/\lambda_i$ in

$$R(P,K,\mu^1/\lambda_2\cdots\lambda_s,\ldots,\mu^{s-1}/\lambda_s,\mu^s,\\ \xi_1\cdots\xi_k\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{s+1},\ldots,\xi_k\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{s+k},\\ \lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{s+k+1},\ldots,\lambda_1^{-1}\cdots\lambda_s^{-1}\mu^{\rho}), \tag{23}$$

on the right-hand side of (11). Suppose first that $G\in\mathcal{F}_2$. Accordingly, to have an equality in (21) it is necessary, from the lemma, that $G_1,\ldots,G_m\in\mathcal{F}_2$, and that there are no subdiagrams $g\in\mathcal{N}$, such that $g\in\mathcal{H}_2\cup\mathcal{F}_2$, and we have [for $d(G)\geqslant 0$]

$$\operatorname*{degr}_{\eta_r}(-T_G)I_{\overline{G}}\prod_{i=1}^{m}F_{G_i}(\mathcal{N})\leqslant d(G)-\sigma(G) \tag{24}$$
$$= d(G)-4L(G).$$

A similar analysis applied to each G_i also yields

$$\operatorname*{degr}_{\eta_r} I_{\overline{G}}\prod_{i=1}^{m}F_{G_i}(\mathcal{N})\leqslant d(G')-4L(G'), \tag{25}$$

where G/G' is independent of η_r and $G'\supset(G_1\cup\cdots\cup G_m)$. Hence from (24) and (25) we have

$$\operatorname*{degr}_{\eta_r} F_G(\mathcal{N})\leqslant d(G)-4L(G). \tag{26}$$

Now suppose that $G\in\mathcal{H}_2$. To this end, we note that in order to have an equality in (19) with g in it replaced by G, and the g_i replaced by the G_i, it is necessary that the $G_i\in\mathcal{F}_2$, and that there are no g in $\mathcal{N}$ with $g'\nsubseteq G_i$, such that $g'\in\mathcal{H}_2$. And if so, we have [for $d(G)\geqslant 0$], with $G_i\in\mathcal{F}_2$, $G\in\mathcal{H}_2$

$$\operatorname*{degr}_{\eta_r}(-T_G)I_{\overline{G}}\prod_{i=1}^{m}F_{G_i}(\mathcal{N})<d(G)-4\sum_{i=1}^{m}L(G_i). \tag{27}$$

Let $G'/(G_1\cup\cdots\cup G_n)$ be the subdiagram of G such that all the lines in G/G' (if not empty) depend on η_r. Then

$$\operatorname*{degr}_{\eta_r} I_{\overline{G}}\prod_{i=1}^{m}F_{G_i}(\mathcal{N})\leqslant d(G')-4L(\overline{G}')-4\sum_{i=1}^{m}L(G_i). \tag{28}$$

where $\overline{G}'=G'/(G_1\cup\cdots\cup G_n)$. With G/G' independent of $1/\lambda_i$, we use the condition $d(G')<d(G)$ stated in the theorem to conclude that

$$\operatorname*{degr}_{\eta_r} F_G(\mathcal{N})<d(G)-4\sum_{i=1}^{m}L(G_i), \tag{29}$$

for $d(G)\geqslant 0$ or $d(G)<0$. If none of the k^l in G depend on η_r then we may use the bound

$$\operatorname*{degr}_{\eta_r} R_\epsilon<d(G')-4L(G')<d(G), \tag{30}$$

where G' is a subdiagram of G such that all the lines in G/G' (if not empty) are independent of η_r, and according to one of the statements of the theorem $d(G')<d(G)$ for $G'\nsubseteq G$.

B. Dimensional analysis in reference to a parameter ξ_i, $i\in[1,\ldots,k]$

If all the external momenta of G are independent of a parameter η_r, then we conclude directly from (22a) and (22b),

$$\operatorname*{degr}_{\eta_r} F_G(\mathcal{N})\leqslant -1-\sigma(G), \tag{31}$$

if some of the k^l of G depend on η_r, and

$$\operatorname*{degr}_{\eta_r} R_\epsilon\leqslant -1, \tag{32}$$

directly from (7), if all the k^l of G are independent of the parameter η_r. If some of the external momenta of G depend on η_r, then quite generally we may find a subdiagram G' such that

$$\operatorname*{degr}_{\eta_r} F_G(\mathcal{N})\leqslant d(G')-\sigma(G). \tag{33}$$

We denote the integral multiplying the factor $(\lambda_1\cdots\lambda_s)^{d(G)+4m-2N}$ on the right-hand side of (11), by $I_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}$. Due to the presence of the term $1/\lambda_1^2\cdots\lambda_s^2$ in $[(1/\lambda_1^2\cdots\lambda_s^2)+\Sigma_{i=1}^m p_{Ei}^2]^{-N}$ on the right-hand side of (11), we note that the maximum dimensionality of $I_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}$ will be attained if all the components of P depend on η_r, for N arbitrarily large. And from (26), (29) we then conclude[5,4]

$$\operatorname*{degr}_{(1/\lambda_i)} I_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}\leqslant d(G)-2N+4m. \tag{34}$$

On the other hand, from (31), (32), we conclude that if none of the components in P depend on the parameter η_r, then

$$\operatorname*{degr}_{\xi_i} I_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}\leqslant -1. \tag{35}$$

If some of the external components in P depend on η_r, then from (33) (see Ref. 5)

$$\operatorname*{degr}_{\xi_i} I_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}\leqslant d+u-2N, \tag{36}$$

where d is some finite number that may determined from the structure of the graph G and is of no importance here, and u denotes the number of components in P depending on η_r. The interesting thing to note in (36) is that, the positive integer N may be chosen arbitrarily *large*, and in particular such that $d+u+1<2N$. Accordingly, from (35) and (36), we may conclude that

$$\operatorname*{degr}_{\xi_i} I_{\xi_1,\ldots,\xi_k;\lambda_1,\ldots,\lambda_s}\leqslant -1. \tag{37}$$

for N arbitrarily large. All told, we then have from (34), (37), and the estimate (11) [see also (10)],

$$|\hat{T}_{\xi_1,\dots,\xi_k;\lambda_1,\dots,\lambda_k}(f)| < C\frac{1}{\xi_1\cdots\xi_k}\mathscr{P}(\ln\xi_1,\dots,\ln\xi_k) \times(\lambda_1\cdots\lambda_s)^{d(G)+4m-2N}\prod_{i=1}^{s}\left(\frac{1}{\lambda_i}\right)^{d(G)-2N+4m}, \quad (38)$$

or

$$|\hat{T}_{\xi_1,\dots,\xi_k;\lambda_1,\dots,\lambda_k}(f)| \leqslant C\frac{1}{\xi_1\cdots\xi_k}\mathscr{P}(\ln\xi_1,\dots,\ln\xi_k)\to 0, \quad (39)$$

for $\xi_1,\dots,\xi_k\to\infty$, $\lambda_1,\dots,\lambda_s\to 0$, where $\mathscr{P}(\ln\xi_1,\dots,\ln\xi_k)$ is some polynomial in $\ln\xi_1,\dots,\ln\xi_k$, and no logarithmic growth occurs in (38), (39) in the parameters $1/\lambda_i$, $i=1,\dots,s$. This completes the proof of the theorem.

As an example, consider the electron self-energy graph in any order without photon self-energy insertions. Let G denote the graph in question, with m denoting the electron mass, and μ denoting a photon mass. We note that the following conditions in the theorem are satisfied: (i) there are no proper, connected, and divergent subdiagrams $g\subset G$ (including G itself) consisting solely of photon lines; and (ii) for any subdiagram $G'\nsubseteq G$ such that G/G' consists solely of photon lines we necessarily have $d(G')<d(G)\,(=1)$. Hence the statement in (39) follows if we scale m by ξ, and μ by λ, and take the limits $\xi\to\infty$, $\lambda\to 0$. A slightly more involved analysis establishes the existence of the above limits even in the presence of photon self-energy insertions.

ACKNOWLEDGMENT

This work was supported by the Department of National Defence Award under CRAD No. 3610-637:F4122.

APPENDIX

We discuss the proof[2] (see also references therein) of Lemma 1. We write $A(P,K,\mu,\epsilon)=\Sigma_a\,\epsilon^a A_a(P,K,\mu)$, and hence in an obvious notation $T^{(\epsilon)}(f)=\Sigma_a\,\epsilon^a T^{(\epsilon)}_a(f)$, where to simplify the notation we have suppressed the dependence on the parameters $\xi_1,\dots,\xi_k,\lambda_1,\dots,\lambda_s$ in the latter. We give only the essential details here. By introducing Feynman parameters we may write

$$T^{(\epsilon)}_a(f)=\int_{\mathbf{R}^{4m}}dP\,f(P)\int_D d\alpha\,N^a(\alpha,P,\mu,\epsilon)[G_\epsilon(\alpha,P,\mu)]^{-t}, \quad (A1)$$

where

$$G_\epsilon(\alpha,P,\mu)=pUp+M^2-i\epsilon(\mathbf{p}\cdot U\mathbf{p}+M^2), \quad (A2)$$

$$\mathrm{M}^2=\sum_{l=1}^{L}\alpha_l\,\mu_l, \quad (A3)$$

and U is rational in α and may be extended to a continuous function[6] everywhere in $D=\{\alpha=(\alpha_1,\dots,\alpha_L),\ \alpha_i\geqslant 0,\ \Sigma_{i=1}^L\alpha_i=1\}$. Here, $N^a(\alpha,P,\mu,\epsilon)$ is rational in α, and is a polynomial in its remaining arguments, and for those μ^j, which we do not scale to zero, it may be also a polynomial in these $(\mu^j)^{-1}$. t is some positive integer. We use the following identity[2]:

$$[G_\epsilon(\alpha,P,\mu)]^{-t}=-[\mathbf{p}\cdot U\mathbf{p}+M^2]^2 t(t+1)\times\int_\epsilon^1 d\gamma_1\int_{\gamma_1}^1 d\gamma[G_\gamma(\alpha,P,\mu)]^{-t-2} + [\mathbf{p}\cdot U\mathbf{p}+M^2]i\epsilon t\,[G_1(\alpha,P,\mu)]^{-t-1} + [G_1(\alpha,P,\mu)]^{-t}. \quad (A4)$$

By substituting each term on the right-hand side of (A4) in turn for $[G_\epsilon]^{-t}$ in (A1), we generate three terms $T^{(\epsilon)}_a(f)_1$, $T^{(\epsilon)}_a(f)_2, T^{(\epsilon)}_a(f)_3$, respectively, and $T^{(\epsilon)}_a(f)=T^{(\epsilon)}_a(f)_1+T^{(\epsilon)}_a(f)_2+T^{(\epsilon)}_a(f)_3$. Because of the ϵ factor in the denominators in $T^{(\epsilon)}_a(f)_2$ and $T^{(\epsilon)}_a(f)_3$ are set equal to 1, no $\epsilon\to+0$ limit problems arise in $T^{(\epsilon)}_a(f)_2$, $T^{(\epsilon)}_a(f)_3$ and one readily bounds

$$|T^{(\epsilon)}_a(f)_i|\leqslant C_i\int_{\mathbf{R}^{4m}}dP\,|f(P)|\int_{\mathbf{R}^{4n}}dK\,|A_a(P,K,\mu)| \times\prod_{l=1}^{L}[Q^2_{l E}+\mu_l^2]^{-1},\quad 0\leqslant\epsilon<1,\quad i=2,3. \quad (A5)$$

To handle $T^{(\epsilon)}_a(f)_1$, we introduce (Ref. 2, and references therein) a $\mathscr{C}^\infty$ function $\chi(x):0\leqslant\chi(x)\leqslant 1$, such that $\chi(x)=0$ for $x<\frac{1}{3}$, and $\chi(x)=1$ for $x>\frac{2}{3}$. We set $x=p^0\,Up^0/[\mathbf{p}\cdot U\mathbf{p}+M^2]$. We then write $T^{(\epsilon)}_a(f)_1=T^{(\epsilon)}_a(f)_{11}+T^{(\epsilon)}_a(f)_{12}$, where $T^{(\epsilon)}_a(f)_{11}$ is nothing but $T^{(\epsilon)}_a(f)_1$ with the integrand in the latter simply multiplied by $[1-\chi(x)]$, and $T^{(\epsilon)}_a(f)_{12}$ is nothing but $T^{(\epsilon)}_a(f)_1$ with the integrand in the latter simply multiplied by $\chi(x)$. Again due to the presence of the multiplicative factor $[1-\chi(x)]$ in the integrand in $T^{(\epsilon)}_a(f)_{11}$ no $\epsilon\to+0$ limit problems arise in the latter and we have a bound for $|T^{(\epsilon)}_a(f)_{11}|$ similar to the one in (A5). To study the integral $T^{(\epsilon)}_a(f)_{12}$, which includes the multiplicative factor $\chi(x)$ in its integrand we use the identity

$$[G_\gamma(\alpha,P,\mu)]^{-t-2} =\frac{(-\frac{1}{2})^{t+1}}{(t+1)!}\left[(p^0\,Up^0)^{-1}\sum_{i=1}^{m}p_i^0\frac{\partial}{\partial p_i^0}\right]^{t+1} \times[G_\gamma(\alpha,P,\mu)]^{-1}, \quad (A6)$$

for $[G_\gamma(\alpha,P,\mu)]^{-t-2}$ in $T^{(\epsilon)}_a(f)_{12}$. Upon integration by parts in P, and using the vanishing property of $f(P)$, together with all of its derivatives, at infinity, we obtain

$$T^{(\epsilon)}_a(f)_{12}=C''\sum_c\int_{\mathbf{R}^{4m}}dP\int_\epsilon^1 d\gamma_1\int_{\gamma_1}^1 d\gamma(p^0)^c N^a_c(\alpha,\mathbf{P},\mu,\epsilon) \times[G_\gamma(\alpha,P,\mu)]^{-1}[\mathbf{p}\cdot U\mathbf{p}+M^2]^2(p^0\,Up^0)^{-t-1} \times\sum_i\chi^i_c(p^0,x,\alpha)L^i_c(P), \quad (A7)$$

where the $\chi^i_c(p^0,x,\alpha)$ may be bounded, in absolute value, by a polynomial in p^0 independent of α, and due to the factor $\chi(x)$, they vanish for $x<\frac{1}{3}$. We have $L^i_c\in\mathscr{S}(\mathbf{R}^{4m})$, and $N^a(\alpha,P,\mu,\epsilon)=\Sigma_c(p^0)^c N^a_c(\alpha,\mathbf{P},\mu,\epsilon)$. The factor $(p^0)^c$ multiplying $N^a_c(\alpha,\mathbf{P},\mu,\epsilon)$ in (A7) survives, after the partial integrations, due to the presence of the dimensionless operator $(p^0_i\partial/\partial p^0_i)$ in (A6). From the Lagrange interpolating formula,[6,7] we also have

$$|(p^0)^c N_c^a(\alpha,\mathbf{P},\mu,\epsilon)| \leqslant \sum_{j=1}^{|c|} |F_c(\rho_j)||N^a(\alpha,\mathbf{P},\rho_j P^0,\mu,\alpha)|, \tag{A8}$$

where $|c|$ is the degree of the polynomial $N(\alpha,P,\mu,\alpha)$ in p^0, and $\rho_1,...,\rho_{|c|}$ are distinct nonvanishing numbers. The $F_c(\rho_j)$ are some finite functions of ρ_j, respectively. From (A8) and the estimate

$$\frac{\mathbf{p}\cdot U\mathbf{p} + \rho_j^2 p^0 U p^0 + M^2}{\mathbf{p}\cdot U\mathbf{p} + p^0 U p^0 + M^2} \leqslant 1 + |\rho_j^2 - 1|, \tag{A9}$$

we obtain a bound for $T_a^{(\epsilon)}(f)_{12}$ similar to the one in (A5) with $f(P)$ replaced by $[1 + \Sigma_{i=1}^m p_{Ei}^2]^{-N}$, with N an arbitrary large positive integer, by finally making the change of variables $\rho_j p^0 \to p^0$. The result in (10) follows from the Lebesgue dominated convergence theorem, and the upper bound in (11) occurs from finally making the transformation of variables $K \to \lambda_1 \cdots \lambda_s K$, $P \to \lambda_1 \cdots \lambda_s P$.

[1]E. B. Manoukian, J. Math. Phys. 25, 1519 (1984).
[2]E. B. Manoukian, "Distributional zero-mass limit of renormalized Feynman amplitudes in Minkowski space," to appear in Nuovo Cimento.
[3]E. B. Manoukian, J. Phys. G 7, 1149 (1981).
[4]S. Weinberg, Phys. Rev. 118, 838 (1960).
[5]E. B. Manoukian, J. Phys. G 7, 1159 (1981).
[6]Y. Hahn and W. Zimmermann, Commun. Math. Phys. 10, 330 (1968); W. Zimmermann, Commun. Math. Phys. 11, 1 (1968).
[7]E. Isaacson and H. B. Keller, *Analysis of Numerical Methods* (Wiley, New York, 1966).

Fierz identities for real Clifford algebras and the number of supercharges

M. Baake and P. Reinicke
Physikalisches Institut, Bonn University, D-5300 Bonn 1, Bonn, Federal Republic of Germany
V. Rittenberg
Département de Physique Théorique, Université de Genève, 1211 Genève 4, Switzerland

(Received 18 October 1984; accepted for publication 7 December 1984)

One considers supersymmetric gauge theories in quantum mechanics with the bosons and fermions belonging to the adjoint representation of the gauge group. One shows that the supersymmetry constraints are related to the existence of certain Fierz identities for real Clifford algebras. These identities are valid when one has 2, 4, 8, and 16 supercharges.

I. INTRODUCTION

Supersymmetric quantum mechanics[1] was extended recently[2] to gauge theories. One considers a system containing bosonic degrees of freedom (coordinates ϕ_a and momenta π_a) and fermionic degrees of freedom Λ_μ

$$[\phi_a,\pi_b] = i\delta_{ab}, \quad \{\Lambda_\mu,\Lambda_\nu\} = 2\delta_{\mu\nu}$$
$$(\phi_a^+ = \phi_a, \quad \pi_b^+ = \pi_b, \quad \Lambda_\mu^+ = \Lambda_\mu). \tag{1}$$

The Lie algebra corresponding to the gauge group G is

$$[L_A,L_B] = if_{ABC}L_C \quad (L_A^+ = L_A), \tag{2}$$

with f_{ABC} totally antisymmetric. The system is described by a Hamiltonian H,

$$H = H(\phi_a,\pi_b,\Lambda_\mu), \tag{3}$$

invariant under the group G,

$$[H,L_A] = 0, \tag{4}$$

and one imposes the Gauss law considering only states $|\psi\rangle$ satisfying the condition

$$L_A|\psi\rangle = 0. \tag{5}$$

The system has M-extended supersymmetry[2] if one has M Hermitian supercharges Q_α ($\alpha = 1,\dots,M$) satisfying the algebra

$$\{Q_\alpha,Q_\beta\} = 2(\delta_{\alpha\beta}H + U^A_{\alpha\beta}L_A),$$
$$[H,Q_\alpha] = iD^A_\alpha L_A, \quad [L_A,Q_\alpha] = 0. \tag{6}$$

In Eqs. (6), $U^A_{\alpha\beta}$ and D^A_α are operators (they depend on the operators ϕ_a, π_a, and Λ_μ). Thus Eqs. (6) do not define a superalgebra. If, however, one restricts oneself to the states $|\psi\rangle$ satisfying Eq. (5) one notices that one is back to supersymmetric quantum mechanics[1] with M supercharges.

II. SPECIAL SOLUTIONS AND REAL CLIFFORD ALGEBRAS

Various examples of Hamiltonians satisfying the algebra (6) can be found in Ref. 2. Here we would like to discuss some properties of a special class of solutions. We assume that we have p multiplets of bosons belonging to the adjoint representation of G,

$$\phi_{Am},\pi_{Am} \quad (m = 1,2,\dots,p), \tag{7}$$

and M multiplets of fermions belonging also to the adjoint representation of G,

$$\Lambda_{A\alpha} \quad (\alpha = 1,2,\dots,M). \tag{8}$$

Notice that the number of fermionic multiplets is taken equal to the number of supercharges but the number of bosonic multiplets is left free.

The generators L_A of the Lie algebra (2) are

$$L_A = f_{ABC}(\phi_{Bm}\pi_{Cm} - (i/4)\Lambda_{B\alpha}\Lambda_{C\alpha}). \tag{9}$$

We consider the following ansatz for the supercharges:

$$Q_\alpha = (1/\sqrt{2})(\Gamma^m_{\alpha\beta}\Lambda_{A\beta}\pi_{Am} + gf_{ABC}\Sigma^{mn}_{\alpha\beta}\Lambda_{A\beta}\phi_{Bm}\phi_{Cn}). \tag{10}$$

Here G is the coupling constant, and the Γ^m are real (the matrix elements are real numbers), symmetric, $M\times M$ matrices satisfying the *real* Clifford algebra

$$\{\Gamma^m,\Gamma^n\} = 2\delta^{mn} \quad (m,n = 1,2,\dots,p) \tag{11}$$

and

$$\Sigma^{mn} = \tfrac{1}{4}[\Gamma^m,\Gamma^n]. \tag{12}$$

Notice the relation

$$[\Sigma^{mn},\Gamma^k] = \delta^{nk}\Gamma^m - \delta^{mk}\Gamma^n. \tag{13}$$

We now compute the anticommutators of the supercharges:

$$\begin{aligned}\{Q_\alpha,Q_\beta\} = \delta_{\alpha\beta}(&\pi_{Am}\pi_{Am}\\ &+ (g^2/2)f_{ABC}f_{AB'C'}\phi_{Bm}\phi_{Cn}\phi_{B'm}\phi_{C'n})\\ &+ 2g\Gamma^m_{\alpha\beta}\phi_{Am}f_{ABC}\phi_{Bn}\pi_{Cn}\\ &+ i(g/2)F^m_{\alpha\beta;\alpha'\beta'}f_{ABC}\phi_{Am}\Lambda_{B\alpha'}\Lambda_{C\beta'},\end{aligned} \tag{14}$$

where

$$\begin{aligned}F^m_{\alpha\beta;\alpha'\beta'} = \Gamma^n_{\alpha\beta'}\Sigma^{nm}_{\beta\alpha'} + \Gamma^n_{\beta\beta'}\Sigma^{nm}_{\alpha\alpha'}\\ + \Gamma^n_{\alpha\alpha'}\Sigma^{nm}_{\beta\beta'} + \Gamma^n_{\beta\alpha'}\Sigma^{nm}_{\alpha\beta'}.\end{aligned} \tag{15}$$

Equation (14) is of the required form (6) if the following Fierz identity is valid:

$$\begin{aligned}\Gamma^n_{\alpha\beta'}\Sigma^{nm}_{\beta\alpha'} + \Gamma^n_{\beta\beta'}\Sigma^{nm}_{\alpha\alpha'} + \Gamma^n_{\alpha\alpha'}\Sigma^{nm}_{\beta\beta'} + \Gamma^n_{\beta\alpha'}\Sigma^{nm}_{\alpha\beta'}\\ = \delta_{\alpha\beta}\Gamma^m_{\alpha'\beta'} - \delta_{\alpha'\beta'}\Gamma^m_{\alpha\beta}\\ (\alpha,\beta = 1,2,\dots,M;\ m,n = 1,2,\dots,p).\end{aligned} \tag{16}$$

Let us pause to discuss some properties of the real Clifford algebra (11). If the algebra has p generators its *irreducible* $r\times r$ representations have a mod 8 periodicity[3] and the values of r are given in Table I. One can write these irreducible representations as tensor products of Pauli matrices as shown in the Appendix. We now turn to the Fierz identities given by Eq. (16). A necessary condition for their existence is obtained contracting the indices α and β in Eq. (16), using Eq. (13), and (excluding from now on the trivial case $p = 1$)

$$\operatorname{tr}\Gamma^m = 0, \tag{17}$$

one obtains

$$2(p-1) = M. \tag{18}$$

 0022-2488/85/051070-02$02.50

TABLE I. Irreducible representations ($r\times r$) of the real Clifford algebras with p generators.

p(mod 8)	r
1	$2^{(p-1)/2}$
2	$2^{p/2}$
3	$2^{(p+1)/2}$
4	$2^{(p+2)/2}$
5	$2^{(p+1)/2}$
6	$2^{(p+2)/2}$
7	$2^{(p+1)/2}$
8	$2^{p/2}$

Now inspecting Table I one finds that the necessary condition (18) is verified only if $p = 2$, 3, 5, and 9 and the representation is irreducible, namely $M = 2$, 4, 8, and 16, respectively. That the necessary condition is also sufficient was verified by brute force using the representations of the Γ^m matrices given in the Appendix. (Since the Clifford algebra and the Fierz identities are left invariant by orthogonal transformations we can choose our basis at will.) In this way for a given number of supercharges M the number of bosonic multiplets p is fixed.

We now use the Fierz identities (16) in Eq. (14) and obtain

$$\{Q_\alpha,Q_\beta\} = 2(H\delta_{\alpha\beta} + g\Gamma^m_{\alpha\beta}\phi_{Am}L_A), \tag{19}$$

where

$$\begin{aligned} H &= \tfrac{1}{2}\pi^2_{Am} + (g^2/4)(f_{ABC}\phi_{Bm}\phi_{Cn})^2 \\ &\quad + i(g/4)\Gamma^m_{\alpha\beta}f_{ABC}\phi_{Am}\Lambda_{B\alpha}\Lambda_{C\beta}, \end{aligned} \tag{20}$$

and one also obtains

$$\begin{aligned} [H,Q_\alpha] &= -\frac{ig}{\sqrt{2}}\sum_{\beta=1}^{M}\sum_{A}\sum_{m=1}^{p}\Gamma^m_{\alpha\alpha}\Gamma^m_{\alpha\beta}\Lambda_{A\beta}L_A \\ &= (-ig/\sqrt{2})\Lambda_{A\alpha}L_A. \end{aligned} \tag{21}$$

The last equality was obtained using the representations of the Γ^m matrices given in the Appendix.[4]

III. CONCLUSION

Let us sum up our results. We have found the new Fierz identities for real Clifford algebras given by Eq. (16). These identities are valid only for $M = 2$ ($p = 2$), $M = 4$ ($p = 3$), $M = 8$ ($p = 5$), and $M = 16$ ($p = 9$). Using these identities one obtains the gauge supersymmetric Hamiltonians (20). It is easy to check that for $M = 4$, 8, and 16 one obtains precisely the dimensionally reduced $N = 1$, 2, and 4 extended $d = 4$ supersymmetric gauge field theories.[2] Notice that one gets here at most 16 supercharges out of purely algebraic properties and not by any spin counting as one does in field theory.

A last observation is in order.[5] Let us consider the case $M = 16$ and Eq. (19). Looking at the expressions of the Γ^m matrices given in (A6) one sees that one can choose eight out of the 16 supercharges such that

$$\begin{aligned} \{Q_\mu,Q_\nu\} &= 2(H\delta_{\mu\nu} + g\Gamma^1_{\mu\nu}\phi_{A1}L_A) \\ &= 2H'\delta_{\mu\nu} \quad (\mu,\nu = 1,2,\ldots,8), \\ [H',Q_\mu] &= 0, \end{aligned} \tag{22}$$

where

$$H' = H + g\phi_{A1}L_A. \tag{23}$$

The Hamiltonian H' and the supercharges Q_μ are the generators of the superalgebra of quantum mechanics[1] and this is the first example we know of where one has a Hamiltonian at most bilinear in the fermionic operators commuting with eight supercharges.[6]

ACKNOWLEDGMENTS

The authors would like to thank R. Flume for many discussions.

One of us (V.R.) was supported by the Swiss National Foundation.

APPENDIX: REPRESENTATIONS OF THE REAL CLIFFORD ALGEBRAS AS TENSOR PRODUCTS OF PAULI MATRICES

The $r\times r$ irreducible representations of the real Clifford algebras

$$\{\Gamma^m,\Gamma^n\} = 2\delta^{mn} \quad (m,n = 1,2,\ldots,p) \tag{A1}$$

can be written in terms of Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A2}$$

We give the explicit construction up to $p = 9$.

(a) For $r = 2, p = 2$,

$$\Gamma^1 = \sigma_z, \quad \Gamma^2 = \sigma_x. \tag{A3}$$

(b) For $r = 4, p = 3$,

$$\Gamma^1 = \sigma_z \otimes 1, \quad \Gamma^2 = \sigma_y \otimes \sigma_y, \quad \Gamma^3 = \sigma_x \otimes 1. \tag{A4}$$

(c) For $r = 8, p = 4$ and 5,

$$\begin{aligned} &\Gamma^1 = \sigma_z \otimes 1 \otimes 1, \quad \Gamma^2 = \sigma_y \otimes 1 \otimes \sigma_y, \\ &\Gamma^3 = \sigma_x \otimes \sigma_z \otimes 1, \quad \Gamma^4 = \sigma_x \otimes \sigma_y \otimes \sigma_y, \\ &\Gamma^5 = \sigma_x \otimes \sigma_x \otimes 1. \end{aligned} \tag{A5}$$

(d) For $r = 16, p = 6$, 7, 8, and 9,

$$\begin{aligned} &\Gamma^1 = \sigma_z \otimes 1 \otimes 1 \otimes 1, \quad \Gamma^2 = \sigma_y \otimes 1 \otimes 1 \otimes \sigma_y, \\ &\Gamma^3 = \sigma_y \otimes \sigma_z \otimes \sigma_y \otimes \sigma_z, \quad \Gamma^4 = \sigma_y \otimes \sigma_z \otimes \sigma_y \otimes \sigma_x, \\ &\Gamma^5 = \sigma_x \otimes \sigma_z \otimes 1 \otimes 1, \quad \Gamma^6 = \sigma_x \otimes \sigma_y \otimes 1 \otimes \sigma_y, \\ &\Gamma^7 = \sigma_x \otimes \sigma_x \otimes \sigma_z \otimes 1, \quad \Gamma^8 = \sigma_x \otimes \sigma_x \otimes \sigma_y \otimes \sigma_y, \\ &\Gamma^9 = \sigma_x \otimes \sigma_x \otimes \sigma_x \otimes 1. \end{aligned} \tag{A6}$$

[1] H. Nicolai, Phys. A **9**, 1479 (1976); E. Witten, Nucl. Phys. B **188**, 513 (1981).

[2] R. Flume, Preprint Bonn-HE-84-15; V. Rittenberg and S. Yankielowicz, Preprint Bonn-HE-84-16; Mark Claudson and Martin B. Halpern, preprint Berkeley UCB-PTH-84/10.

[3] See, for example, R. Coquereaux, Phys. Lett. B **115**, 389 (1982).

[4] One can show that the Fierz identities, $\Sigma^p_{m=1}\Gamma^m_{\alpha\alpha}\Gamma^m_{\alpha\beta} = \delta_{\alpha\beta}$, are valid under the same conditions as the Fierz identities (16).

[5] This observation belongs to R. Flume.

[6] M. de Crombrugghe and V. Rittenberg, Ann. Phys. **151**, 99 (1983).

Conformal symmetry and constants of motion

V. Hussin and S. Sinzinkayo[a)]
Université de Liège, Physique Théorique et Mathématique, Institut de Physique au Sart Tilman, Bâtiment B.5, B-4000 Liège 1, Belgium

(Received 6 August 1984; accepted for publication 28 September 1984)

Conformal space-time symmetries of electromagnetic fields and potentials are studied in order to determine whole sets of constants of motion when spin-$\frac{1}{2}$ charged particles interact with such external electromagnetic fields. The relativistic and nonrelativistic contexts are discussed through the Hamiltonian formalism. In particular, the interesting example of the magnetic monopole field is recovered and related to the recent works of Jackiw and D'Hoker–Vinet.

I. INTRODUCTION

Dynamical symmetry[1] and supersymmetry[2] of the magnetic monopole have recently been studied, with the aim of understanding fundamental physical processes. In particular, the determination of the whole set of constants of motion characterizing the physical particle in interaction is an essential step in these approaches and, more generally, in every physical situation.

The problem[1,2] of a spin-$\frac{1}{2}$ particle with gyromagnetic ratio 2, in the presence of a Dirac monopole,[3] is nothing else than a particular case of a (nonrelativistic) particle interacting with a nonconstant external electromagnetic field. Thus it is a particular case of a recent study[4] leading to the determination of whole sets of constants of motion. In that contribution, Beckers and Hussin[4] have shown the interest of the symmetry properties of external (constant or not) electromagnetic fields and of their gauge fields—the electromagnetic potentials—in both relativistic and nonrelativistic contexts. These symmetry properties enter into Lagrangian as well as Hamiltonian formalisms. From a physical[4] or a geometrical[5] point of view, the importance of the symmetries of the associated gauge fields has been especially pointed out.

Let us just recall[4] here that constant electromagnetic fields $F \equiv (\mathbf{E},\mathbf{B})$ always admit the largest symmetry[6,7] inside the Poincaré[8] or Galilei[9] groups. Moreover, constant or nonconstant F's interacting with charged particles lead to specific sets of constants of motion, depending on compensating[10] gauge functions and on symmetry properties[11] of the associated gauge fields. Then the operators corresponding to these whole sets generate isomorphic algebras, which are central extensions[12] by $\mathbb{R}$ of the F-symmetry algebra. The exact symmetries[11] on the gauge fields yield informations about constants of motion, which are unchanged with respect to the free case.

Here we want to show that this method[4] is not limited to (relativistic) Poincaré or (nonrelativistic) Galilean invariances: it can be extended to conformal[13] or Schrödinger[14–16] (i.e., nonrelativistic and "conformal") symmetries so that the specific application handled by Jackiw[1] and D'Hoker and Vinet[2] enters in our developments when Schrödinger invariance is considered. This is one particular aim of the present paper.

Consequently, let us put the problem of determining the constants of motion of a physical system interacting with an arbitrary electromagnetic field at the level of (relativistic or nonrelativistic) conformal symmetries. Physically speaking, the particles admitting conformal symmetries have to be massless from a relativistic point of view, and we know the interest of such a limit in connection with particles studied in the domain of very high energies. Otherwise, from a nonrelativistic point of view, the mass parameter still is an invariant under "conformal" or Schrödinger transformations.

In Sec. II, we will recall some elements on conformal space-time symmetries in the relativistic context and the well-known invariance conditions[17] on electromagnetic fields and potentials. Parallel considerations in the nonrelativistic context will also be given, and the corresponding invariance conditions will be established under Schrödinger transformations. Section III will be devoted to constant electromagnetic fields admitting extended Poincaré symmetries (for example, Weyl symmetries[18]) and we will give a new set of constants of motion. In Sec. IV, the physical cases of nonconstant electromagnetic fields will be studied with respect to conformal symmetries in both relativistic and nonrelativistic contexts. The specific example of the magnetic monopole field will be recovered and discussed in connection with the Jackiw[1] and D'Hoker–Vinet[2] results, leading to a $SO(3) \otimes SO(2,1)$ symmetry. In particular, we will find a special interest in the study of the famous Wu–Yang[19] potentials in order to get the constants of motion. A parallel and new discussion of $SO(2,1) \otimes SO(2,1)$ symmetry will also be enhanced. The whole sets of constants of motion will be determined in each case and comments connecting relativistic and nonrelativistic considerations will be given.

II. RELATIVISTIC AND NONRELATIVISTIC CONFORMAL SYMMETRIES AND INVARIANCE CONDITIONS

First, let us recall (Sec. II A) some elements on the conformal space-time symmetry in the relativistic context and the well-known[17] invariance conditions on electromagnetic fields and potentials. Second, let us consider (Sec. II B) the nonrelativistic context (when the Galilei group has been extended to the Schrödinger one[14–16]) and let us establish the corresponding invariance conditions on electromagnetic fields and potentials in the magnetic limit.[20]

A. Relativistic conformal symmetry

We know that the action of the conformal group[13] on space-time events x reads in infinitesimal form

$$x \to x': \; x'^{\mu} = x^{\mu} - \xi^{\mu} \;\; (\mu = 0,1,2,3),$$

[a)] Boursier Agence Generale de Cooperation et Developpement.

 0022-2488/85/051072-05$02.50

with

$$\xi^\mu = -a^\mu - \omega^\mu{}_\nu x^\nu - \rho x^\mu - 2(c\cdot x)x^\mu + c^\mu x^2, \tag{2.1}$$

where the infinitesimal parameters $a^\mu, \omega^{\mu\nu}(= -\omega^{\nu\mu}), \rho$, and c^μ refer, respectively, to space-time translations, homogeneous Lorentz transformations, dilatations, and special conformal transformations. The conformal Lie algebra is then generated by the 15 associated infinitesimal operators $P^\mu, M^{\mu\nu}, D$, and C^μ satisfying the nonzero commutation relations[13]

$$\begin{aligned}
[M^{\mu\nu}, M^{\rho\sigma}] &= i(g^{\mu\sigma}M^{\nu\rho} + g^{\nu\rho}M^{\mu\sigma} \\
&\quad - g^{\mu\rho}M^{\nu\sigma} - g^{\nu\sigma}M^{\mu\rho}), \\
[M^{\mu\nu}, P^\rho] &= i(g^{\nu\rho}P^\mu - g^{\mu\rho}P^\nu), \\
[M^{\mu\nu}, C^\rho] &= i(g^{\nu\rho}C^\mu - g^{\mu\rho}C^\nu), \\
[P^\mu, D] &= iP^\mu, \quad [C^\mu, D] = -iC^\mu, \\
[P^\mu, C^\nu] &= 2i(g^{\mu\nu}D - M^{\mu\nu}).
\end{aligned} \tag{2.2}$$

The invariance conditions on the tensor fields $F \equiv \{F_{\mu\nu}\}$ (the electromagnetic field) and $A \equiv \{A_\mu\}$ (the electromagnetic potential) can then be expressed by the annulation of their Lie derivative with respect to the vector fields $X = \xi^\mu \partial_\mu (\partial_\mu = \partial/\partial x^\mu)$,

$$L_X F = 0, \quad \text{and} \quad L_X A = 0. \tag{2.3}$$

Explicitly these conditions are

$$\begin{aligned}
L_X F_{\mu\nu}(x) &= \xi^\alpha \partial_\alpha F_{\mu\nu}(x) + (\partial_\mu \xi^\alpha) F_{\alpha\nu}(x) \\
&\quad + (\partial_\nu \xi^\alpha) F_{\mu\alpha}(x) = 0
\end{aligned} \tag{2.4}$$

and

$$L_X A_\mu(x) = \xi^\alpha \partial_\alpha A_\mu(x) + (\partial_\mu \xi^\alpha) A_\alpha(x) = 0. \tag{2.5}$$

Moreover, A being a four-potential associated with a given electromagnetic field $F (F = dA)$ which admits the symmetry algebra $\mathbb{G}_F$, we have

$$L_X A_\mu(x) = \partial_\mu W_X(x), \quad \forall X \in \mathbb{G}_F, \tag{2.6}$$

where W is called a compensating gauge transformation[10] associated with A and satisfies the property[5]

$$\begin{aligned}
&L_X W_{X'}(x) - L_{X'} W_X(x) - W_{[X,X']}(x) = c(X, X'), \\
&\forall X, \ X' \in \mathbb{G}_F.
\end{aligned} \tag{2.7}$$

Then c is a skew-symmetric mapping from $\mathbb{G}_F \times \mathbb{G}_F$ into $\mathbb{R}$. In fact, it is a two-cocycle[21] of $\mathbb{G}_F$ for the trivial representation of $\mathbb{R}$ and it is defined up to a two-cobord. So it refers to a cohomology class $c(F) = [c] \in H^2(\mathbb{G}_F, \mathbb{R}, 0)$ of the second cohomology space of the trivial representation of $\mathbb{G}_F$ on $\mathbb{R}$. Because of the one-to-one correspondence[21] between $H^2(\mathbb{G}_F, \mathbb{R}, 0)$ and the central extensions[12] of $\mathbb{G}_F$ by $\mathbb{R}$, $c(F)$ characterizes the extensions of the symmetry algebra of the field F.

B. Nonrelativistic conformal or Schrödinger symmetry

The Schrödinger group[15] S of the Newtonian space-time is a 12-parameter Lie group containing, beside the Galilei group $\mathscr{G}$, the group of dilatations and a one-parameter group of transformations, the so-called expansions, which are similar to the special conformal transformations in the relativistic context. The action of S on the events $x \equiv (t, \mathbf{r})$ is then given in infinitesimal form by[14–16]

$$\begin{aligned}
t' &= t + \alpha t + ct^2 + b, \\
\mathbf{r}' &= \mathbf{r} + \boldsymbol{\theta} \times \mathbf{r} - \mathbf{v}t + \tfrac{1}{2}\alpha \mathbf{r} + ct\mathbf{r} + \mathbf{a},
\end{aligned} \tag{2.8}$$

where b and $\mathbf{a}$ refer to time and space translations, $\boldsymbol{\theta}$ to spatial rotations, $\mathbf{v}$ to pure Galilean transformations, α to dilatations, and c to expansions. If we denote by $H, \mathbf{P}, \mathbf{J}, \mathbf{K}, D$, and C the generators of the corresponding Schrödinger algebra $\mathscr{S}$, we have the commutators

$$\begin{aligned}
&[J_k, J_l] = i\epsilon_{klm} J_m, \quad [J_k, P_l] = i\epsilon_{klm} P_m, \\
&[J_k, K_l] = i\epsilon_{klm} K_m, \\
&[K_k, P_l] = 0, \quad [K_k, H] = -iP_k, \\
&[P_k, P_l] = [K_k, K_l] = [H, J_k] = [H, P_k] = 0, \\
&[D, P_k] = -\frac{i}{2}P_k, \quad [D, K_k] = \frac{i}{2}K_k, \quad [D, H] = -iH, \\
&[C, P_k] = -iK_k, \quad [C, H] = -2iD, \quad [C, D] = -iC, \\
&[D, J_k] = [C, J_k] \\
&\qquad = [C, K_k] = 0 \quad (k, l, m = 1, 2, 3).
\end{aligned} \tag{2.9}$$

The transformation laws of electromagnetic fields and potentials under Galilean transformations have been given[20] in the so-called magnetic limit. Under Schrödinger transformations, these laws can easily be generalized. Indeed, using Eq. (2.8) we deduce the action on the derivatives ∂_t and $\boldsymbol{\nabla}$ and, using the correspondence between the potentials $(V, \mathbf{A})$ and $(\partial_t, -\boldsymbol{\nabla})$, we get

$$\begin{aligned}
V'(t', \mathbf{r}') &= (1 - (\alpha + 2ct))V(t, \mathbf{r}) \\
&\quad - (\mathbf{v} - c\mathbf{r}) \cdot \mathbf{A}(t, \mathbf{r}), \\
\mathbf{A}'(t', \mathbf{r}) &= (1 - \tfrac{1}{2}(\alpha + 2ct))\mathbf{A}(t, \mathbf{r}) + \boldsymbol{\theta} \times \mathbf{A}(t, \mathbf{r}).
\end{aligned} \tag{2.10}$$

The electric $\mathbf{E}$ and magnetic $\mathbf{B}$ fields still being related to V and $\mathbf{A}$ by the definitions

$$\mathbf{E} = -\boldsymbol{\nabla} V - \partial_t \mathbf{A}, \quad \mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A},$$

we obtain the transformation laws of $\mathbf{E}$ and $\mathbf{B}$ in the magnetic limit

$$\begin{aligned}
\mathbf{E}'(t', \mathbf{r}') &= (1 - \tfrac{3}{2}(\alpha + 2ct))\mathbf{E}(t, \mathbf{r}) \\
&\quad + \boldsymbol{\theta} \times \mathbf{E}(t, \mathbf{r}) + (\mathbf{v} - c\mathbf{r}) \times \mathbf{B}(t, \mathbf{r}), \\
\mathbf{B}'(t', \mathbf{r}') &= (1 - (\alpha + 2ct))\mathbf{B}(t, \mathbf{r}) + \boldsymbol{\theta} \times \mathbf{B}(t, \mathbf{r}).
\end{aligned} \tag{2.11}$$

Invariance conditions on $(V, \mathbf{A})$ or on $(\mathbf{E}, \mathbf{B})$ can then be deduced. Explicitly we get

$$\begin{aligned}
&(\alpha + 2ct)V + (\mathbf{v} - c\mathbf{r}) \cdot \mathbf{A} + \mathscr{D} V = 0, \\
&\tfrac{1}{2}(\alpha + 2\text{ct})\mathbf{A} - \boldsymbol{\theta} \times \mathbf{A} + \mathscr{D} \mathbf{A} = 0,
\end{aligned} \tag{2.12}$$

and

$$\begin{aligned}
&\tfrac{3}{2}(\alpha + 2ct)\mathbf{E} - \boldsymbol{\theta} \times \mathbf{E} - (\mathbf{v} - c\mathbf{r}) \times \mathbf{B} + \mathscr{D} \mathbf{E} = 0, \\
&(\alpha + 2ct)\mathbf{B} - \boldsymbol{\theta} \times \mathbf{B} + \mathscr{D} \mathbf{B} = 0,
\end{aligned} \tag{2.13}$$

where

$$\begin{aligned}
\mathscr{D} &= b\partial_t + \mathbf{a} \cdot \boldsymbol{\nabla} + \boldsymbol{\theta} \cdot (\mathbf{r} \times \boldsymbol{\nabla}) - t\mathbf{v} \cdot \boldsymbol{\nabla} + \alpha(t\partial_t + \tfrac{1}{2}\mathbf{r} \cdot \boldsymbol{\nabla}) \\
&\quad + ct(t\partial_t + \mathbf{r} \cdot \boldsymbol{\nabla}).
\end{aligned} \tag{2.14}$$

III. CONSTANT ELECTROMAGNETIC FIELDS AND CONSTANTS OF MOTION

The symmetry properties of constant and uniform electromagnetic fields are well known under the Poincaré[6] as well as under the Galilei[7] groups. In the relativistic context, two kinds of fields have to be considered; the so-called parallel ($F_{\parallel}$) and orthogonal ($F_{\perp}$) fields. If the Poincaré symmetry is extended to the conformal one, the field $F_{\parallel}$ does not admit a kinematical algebra larger than the six-dimensional one in the Poincaré case so that no supplementary interesting feature arises. On the contrary, when the Poincaré symmetry is extended by dilatations, (i.e., Weyl symmetries[22]), it has been shown[18] that the field $F_{\perp}$, chosen such as

$$\mathbf{E} = (E,0,0), \quad \mathbf{B} = (0,E,0), \tag{3.1}$$

admits a kinematical algebra of dimension 7

$$\mathbb{G}^W_{F_\perp} = \{\mathscr{A}^1 = J^1 + K^2,\ \mathscr{A}^2 = J^2 - K^1,\ 2K^3 - D, P^\mu\}. \tag{3.2}$$

The generator $2K^3-D$ is then a new generator with respect to the Poincaré case and it is easy to convince ourselves that, under the whole conformal symmetry, the $F_{\perp}$-kinematical algebra is again $\mathbb{G}^W_{F_\perp} \equiv (3.2)$.

Moreover, with the field $F_\perp \equiv (3.1)$, we can associate a four-potential (the gauge-symmetrical potential, for example)

$$V_\perp = -\tfrac{1}{2}Ex, \quad \mathbf{A}_\perp = \tfrac{1}{2}E(z-t,0,-x), \tag{3.3}$$

which admits a symmetry algebra of maximal dimension[11] inside the one of the field $F_\perp$. We effectively have the algebra

$$\mathbb{G}_{A_\perp} = \{\mathscr{A}^1, \mathscr{A}^2, 2K^3 - D, P^2, P^0 - P^3\}, \tag{3.4}$$

which evidently contains the symmetry algebra (see Ref. 11) $\mathbb{G}^P_{A_\perp}$. Let us also notice that for the symmetries of $F_\perp$ which are not in $\mathbb{G}_{A_\perp}$, we obtain the compensating gauge transformations through Eq. (2.6)

$$W_{P^1}(x) = \tfrac{1}{2}E(t-z), \quad W_{P^0+P^3}(x) = -Ex. \tag{3.5}$$

This leads to the existence of a nontrivial extension (the physical one) of $\mathbb{G}^W_{F_\perp}$ by $\mathbb{R}$, which is characterized by the cocycle c satisfying Eq. (2.7) and defined by

$$c(P^0,P^1) = E = c(P^3,P^1). \tag{3.6}$$

Let us notice that an extension of $\mathbb{G}_{F_\perp}$ already appears in the Poincaré context.[6,11]

The preceding results, restricted to the Poincaré context, have been used[4] in the search for the constants of motion associated with the description of charged particles interacting with the external field $F_\perp$. They can be extended to the conformal context provided that the particles are massless. Following the above method,[4] the whole set of constants of motion is then obtained as follows:

$$\begin{aligned}
&\pi^0 = H_D - eV, \quad \boldsymbol{\pi} = \mathbf{p} - e\mathbf{A},\\
&\mathscr{A}^1 = (t-z)p^2 - y(H_D - p^3) + \Sigma^1 + i\alpha^2/2,\\
&\mathscr{A}^2 = -(t-z)p^1 + x(H_D - p^3) + \Sigma^2 - i\alpha^1/2,\\
&2K^3 - D = \mathbf{r}\cdot\mathbf{p} - tH_D - 2zH_D + 2tp^3 + i\alpha^3,
\end{aligned} \tag{3.7}$$

where H_D is the Dirac Hamiltonian for spin-$\frac{1}{2}$ particles in interaction with the $F_\perp$ field

$$H_D = \boldsymbol{\alpha}\cdot(\mathbf{p} + e\mathbf{A}) - eV \quad (\mathbf{p} = -i\nabla) \tag{3.8}$$

and $\boldsymbol{\Sigma} = -(i/2)(\alpha^2\alpha^3, \alpha^3\alpha^1, \alpha^1\alpha^2)$. The operators (3.7) give a specific realization of the generators of the nontrivial physical extension of $\mathbb{G}^W_F$ by $\mathbb{R}$ corresponding to the cocycle $c \equiv (3.6)$. Such a realization depends on the form of the potential $A \equiv (3.3)$ associated with $F_\perp$, and only the operators corresponding to the symmetries of A are effectively unchanged[4] with respect to the free case.

Let us end this section by noticing that in the nonrelativistic context, the search for invariance properties on constant and uniform fields throughout the Schrödinger symmetry does not give new interesting results. Here we have only the Galilean conclusions.[7]

IV. ELECTROMAGNETIC FIELDS INVARIANT UNDER MAXIMAL SUBALGEBRAS OF THE CONFORMAL ALGEBRA AND CONSTANTS OF MOTION

By noticing that the whole conformal algebra admits only the trivial invariant electromagnetic field $F = 0$, Beckers, Harnad, Perroud, and Winternitz[17] have searched for electromagnetic fields invariant under maximal conformal subalgebras. They have shown that, among these, only two admit nontrivial invariant electromagnetic fields: the $so(3) \oplus so(2,1)$ and $so(2,1) \oplus so(2,1)$ algebras, each one being of dimension 6.

Let us discuss both cases in order to associate with each invariant field some electromagnetic potentials, to determine their symmetry properties as well as the corresponding compensating gauge transformations. Then, by the Beckers–Hussin method,[4] we will be able to determine if the symmetry algebra of the field under consideration admits trivial or nontrivial physical extensions by $\mathbb{R}$ and to obtain the constants of motion associated with the description of charged particles interacting with such electromagnetic fields.

A. $so(3) \oplus so(2,1)$

In the relativistic context, we can choose the basis[17]

$$\mathbb{G}_1 \equiv \{\mathbf{J}, P^0, D, C^0\} \tag{4.1}$$

for such an algebra. The more general invariant electromagnetic field is then the field $F \equiv (\mathbf{E},\mathbf{B})$ characterized by[17]

$$\mathbf{E} = E(\mathbf{r}/r^3), \quad \mathbf{B} = B(\mathbf{r}/r^3), \tag{4.2}$$

where $r^2 = x^2 + y^2 + z^2$ and E,B are real numbers. Such an interesting physical field F may be interpreted as the one of a static point electric charge and of a point magnetic charge—a magnetic monopole—located at the origin of the coordinate system in $\mathbb{R}^3$ (where the field is singular) or, in other words, as the Coulomb field and the magnetic monopole field. The two charges are proportional to E and B, respectively. Let us notice that the subalgebra $\{\mathbf{J},P^0\} \subset \mathbb{G}_1$ corresponds to a CEPS,[23] i.e., a connected electromagnetic Poincaré subalgebra, already admitting the invariant field (4.2).

Now, from the approach of Wu and Yang,[19] we can consider two alternative forms of singular potentials $A_\pm = (V_\pm, \mathbf{A}_\pm)$ with

$$\begin{aligned}
&V_\pm = E/r, \quad A^1_\pm = \mp By/r(r \pm z),\\
&A^2_\pm = \pm Bx/r(r \pm z), \quad A^3_\pm = 0.
\end{aligned} \tag{4.3}$$

We know[19] that $\mathbf{A}_+(\mathbf{A}_-)$ has string singularities on the negative (positive) z axis and that the field $F \equiv (4.2)$ can be described by using the two coordinate domains and potentials

$A_\pm$. In the overlapping region of the two domains, the potentials are related by the gauge transformation[19]

$$\lambda = 2B \tan^{-1}(y/x). \tag{4.4}$$

Using Eq. (2.5) let us determine the symmetry properties of the two potentials $A_\pm$ with the help of spherical coordinates (r,θ,ϕ). We easily find the symmetry algebra

$$\mathbb{G}_{A_\pm} = \{J^3, P^0, D\} \subset \mathbb{G}_1. \tag{4.5}$$

The complementary symmetries of the field (4.2) lead through Eq. (2.6) to

$$W_{J^1}^{\pm} = Bx/(r \pm z), \quad W_{J^2}^{\pm} = By/(r \pm z), \quad W_{C^0} = 2Er, \tag{4.6}$$

and we notice that the compensating gauge transformations $W_{J^1}^{\pm}$ and $W_{J^2}^{\pm}$ satisfy the relations

$$L_{J^1} W_{J^2}^{\pm} - L_{J^2} W_{J^1}^{\pm} = \pm B. \tag{4.7}$$

So, by comparison with Eq. (2.7) we have to choose $W_{J^3}^{\pm} = \pm B$. Such a choice does not alter the invariance of $A_\pm$ under J^3 but gives the well-known fact that the algebra so(3) admits only trivial extension by $\mathbf{R}$ ($[c] = 0$). Moreover, it is easy to show that the cocycle c is zero for all the elements of $\mathbb{G}_1$ so that the algebra $\mathbb{G}_1$ admits only trivial physical extensions by $\mathbf{R}$.

The symmetry properties of the field (4.2) and of the associated potentials (4.3) are of special interest for the search for constants of motion.[4] These constants can be determined in the Lagrangian formalism using the Noether theorem,[4] but let us here study the Hamiltonian approach. If we consider the Hamiltonian $H_\mathrm{D} \equiv (3.8)$ the whole set of constants of motion is easily obtained using the Beckers–Hussin method.[4] We effectively get

$$\begin{aligned} \mathscr{J} &= \mathbf{r} \times \mathbf{p} + \mathbf{\Sigma} + e\mathbf{W}_J^{\pm}, \\ P^0 &= H_\mathrm{D}, \quad D = tH_\mathrm{D} - \mathbf{r}\cdot\mathbf{p} + 3i/2, \\ C^0 &= 2tD - (t^2 - r^2)H_\mathrm{D} - i\boldsymbol{\alpha}\cdot\mathbf{r} + eW_{C^0}, \end{aligned} \tag{4.8}$$

with

$$\mathbf{W}_J^{\pm} = (W_{J^1}^{\pm}, W_{J^2}^{\pm}, W_{J^3}^{\pm}). \tag{4.9}$$

Let us first remark that in the expression of D, the term $3i/2$ corresponds to the scale dimension[13] of the Dirac spinor field. In particular, it is necessarily added in order to get the annulation of the total time derivative of C^0. Second, the quantities (4.8) can also be written as

$$\mathbf{J} = \mathbf{r} \times \mathbf{\Pi} + \mathbf{\Sigma} + eB(\mathbf{r}/r) \quad (\mathbf{\Pi} = \mathbf{p} + e\mathbf{A}), \tag{4.10a}$$

$$P^0 = \boldsymbol{\alpha}\cdot\mathbf{\Pi} - e(E/r), \tag{4.10b}$$

$$D = t(\boldsymbol{\alpha}\cdot\mathbf{\Pi}) - \mathbf{r}\cdot\mathbf{\Pi} + 3i/2 - e(Et/r), \tag{4.10c}$$

$$C^0 = 2tD - (t^2 - r^2)(\boldsymbol{\alpha}\cdot\mathbf{\Pi}) - i\boldsymbol{\alpha}\cdot\mathbf{r} + e[E(t^2 + r^2)/r]. \tag{4.10d}$$

The last terms of these expressions are nonkinematical contributions[1] to the constants of motion. For the expression (4.10a) we recover the famous Poincaré result[24] with the contribution $eB(\mathbf{r}/r)$. Finally, among the constants (4.8) generating the so(3) ⊕ so(2,1) algebra, we find, as expected, that the only operators unchanged with respect to the free case are those associated with the symmetry generators (4.5) of the potentials, although the operator $\mathscr{J}^3$ is slightly modified by an additional constant in order to maintain the current so(3) structure.

Let us now consider the nonrelativistic context. Indeed, the Schrödinger algebra $\mathscr{S}$ also admits a so(3) ⊕ so(2,1) subalgebra[14] given here by

$$\mathbb{G}_1' \equiv \{\mathbf{J}, H, D, C\}. \tag{4.11}$$

It is then easy to show by using Eqs. (2.5) that the field (4.2) is invariant under $\mathbb{G}_1'$ if and only if $E = 0$. This physical situation is very interesting: it corresponds to the purely magnetic monopole field so often discussed in the literature. As a comment, let us however mention that the field $F \equiv (\mathbf{E} = 0, \mathbf{B} = B(\mathbf{r}/r^3))$ is not the more general electromagnetic field invariant under $G_1' \equiv (4.11)$. In fact, by Eq. (2.5) we get

$$\mathbf{E} = E(\mathbf{r}/r^4), \quad \mathbf{B} = B(\mathbf{r}/r^3), \tag{4.12}$$

where we notice that the electric field derives from a scalar potential of the type

$$V = E/2r^2 = \lambda(1/r^2). \tag{4.13}$$

If we limit ourselves to the purely magnetic monopole field, the whole set of constants of motion can evidently be determined when nonrelativistic spin-$\frac{1}{2}$ particles interact with such a field. We immediately recover the recent results of Jackiw[1] and D'Hoker–Vinet[2] with the Pauli Hamiltonian

$$H_P = (1/2m)\mathbf{\Pi}^2 - (e/2m)\mathbf{B}\cdot\boldsymbol{\sigma}, \tag{4.14}$$

where $\sigma \equiv (\sigma^1, \sigma^2, \sigma^3)$ are the usual Pauli matrices. This set of constants of motion is

$$\begin{aligned} \mathscr{J} &= \mathbf{r} \times \mathbf{\Pi} + i(\boldsymbol{\sigma}/2) + eB(\mathbf{r}/r), \\ H &= H_P, \quad D = tH_P - \tfrac{1}{2}\mathbf{r}\cdot\mathbf{p}, \\ C &= 2tD - t^2 H_P - \tfrac{1}{2}m\mathbf{r}^2. \end{aligned} \tag{4.15}$$

The associated operators generate a Lie algebra isomorphic to $\mathbb{G}_1'$ and their explicit forms can effectively be obtained[4] through the use of the potentials (4.3) with $E = 0$. Let us also notice that, if necessary, in connection with Eqs. (4.12) and (4.13), we can add a $1/r^2$ term in the Pauli Hamiltonian without changing the dynamical symmetry.[1,2]

In conclusion, in both relativistic and nonrelativistic contexts, the interaction with a magnetic monopole field gives rise to a dynamical symmetry isomorphic to so(3) ⊕ so(2,1). The realization (4.15) of the dilatation D and expansion C generators are slightly different with respect to the relativistic case but this is due to the transformation law (2.8) where, for example, the effect of the dilatation is not the same on time and space coordinates.

B. so(2,1) ⊕ so(2,1)

In the relativistic context, a basis for such an algebra is given by

$$\mathbb{G}_2 = \{J^3, K^1, K^2, P^3, D, C^3\}, \tag{4.16}$$

and the more general invariant electromagnetic field takes the form[17]

$$\begin{aligned} \mathbf{E} &= (1/\lambda^3)(-My, Mx, Nt), \\ \mathbf{B} &= (1/\lambda^3)(Ny, -Nx, Mt), \end{aligned} \tag{4.17}$$

where $\lambda = |t^2 - x^2 - y^2|^{1/2}$ and M, N are arbitrary real numbers. The field singularities occur when $t^2 = x^2 + y^2$, that is, on a cylinder along the z axis, whose radius is increasing with the velocity of light.[17] Let us also notice that the subalgebra $\{J^3, K^1, K^2, P^3\}$ is once again a CEPS[23] so that such a field has already been determined.[23]

The differences between this and the preceding cases are essentially due to the fact that here we are dealing with an so(2,1) noncompact subalgebra of the Lorentz algebra, while so(3) is compact. Then we have to use specific "spherical coordinates"[25] in the hyperboloids

$$t^2 - x^2 - y^2 = \pm \lambda^2.$$

That is, for the $+\lambda^2$ case

$$t = \lambda \cosh \alpha, \quad x = \lambda \sinh \alpha \cos \phi,$$
$$y = \lambda \sinh \alpha \sin \phi, \quad z = z,$$

or for the $-\lambda^2$ case

$$t = \lambda \sinh \alpha, \quad x = \lambda \cosh \alpha \cos \phi,$$
$$y = \lambda \cosh \alpha \sin \phi, \quad z = z.$$

So in a way parallel to the so(3) ⊕ so(2,1) case where the potentials were given by Eq. (4.3), we can construct some potentials associated with the field (4.16). For example, according to the preceding remarks we get the two potentials

$$A'_{\pm} = (V'_{\pm}, \mathbf{A}'_{\pm}) \text{ where}$$
$$V'_{\pm} = 0, \quad A'^{1}_{\pm} = -My/\lambda\,(t \pm \lambda), \qquad A'^{2}_{\pm} = Mx/\lambda\,(t \pm \lambda), \quad A'^{3}_{\pm} = N\lambda^{-1}, \tag{4.18}$$

which have singularities on the negative or positive t axis. The potentials A'_{+} and A'_{-} are well-defined and regular in their domain. The field $F \equiv$(4.17) can be described by using the two coordinate domains and potentials $A'_{\pm}$. In the overlapping region, the potentials are related by

$$A'_{+} = A'_{-} + d\lambda', \quad \lambda' = 2M \tan^{-1}(y/x). \tag{4.19}$$

The symmetry algebra of these potentials is

$$G_{A'} = \{J^3, P^3, D\},$$

and the compensating gauge transformations satisfying Eq. (2.6) are

$$W^{\pm}_{J^3} = \pm M, \quad W^{\pm}_{K^1} = -My/(\lambda \pm t), \qquad W^{\pm}_{K^2} = Mx/(\lambda \pm t), \quad W_{P^3} = 0, \quad W_D = 0, \quad W_{C^3} = 2N\lambda. \tag{4.20}$$

With these quantities, it is easy to show from Eq. (2.7) that the cocycle c is zero and consequently that the symmetry algebra of the field $F \equiv$(4.17) admits only trivial extensions by $\mathbb{R}$.

Here also the constants of motion for the Hamiltonian (3.8) are easily obtained as the quantities

$$\begin{aligned}
\mathscr{J}^3 &= (\mathbf{r}\times\mathbf{p})^3 + \Sigma^3 + eW^{\pm}_{J^3},\\
K^1 &= tp^1 - xH_D + i\alpha^1/2 + eW^{\pm}_{K^1},\\
K^2 &= tp^2 - yH_D + i\alpha^2/2 + eW^{\pm}_{K^2},\\
P^3 &= p^3, \quad D = tH_D - \mathbf{r}\cdot\mathbf{p} + 3i/2,\\
C^3 &= 2zD - (t^2 - \mathbf{r}^2)p^3 - i\alpha^3 t + 2(\mathbf{\Sigma}\times\mathbf{r})^3 + eW_{C^3}.
\end{aligned} \tag{4.21}$$

The associated operators evidently generate the Lie algebra so(2,1) ⊕ so(2,1).

Let us finally notice that the Schrödinger algebra $\mathscr{S}$ does not contain a so(2,1) ⊕ so(2,1) algebra so that we do not have to study the nonrelativistic context here.

ACKNOWLEDGMENTS

We want to thank very heartily Professor J. Beckers for stimulating discussions and his kind interest.

One of the authors (V.H.) wants to thank the Belgian Fonds National de la Recherche Scientifique for its financial support.

[1] R. Jackiw, Ann. Phys. (NY) **129**, 183 (1980).
[2] E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984); "Dynamical Supersymmetry of the Magnetic Monopole and the $1/r^2$-potential," (to be published).
[3] P. A. M. Dirac, Proc. Roy. Soc. London, Ser. A **137**, 60 (1931).
[4] J. Beckers and V. Hussin, Phys. Rev. D **29**, 2814 (1984).
[5] J. Beckers and V. Hussin, J. Math. Phys. **26**, 925 (1985).
[6] H. Bacry, Ph. Combe, and J. L. Richard, Nuovo Cimento A **67**, 267 (1970).
[7] H. Bacry, Ph. Combe, and J. L. Richard, Nuovo Cimento A **70**, 289 (1970).
[8] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[9] J. M. Levy-Leblond, *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. 2, p. 221; J. Math. Phys. **4**, 776 (1963).
[10] A. Janner and T. Janssen, Physica **53**, 1 (1971).
[11] J. Beckers and V. Hussin, J. Math. Phys. **24**, 1286 (1983).
[12] V. Bargmann, Ann. Math. **59**, 1 (1954); M. Hammermesh, *Group Theory* (Addison–Wesley, Reading, MA, 1962).
[13] H. A. Kastrup, Ann. Phys. Leipzig **9**, 388 (1962); G. Mack and A. Salam, Ann. Phys. (NY) **53**, 174 (1969).
[14] C. R. Hagen, Phys. Rev. D **5**, 377 (1972).
[15] U. Niederer, Helv. Phys. Acta **45**, 802 (1972).
[16] P. Roman, J. J. Aghasi, R. M. Santilli, and P. L. Huddleston, Nuovo Cimento A **12**, 185 (1972).
[17] J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[18] S. Sinzinkayo, Nuovo Cimento B **79**, 50 (1984).
[19] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975); Nucl. Phys. B **107**, 365 (1976); Phys. Rev. D **14**, 437 (1976).
[20] M. Le Bellac and J. M. Levy-Leblond, Nuovo Cimento B **14**, 217 (1973).
[21] L. Michel, *Application of Mathematics to Problems in Theoretical Physics*, (Gordon and Breach, New York, 1967).
[22] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1615 (1975).
[23] Ph. Combe and P. Sorba, Physica A **80**, 271 (1975).
[24] H. Poincaré, C. R. Acad. Sci. Paris **123**, 530 (1896).
[25] P. Winternitz, "Expansions of relativistic scattering amplitudes and harmonic analysis on the Lorentz group," in 1970–71 Lecture Series, edited by D. E. Myers (Univ. Arizona, 1971).

Remarks about the nilpotent symmetries of quantum gauge field theories

F. A. Kaempffer
Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada V6T 2A6

(Received 26 June 1984; accepted for publication 21 December 1984)

A conjugation symmetric representation of Becchi–Rouet–Stora (BRS) and anti-BRS operations is presented which differs from the usual versions by terms cubic in the ghost fields. Some consequences of requiring quantum gauge field theories to be invariant under both BRS and anti-BRS transformations are examined.

I. INTRODUCTION

It is now well-known that invariance under local gauge transformations,[1] which is the very *raison d'être* for compensating gauge potentials[2] in classical field theories, but which is impaired by gauge-fixing conditions (needed before canonical quantization can proceed) and subsequent introduction of Faddeev–Popov ghosts[3,4] (needed to salvage unitarity of the scattering operator[5]) is in quantum gauge field theories superseded by invariance under the global and nilpotent Becchi–Rouet–Stora[6] (BRS) and anti-BRS[7,8] transformations.

However, the published versions of these transformations, as applied to constituent fermion source fields ψ, gauge potentials[9] A^A_a, auxiliary boson fields b^A, and ghost–antighost pairs C^A, $\overline{C}^A$, do not exhibit the conjugation symmetry one would expect if, as persuasively advocated by Beaulieu and Thierry-Meg,[10] BRS and anti-BRS operations are conceived as exterior differentiations in the sense of Cartan when applied to ghost fields C^A, $\overline{C}^A$ that are Hermitian conjugates. For example, in the work of Ojima,[11] which employs ghosts that are not Hermitian conjugates,[12] the auxiliary fields b^A are BRS but not anti-BRS invariants, and in the work of Beaulieu and Thierry-Meg,[13] which properly employs Hermitian conjugate ghosts, the auxiliary boson fields are not Hermitian, and the definition of BRS and anti-BRS operations is not manifestly conjugation symmetric.

Motivated by a desire to examine the consequences of raising the requirement of BRS and anti-BRS invariance to the level of a fundamental principle, as promulgated by Kugo and Uehara[14] and by Beaulieu and Thierry-Meg,[15] the results reported in this paper emerged during an effort aimed at constructing all BRS and anti-BRS invariant actions of the Yang-Mills type, including those that lead to theories which are not renormalizable by power counting. For that purpose it is useful to have available a manifestly conjugation symmetric representation of the nilpotent BRS and anti-BRS operators, enabling one to enumerate in a straightforward and trenchant manner all polynomials, of given canonical scale dimension and given relativistic covariance, that are BRS and anti-BRS invariants.

Definitions of BRS and anti-BRS operations meeting these requirements are presented (Sec. II). They differ from the usual versions by terms which are cubic in the ghost fields. Application of the ensuing formalism to the most general renormalizable, conjugation symmetric, and canonically quantizable Yang-Mills action (Sec. III) which retains invariance under global ghost phase transformations, classical gauge transformations, BRS and anti-BRS transformations, yields definitions of the corresponding Noether currents without any spurious terms[16] that destroy their conjugation symmetry, and enables one to give remarkably simple proofs of the conservation of ghost charge (Sec. IV), internal charges (Sec. V), and BRS and anti-BRS charge (Sec. VI), which, respectively, generate[17] these global transformations.

When trying to define BRS and anti-BRS operations for the gravitational field one encounters the question of what constitutes the gauge group in that case. An obvious choice is the group of infinitesimal coordinate transformations which may be looked upon as a kind of local "gauge transformations" in the context of Einstein's theory of gravitation. Although the corresponding BRS operations, introduced by Nakanishi,[18] can be augmented by anti-BRS operations and cast in conjugation symmetric form (Sec. VII), they do not commute with differentiations. This defect is avoided by the transformations of Delbourgo and Medrano[19] which are the BRS analog of the Lie derivatives of tensors. Since one has the option of looking upon the Lie derivative as an alternative definition of what is meant by "gauge transformation" in the context of general relativity, as has been demonstrated by Nishijima and Okawa,[20] and since the corresponding BRS operations can also be augmented by anti-BRS operations and cast in conjugation symmetric form (Sec. VIII), the requirement that any quantum theory of gravitation be invariant under both BRS and anti-BRS transformations, in the sense of Delbourgo and Medrano, can be used to bring out the close analogy between Weyl gravity and gauge theories of the Yang–Mills type (Sec. IX).

II. DEFINITION OF BRS AND ANTI-BRS OPERATIONS

For a Yang–Mills type gauge theory based on a compact Lie group with generators S_A, structure constants f_{ABC}, and coupling parameter g, let BRS and anti-BRS operations be defined by[21]

$$\begin{aligned}
&\Delta\psi = igC^AS_A\psi, \quad \Delta\bar\psi = ig\bar\psi C^AS_A,\\
&\Delta A^A_a = \partial_a C^A - gf^A{}_{BC}C^BA^C_a \equiv D^{AB}_a C_B,\\
&\Delta C^A = -\tfrac{1}{2}gf^A{}_{BC}C^BC^C, \quad \Delta\overline{C}^A = b^A - \tfrac{1}{2}gf^A{}_{BC}\overline{C}^BC^C,\\
&\Delta b^A = \tfrac{1}{2}gf^A{}_{BC}b^BC^C - (g^2/4)f^A{}_{BC}f^C{}_{DE}C^B\overline{C}^DC^E
\end{aligned} \tag{2.1}$$

and

$$\begin{aligned}
&\overline{\Delta}\bar\psi = ig\bar\psi\overline{C}^AS_A, \quad \overline{\Delta}\psi = ig\overline{C}^AS_A\psi, \quad \overline{\Delta}A^A_a = D^{AB}_a\overline{C}_B,\\
&\overline{\Delta C}^A = -\tfrac{1}{2}gf^A{}_{BC}\overline{C}^B\overline{C}^C,\\
&\overline{\Delta}C^A = -b^A - \tfrac{1}{2}gf^A{}_{BC}\overline{C}^BC^C,\\
&\overline{\Delta}b^A = \tfrac{1}{2}gf^A{}_{BC}b^B\overline{C}^C + (g^2/4)f^A{}_{BC}f^C{}_{DE}\overline{C}^BC^D\overline{C}^E,
\end{aligned} \tag{2.2}$$

where Δ and $\overline{\Delta}$ are exterior differentiation operators whose

 0022-2488/85/051077-06$02.50

action on any polynomial in the fields is governed by the graded Leibniz rule

$$\Delta(XY) = (\Delta X)Y \pm X(\Delta Y), \tag{2.3}$$

with the minus sign when X contains an odd number of anticommuting fields such as $C, \overline{C}$.

With the Hermiticity assignments $\psi^+ = \bar\psi\gamma^4$, $A_a^{A+} = A_a^A$, $C_A^+ = \overline{C}_A$, $\overline{C}_A^+ = C_A$, $b_A^+ = b_A$, $\partial_a^+ = \partial_a$, $f_{ABC}^+ = f_{ABC}$, and $\Delta^+ = \underleftarrow{\bar\Delta}$, $\bar\Delta^+ = \underleftarrow{\Delta}$, where $X\underleftarrow{\bar\Delta} = \pm\bar\Delta X$, $X\underleftarrow{\Delta} = \pm\Delta X$ with the minus sign when X contains an odd number of anticommuting fields, one obtains the operation $\bar\Delta$ from the operation Δ by conjugation, and vice versa, keeping in mind that conjugation reverses the order of terms. The operations Δ and $\bar\Delta$ anticommute and are nilpotent,

$$\Delta\bar\Delta + \bar\Delta\Delta = \Delta\Delta = \bar\Delta\bar\Delta = 0, \tag{2.4}$$

as is easily verified with repeated use of the Jacobi identity

$$f_{ABC}f^C{}_{DE}(X^BY^DZ^E + X^DY^EZ^B + X^EY^BZ^D) = 0.$$

Since these definitions of Δ and $\bar\Delta$, as applied to fermion fields ψ and gauge potentials A_a^A, are identical with the usual versions, the invariance of the classical Yang–Mills action $L_{\rm YM}$ under both Δ and $\bar\Delta$ remains inviolate. The essential difference between the definitions (2.1) and (2.2) and the usual versions resides in the terms cubic in $C, \overline{C}$ which are necessary to ensure conjugation symmetry without violating the relations (2.4).

The nilpotency of Δ and $\bar\Delta$ guarantees that application of the product $\Delta\bar\Delta$ to any polynomial in the fields will yield an invariant under both Δ and $\bar\Delta$. If the usual canonical scale dimensions $(-\frac{3}{2}, -1, -2, -1)$ are assigned to the fields (ψ, A_a^A, b^A, C^A), the operations Δ and $\bar\Delta$ each carry the scale dimension -1 by their definitions. Therefore, the most general nonvanishing scalar, of scale dimension -4 and invariant under both Δ and $\bar\Delta$, obtainable in this fashion is $\Delta\bar\Delta(A_a^AA_A^a - \xi\overline{C}^AC_A)$, where ξ is the conventional gauge-fixing parameter, as an examination of all possible scalar polynomials of scale dimension -2 will show.

III. CANONICAL QUANTIZATION

The foregoing definitions yield as the most general scalar action, of scale dimension -4 and invariant under both Δ and $\bar\Delta$, with a convention of sign and an overall factor absorbed in the definition of fields, and with $F_{ab}^A = \partial_bA_a^A - \partial_aA_b^A - gf^A{}_{BC}A_a^BA_b^C$,

$$\begin{aligned} L &= L_{YM} + \tfrac{i}{2}\bar\Delta\Delta(A_a^AA_A^a - \xi\overline{C}^AC_A) \\ &= i\bar\psi\gamma^a(\partial_a - igA_a^AS_A)\psi - \tfrac{1}{4}F_{ab}^AF_A^{ab} - A_A^a\,\partial_ab^A \\ &\quad + \tfrac{i}{2}([\partial^a\overline{C}_A, D_a^{AB}C_B]_- - [\partial^aC_A, D_a^{AB}\overline{C}_B]_-) \\ &\quad + \tfrac{1}{2}\xi(b_Ab^A + (g^2/4)f_{ABC}\overline{C}^BC^Cf^A{}_{DE}\overline{C}^DC^E). \end{aligned} \tag{3.1}$$

Upon variation with respect to the field variables, keeping in mind the graded Leibniz rule for differentiation with respect to anticommuting fields, one obtains the field equations

$$i\gamma^a(\partial_a - igA_a^AS_A)\psi = 0, \tag{3.2}$$

$$\begin{aligned} &D_{AB}^bF_{ab}^B + g\bar\psi\gamma_aS_A\psi - \partial_ab_A - \tfrac{i}{2}gf_{ABC} \\ &\quad\times[(\partial_a\overline{C}^B)C^C - (\partial_aC^B)\overline{C}^C] = 0, \end{aligned} \tag{3.3}$$

$$\partial_aA_A^a + \xi b_A = 0, \tag{3.4}$$

$$\begin{aligned} &\partial_a(D_{AB}^a\overline{C}^B) - \tfrac{1}{2}gf_{ABC}(\partial^aA_a^B)\overline{C}^C \\ &\quad + (\xi g^2/4)f_{ABC}f^C{}_{DE}\overline{C}^BC^D\overline{C}^E = 0, \end{aligned} \tag{3.5}$$

$$\begin{aligned} &\partial_a(D_{AB}^aC^B) - \tfrac{1}{2}gf_{ABC}(\partial^aA_a^B)C^C \\ &\quad - (\xi g^2/4)f_{ABC}f^C{}_{DE}C^B\overline{C}^DC^E = 0, \end{aligned} \tag{3.6}$$

which exhibit the desired conjugation symmetry. In particular, Eqs. (3.5) and (3.6) are, respectively, the $\bar\Delta$ and Δ transforms of Eq. (3.4), $\partial_a(\bar\Delta A_A^a) + \xi\bar\Delta b_A = 0$ and $\partial_a(\Delta A_A^a) + \xi\Delta b_A = 0$.

Canonical quantization can now proceed as usual by introduction of canonical momenta

$$\begin{aligned} &\pi_\psi \equiv \frac{\partial L}{\partial\dot\psi} = -i\bar\psi\gamma^4, \quad P_A^a \equiv \frac{\partial L}{\partial\dot A_a^A} = F_A^{4a} \quad (a = 1, 2, 3), \\ &p_A \equiv \frac{\partial L}{\partial\dot b^A} = -A_A^4, \\ &\Pi_A \equiv \frac{\partial L}{\partial\dot C^A} = -\dot{\overline{C}}_A + \tfrac{1}{2}gf_{ABC}\overline{C}^BA_4^C, \\ &\overline{\Pi}_A \equiv \frac{\partial L}{\partial\dot{\overline{C}}^A} = \dot C_A - \tfrac{1}{2}gf_{ABC}C^BA_4^C, \end{aligned} \tag{3.7}$$

and imposition of equal-time commutators

$$\begin{aligned} &[\pi_\psi^\beta(\mathbf{x}, t), \psi_\alpha(\mathbf{y}, t)]_+ = -i\delta_\alpha^\beta\delta(\mathbf{x} - \mathbf{y}), \\ &[P_A^b(\mathbf{x}, t), A_a^B(\mathbf{y}, t)]_- = -i\delta_a^b\delta_A^B\delta(\mathbf{x} - \mathbf{y}), \\ &[p_A(\mathbf{x}, t), b^B(\mathbf{y}, t)]_- = -i\delta_A^B\delta(\mathbf{x} - \mathbf{y}), \\ &[\Pi_A(\mathbf{x}, t), C^B(\mathbf{y}, t)]_+ = -i\delta_A^B\delta(\mathbf{x} - \mathbf{y}), \\ &[\overline{\Pi}_A(\mathbf{x}, t), \overline{C}^B(\mathbf{y}, t)]_+ = -i\delta_A^B\delta(\mathbf{x} - \mathbf{y}). \end{aligned} \tag{3.8}$$

IV. THE CONSERVATION OF GHOST CHARGE

Since the action (3.1) is invariant under the global "ghost phase" transformation[22]

$$\delta C^A = i\omega C^A, \quad \delta\overline{C}^A = -i\omega\overline{C}^A, \tag{4.1}$$

with an infinitesimal parameter ω, there exists an associated Noether current

$$\begin{aligned} J^a = i[&(\partial^a\overline{C}_A - \tfrac{1}{2}gf_{ABC}\overline{C}^BA^{Ca})C^A \\ &+ (\partial^aC_A - \tfrac{1}{2}gf_{ABC}C^BA^{Ca})\overline{C}^A] \end{aligned} \tag{4.2}$$

whose timelike component defines the "ghost charge"

$$Q = \int J^4\,d^3\mathbf{y} = -i\int(\Pi_AC^A - \overline{\Pi}_A\overline{C}^A)d^3\mathbf{y}, \tag{4.3}$$

which generates the transformation by

$$\begin{aligned} &[i\omega Q, C^A(x)]_- = i\omega C^A(x), \\ &[i\omega Q, \overline{C}^A(x)]_- = -i\omega\overline{C}^A(x). \end{aligned} \tag{4.4}$$

The conservation of ghost charge is most easily derived by noting that J^a can be written

$$J^a = i[\Delta(A_A^a\overline{C}^A) + \bar\Delta(A_A^aC^A)], \tag{4.5}$$

so that, invoking the field equation (3.4),

$$\begin{aligned} \partial_aJ^a = i\{&\Delta(A_A^a\,\partial_a\overline{C}^A) + \bar\Delta(A_A^a\,\partial_aC^A) \\ &- \xi[\Delta(b_A\overline{C}^A) + \bar\Delta(b_AC^A)]\} = 0, \end{aligned} \tag{4.6}$$

because, owing to the definitions (2.1) and (2.2), one has separately

$$\Delta (A_A^a\, \partial_a \overline{C}^A) + \overline{\Delta} (A_A^a\, \partial_a C^A) = 0 \tag{4.7}$$

and

$$\Delta (b_A \overline{C}^A) + \overline{\Delta} (b_A C^A) = 0. \tag{4.8}$$

V. THE CONSERVATION OF INTERNAL CHARGES

Since the action (3.1) retains the invariance under the classical global gauge transformations

$$\begin{aligned} &\delta_A \psi = i g S_A \psi, \quad \delta_A \bar{\psi} = -i g \bar{\psi} S_A, \\ &\delta_A A_C^a = -g f_{ABC} A^{Ba}, \quad \delta_A b_C = -g f_{ABC} b^B, \\ &\delta_A C_C = -g f_{ABC} C^B, \quad \delta_A \overline{C}_C = -g f_{ABC} \overline{C}^B, \end{aligned} \tag{5.1}$$

with infinitesimal parameters g, there exist associated Noether currents

$$\begin{aligned} J_A^a = & -\bar{\psi}\gamma^a S_A \psi - f_{ABC} [A_b^B F^{Cab} - b^B A^{Ca} \\ & + (\partial^a \overline{C}^C - \tfrac{1}{2} g f^{CDE} \overline{C}_D A_E^a) C^B \\ & - (\partial^a C^C - \tfrac{1}{2} g f^{CDE} C_D A_E^a) \overline{C}^B], \end{aligned} \tag{5.2}$$

whose timelike components define the "internal charges"

$$\begin{aligned} Q_A = \int J_A^4\, d^3\mathbf{y} = \int [& -i\pi_\psi S_A \psi - f_{ABC} (A_a^B P^{Ca} \\ & + b^B p^C - \Pi^C C^B - \overline{\Pi}^C \overline{C}^B)] d^3\mathbf{y}, \end{aligned} \tag{5.3}$$

which generate the transformations by

$$[i g Q_A, (\cdots)]_- = \delta_A (\cdots). \tag{5.4}$$

The conservation of internal charges is most easily derived by noting that, owing to the field equations (3.3), one has the "Maxwell" equations[23]

$$g J_A^a = \partial_b F_A^{ab} + \overline{\Delta}\Delta A_A^a, \tag{5.5}$$

so that, invoking again (3.4),

$$g\, \partial_a J_A^a = \overline{\Delta}\Delta\, \partial_a A_A^a = \xi \Delta \overline{\Delta} b_A = 0 \tag{5.6}$$

by straightforward application of definitions (2.1) and (2.2).

VI. THE CONSERVATION OF BRS AND ANTI-BRS CHARGE

The Noether current associated with the invariance of action (3.1) under the global BRS transformation Δ,

$$\begin{aligned} J_\Delta^a \equiv & \frac{\partial L}{\partial(\partial_a \psi)} \Delta \psi + \frac{\partial L}{\partial(\partial_a A_b^A)} \Delta A_b^A + \frac{\partial L}{\partial(\partial_a b^A)} \Delta b^A \\ & + \frac{\partial L}{\partial(\partial_a C^A)} \Delta C^A + \frac{\partial L}{\partial(\partial_a \overline{C}^A)} \Delta \overline{C}^A \\ = & g \bar{\psi} \gamma^a C^A S_A \psi + F_A^{ab} D_b^{AB} C_B + b^A D_{AB}^a C^B \\ & + \tfrac{1}{2} g f_{ABC} [(\partial^a \overline{C}^A) C^B C^C - (\partial^a C^A) \overline{C}^B C^C] \\ & + (g^2/4) f^A{}_{BC} f^C{}_{DE} A_A^a \overline{C}^B C^D C^E, \end{aligned} \tag{6.1}$$

enables one to define a nilpotent "BRS charge"

$$\begin{aligned} Q_\Delta = \int J_\Delta^4\, d^3\mathbf{y} = \int [& i g \pi_\psi C^A S_A \psi \\ & + P_A^a D_a^{AB} C_B + b_A \overline{\Pi}^A] \\ & + \tfrac{1}{2} g f_{ABC} (p^A b^B C^C - \Pi^A C^B C^C - \overline{\Pi}^A \overline{C}^B C^C) \\ & - (g^2/4) f_{ABC} f^C{}_{DE}\, p^A C^B \overline{C}^D C^E] d^3\mathbf{y}, \end{aligned} \tag{6.2}$$

which generates the transformations (2.1) through the scheme[24]

$$[iQ_\Delta, (\cdots)]_\mp = \Delta (\cdots) \tag{6.3}$$

with (commutator, anticommutator) if $(\cdots)$ stands for a (boson, fermion) field variable. Similarly, invariance of (3.1) under the anti-BRS transformation $\overline{\Delta}$ yields the Noether current

$$\begin{aligned} J_{\overline{\Delta}}^a = & g \bar{\psi} \gamma^a \overline{C}^A S_A \psi + F_A^{ab} D_b^{AB} \overline{C}_B + b^A D_{AB}^a \overline{C}^B \\ & + \tfrac{1}{2} g f_{ABC} [(\partial^a \overline{C}^A) \overline{C}^B C^C - (\partial^a C^A) \overline{C}^B \overline{C}^C] \\ & - (g^2/4) f^A{}_{BC} f^C{}_{DE} A_A^a C^B \overline{C}^D \overline{C}^E \end{aligned} \tag{6.4}$$

and the nilpotent "anti-BRS charge"

$$\begin{aligned} Q_{\overline{\Delta}} = \int [& i g \pi_\psi \overline{C}^A S_A \psi + P_A^a D_a^{AB} \overline{C}_B - b_A \Pi^A \\ & + \tfrac{1}{2} g f_{ABC} (p^A b^B \overline{C}^C - \overline{\Pi}^A \overline{C}^B \overline{C}^C - \Pi^A \overline{C}^B C^C) \\ & + (g^2/4) f_{ABC} f^C{}_{DE}\, p^A \overline{C}^B C^D \overline{C}^E] d^3\mathbf{y} \end{aligned} \tag{6.5}$$

which generates the transformation (2.2) by

$$[iQ_{\overline{\Delta}}, (\cdots)]_\mp = \overline{\Delta} (\cdots). \tag{6.6}$$

The field equations (3.3), used in conjunction with the identity

$$(D_b^{AB} F_B^{ab}) C_A + F_A^{ab} D_b^{AB} C_B = \partial_b (F_A^{ab} C^A) \tag{6.7}$$

show that, up to a divergence, the BRS current J_Δ^a is the BRS transform of the ghost current iJ^a defined in (4.5),

$$J_\Delta^a = i\Delta J^a + \partial_b (F_A^{ab} C^A). \tag{6.8}$$

Similarly one finds

$$J_{\overline{\Delta}}^a = -i\overline{\Delta} J^a + \partial_b (F_A^{ab} \overline{C}^A). \tag{6.9}$$

Thus, the conservation of BRS and anti-BRS charge,

$$\partial_a J_\Delta^a = \partial_a J_{\overline{\Delta}}^a = 0, \tag{6.10}$$

is a consequence of the conservation of ghost charge $\partial_a J^a = 0$.

VII. THE TRANSFORMATION OF NAKANISHI

The transformation of Nakanishi,

$$\Delta T^{\cdot\cdot a \cdot\cdot}{}_{\cdot\cdot b \cdot\cdot} = T^{\cdot\cdot c \cdot\cdot}{}_{\cdot\cdot b \cdot\cdot} C_c{}^a - C_b{}^c T^{\cdot\cdot a \cdot\cdot}{}_{\cdot\cdot c \cdot\cdot}, \tag{7.1}$$

which is the BRS analog of the general infinitesimal coordinate transformation $\delta T = T'(x') - T(x)$ applied to a tensor, can be augmented by the anti-BRS transformation

$$\overline{\Delta} T^{\cdot\cdot a \cdot\cdot}{}_{\cdot\cdot b \cdot\cdot} = T^{\cdot\cdot c \cdot\cdot}{}_{\cdot\cdot b \cdot\cdot} \overline{C}_c{}^a - \overline{C}_b{}^c T^{\cdot\cdot a \cdot\cdot}{}_{\cdot\cdot c \cdot\cdot}, \tag{7.2}$$

and cast in conjugation symmetric form by writing

$$\begin{aligned} &\Delta C_a{}^b = -C_a{}^c C_c{}^b, \\ &\Delta \overline{C}_a{}^b = b_a{}^b - \tfrac{1}{2}(\overline{C}_a{}^c C_c{}^b - \overline{C}_c{}^b C_a{}^c), \\ &\Delta b_a{}^b = \tfrac{1}{2}(b_a{}^c C_c{}^b - b_c{}^b C_a{}^c) + \tfrac{1}{4}(\overline{C}_a{}^d C_d{}^c C_c{}^b + \overline{C}_c{}^b C_d{}^c C_a{}^d). \end{aligned} \tag{7.3}$$

and

$$\begin{aligned} &\overline{\Delta} \overline{C}_a{}^b = -\overline{C}_a{}^c \overline{C}_c{}^b, \\ &\overline{\Delta} C_a{}^b = -b_a{}^b - \tfrac{1}{2}(\overline{C}_a{}^c C_c{}^b - \overline{C}_c{}^b C_a{}^c), \\ &\overline{\Delta} b_a{}^b = \tfrac{1}{2}(b_a{}^c \overline{C}_c{}^b - b_c{}^b \overline{C}_a{}^c) \\ &\qquad + \tfrac{1}{4}(\overline{C}_a{}^d \overline{C}_d{}^c C_c{}^b + \overline{C}_c{}^b \overline{C}_d{}^c C_a{}^d). \end{aligned} \tag{7.4}$$

Once again, the transformation formulas for the Hermitian auxiliary fields $b_a{}^b$ must contain the terms cubic in the anticommuting ghost fields $C_a{}^b$ and their Hermitian conjugates $\overline{C}_a{}^b$ to ensure validity of the nilpotency relations (2.4). The Nakanishi operations Δ and $\overline{\Delta}$ do not commute with the operations of differentiation,

$$\Delta\partial_b - \partial_b\Delta = -C_b{}^c\partial_c, \quad \overline{\Delta}\partial_b - \partial_b\overline{\Delta} = -\overline{C}_b{}^c\partial_c. \tag{7.5}$$

These commutation relations are most easily derived, using the transformation property of the affinities

$$\Delta\Gamma^c_{ab} = \Gamma^d_{ab}C_d{}^c - C_a{}^d\Gamma^c_{db} - C_b{}^d\Gamma^c_{ad} - \partial_b C_a{}^c, \tag{7.6}$$

by requiring the covariant derivative $\nabla_b T_a = \partial_b T_a - \Gamma^c_{ab}T_c$ to transform as a tensor under Δ and $\overline{\Delta}$ as prescribed in (7.1) and (7.2).

The ghost fields $C_a{}^b$ themselves are not tensors. The first equation (7.3), compared with (7.1), indicates the $C_a{}^b$ transform as a set of four ($b = 1, 2, 3, 4$) covariant vectors $C_a{}^{(b)}$. The required invariance[25] of $\partial_a\,\partial_b - \partial_b\,\partial_a = 0$ under both Δ and $\overline{\Delta}$ subjects the ghost fields to the constraints

$$m^c_{ab} \equiv \partial_a C_b{}^c - \partial_b C_a{}^c = 0, \quad \overline{m}^c_{ab} \equiv \partial_a \overline{C}_b{}^c - \partial_b \overline{C}_a{}^c = 0, \tag{7.7}$$

which can be satisfied by writing

$$C_a{}^b = \partial_a C^b, \quad \overline{C}_a{}^b = \partial_a \overline{C}^b, \tag{7.8}$$

with four anticommuting scalar ghost fields C^b and their Hermitian conjugates $\overline{C}^b$. If one introduces similarly four Hermitian auxiliary fields b^b by writing

$$b_a{}^b = \partial_a b^b + \tfrac{1}{2}(\overline{C}_a{}^c C_c{}^b + \overline{C}_c{}^b C_a{}^c), \tag{7.9}$$

then the transformation formulas (7.3) and (7.4) are obtainable from

$$\begin{aligned} &\Delta C^a = 0, \quad \Delta\overline{C}^a = b_b, \quad \Delta b^a = 0, \\ &\overline{\Delta C}^a = 0, \quad \overline{\Delta} C^a = -b^a, \quad \overline{\Delta} b^a = 0, \end{aligned} \tag{7.10}$$

keeping in mind that the definitions (7.9) subject the auxiliary fields $b_a{}^b$ to the constraints

$$\begin{aligned} n^c_{ab} \equiv\; &\partial_a\,[\,b_b{}^c - \tfrac{1}{2}(\overline{C}_b{}^d C_d{}^c + \overline{C}_d{}^c C_b{}^d)] \\ &- \partial_b\,[\,b_a{}^c - \tfrac{1}{2}(\overline{C}_a{}^d C_d{}^c + \overline{C}_d{}^c C_a{}^d)] = 0. \end{aligned} \tag{7.11}$$

VIII. THE TRANSFORMATION OF DELBOURGO AND MEDRANO

The operation of Delbourgo and Medrano,

$$\Lambda \equiv \Delta - C^a\,\partial_a, \tag{8.1}$$

which is the BRS analog of the Lie derivative $\lambda T = T'(x) - T(x)$ applied to a tensor, commutes with differentiation operations, can also be augmented by the anti-BRS transformation

$$\overline{\Lambda} \equiv \overline{\Delta} - \overline{C}^a\,\partial_a, \tag{8.2}$$

and case in conjugation symmetric form by writing

$$\begin{aligned} &\Lambda C^a = -C^b\,\partial_b C^a, \quad \Lambda\overline{C}^a = b^a - C^b\,\partial_b\overline{C}^a, \\ &\Lambda b^a = -C^b\,\partial_b b^a, \quad \overline{\Lambda C}^a = -\overline{C}^b\,\partial_b\overline{C}^a, \\ &\overline{\Lambda} C^a = -b^a - \overline{C}^b\,\partial_b C^a, \quad \overline{\Lambda} b^a = -\overline{C}^b\,\partial_b b^a, \end{aligned} \tag{8.3}$$

without loss of the nilpotency properties

$$\Lambda\Lambda = \overline{\Lambda\Lambda} = \Lambda\overline{\Lambda} + \overline{\Lambda}\Lambda = 0. \tag{8.4}$$

Possible candidates for the ghost charge conserving gauge-fixing action, of scale dimension -2 and linear in the Goldberg[26] variable $\tilde{g}^{ab} = g^{1/2}g^{ab}$, which must be added to the classical action $L_{\mathrm{E}} = g^{1/2}R$ of Einstein gravity for the purpose of canonical quantization, are $\tilde{g}^{ab}\,\partial_a\,\partial_b$, $\tilde{g}^{ab}(\partial_a\overline{C}^c)(\partial_b C_c)$, etc. The requirement of invariance under both Λ and $\overline{\Lambda}$ selects uniquely the linear combination

$$L' = \tilde{g}^{ab}\,[(\partial_a b_b) - (\partial_a\overline{C}^c)(\partial_b C_c)] \tag{8.5}$$

which transforms as $\Lambda L' = -\partial_a(C^a L')$ and $\overline{\Lambda} L' = -\partial_a(\overline{C}^a L')$, so that the corresponding action integral is indeed invariant under both Λ and $\overline{\Lambda}$. To this may be added, as in the Yang–Mills case, the manifestly invariant action $\overline{\Lambda}\Lambda\overline{C}^a C_a$ with a conventional gauge fixing parameter ξ, resulting in the total action

$$L = L_{\mathrm{E}} + L' - \tfrac{1}{2}\xi\overline{\Lambda}\Lambda\overline{C}^a C_a, \tag{8.6}$$

which has the desired symmetry properties.

However, the resulting conjugation symmetric version of quantized Einstein gravity suffers from the well-known proliferation of infinities which are not amenable to renormalization[27] and will, therefore, not be pursued here.

IX. WEYL GRAVITY

Of current interest[28] is Weyl gravity,[29] characterized by the action function[30]

$$L_{\mathrm{W}} = g^{1/2}(\alpha R^2 + \beta R^{ab}R_{ab}), \tag{9.1}$$

because it may be renormalizable,[31] and because there are reasons to believe that Einstein gravity may emerge from it through a low energy effective action induced by quantum fluctuations, or by some process of spontaneous symmetry breaking.[32] Owing to the well-known topological invariant formed with $g^{1/2}(R^2 - 4R^{ab}R_{ab} + R^{abcd}R_{abcd})$ the case $\beta = -4\alpha$ amounts to using the square of the Riemann tensor as action, and the case $\beta = -3\alpha$ yields the conformally invariant theory flowing from the action formed with the square of the Weyl tensor.

The close analogy[33] of L_{W} with the classical Yang–Mills action suggests exploiting this analogy by considering the corresponding gauge-fixing and ghost action

$$L_{gf} = \tfrac{1}{2}\overline{\Lambda}\Lambda\,[\,\tilde{g}^{ab}\Gamma^c_{ad}\Gamma^d_{bc} - \xi\,g^{1/2}(\partial_a\overline{C}^b)(\partial_b C^a)], \tag{9.2}$$

which is manifestly invariant under both Λ and $\overline{\Lambda}$, and has the correct scale dimension -4. If $C_a{}^b$, $\overline{C}_a{}^b$, $b_a{}^b$ are taken as independent variables, the constraints (7.7) and (7.11) must be taken into account by adding the constraining action

$$\begin{aligned} L_c &= \tfrac{1}{2}(\overline{M}{}^{ab}_c m^c_{ab} + \overline{m}^c_{ab}M^{ab}_c + N^{ab}_c n^c_{ab}) \\ &= -(\partial_a\overline{M}{}^{ab}_c)C_b{}^c - \overline{C}_b{}^c(\partial_a M^{ab}_c) \\ &\quad - (\partial_a N^{ab}_c)[\,b_b{}^c - \tfrac{1}{2}(\overline{C}_b{}^d C_d{}^c + \overline{C}_d{}^c C_b{}^d)] + \text{div.} \end{aligned} \tag{9.3}$$

(where div. stands for divergence) with anticommuting Lagrangian multiplier fields $\overline{M}{}^{ab}_c = -\overline{M}{}^{ba}_c$, $M^{ab}_c = -M^{ba}_c$, and commuting fields $N^{ab}_c = -N^{ba}_c$ transforming so that the action integral formed with L_c retains the invariance under both Λ and $\overline{\Lambda}$.

Alternatively, if C^a, $\overline{C}^a$, and b^a are taken as independent variables, one pays the price for the absence of L_c in form of higher-order field equations, but which turn out to be integrable at least once immediately. Thus, the properly symmetrized expression obtained from (9.2) upon evaluation of the operation $\bar{A}A$

$$\begin{aligned} L_{\mathrm{Gf}} = A^{bc}_{d}[\partial_b\,\partial_c b^d &+ (\partial_b\,\partial_c\bar{C}^a)(\partial_a C^d) \\ &+ (\partial_a\bar{C}^d)(\partial_b\,\partial_c C^a)] \\ &+ (\partial_b\,\partial_c\bar{C}^{(d})\,\tilde{g}^{a)(b}(\partial_a\,\partial_d C^{c)}) \\ &+ \tfrac{1}{2}\xi\, g^{1/2}\{(\partial_a b^b)(\partial_b b^a) \\ &+ (\partial_a b^b)[(\partial_c\bar{C}^a)(\partial_b C^c) \\ &+ (\partial_b\bar{C}^c)(\partial_c C^a)]\} + \text{div.}, \end{aligned} \qquad (9.4)$$

yields, upon variation with respect to b^a, the analog of the gauge-fixing equation (3.4),

$$\partial_b(\partial_c A^{bc}_a - \xi\, g^{1/2} b_a{}^b) = 0, \quad \text{with}\, A^{bc}_a = \tilde{g}^{d(b}\Gamma^{c)}_{da}, \qquad (9.5)$$

and variation with respect to $\bar{C}^a$ (or C^a) yields the analog of the ghost field equation (3.6) [or (3.5)] which, after some calculation, may be cast in the form

$$A\,\partial_b(\partial_c A^{bc}_a - \xi\, g^{1/2} b_a{}^b) = 0 \qquad (9.6)$$

(or its conjugate) to exhibit the fact that it is the A (or $\bar{A}$) transform of (9.5). The gauge-fixing equation (9.5) integrated once reads

$$\partial_c A^{bc}_a - \xi\, g^{1/2} b_a{}^b = \partial_c Q^{bc}_a \qquad (9.7)$$

with an antisymmetric, but otherwise arbitrary, field $Q^{bc}_a = -Q^{cb}_a$. The field equation (9.6) integrated once,

$$A(\partial_c A^{bc}_a - \xi\, g^{1/2} b_a{}^b) = \partial_c P^{bc}_a \qquad (9.8)$$

with an anticommuting and antisymmetric field $P^{bc}_a = -P^{cb}_a$, will remain the A transform of (9.7) provided Q transforms as

$$A\,\partial_c Q^{bc}_a = \partial_c P^{bc}_a. \qquad (9.9)$$

The nilpotency of A requires then that P transform as

$$A\,\partial_c P^{bc}_a = 0. \qquad (9.10)$$

On the other hand, if $C_a{}^b$, $\bar{C}_a{}^b$, $b_a{}^b$ are the independent variables, the action function $L_{\mathrm{Gf}} + L_c$ yields, upon variation with respect to $b_b{}^a$, the gauge-fixing equation

$$\partial_c A^{bc}_a - \xi\, g^{1/2} b_a{}^b = \partial_c N^{bc}_a, \qquad (9.11)$$

and the field equation obtained upon variation with respect to $\bar{C}_b{}^a$, again after some calculation, may be cast in the form

$$\begin{aligned} &A(\partial_c A^{bc}_a - \xi\, g^{1/2} b_a{}^b) \\ &\quad = \partial_c[(\partial_d N^{cd}_a)C^b - (\partial_d N^{bd}_a)C^c - M^{bc}_a]. \end{aligned} \qquad (9.12)$$

Therefore, Eq. (9.12) is the A transform of (9.11) provided

$$A\,\partial_c N^{bc}_a = +\partial_c[(\partial_d N^{cd}_a)C^b - (\partial_d N^{bd}_a)C^c - M^{bc}_a]. \qquad (9.13)$$

The requirement of nilpotency, $A^2\partial_c N^{bc}_a = 0$, furnishes now the transformation formula

$$A\,\partial_c M^{bc}_a = -\partial_c[(\partial_d M^{cd}_a)C^b - (\partial_d M^{bd}_a)C^c] \qquad (9.14)$$

which satisfies the nilpotency $A^2\partial_c M^{bc}_a = 0$ identically. From the conjugate of Eq. (9.11) follows that the A transform of $\bar{M}^{bc}_a$ is also of the form

$$A\,\partial_c\bar{M}^{bc}_a = \partial_c X^{bc}_a \qquad (9.15)$$

with

$$X^{bc}_a = -X^{cb}_a. \qquad (9.16)$$

These transformation formulas (and their conjugates) are now sufficient to demonstrate, by a straightforward calculation, the invariance under A (and $\bar{A}$) of the action integral formed with L_c.

To obtain the remaining gravitational field equations one may treat the connections Γ^a_{bc} and the metric g_{ab} as initially independent variables, and then impose the metric compatibility

$$\partial_a g_{bc} = \Gamma^d_{ba} g_{dc} + \Gamma^d_{ca} g_{db} \qquad (9.17)$$

as a constraint, because the derivative

$$\begin{aligned} \frac{\partial L_W}{\partial(\partial_b\Gamma^d_{ca})} &= 2g^{1/2}[\alpha R(g^{ca}\delta^b_d - g^{b(c}\delta^{a)}_d) \\ &\quad + \beta(R^{ca}\delta^b_d - R^{b(c}\delta^{a)}_d)] \equiv F_d{}^{cab} \end{aligned} \qquad (9.18)$$

may be looked upon as a field tensor density satisfying the antisymmetry

$$F_d{}^{(cab)} = 0, \qquad (9.19)$$

enabling one to cast the field equations $(\delta L/\delta\Gamma^d_{ca}) = 0$ in the form

$$\partial_b F_d{}^{cab} = \frac{\partial L}{\partial\Gamma^d_{ca}}. \qquad (9.20)$$

Within the context of Weyl gravity this definition of the field tensor has exactly the same physical meaning as the corresponding field tensor $F_A{}^{ab}$ of Yang–Mills theories provided

$$\begin{aligned} &g^{-1/2}F_d{}^{cab}F^d{}_{cab} \\ &\quad = \text{const}\, g^{1/2}[(\alpha/\beta)(4\alpha + 2\beta)R^2 + \beta R^{ab}R_{ab}] \end{aligned} \qquad (9.21)$$

is proportional to L_W. This condition singles out the special case $\beta = -4\alpha$. Thus the square of the Riemann tensor furnishes the only *exact* analog of the classical Yang–Mills action.

Although it is formally possible to derive a "field tensor" density in the case of Einstein gravity from $L_{\mathrm{E}} = g^{1/2}R$ by the prescription

$$\text{“}F\text{”}_d{}^{cab} = \frac{\partial L_{\mathrm{E}}}{\partial(\partial_b\Gamma^d_{ca})} = g^{1/2}(g^{ca}\delta^b_d - g^{b(c}\delta^{a)}_d), \qquad (9.22)$$

such a definition does not have the same physical significance in this case because $g^{-1/2}\text{“}F\text{”}_d{}^{cab}\,\text{“}F\text{”}^d{}_{cab} = \text{const}\, g^{1/2}$, and because the metric compatibility (9.17) turns the field equations $(\delta L_{\mathrm{E}}/\delta\Gamma^d_{ca}) = 0$, written as

$$\begin{aligned} \partial_b\text{“}F\text{”}_d{}^{cab} &= \frac{\partial L_{\mathrm{E}}}{\partial\Gamma^d_{ca}} \\ &= g^{1/2}(g^{ef}\Gamma^{(c}_{ef}\delta^{a)}_d + g^{ca}\Gamma^e_{ed} \\ &\quad - g^{ec}\Gamma^a_{ed} - g^{ea}\Gamma^c_{ed}), \end{aligned} \qquad (9.23)$$

into identities, whereas the physical content of the theory resides solely in Einsteins's field equations $(\delta L_{\mathrm{E}}/\delta g_{ab}) = g^{1/2}(\frac{1}{2}R\, g^{ab} - R^{ab}) = 0$.

ACKNOWLEDGMENT

This work has been supported by the award of an Izaak Walton Killam Memorial Senior Fellowship.

[1]F. London, Z. Phys. **42**, 375 (1927).
[2]C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[3]R. P. Feynman, Acta Phys. Pol. **24**, 697 (1963).
[4]L. D. Faddeev and V. Popov, Phys. Lett. B **25**, 29 (1967).
[5]Although it is often stated [see, e.g., N. Nakanishi, Progr. Theor. Phys. **35**, 1111 (1966); J. C. Taylor, CERN Rep. **83-05**, 123 (1983)] that in Abelian gauge theories such as QED, ghosts are not needed, J. Ambjørn and R. Hughes, Preprint CALT-68-944, 1982, have demonstrated that even in QED inclusion of ghosts is necessary in noncovariant situations involving static boundary conditions.
[6]C. Becchi, A. Rouet, and R. Stora, Ann. Phys. **98**, 287 (1976).
[7]C. Curci and R. Ferrari, Nuovo Cimento **32**, 151 (1976); Phys. Lett. B **63**, 91 (1976).
[8]I. Ojima, Progr. Theor. Phys. **64**, 625 (1980).
[9]The capital superscript is an internal symmetry label ($A = 1, ..., N$), and the lower case subscript is a Minkowski label ($a = 1, ..., 4$)]. The metric $\delta_{ab} = \text{diag}(-1, \ -1, \ -1, \ +1)$ and the Levi–Civita density $\epsilon^{1234} = -\epsilon_{1234} = -1$ are used throughout this paper.
[10]L. Beaulieu and J. Thierry-Meg, Nucl. Phys. B **197**, 477 (1982).
[11]Reference 8, Eqs. (2.5) and (2.6).
[12]See the discussion in T. Kugo and I. Ojima, Progr. Theor. Phys. Suppl. **66**, Chap. I (1979).
[13]Reference 10, Eqs. (2.1) and (2.6).
[14]T. Kugo and S. Uehara, Nucl. Phys. B **197**, 378 (1982).
[15]Reference 10, Sec. 7.
[16]T. Kugo and I. Ojima, Phys. Lett. B. **73**, 459 (1978).
[17]The existence of these generators was assumed by C. Curci and R. Ferrari, Nuovo Cimento A **35**, 273 (1976).
[18]N. Nakanishi, Progr. Theor. Phys. **59**, 972 (1978).
[19]R. Delbourgo and M. R. Medrano, Nucl. Phys. B **110**, 467 (1976).
[20]K. Nishijima and M. Okawa, Progr. Theor. Phys. **60**, 272 (1978).
[21]After submission of this paper it was pointed out to the author that Eq. (2.1) is identical in content with Eq. (3.15) of J. Thierry-Meg and Y. Ne'eman, Nuovo Cimento A **71**, 104 (1982).
[22]In the work of Kugo and Ojima (Ref. 12, p. 17), which employs ghosts that are not Hermitian conjugates, the conservation of ghost charge is attributed to an invariance under "ghost scale" transformations which appear to be at variance with assigning both ghost and antighost fields the same canonical scale dimension.
[23]I. Ojima, Nucl. Phys. B **143**, 340 (1978).
[24]Owing to the nilpotency of Δ one may dispense with an infinitesimal parameter.
[25]Y. M. Cho, Phys. Rev. D **14**, 2521 (1976).
[26]J. N. Goldberg, Phys. Rev. **111**, 315 (1958).
[27]B. S. DeWitt, Phys. Rev. **162**, 1239 (1967).
[28]N. H. Barth and S. M. Christensen, Phys. Rev. D **28**, 1876 (1983).
[29]H. Weyl, Ann. Phys. Leipzig **59**, 101 (1919).
[30]W. Pauli, Verh. Dtsch. Phys. Ges. **21**, 742 (1919). In Pauli's paper only the case $\beta = -4\alpha$ is considered. In this paper any theory flowing from the general linear combination (9.1) will be called "Weyl gravity."
[31]K. R. Stelle, Phys. Rev. D **16**, 953 (1977).
[32]S. Adler, Rev. Mod. Phys. **54**, 729 (1982); A. Zee, Phys. Lett. B **109**, 183 (1982); Ann Phys. **151**, 431 (1983).
[33]H. Kikugawa and J. Kikugawa, Progr. Theor. Phys. **70**, 625 (1983).

Lagrangian dynamics on higher-dimensional spaces with applications to Kaluza–Klein theories[a)]

C-C. Chiang, S-C. Lee, and G. Marmo[b)]
Institute of Physics, Academia Sinica, Taipei, Taiwan, Republic of China

(Received 30 July 1984; accepted for publication 30 November 1984)

We review and apply the method of Lagrangian dynamics to particle motion in higher-dimensional spaces. We discuss in detail the case of a Kaluza–Klein theory with coset spaces as fiber. While the total metric we use in general need not allow for Killing vectors, we require that the restriction to the fiber does. We find that for general Jordan–Thiry scalar fields, the geodesic motion in total space is not describable in terms of particle motion in the base manifold with the usual internal charges. The cases when this is possible are discussed. We also consider the geodesic motion in the Sorkin–Gross–Perry Kaluza–Klein monopoles. We find all the conserved quantities and the equations can be integrated by quadrature.

I. INTRODUCTION

The success of special relativity provides a unified picture of electric and magnetic fields, and the quest for unification of gravity and electromagnetism started very soon with Weyl[1] after gravity was relativized. Since then there has been constant interest in setting a mathematical framework and presenting new perspectives for the unification schemes.[2–20] The recent success of the electroweak unification and the attempt at grand unification has triggered a renewed interest in the generalization of the original suggestion by Kaluza to unify gravitation and electromagnetism.[3] In this approach the carrier space for the unification is usually a fiber bundle over space-time and fibers are taken to be diffeomorphic to a group G (see Refs. 21–23) or a homogeneous space G/H (see Refs. 24 and 25).

There are mainly two approaches to the generalized Kaluza–Klein theories. The first generalizes the old idea of cylindricity[10] by requiring the metric tensor of the total space to admit the group G as a group of isometries. The second requires G to be a group of isometries for the metric tensor restricted to tangent vectors along the fiber.[24–26] In this context we would like to study the geodesic motions on bundles over space-time with homogeneous space as fibers. Previous works[27] on geodesic motions considered only the case when the fiber is diffeomorphic to a group G. These will emerge as particular cases of ours.

Usually, in the existing literature, the bundle space is assumed to be a product bundle. Apart from technicalities, the relevant difference between product bundles and nontrivial bundles relies on the fact that, in principle bundles, for instance, the group which can act is $G\times G$ (left and right action) while in nontrivial ones it usually reduces to a subgroup of $G\times G$. The fact that all grand unified theories predict the existence of monopoles on the one hand and the existence of monopole solutions in Kaluza–Klein theories on the other hand[28] suggests that the situation where the bundle structure is nontrivial must be investigated. Such investigation obviously calls for a coordinate-free treatment if it is possible. In this paper, we shall not assume our bundle to be trivial, and we shall use the intrinsic differential calculus—Lie derivatives and exterior derivatives. There are now quite a few books where classical mechanics has been dealt with in this language.[29–32] With a few exceptions, the emphasis is usually on symplectic mechanics on the phase space (cotangent bundle). We are here mainly interested in the Lagrangian formalism. For this reason, and because our presentation might be useful from the point of view of doing computations, we will indulge a little bit in presenting Lagrangian mechanics on Lie groups and coset spaces.

With more conventional language, motions of classical particles with internal structure have been thoroughly investigated (see for instance Balachandran *et al.*[33] and references therein). A comprehensive treatment of the five-dimensional case, which has been very much inspiring for us, was given by Lichnerowicz.[34] Classical motions of particles with internal structure have also been investigated in the framework of presymplectic mechanics and degenerate Lagrangian formalisms.[35] The spirit is, however, different from the Kaluza–Klein approach.

The paper is organized as follows. The next section deals with Lagrangian dynamics on Lie groups. It is used mainly to fix the notations. In Sec. III, we consider the geodesic motion on Lie groups. To obtain the invariants of the motion and because of their importance to quantization and other applications, we briefly discuss the Poisson brackets and the momentum map in Sec. IV. In Sec. V we consider geodesic motion on coset spaces and in Sec. VI we deal with geodesic motion on fiber bundles with homogeneous fibers. Finally, we treat somewhat in detail the geodesic motion in the recently discovered Abelian Kaluza–Klein monopoles[28] in Sec. VII. Some remarks on our results are given in Sec. VIII.

II. LAGRANGIAN DYNAMICS ON A LIE GROUP

We assume that the configuration space of our dynamical system is a Lie group G. For instance, this is the case for the motion of a rigid body [$G = \mathrm{SO}(3)$] and for the motion of the "internal degrees of freedom" of a classical particle with internal structure.

The Lagrangian $\mathscr{L}$ is a function on the tangent bundle TG. Euler–Lagrange equations in a "natural chart" for TG, say, $(\xi^i, \dot\xi^i)$: $TU \to R^{2n}$, $U\subset G$, have the form

[a)] Contents of this paper have been reported at the Workshop on Particles and Fields at the National Taiwan University, Taipei, March 1984.

[b)] Permanent address: Istituto di Fisica Teorica, dell'Università Mostra d'Oltremare, pad. 19-80125, Napoli, Italy.

 0022-2488/85/051083-10$02.50

$$\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{\xi}^i}-\frac{\partial \mathscr{L}}{\partial \xi^i}=0, \quad i\in\{1,\ldots,n=\dim G\}. \tag{2.1}$$

A simple algebraic manipulation gives

$$\frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial \dot{\xi}^i}d\xi^i\right)-d\mathscr{L}=0. \tag{2.2}$$

Let

$$\theta_{\mathscr{L}}\equiv\frac{\partial \mathscr{L}}{\partial \dot{\xi}^i}d\xi^i, \tag{2.3}$$

and replacing d/dt by the Lie derivative L_Δ along a second-order vector field Δ (see Ref. 32),

$$\Delta=\dot{\xi}^i\frac{\partial}{\partial \xi^i}+\Delta^i\frac{\partial}{\partial \dot{\xi}^i}, \tag{2.4}$$

which describes the dynamics, we get the intrinsic formulation

$$L_\Delta\,\theta_{\mathscr{L}}-d\mathscr{L}=0. \tag{2.5}$$

Equations (2.1) are then nothing but the projection of this one-form along coordinate directions

$$\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{\xi}^i}-\frac{\partial \mathscr{L}}{\partial \xi^i}=i_{\partial/\partial\xi^i}(L_\Delta\,\theta_{\mathscr{L}}-d\mathscr{L}). \tag{2.6}$$

Notice that Eq. (2.5) is globally defined but Eqs. (2.1) are only locally defined.

A Lie group as well as any parallelizable manifold has a global basis of "directions." A particularly useful choice is a basis of left-invariant vector fields (infinitesimal generators of right translations) or right-invariant vector fields. Let us denote

$$Y_1,\ldots,Y_n;\quad \theta^1,\ldots,\theta^n$$

as a basis of left-invariant vector fields and the dual one-forms such that

$$\theta^a(Y_b)=\delta^a{}_b.$$

Correspondingly, we shall write $\hat{Y}_a$, $\hat{\theta}^a$ for the right-invariant vector fields and the dual one-forms.

To use the machinery of Lie derivatives and exterior differential calculus, it is appropriate to lift vector fields Y_a on G to vector fields on TG. An infinitesimal transformation $\delta\xi$ on G defines an infinitesimal transformation $\delta\xi$ and $\delta\dot{\xi}=(d/dt)\delta\xi$ on TG. Correspondingly, any vector field Y_a on G lifts to a vector field $\dot{Y}_a$ on TG. Vector fields Y_a and $\dot{Y}_a$ can be expressed in terms of the local coordinates

$$Y_a=Y^i_a(\xi)\frac{\partial}{\partial \xi^i}, \tag{2.7}$$

$$\dot{Y}_a=Y^i_a(\xi)\frac{\partial}{\partial \xi^i}+\frac{d}{dt}Y^i_a(\xi)\frac{\partial}{\partial \dot{\xi}^i}$$

$$=Y^i_a(\xi)\frac{\partial}{\partial \xi^i}+\partial\frac{Y^i_a}{\partial \xi^i}\dot{\xi}^i\frac{\partial}{\partial \dot{\xi}^i}. \tag{2.8}$$

From Eq. (2.8) it is easy to verify that the lift $\{\dot{Y}_a\}$ of the basis $\{Y_a\}$ satisfies the same Lie algebra as $\{Y_a\}$.

We can now project the Euler–Lagrange equation (2.5) along the direction $\dot{Y}_a$ and have

$$i_{\dot{Y}_a}L_\Delta\,\theta_{\mathscr{L}}=i_{\dot{Y}_a}d\mathscr{L}. \tag{2.9}$$

From Eqs. (2.4) and (2.8), we see that the commutator $[\Delta,\dot{Y}_a]$ for any second-order vector field Δ and any lifted vector $\dot{Y}_a$ do not have $\partial/\partial\dot{\xi}^i$ components. In particular

$$i_{[\Delta,\dot{Y}_a]}\,\theta_{\mathscr{L}}=0. \tag{2.10}$$

Using Leibnitz's rule for Lie derivatives, Eq. (2.10), and Cartan's identity

$$L_x=i_x d+d i_x, \tag{2.11}$$

we obtain from Eq. (2.9)

$$L_\Delta(i_{\dot{Y}}\,\theta_{\mathscr{L}})=L_{\dot{Y}_a}\mathscr{L}. \tag{2.12}$$

Finally, defining

$$i_{\dot{Y}_a}\theta_{\mathscr{L}}=P_a=\frac{\partial \mathscr{L}}{\partial \dot{\xi}^i}Y^i_a=P_i\,Y^i_a, \tag{2.13}$$

we can write the equations of motion in the form

$$\frac{d}{dt}P_a=L_{\dot{Y}_a}\mathscr{L}. \tag{2.14}$$

III. GEODESIC MOTIONS ON TG

With the preliminary discussion on the Lagrangian dynamics on Lie groups, we will consider a particular situation in which a Lagrangian is associated with a metric tensor g on G (see Refs. 36–39).

On the tangent bundle $TG\xrightarrow{\pi}G$, we can associate functions on TG with one-forms on G in a natural way. With $(\xi,\dot{\xi})\in T_\xi G$ and $\theta\in\mathfrak{X}^*(G)$, where $\mathfrak{X}^*(G)$ is the space of one-forms, we define

$$\dot{\theta}(\xi,\dot{\xi})\equiv i_{\dot{\xi}}\theta(\xi). \tag{3.1}$$

This can be written in the coordinate independent form

$$\dot{\theta}=i_\Delta\pi^*\theta, \tag{3.2}$$

where Δ is an arbitrary second-order vector field. To avoid heavy notations, we shall drop π^* in what follows, and with abuse of notation write

$$\dot{\theta}=i_\Delta\,\theta. \tag{3.3}$$

For the metric g, we may write, using the basis (2.7),

$$g=g_{ab}\,\theta^a\otimes\theta^b. \tag{3.4}$$

We then define the associated Lagrangian

$$\mathscr{L}_g\equiv\frac{1}{2}g_{ab}\,\dot{\theta}^a\dot{\theta}^b=\tfrac{1}{2}\pi^*g(\Delta,\Delta). \tag{3.5}$$

We will also drop π^* in front of g to simplify the notations. We can now compute $\theta_{\mathscr{L}_g}$ according to (2.3) to get

$$\theta_{\mathscr{L}_g}=g(\Delta,\cdot), \tag{3.6}$$

so that, from Eq. (2.13)

$$P_a=g(\Delta,\dot{Y}_a)=g(\Delta,Y_a)=g_{ab}\,\dot{\theta}^b. \tag{3.7}$$

The equations of motion are then

$$\frac{dP_a}{dt}=\frac{1}{2}L_{\dot{Y}_a}g(\Delta,\Delta)=\frac{1}{2}(L_{Y_a}g)(\Delta,\Delta), \tag{3.8}$$

where we have used the remark above Eq. (2.10) to get the second equality.

If the metric is right invariant, then the P_a are constants of motion. On the other hand, suppose the metric is left invariant, then it is easy to see that g_{ab} in Eq. (3.4) is constant. From (A5), we find that

$$L_{Y_a} g = -g_{ce} C^e{}_{ab} (\theta^b \otimes \theta^c + \theta^c \otimes \theta^b)$$
$$\equiv -C_{cab} (\theta^b \otimes \theta^c + \theta^c \otimes \theta^b). \tag{3.9}$$

In particular, g is bi-invariant if C_{cab} is totally antisymmetric. The equations of motion become

$$\Psi P_a = -C_{cab} \dot\theta^b \dot\theta^c = -g_{ae} C^{ceb} P_b P_c, \tag{3.10}$$

or equivalently,

$$\ddot\theta^a + g^{ae} C_{ceb} \dot\theta^b \dot\theta^c = 0, \tag{3.11}$$

from which we can read off the connection components in the (noncoordinate) basis Y_a. Note that if g is bi-invariant, P_a is also conserved.

IV. POISSON BRACKETS OF MOMENTA AND THE MOMENTUM MAP

We would like to consider the Poisson brackets of momenta[40] defined in (2.13). From (A17), we have

$$[Y_a, Y_b] = C^c{}_{ab} Y_c, \tag{4.1}$$

and we notice that the lifts $\dot Y_a$ have the same commutator structure as Y_a.

When a symplectic structure, say $\omega_{\mathscr{L}}$ (a closed nondegenerate two-form), is available on some manifold M, it is possible to define the associated Poisson brackets on functions in the following way. For any function $f \in F(TM)$, consider the vector field X_f defined by

$$i_{X_f} \omega_{\mathscr{L}} = -df. \tag{4.2}$$

The Poisson brackets of $f, g \in F(TM)$ are then defined by

$$\{f,g\} = \omega_{\mathscr{L}}(X_f, X_g). \tag{4.3}$$

The vector fields X_f are called Hamiltonian vector fields and are generators of canonical transformations.

In our case,

$$\omega_{\mathscr{L}} = d\theta_{\mathscr{L}}, \tag{4.4}$$

and is nondegenerate if

$$\left|\left| \frac{\partial^2 \mathscr{L}}{\partial \dot\xi^i \partial \dot\xi^j} \right|\right| \neq 0. \tag{4.5}$$

To define Poisson brackets for all functions on the group, we need the condition (4.5). Notice that the lifted action of G in terms of $\{\dot Y_a\}$ need not be Hamiltonian and we cannot, therefore, set

$$\{P_a, P_b\} = \omega_{\mathscr{L}}(\dot Y_a, \dot Y_b) \tag{4.6}$$

to get a Poisson bracket which satisfies the Jacobi identity. Instead, we have to define a new lifted action of G given by

$$i_{\overline{Y}_a} \omega_{\mathscr{L}} = -dP_a = -d i_{Y_a} \theta_{\mathscr{L}}. \tag{4.7}$$

(See Ref. 41.) Using (2.11), we find

$$i_{\overline{Y}_a} d\theta_{\mathscr{L}} = L_{\overline{Y}_a} \theta_{\mathscr{L}} - d i_{Y_a} \theta_{\mathscr{L}} = -d i_{Y_a} \theta_{\mathscr{L}}. \tag{4.8}$$

Because $\overline{Y}_a$ is a lifting of Y_a, we have

$$i_{\overline{Y}_a} \theta_{\mathscr{L}} = i_{Y_a} \theta_{\mathscr{L}}, \tag{4.9}$$

hence (4.8) implies

$$L_{\overline{Y}_a} \theta_{\mathscr{L}} = 0. \tag{4.10}$$

Now we have

$$L_{\overline{Y}_a}(i_{\overline{Y}_b} \theta_{\mathscr{L}}) = i_{[\overline{Y}_a, \overline{Y}_b]} \theta_{\mathscr{L}}$$
$$= i_{[\overline{Y}_a, \overline{Y}_b]} \theta_{\mathscr{L}} = C^c{}_{ab} P_c. \tag{4.11}$$

On the other hand,

$$L_{\overline{Y}_a}(i_{\overline{Y}_b} \theta_{\mathscr{L}}) = L_{\overline{Y}_a} P_b = i_{\overline{Y}_a} dP_b$$
$$= -i_{\overline{Y}_a} i_{\overline{Y}_b} \omega_{\mathscr{L}} = \{P_a, P_b\}. \tag{4.12}$$

Hence we get

$$\{P_a, P_b\} = C^c{}_{ab} P_c. \tag{4.13}$$

Let $\{t_a\}$ be a basis of the Lie algebra g_G of G, and let e^a be the dual basis of g_G^*. We define the momentum map[30] μ: $TG \to g_G^*$ by

$$\mu(\xi, \dot\xi) \equiv P_a(\xi, \dot\xi) e^a. \tag{4.14}$$

The vectors $t_a \in g_G$ can be thought of as linear functions on g_G^*. In particular, they define a coordinate system for g_G^*. On this space, we have a natural Poisson bracket given by[42–45]

$$\{f,g\}_{g_G^*} \equiv \frac{\partial f}{\partial t_a} \frac{\partial g}{\partial t_b} \{t_a, t_b\} \equiv \frac{\partial f}{\partial t_a} \frac{\partial g}{\partial t_b} C^c{}_{ab} t_c. \tag{4.15}$$

Suppose the $\{t_a\}$ have been so chosen that they correspond to Y_a. Then we have, from (4.13)–(4.15),

$$\mu^* t_a = P_a, \quad \mu^* \{t_a, t_b\}_{g_G^*} = \{P_a, P_b\}_{\omega_{\mathscr{L}}}. \tag{4.16}$$

To relate dynamical evolution on TG with dynamical evolution on g_G^*, we need the following definition.

Given a smooth map

$$\phi: M \to N, \tag{4.17}$$

we say that $X \in \mathfrak{X}(M)$ is ϕ related to $Y \in \mathfrak{X}(N)$ if

$$L_X \phi^* f = \phi^* L_Y f, \quad \forall f \in F(N). \tag{4.18}$$

By looking back at the dynamics on TG defined by $\mathscr{L}_g$ in (3.5), it is easy to see from (3.10) that the dynamical vector field Δ on TG is μ related to the dynamical vector field $\widetilde\Delta$ on g_G^* defined by

$$L_{\widetilde\Delta} t_a = -g_{ae} C^{ceb} t_b t_c, \tag{4.19}$$

which implies

$$\widetilde\Delta = -g_{ae} C^{ceb} t_b t_c \frac{\partial}{\partial t_a}. \tag{4.20}$$

When the map (4.17) is a submersion onto, we say that $X \in \mathfrak{X}(M)$ is projectable[44] if we can define a vector field $\widetilde X$ which is ϕ related to X. For the action of G on TG defined by $\{\overline{Y}_a\}$, it is easy to see from (4.11) that every $\overline{Y}_a$ is μ projectable onto $\widetilde Y_a$ with

$$\widetilde Y_a = C^c{}_{ab} t_c \frac{\partial}{\partial t_b}. \tag{4.21}$$

It follows that

$$\widetilde\Delta = t^a \widetilde Y_a. \tag{4.22}$$

Since $\widetilde Y_a$ are the generators of the adjoint action, any adjoint-invariant polynomial in g_G^* gives rise to constants of motion for Δ when pulled back to TG through μ.

V. LAGRANGIAN DYNAMICS ON *G/H*

Let H be a closed subgroup of G which acts on G on the left. Let G/H be the right coset space. Then G/H is a differentiable manifold and the natural projection π

$$\pi: G \to G/H, \quad g \to H \cdot g \tag{5.1}$$

is smooth.

As before, we write Y_a, $\hat{Y}_a$ for left- and right-invariant vector fields on G, respectively. We shall use $\bar{a}, \bar{b},\ldots$ for indices relating to the subgroup H.

By using the fact that right-invariant vector fields and left-invariant vector fields commute, we can easily show that left-invariant vector fields are π projectable onto G/H. Indeed, right-invariant vector fields $\hat{Y}_{\bar{a}}$ are infinitesimal generators for the left action of H on G inducing the projection π. In particular, functions $f \in F(G/H)$ enjoy the property

$$L_{\hat{Y}_{\bar{a}}} \pi^* f = 0. \tag{5.2}$$

Thus for any function $f \in F(G/H)$, the function $L_{Y_a} \pi^* f$ is such that

$$L_{\hat{Y}_{\bar{a}}} (L_{Y_b} \pi^* f) = 0. \tag{5.3}$$

Hence there exists $h \in F(G/H)$ such that

$$L_{Y_a} \pi^* f = \pi^* h. \tag{5.4}$$

The projected vector field $\tilde{Y}_a$ is then defined by the relation

$$L_{\tilde{Y}_a} f = h. \tag{5.5}$$

On the other hand,

$$L_{\hat{Y}_{\bar{a}}} (L_{\hat{Y}_b} \pi^* f) = C^c{}_{\bar{a}b} L_{\hat{Y}_c} \pi^* f. \tag{5.6}$$

Hence $\hat{Y}_b$ is projectable iff

$$C^i{}_{\bar{a}b} = 0, \quad \forall \bar{a}, i, \tag{5.7}$$

where i is the index for the complement of the subalgebra g_H in g_G. When (5.7) is satisfied, $\hat{Y}_b$ are infinitesimal generators of the normalizer of H in G, and their projections on G/H are G-invariant vector fields.

Let g be a nondegenerate metric on G satisfying

$$L_{\hat{Y}_{\bar{a}}} g = 0, \tag{5.8}$$

and let

$$g_{ab} \equiv g(Y_a, Y_b). \tag{5.9}$$

Then g_{ab} is projectable. Denote its inverse by g_{ab}. Then the tensor

$$K \equiv g^{ab} \, Y_a \otimes Y_b \tag{5.10}$$

is projectable with projection

$$\tilde{K} = g^{ab} \tilde{Y}_a \otimes \tilde{Y}_b. \tag{5.11}$$

Since $\tilde{K}$ is nondegenerate, we can define the "projection" of g, denoted as $\tilde{g}$, to be the inverse of $\tilde{K}$. Explicitly, let $\hat{e}^i$ be a basis of one-forms on G/H. Let

$$\tilde{g}^{ij} \equiv \tilde{K}(\hat{e}^i, \hat{e}^j), \quad \tilde{g}_{ik} \, \tilde{g}^{kj} = \delta_i{}^j. \tag{5.12}$$

Then

$$\tilde{g} = \tilde{g}_{ij} \hat{e}^i \otimes \hat{e}^j. \tag{5.13}$$

If the $\{\hat{e}^i\}$ are the basis one-forms defined in (A30), then we have, from (A28) and (5.12),

$$\tilde{g}^{ij} = g^{ab} D^i{}_a(\sigma(y)) D^j{}_b(\sigma(y)). \tag{5.14}$$

In case g is right invariant, the projection $\tilde{g}$ will be G invariant. The condition for g to be right invariant is that

$$g^{ab} D^c{}_a D^e{}_b = \text{const.} \tag{5.15}$$

This can easily be seen from (A11) and (5.10). In particular, $\tilde{g}^{ij}$, $\tilde{g}_{ij}$ are constant, as we found in (A52).

Note that if g is bi-invariant, g^{ab} itself is constant. Since g is projectable when (5.8) is satisfied, the dynamics determined by the Lagrangian defined by g, as in (3.5), are also projectable.

Instead of projecting the dynamics from G, we shall work directly in G/H, with the Lagrangian

$$\mathscr{L} = \tfrac{1}{2} \tilde{g}(\Delta, \Delta) = \tfrac{1}{2} \tilde{g}_{ij} \, \dot{\hat{e}}^i \, \dot{\hat{e}}^j, \tag{5.16}$$

where $\tilde{g}$ is assumed to be invariant under G. Since g is invariant, the momenta

$$P_a = \theta_{\mathscr{L}}(\tilde{Y}_a) = \tilde{g}(\Delta, \tilde{Y}_a) \tag{5.17}$$

are constants of motion.

To deal with the projectability of the dynamics later on, it is convenient to define the following momenta:

$$\hat{P}_i \equiv \theta_{\mathscr{L}}(\hat{\tilde{Y}}_i) = \tilde{g}(\Delta, \hat{\tilde{Y}}_i) = \tilde{g}_{ij} \, \dot{\hat{e}}^j. \tag{5.18}$$

The Euler–Lagrange equation (2.5), projected along $\hat{\tilde{Y}}_i$, becomes

$$\dot{\hat{P}}_i = \tfrac{1}{2} (L_{\hat{\tilde{Y}}_i} \tilde{g})(\Delta, \Delta). \tag{5.19}$$

Using (A40), we can easily compute the Lie derivatives in (5.19) and obtain

$$\dot{\hat{P}}_i = (-\tilde{g}_{im} C^{jmk} + \tilde{g}^{km} H_m{}^{\bar{a}} C^j{}_{\bar{a}i}) \hat{P}_j \, \hat{P}_k. \tag{5.20}$$

In case the subgroup H is trivial, $H_m{}^{\bar{a}}$ vanishes, and the index i coincides with index a. We then get back Eq. (3.10).

Since P_a are conserved, the momentum map (4.14) projects Δ onto $\tilde{\Delta} = 0$. If the normalizer N of H is nontrivial, then the group N/H acts on G/H, and we can have a nontrivial momentum map from $T(G/H)$ onto $g^*_{N/H}$. With all this preparation, we are now ready to consider geodesic motions on fiber bundles over space-time with homogeneous fibers.

VI. GEODESIC MOTIONS ON FIBER BUNDLES WITH HOMOGENEOUS FIBERS

We start with a $4+N$-dimensional space E with a metric $\bar{g}$ given in terms of local coordinates by[18,24,26]

$$\bar{g} = g_{\mu\nu}(x) dx^\mu \otimes dx^\nu + \phi_{mn}(x,y) \Sigma^m \otimes \Sigma^n, \tag{6.1}$$

where

$$\Sigma^m \equiv dy^m - \tilde{Y}_a{}^m(y) A_\mu{}^a(x) dx^\mu. \tag{6.2}$$

The $\tilde{Y}_a$ are the vector fields that generate the group action on E. With this, E has a bundle structure $\pi: E \to M$. We shall assume the fiber is diffeomorphic to G/H for some closed subgroup H of G. The $g_{\mu\nu}$ can be thought of as a metric on the base manifold, and $\tilde{g}$, defined by

$$\tilde{g} = \phi_{mn}(x,y) dy^m \otimes dy^n, \tag{6.3}$$

can be considered as a metric on the fiber at x. In earlier works by many authors, the $\tilde{Y}_a$ were usually assumed to be Killing vectors of $\bar{g}$. In what follows we assume that the $\tilde{Y}_a$ are Killing vectors of $\tilde{g}$, not for the full metric $\bar{g}$.

Although it is desirable to work with the $\tilde{Y}_a$, which are globally defined as we saw in the previous section, it is often convenient to introduce the vector fields $\hat{\tilde{Y}}_i$ and their dual form $\hat{e}^i$, which are only locally defined on the fiber. The invariant metric on the fiber can then be written as

$$\tilde{g} = \tilde{g}_{mn}(x)\hat{e}^m(y)\hat{e}^n(y). \tag{6.4}$$

The advantage of using this "moving frame" is that $\tilde{g}_{mn}$ depends only on x. The total metric can be written as

$$\bar{g} = g_{\mu\nu}(x)dx^\mu \otimes dx^\nu + \tilde{g}_{mn}(x)\theta^m \otimes \theta^n, \tag{6.5}$$

where now

$$\theta^m \equiv \hat{e}^m(y) + D^m{}_a(y)A^a(x). \tag{6.6}$$

Recall that $D^m{}_a$ is defined in terms of a local section σ: $G/H \to G$. Here we shall suppress σ and write y for $\sigma(y)$.

As usual, the Lagrangian function on TE is taken to be

$$\mathscr{L}_{\bar{g}} = \tfrac{1}{2}\bar{g}(\Delta,\Delta) = \tfrac{1}{2}g(\Delta,\Delta) + \tfrac{1}{2}\phi_{mn}\dot{\Sigma}^m\dot{\Sigma}^n, \tag{6.7}$$

and we have

$$\theta_{\mathscr{L}} = \bar{g}(\Delta,\cdot) = g(\Delta,\cdot) + \phi_{mn}\Sigma^m\dot{\Sigma}^n. \tag{6.8}$$

As before, we define

$$P_a = i_{\tilde{Y}_a}\theta_{\mathscr{L}} = \phi_{mn}\Sigma^m(\tilde{Y}_a)\Sigma^n(\Delta) = \phi_{mn}\tilde{Y}_a{}^m\dot{\Sigma}^n. \tag{6.9}$$

It is convenient to rewrite (6.7)–(6.9) in terms of the metric $\tilde{g}$:

$$2\mathscr{L}_{\bar{g}} = g(\Delta,\Delta) + \tilde{g}(\Delta',\Delta'), \tag{6.10}$$

$$P_a = \tilde{g}(\tilde{Y}_a,\Delta'), \tag{6.11}$$

$$\theta_{\mathscr{L}} = g(\Delta,\cdot) + \tilde{g}(\Delta',\cdot) - P_a A^a, \tag{6.12}$$

where

$$\Delta' \equiv \Delta - \tilde{Y}_a A^a(\Delta). \tag{6.13}$$

Using the fact that $\tilde{g}$ is invariant, we get

$$L_{\tilde{Y}_a}\tilde{g} = \tilde{g}(-[\tilde{Y}_a,\tilde{Y}_b]A^b(\Delta),\Delta'). \tag{6.14}$$

Hence from (A14) and (2.12), we get

$$\dot{P}_a = -C^c{}_{ab}A^b(\Delta)P_c = -C^c{}_{ab}A_\mu{}^b(x)P_c\dot{x}^\mu. \tag{6.15}$$

This, being "Wong's equation" for the internal momentum P_a (see Ref. 46), suggests the identification of P_a with the internal charge or isospin. It is part of the geodesic equations in the total space E.

To obtain the other geodesic equations, let us define

$$e_\mu \equiv \partial_\mu + A_\mu{}^a\tilde{Y}_a, \tag{6.16}$$

$$P_\mu = i_{\dot{e}_\mu}\theta^L = g_{\mu\nu}\dot{x}^\nu, \tag{6.17}$$

and note that

$$[e_\mu,e_\nu] = F_{\mu\nu}{}^a\tilde{Y}_a, \tag{6.18}$$

$$[e_\mu,\tilde{Y}_a] = A_\mu{}^bC^c{}_{ba}\tilde{Y}_c, \tag{6.19}$$

$$[e_\mu,\hat{\tilde{Y}}_i] = A_\mu{}^a\omega_a{}^{\bar{a}}C^j{}_{\bar{a}i}\hat{\tilde{Y}}_j, \tag{6.20}$$

$$\bar{g}(e_\mu,e_\nu) = g_{\mu\nu}(x), \tag{6.21a}$$

$$\bar{g}(e_\mu,Y_a) = 0, \tag{6.21b}$$

where

$$F_{\mu\nu}{}^a = \partial_\mu A_\nu{}^a - \partial_\nu A_\mu{}^a + C^a{}_{bc}A_\mu{}^bA_\nu{}^c. \tag{6.22}$$

From (6.20), we find

$$L_{e_\mu}\hat{e}^m = -A_\mu{}^a\omega_a{}^{\bar{a}}C^m{}_{\bar{a}i}\hat{e}^i - D^m{}_a\,\partial_\nu A_\mu{}^a\,dx^\nu. \tag{6.23}$$

Hence, using (6.4) we get

$$\begin{aligned} L_{e_\mu}\tilde{g} &= \partial_\mu\tilde{g}_{mn} - A_\mu{}^a\omega_a{}^{\bar{a}}(\tilde{g}_{in}C^i{}_{\bar{a}m} + \tilde{g}_{im}C^i{}_{\bar{a}n})\hat{e}^m\otimes\hat{e}^n \\ &\quad - \tilde{g}_{mn}D^m{}_a\,\partial_\nu A_\mu{}^a(dx^\nu\otimes\hat{e}^n + \hat{e}^n\otimes dx^\nu) \\ &= \partial_\mu\tilde{g}_{mn}\hat{e}^m\otimes\hat{e}^n + \partial_\nu A_\mu{}^a[dx^\nu\otimes\tilde{g}(\tilde{Y}_a,\cdot) \\ &\quad + \tilde{g}(\tilde{Y}_a,\cdot)\otimes dx^\nu], \end{aligned} \tag{6.24}$$

where we used (A51) in obtaining the second equality.

Now we can compute the Lie derivative of $\mathscr{L}_{\bar{g}}$ along $\dot{e}_\mu$. We get

$$L_{\dot{e}_\mu}\mathscr{L}_{\bar{g}} = \tfrac{1}{2}g_{\nu\lambda,\mu}\dot{x}^\nu\dot{x}^\lambda - F_{\mu\nu}{}^a\dot{x}^\nu P_a - \tfrac{1}{2}\partial_\mu\tilde{g}^{mn}\hat{P}_m\hat{P}_n, \tag{6.25}$$

where we have used

$$\hat{e}^m(\Delta') = \tilde{g}^{mn}\hat{P}_n, \quad \hat{P}_n \equiv -P_a(D^{-1})^a{}_n = \tilde{g}(\hat{\tilde{Y}}_n,\Delta'). \tag{6.26}$$

Combining (6.17) and (6.25), we get the geodesic equations

$$\ddot{x}^\mu + \Gamma_{\nu\lambda}{}^\mu\dot{x}^\nu\dot{x}^\lambda = P_a F_{\nu\lambda}{}^a\dot{x}^\nu g^{\lambda\mu} - \tfrac{1}{2}\partial^\mu\tilde{g}^{mn}\hat{P}_m\hat{P}_n, \tag{6.27}$$

where

$$\Gamma_{\nu\lambda}{}^\mu = \tfrac{1}{2}g^{\mu\sigma}(g_{\sigma\nu,\lambda} + g_{\sigma\lambda,\nu} - g_{\nu\lambda,\sigma}). \tag{6.28}$$

It is also easy to get the equations of motion for $\hat{P}_i$. We find

$$\begin{aligned}\dot{\hat{P}}_i &= A_\mu{}^a\omega_a{}^{\bar{a}}C^j{}_{\bar{a}i}\hat{P}_j\dot{x}^\mu \\ &\quad + [-\tilde{g}_{im}C^{jmk} + \tilde{g}^{km}H_m{}^{\bar{a}}C^j{}_{\bar{a}i}]\hat{P}_j\hat{P}_k. \end{aligned} \tag{6.29}$$

This should be compared with Eq. (5.20). Note that C^{ijk}, in general, depends on x since indices are raised and lowered, respectively, by $\tilde{g}^{ij}$ and $\tilde{g}_{ij}$, which depend on x.

Up to now, our treatment has been quite general. We have not assumed G/H to be reductive nor have we assumed G to be semisimple. As is seen from Eq. (6.27), the geodesic equations, in general, are not "projectable" onto the base manifold M and the internal charge space, represented by P_a's. They have explicit y dependence. We shall now discuss two special cases. First let us consider the case when $\tilde{g}$ is independent of x, i.e.,

$$\partial_\mu\tilde{g}_{mn}(x) = 0. \tag{6.30}$$

In this case, the geodesic equations become

$$\begin{aligned}\dot{P}_a &= -C^c{}_{ab}A^b(\Delta)P_c, \\ x^\mu &+ \Gamma_{\nu\lambda}{}^\mu\dot{x}^\nu\dot{x}^\lambda = P_a F_{\nu\mu}{}^a\dot{x}^\lambda g^{\lambda,\mu}.\end{aligned} \tag{6.31}$$

They have the same form as Wong's equations[46] in the presence of gravity. They can be thought of as equations on the variables $(x^\mu, \dot{x}^\mu, Pa)$.

If the bundle $\pi: E \to M$ is trivial, the equations of motion can be projected onto $g_G^* \times TM$ by using the projection

$$(\mu,\mathbb{1}): TE \to g_G^* \times TM, \quad (Y,\dot{Y},x,\dot{x}) \to (P_a,x,\dot{x}).$$

The projected dynamics can be expanded in terms of the generators of the coadjoint representation with the vector field

$$\tilde{\Delta} = A^a(\Delta)C^c{}_{ab}\,t_c\frac{\partial}{\partial t_b}. \tag{6.32}$$

Again, this implies that adjoint-invariant polynomials in the P_a's will be constants of motion for Δ. The vector variable P_a precesses without leaving the coadjoint orbit on which it started. If the bundle is nontrivial (for instance, the Kaluza–Klein monopole in five dimensions[28]), the momentum map allows us to define a momentum bundle. This is the bundle associated with the usual construction to build up associated bundles (see, for instance, the last reference in Ref. 31) by using the equivariance of the momentum map μ. The case of a charged particle moving in the external field of a non-Abelian monopole has been analyzed in some details by Balachandran *et al.*[47]

Next, let us consider the case when the subgroup H is trivial. In this case, the fiber is the group G. The geodesic equations are

$$\begin{aligned} \dot{P}_a &= -C^c{}_{ab} A_\mu{}^b(x) P_c \dot{x}^\mu, \\ \dot{\hat{P}}_a &= -\tilde{g}_{ae} C^{bec} \hat{P}_b \hat{P}_c, \\ \ddot{x} + \Gamma_{\nu\lambda}{}^\mu \dot{x}^\nu \dot{x}^\lambda &= P_a F_{\nu\lambda}{}^a \dot{x}^\nu \tilde{g}^{\lambda\mu} - \tfrac{1}{2} \partial^\mu \tilde{g}^{ab}(x) \hat{P}_a \hat{P}_b. \end{aligned} \tag{6.33}$$

If we specify $(x^\mu, \dot{x}^\mu, P_a, \hat{P}_a)$ at some initial time, these equations can be solved for a later time. The dynamics in the total space E describe the motion of a particle in a gravitational field with internal charges $(P_a, \hat{P}_a)$ which couple to an external Yang–Mills field and the scalar field $\tilde{g}^{ab}$. The scalar fields are sometimes considered as providing a "dielectric" medium in which the particle moves.[12,13,15] In this respect, let us recall that in the case of the motion of a rigid body, P_a is the angular momentum in the "lab" frame, which is conserved, while $\hat{P}_a$ is the angular momentum in the "co-moving frame," which processes. Here we may also try to interpret $\hat{P}_a$ as being the non-Abelian charge seen in the comoving frame. It is difficult to assess various interpretations without considering the field part of the dynamics. These are currently under investigation and will be discussed elsewhere.

Besides the two cases we have discussed, there are other cases where the equations of motion in E are projectable onto $TM \times g^*_G$. For example, assume that $\tilde{g}$ satisfies

$$\tilde{g}^{ij}(x) = \phi(x) k^{ij}, \tag{6.34}$$

where k^{ij} are constants. Then

$$\partial^\mu \tilde{g}^{ij} = (\phi^{-1} \partial^\mu \phi) \tilde{g}^{ij}, \tag{6.35}$$

so that (6.27) becomes

$$\ddot{x}^\mu + \Gamma_{\nu\lambda}{}^\mu \dot{x}^\nu \dot{x}^\lambda = P_a F_{\nu\lambda}{}^a \dot{x}^\nu g^{\lambda\mu} - \tfrac{1}{2}(\phi^{-1} \partial^\mu \phi) P^2, \tag{6.36}$$

where

$$P^2 = g^{ab} P_a P_b = \tilde{g}^{ij} \hat{P}_i \hat{P}_j, \tag{6.37}$$

and we have used (5.14), (A28), and the definitions of P_a and $\hat{P}_i$.

Another case, where equations of motion can be expressed in terms of $(x, \dot{x}, P_a)$, arises when the group G is Abelian. A particular example of this situation when $G = \mathrm{U}(1)$ is discussed in the following section.

VII. ABELIAN KALUZA–KLEIN MONOPOLES

In this section we shall consider an example in which the bundle $\pi: E \to M$ is nontrivial, namely, the Abelian Kaluza–Klein monopole discussed by Sorkin and by Gross and Perry.[28] Let us write their monopole solution in the following form:

$$\begin{aligned} \bar{g} = & -dt \otimes dt + (1 + m/\rho)^2 dr \otimes dr \\ & + 4\left(\frac{2mr}{\rho + m}\right)^2 \left(\mathrm{Tr}\,\frac{\sigma^3}{2i} s^{-1}\, ds\right)^2 \\ & - 2r^2\, \mathrm{Tr}\left(\frac{\sigma_i}{2i}\, d\hat{x}^i\right)^2, \end{aligned} \tag{7.1}$$

where

$$\rho^2 = r^2 + m^2, \quad s \in \mathrm{su}(2), \quad \sigma_i \hat{x}^i \equiv s\sigma_3 s^{-1}, \tag{7.2}$$

and σ_i are the Pauli matrices. As usual, we shall use the invariant fields Y_i, $\hat{Y}_i$ and their dual forms θ^i, $\hat{\theta}^i$ so that

$$i_{Y_i} s^{-1}\, ds = (2i)^{-1} \sigma_i, \tag{7.3a}$$

$$i_{\hat{Y}_i} s^{-1}\, ds = -(2i)^{-1} \sigma_i. \tag{7.3b}$$

Let us first show that $\bar{g}$ is invariant under $\hat{Y}_i$. It is clear that

$$L_{\hat{Y}_i} s^{-1}\, ds = 0, \tag{7.4}$$

hence we have only to show that the last term in $\bar{g}$ is invariant. To do this, we compute

$$\begin{aligned} L_{\hat{Y}_i}((2i)^{-1} \sigma_j\, d\hat{x}^j) &= d\,(i_{\hat{Y}_i}((2i)^{-1} \sigma_j\, d\hat{x}^j)) \\ &= d i_{\hat{Y}_i}\, [ds\, s^{-1}, (2i)^{-1} \sigma_j\, \hat{x}^j] \\ &= d\,(-\epsilon_{ijk} \hat{x}^j ((2i)^{-1} \sigma_k)), \end{aligned} \tag{7.5}$$

from which it follows that

$$\begin{aligned} L_{\hat{Y}_i}\, \mathrm{Tr}\left(\frac{1}{2i} \sigma_j\, d\hat{x}^j\right)^2 = \mathrm{Tr}\bigg(& -\epsilon_{ijk} \frac{1}{2i} \sigma_k \frac{1}{2i} \sigma_l \\ & - \epsilon_{ilk} \frac{1}{2i} \sigma_j \frac{1}{2i} \sigma_k \bigg) d\hat{x}^j \otimes d\hat{x}^l = 0. \end{aligned} \tag{7.6}$$

Therefore $\bar{g}$ has the Killing vectors $\hat{Y}_i$. It is also obvious that ∂_t is a Killing vector for $\bar{g}$. The Lagrangian

$$\mathscr{L} \equiv \tfrac{1}{2} \bar{g}(\Delta, \Delta)$$

is then invariant under these Killing vectors.

As usual, we have

$$\theta_{\mathscr{L}} = \bar{g}(\Delta, \cdot),$$

and the conserved momenta P_t, $\hat{P}_i$ associated with the Killing vectors ∂_t, $\hat{Y}_i$ can easily be found. We find

$$\begin{aligned} P_t &= i_{\partial_t} \theta_{\mathscr{L}} = -\dot{t}, \\ \hat{P}_i &= i_{\hat{Y}_i} \theta_{\mathscr{L}} = 2\left(\frac{2mr}{\rho + m}\right)^2 \hat{x}_i \left(\mathrm{Tr}\,\frac{\sigma_3}{2i} s^{-1} \dot{s}\right) - r^2 \epsilon_{ijk}\, \hat{x}^j \dot{\hat{x}}^k. \end{aligned} \tag{7.7}$$

The first merely establishes the proportionality between the coordinate t and the proper time τ. The second one gives us the conserved angular momentum. Notice that

$$\mathrm{Tr}(\sigma_3/2i) s^{-1} \dot{s}) = -\tfrac{1}{2} \dot{\theta}^3, \tag{7.8}$$

so that

$$-\hat{P}_i = (2mr/(\rho + m))^2 \hat{x}_i\, \dot{\theta}^3 + r^2 \epsilon_{ijk}\, \hat{x}^j \dot{\hat{x}}^k. \tag{7.9}$$

We see that the angular momentum contains the familiar radial term and the orbital angular momentum term $\mathbf{r} \times \mathbf{v}$. Also note that the last two terms of $\bar{g}$ can be written as

$$(2mr/(\rho + m))^2 \theta^3 \otimes \theta^3 + r^2(\theta^1 \otimes \theta^1 + \theta^2 \otimes \theta^2), \tag{7.10}$$

hence the metric $\bar{g}$ is also invariant under Y_3. The "charge"

$$P_3 = i_{Y_3} \theta_{\mathscr{L}} = (2mr/(\rho + m))^2 \dot{\theta}^3 \tag{7.11}$$

is therefore conserved. Using (7.11), we may write (7.9) as

$$-\hat{P}_i = P_3 \hat{x}_i + r^2 \epsilon_{ijk} \hat{x}^j \dot{\hat{x}}^k, \tag{7.12}$$

and get

$$\hat{P}^2 = P_3{}^2 + r^4(\dot{x}^i)^2. \tag{7.13}$$

To be complete, we also consider the motion in the radial direction. Let

$$P_r = i_{\partial_r}\, \theta_{\mathscr{L}} = (1 + m/\rho)^2 \dot{r}. \tag{7.14}$$

It is easy to see that

$$2L_{\partial_r} \mathscr{L} = \partial_r \Big(1 + \frac{m}{\rho}\Big)^2 \dot{r}^2 + \partial_r \Big(\frac{2mr}{\rho + m}\Big)^2 (\dot{\theta}^3)^2 + 2r[(\dot{\theta}^1)^2 + (\dot{\theta}^2)^2], \tag{7.15}$$

so that the Euler–Lagrange equation gives

$$2(1 + m/\rho)^2 \ddot{r} + \partial_r (1 + m/\rho)^2 \dot{r}^2 = \Big[\Big(\frac{2mr}{\rho + m}\Big)^{-4} \partial_r \Big(\frac{2mr}{\rho + m}\Big)^2\Big] P_3{}^2 + \frac{2}{r^3} J^2, \tag{7.16}$$

where

$$J^2 \equiv \hat{P}^2 - P_3{}^2,$$

and we have used the fact

$$(\dot{\theta}^1)^2 + (\dot{\theta}^2)^2 = (\dot{x}^i)^2 \tag{7.17}$$

and Eqs. (7.11) and (7.13). Let

$$Y = r + m \ln(\rho + r), \quad \frac{dy}{dr} = 1 + \frac{m}{\rho}. \tag{7.18}$$

Then Eq. (6.48) can be written as

$$\ddot{y} = -\frac{1}{2} \partial y \Big[\Big(\frac{\rho + m}{2mr}\Big)^2 P_3{}^2 + \frac{1}{r^2} J^2\Big]. \tag{7.19}$$

Hence we see that

$$E = \frac{1}{2} \dot{y}^2 + \frac{1}{2} \Big[\Big(\frac{\rho + m}{2mr}\Big)^2 P_3{}^2 + \frac{1}{r^2} J^2\Big] \tag{7.20}$$

is conserved. Equation (7.20) can, of course, be integrated by quadrature:

$$t = \int^r \frac{1 + m/\rho}{\sqrt{2(E - V)}}\, dt. \tag{7.21}$$

Nevertheless, the qualitative feature of E is best seen by analyzing the potential

$$V(r) = \frac{1}{2} \Big[\Big(\frac{\rho + m}{2mr}\Big)^2 P_3{}^2 + \frac{1}{r^2} J^2\Big]. \tag{7.22}$$

We see that

$$\lim_{r \to 0} V(r) \to \frac{1}{2r^2} P^2 + \frac{P_3{}^2}{4m^2}, \quad V(\infty) = \frac{1}{8m^2} P_3{}^2, \tag{7.23}$$

and

$$V'(r) < 0, \quad \text{for all } r. \tag{7.24}$$

Hence

$$E > (8m)^{-2} P_3{}^2. \tag{7.25}$$

The particle will reach origin if

$$E \gg (4m)^{-2} P_3{}^2 \tag{7.26}$$

and $\hat{P} = 0$.

Finally, the quadrature (7.21) gives

$$(2m)^{-1} \sqrt{8m^2 E - P_3{}^2}\, t = \sqrt{(\rho - a)^2 - b^2} + (m + a) \times \ln\big[(\rho - a) + \sqrt{(\rho - a)^2 - b^2}\big] + \text{const}, \tag{7.27}$$

where

$$a = m P_3{}^2/(8m^2 E - P_3{}^2),$$

$$b^2 = \frac{64 m^6 E^2}{(8m^2 E - P_3{}^2)^2} + \frac{4m^2 J^2}{8m^2 E - P_3{}^2} \gg a^2, \tag{7.28}$$

and

$$\rho \gg a + b, \quad \rho \gg m. \tag{7.29}$$

This last condition determines the turning point.

VIII. CONCLUDING REMARKS

The definition of $\hat{P}_i$ in Eq. (5.20) depends on a local section σ; nevertheless it can be shown that the equation itself has a proper transformation property under the change of σ. Similarly, the metric $\bar{g}$ given in Eq. (6.1) in terms of local coordinates depends on a local section or equivalently a local trivialization. It is well known, however, that $\bar{g}$ has a proper transformation property under the change of local trivialization if A_μ transforms like a gauge field. With this understanding, one can also show that the geodesic equations (6.15), (6.27), and (6.29) all have proper transformation properties.

Note that the geodesic equations (6.15) and (6.27), in general, are not projectable in the sense that they cannot be described in terms of $(x^\mu, \dot{x}^\mu, P_a)$ only. However, one may expect that it can be done in terms of $(x^\mu, \dot{x}^\mu, P_a, \hat{P}_i)$. We have seen explicitly how this happens in the case when the fiber is diffeomorphic to a group. In the general case some additional investigation is required.

Finally, when $\bar{g}$ is independent of x, the geodesic motion in terms of $(x^\mu, \dot{x}^\mu, P_a)$ on the manifold, i.e., Wong's equations (6.31), will not distinguish between the various subgroups H. In particular, it will not tell whether the fiber is diffeomorphic to a group or a coset space. This remark also applies to the "generalized Wong's equation" (6.36). The projection onto variables $(x^\mu, \dot{x}^\mu, P_a, \hat{P}_i)$ can be achieved only locally. This seems to suggest that the "external charge," because of the screening effect of the "dielectric medium," does not have a global meaning.

APPENDIX: DIFFERENTIAL CALCULUS ON G/H

Let us consider a Lie group G and the invariant forms on G:

$$s^{-1} ds \equiv \theta^a t_a, \quad ds\, s^{-1} \equiv -\hat{\theta}^a t_a, \tag{A1}$$

where t_a are generators of G.

The dual of θ^a and $\hat{\theta}^a$ are the invariant vector fields Y_a and $\hat{Y}_a$, respectively. We have

$$i_{Y_a}(s^{-1} ds) = t_a, \quad i_{\hat{Y}_a}(ds\, s^{-1}) = -t_a. \tag{A2}$$

Taking the exterior derivative of (A1), we get

$$d\theta^a = -\tfrac{1}{2} C^a{}_{bc}\, \theta^b \wedge \theta^c, \tag{A3}$$

$$d\hat{\theta}^a = -\tfrac{1}{2} C^a{}_{bc}\, \hat{\theta}^b \wedge \hat{\theta}^c, \tag{A4}$$

where

$$[t_b, t_c] = C^a{}_{bc}\, t_a$$

and

$$\theta^b \wedge \theta^c \equiv \theta^b \otimes \theta^c - \theta^c \otimes \theta^b.$$

Taking the derivative i_{Y_a} on both sides of (A3) and using the Cartan relation for the Lie derivative L_{Y_a},

$$L_{Y_a} = i_{Y_a}\, d + d i_{Y_a},$$

we get

$$L_{Y_a}\,\theta^c = -C^c{}_{ab}\,\theta^b. \tag{A5}$$

Similarly from (A4) we get

$$L_{\hat{Y}_a}\,\hat{\theta}^c = -C^c{}_{ab}\,\hat{\theta}^b. \tag{A6}$$

Then (A5) and (A6) imply

$$L_{Y_a}\,Y_b = C^c{}_{ab}\,Y_c, \quad L_{\hat{Y}_a}\,\hat{Y}_b = C^c{}_{ab}\,\hat{Y}_c. \tag{A7}$$

Since Y_a, $\hat{Y}_a$ generate the right and left actions on G, respectively, and since these actions commute, we have

$$L_{Y_a}\,\hat{Y}_b = -L_{\hat{Y}_b}\,Y_a = 0. \tag{A8}$$

By defining the matrices for adjoint representation by

$$\mathrm{ad}(s)t_a \equiv t_b D^b{}_a(s), \quad s \in G, \tag{A9}$$

we get from (A1)

$$\hat{\theta}^a = -\theta^b D^a{}_b(s), \quad \theta^a = -\hat{\theta}^b D^a{}_b(s^{-1}), \tag{A10}$$

which imply

$$\hat{Y}_a = -Y_b D^b{}_a(s^{-1}), \quad Y_a = -\hat{Y}_b D^b{}_a(s). \tag{A11}$$

From (A7), (A8), and (A11) we get

$$\begin{aligned} dD^a{}_b(s) &= -C^a{}_{ce} D^e{}_b(s)\,\hat{\theta}^c \\ &= -C^e{}_{bc} D^a{}_e(s)\,\theta^c. \end{aligned} \tag{A12}$$

Let G/H be the right coset space. We can consider G to be an H bundle over G/H, with H acting on G on the left and with the natural projection π: $G \to G/H$. Under π, we have

$$\pi_*\, Y_a = \tilde{Y}_a, \quad \pi_*\,\hat{Y}_a = 0, \quad t_a \in g_H. \tag{A13}$$

We shall denote the elements in the subalgebra g_H with indices $\bar{a}$, $\bar{b}$,..., etc., and those in the complement of g_H with indices $\bar{i}$, $\bar{j}$, $\bar{k}$,..., etc.

Note that the right-invariant fields $\hat{Y}_a$ in general are not projectable unless the corresponding generator t_a belongs to the subalgebra of the normalizer N or H in G.

From (A7) and (A13), we have

$$(\tilde{Y}_a, \tilde{Y}_b) = C^c{}_{ab}\,\tilde{Y}_c. \tag{A14}$$

Although it is desirable to work with the global quantities $\tilde{Y}_a$, for computational purposes, it is often convenient to introduce a local section σ: $U \subset G/H \to G$, and consider the pullback of the invariant forms. On $\pi^{-1}(U)$, we have the decomposition

$$s = h\cdot\sigma(y), \quad \pi(s) = y. \tag{A15}$$

Then we have

$$s^{-1} ds = \sigma^{-1}(h^{-1}\,dh)\sigma + \sigma^{-1}\,d\sigma, \tag{A16}$$

$$ds\, s^{-1} = dh\cdot h^{-1} + h\,(d\sigma\cdot\sigma^{-1})h, \tag{A17}$$

where, for simplicity in notation, we assume G is a linear group, but the formulas below are valid for a general Lie group.

Since $\hat{Y}_{\bar{a}}$ generates the H action on G, we have

$$i_{\hat{Y}_a}\sigma^{-1}\,d\sigma = 0, \quad i_{\hat{Y}_a}\,d\sigma\cdot\sigma^{-1} = 0. \tag{A18}$$

Also, let X be any vector field on G/H. Then on the section $\sigma(U)$, we have

$$i_{\sigma_* X}\,h^{-1}\,dh = 0, \quad i_{\sigma_* X}\,dh\cdot h^{-1} = 0. \tag{A19}$$

To proceed, we define

$$h^{-1}\,dh \equiv \omega_a{}^{\bar{b}}(s)t_{\bar{b}}\,\theta^a(s) \equiv \omega_a(s)\,\theta^a(s) \equiv \omega, \tag{A20}$$

$$dh\,h^{-1} \equiv -H_a{}^{\bar{b}}(s)t_{\bar{b}}\,\hat{\theta}^a(s) \equiv -H_a(s)\,\hat{\theta}^a(s) \equiv -H. \tag{A21}$$

Taking the derivative $i_{\hat{Y}_{\bar{a}}}$ on (A17) and using (A18) and (A21), we get

$$t_{\bar{a}} = H_{\bar{a}} = -i_{\hat{Y}_{\bar{a}}}(dh\cdot h^{-1}). \tag{A22}$$

Similarly, using $i_{\hat{Y}_i}$ to operate on (A17), we get

$$i_{\hat{Y}_i}(d\sigma\cdot\sigma^{-1}) = -t_i + H_i. \tag{A23}$$

Let $\hat{\tilde{Y}}_i \in T(G/H)$ be such that

$$\sigma^*(d\sigma\cdot\sigma^{-1})(\hat{\tilde{Y}}_i) = -t_i + H_i. \tag{A24}$$

Then we have, from (A22)–(A24),

$$\hat{Y}_i = \sigma_*\,\hat{\tilde{Y}}_i + H_i{}^{\bar{a}}(\sigma)\,\hat{Y}_{\bar{a}}. \tag{A25}$$

From this and (A11), we find, on $\sigma(U)$,

$$Y_a = \sigma_*(-\hat{\tilde{Y}}_i D^i{}_a(\sigma)) - \omega_a{}^{\bar{a}}(\sigma)\hat{Y}_{\bar{a}}, \tag{A26}$$

where we have used (A22) and the relation

$$\omega_a(\sigma) = H_b(\sigma) D^b{}_a(\sigma), \tag{A27}$$

which follows from the definitions (A20) and (A21).

Since $\pi\cdot\sigma$ = identity on G/H, we have, from (A13) and (A26),

$$\tilde{Y}_a(y) = -\hat{\tilde{Y}}_i(y) D^i{}_a(\sigma(y)), \tag{A28}$$

$$\hat{\tilde{Y}}_i = -\tilde{Y}_a D^a{}_i(\sigma(y)^{-1}). \tag{A28$'$}$$

Let $\hat{e}^i$ be the dual form of $\hat{\tilde{Y}}_i$. From (A24), we get

$$\sigma^*(ds\,s^{-1}) = (-t_i + H_i)\hat{e}^i, \tag{A29}$$

or equivalently,

$$\sigma^*(\hat{\theta}^i) = \hat{e}^i, \quad \sigma^*(\hat{\theta}^{\bar{a}}) = -H_i{}^{\bar{a}}\hat{e}^i. \tag{A30}$$

It follows that

$$e^a \equiv \sigma^*(\theta^a) = -\hat{e}^i[D^a{}_i(\sigma^{-1}) - H_i{}^{\bar{a}}(\sigma) D^a{}_{\bar{a}}(\sigma^{-1})] \tag{A31}$$

and

$$e^a(\tilde{Y}_b) = \delta_b{}^a - \omega_b{}^{\bar{a}}(\sigma) D^a{}_{\bar{a}}(\sigma^{-1}). \tag{A32}$$

Next we consider the derivative of $h^{-1}\,dh$ and $dh\cdot h^{-1}$. From (A20) and (A21), we get

$$d\omega_a \wedge \theta^a = (-\tfrac{1}{2}[\omega_b,\omega_c] + \tfrac{1}{2}\omega_a C^a{}_{bc})\theta^b \wedge \theta^c, \tag{A33}$$

$$dH_a \wedge \hat{e}^a = (-\tfrac{1}{2}[H_b,H_c] + \tfrac{1}{2}H_a C^a{}_{bc})\hat{\theta}^b \wedge \hat{\theta}^c. \tag{A34}$$

Evaluating (A33) and (A34) at the invariant fields Y_a and $\hat{Y}_a$, we get

$$L_{Y_a}\omega_b - L_{Y_b}\omega_a + [\omega_a,\omega_b] = \omega_c C^c{}_{ab}, \tag{A35}$$

$$L_{\hat{Y}_a} H_b - L_{\hat{Y}_b} H_a + [H_a,H_b] = H_c C^c{}_{ab}, \tag{A36}$$

$$L_{\hat{Y}_a}\omega_b = 0, \quad L_{Y_a} H_b = 0. \tag{A37}$$

Equation (A35) is easily projected onto G/H. The pullback of (A34) becomes

$$dH_i \wedge \hat{e}^i + H_i\, d\hat{e}^i = \{\tfrac{1}{2}[H_j,H_k] - \tfrac{1}{2} t_{\bar a}(H_j{}^{\bar b} C^{\bar a}{}_{\bar b k} - H_k{}^{\bar b} C^{\bar a}{}_{\bar b j}) + \tfrac{1}{2} t_{\bar a} C^{\bar a}{}_{jk}\} e^j \wedge \hat{e}^k. \qquad \text{(A37')}$$

Next we take the exterior derivative of (A30) and get

$$d\hat{e}^i = -\tfrac{1}{2}(C^i{}_{jk} + f^i{}_{jk})\hat{e}^j \wedge \hat{e}^k, \qquad \text{(A38)}$$

where

$$f^i{}_{jk} \equiv -H_j{}^{\bar a} C^i{}_{\bar a k} + H_k{}^{\bar a} C^i{}_{\bar a j}. \qquad \text{(A39)}$$

Evaluating (A38) at $\tilde{Y}_j$, we get

$$L_{\hat{\tilde{Y}}_j} e^i = -(C^i{}_{jk} + f^i{}_{jk})\hat{e}^k \qquad \text{(A40)}$$

and

$$[\hat{\tilde{Y}}_i, \hat{\tilde{Y}}_j] = (C^k{}_{ij} + f^k{}_{ij})\hat{\tilde{Y}}_k. \qquad \text{(A41)}$$

There are various ways to compute $(\tilde{Y}_a, \hat{\tilde{Y}}_i)$. For example, one may use (A25), (A26), and (A35)–(A37). Instead, we should first derive a useful formula. Take the exterior derivative of the equation

$$\mathrm{Ad}(h)\, t_a = t_b\, D^b{}_a(h).$$

We get

$$dD^a{}_b(h) = -\hat{\theta}^c H_c{}^{\bar a} C^a{}_{\bar a e}\, D^e{}_b(h). \qquad \text{(A42)}$$

Using (A12) and (A42) we find, on $\sigma(U)$,

$$dD^a{}_a(\sigma) = -(C^a{}_{ic}\, D^c{}_b(\sigma) + C^a{}_{c\bar a}\, H_i{}^{\bar a} D^c{}_b(\sigma))\, \hat{\theta}^i. \qquad \text{(A43)}$$

Pulled back by σ, we get on G/H

$$dD^a{}_b(\sigma(y)) = -(C^a{}_{ic}\, D^c{}_b(\sigma(y)) + C^a{}_{c\bar a}\, H_i{}^{\bar a} D^c{}_b(\sigma(y)))\hat{e}^i, \qquad \text{(A44)}$$

which implies

$$L_{\hat{\tilde{Y}}_i} D^a{}_b(\sigma(y)) = -C^a{}_{ic} D^c{}_b(\sigma(y)) + C^a{}_{\bar a c}\, H_i{}^{\bar a} D^c{}_b(\sigma(y)), \qquad \text{(A45)}$$

$$L_{\tilde{Y}_c} D^a{}_b(\sigma(y)) = -C^e{}_{bc} D^a{}_e(\sigma(y)) - C^a{}_{\bar a e}\omega_c{}^{\bar a} D^e{}_b(\sigma(y)). \qquad \text{(A46)}$$

From these equations, we can compute $[\tilde{Y}_a, \tilde{Y}_i]$ and get

$$[\tilde{Y}_a, \tilde{Y}_i] = \omega_a{}^{\bar a} C^j{}_{\bar a i}\, \tilde{Y}_j, \quad [\tilde{Y}_a, \hat{e}^i] = -\omega_a{}^{\bar a} C^i{}_{\bar a j}\, \hat{e}^j. \qquad \text{(A47)}$$

Finally, let us consider invariant metric on G/H. A general metric $\tilde{g}$ can be written

$$\tilde{g} = \tilde{g}_{ij}\hat{e}^i \otimes \hat{e}^j, \quad \tilde{g}_{ij} = \tilde{g}_{ji}. \qquad \text{(A48)}$$

Using (A47), we find

$$L_{\tilde{Y}_a} \tilde{g} = \{L_{\tilde{Y}_a} \tilde{g}_{jk} - \omega_a{}^{\bar a}(\tilde{g}_{ij} C^i{}_{\bar a k} + \tilde{g}_{ik} C^i{}_{\bar a j})\}\hat{e}^j \otimes \hat{e}^k. \qquad \text{(A49)}$$

Hence, $\tilde{g}$ is invariant iff

$$L_{\tilde{Y}_a} \tilde{g}_{jk} = \omega_a{}^{\bar a}[\tilde{g}_{ij} C^i{}_{\bar a k} + \tilde{g}_{ik} C^i{}_{\bar a j}]. \qquad \text{(A50)}$$

Multiply both sides by $D^a{}_{\bar b}(\sigma(y)^{-1})$. By using the fact that $Y_a D^a{}_{\bar b}(\sigma^{-1})$ vanishes, and using (A22) and (A27), we get

$$\tilde{g}_{ij} C^i{}_{\bar a k} + \tilde{g}_{ik} C^i{}_{\bar a j} = 0, \qquad \text{(A51)}$$

and hence

$$L_{\tilde{Y}_a} \tilde{g}_{ij} = 0. \qquad \text{(A52)}$$

That is, g_{ij} is constant on G/H.

Equation (A51) is the condition for the metric to be consistent with the group action. Only when it is satisfied is the metric $\tilde{g}$ independent of the local section σ we use to define it. To see this, consider the invariant metric g on G whose pullback is $\tilde{g}$:

$$g \equiv \tilde{g}_{ij}\, \hat{\theta}^i \otimes \hat{\theta}^j,$$
$$\tilde{g}_{ij} \text{ symmetric, constant,} \quad \sigma^*(g) = \tilde{g}. \qquad \text{(A53)}$$

Then we have

$$L_{\hat{Y}_a}\, g = -(\tilde{g}_{ij} C^i{}_{\bar a k} + \tilde{g}_{ik} C^i{}_{\bar a j})\, \hat{\theta}^j \otimes \hat{\theta}^k. \qquad \text{(A54)}$$

Hence condition (A51), seen to be equivalent to $L_{\hat{Y}_{\bar a}}\, g$, vanishes. That is g, considered as a metric on the complement of the subalgebra g_H in g_G, is $\mathrm{Ad}\, H$ invariant. Thus, when (A51) is satisfied, $\tilde{g}$ is σ independent and is a global quantity despite its appearance in (A48).

[1]H. Weyl, Sitzungsber. Preuss. Akad. Wiss., 465 (1918).
[2]H. Weyl, Ann. Phys. (Leipzig) **59**, 101 (1919).
[3]Th. Kaluza, Zum Unitätsproblem Phys. Berl. Berichte, 966 (1921).
[4]O. Klein, Z. Phys. **37**, 895 (1926).
[5]A. Einstein, Sitzungsber. Preuss. Akad. Wiss., Phys. Math., 23, 26 (1927).
[6]W. Pauli, Ann. Phys. **18**, 305 (1933).
[7]O. Vebleu, *Projektive Relativitätsheorie* (Springer, Berlin, 1933).
[8]O. Klein, "New Theories in Physics," Conference report, Warsaw, 1938.
[9]A. Einstein and P. Bergmann, Ann. Math. **39**, 683 (1938).
[10]P. G. Bergmann, *An Introduction to the Theory of Relativity* (Prentice-Hall, New York, 1942).
[11]P. Jordan, Ann. Phys. (Leipzig), 219 (1947).
[12]A. Lichnerowicz and Y. Thiry, C. R. Acad. Sci. **224**, 529 (1947).
[13]Y. Thiry, C. R. Acad. Sci. **226**, 216, 881 (1948).
[14]P. G. Bergmann, Ann. Math. **49**, 255 (1948).
[15]P. Jordan, Astron. Nachr. **276**, 193 (1948).
[16]T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[17]E. Lubkin, Ann. Phys. (NY) **23**, 233 (1963).
[18]B. S. DeWitt, *Dynamic Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[19]A. Trautmann, Rep. Math. Phys. **1**, 29 (1970).
[20]W. Thirring, Acta Phys. Austrica Suppl. **19**, 439 (1978).
[21]Y. M. Cho, J. Math. Phys. **16**, 235 (1976); L. N. Chang, K. I. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976); A. S. Schwarz, Commun. Math. Phys. **56**, 79 (1977); Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975).
[22]C. A. Orzalesi, Fortschr. Phys. **29**, 413 (1981); C. Destri, C. A. Orzalesi, and P. Rossi, Ann. Phys. (NY) **147**, 321 (1983).
[23]W. Kopczynski, in *Differential Geometric Methods in Mathematical Physics* (Springer, Berlin, 1980).
[24]G. Domokos and S. Kovesi-Domokos, Nuovo Cimento A **44**, 318 (1978); J. F. Luciani, Nucl. Phys. B **135**, 111 (1978); E. Witten, Nucl. Phys. B **186**, 412 (1981); A. Salam and J. Strathdee, Ann. Phys. **141**, 316 (1982); R. Percacci and S. Randjbar-Daemi, J. Math. Phys. **24**, 807 (1983); M. A. Awada, Phys. Lett. B **127**, 415 (1983).
[25]Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, Nucl. Phys. B **127**, 57 (1977); Z. Horvath and L. Palla, Nucl. Phys. B **142**, 327 (1978); Y. M. Cho, "Gauge Theory on Homogeneous Fiber Bundle," CERN Preprint TH3414, 1983; R. Kerner, "Gauge Symmetries in Multiple Fibre Bundles," CERN Preprint TH3669, 1983.
[26]S. Weinberg, Phys. Lett. B **125**, 265 (1983).
[27]R. Kerner, Ann. Inst. Henri Poincaré B **9**, 143 (1968); C. A. Orzalesi and M. Pauri, Nuovo Cimento B **67**, 193 (1982); J. G. Miller, J. Math. Phys. **17**, 643 (1976); D. Moisson and S. J. Orfanidis, Phys. Rev. D **15**, 3608 (1976); N. K. Nielsen, Nucl. Phys. B **167**, 249 (1980).
[28]R. Sorkin, Phys. Rev. Lett. **51**, 87 (1983); D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 29 (1983); S. M. Barr, Phys. Lett. B **129**, 303 (1983).
[29]V. I. Arnold, *Mathematical Methods of Classical Mechanics, Springer Graduate Texts in Mathematics No. 60* (Springer, New York, 1978).

[30]J. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).
[31]S. Sternberg, *Lectures on Differential Geometry* (Prentice–Hall, Englewood Cliffs, NJ, 1963); R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1969); W. Thirring, *Classical Dynamical Systems, A Course in Mathematical Physics* (Springer-Verlag, Berlin, 1978), Vol. 1; C. Von Westenholz, *Differential Forms in Mathematical Physics* (North-Holland, Amsterdam, 1978).
[32]R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin, Reading, MA, 1978).
[33]A. P. Balachandran, G. Marmo, B-S. Skagerstam, and A. Stern, "Gauge Symmetries and Fiber Bundles—Applications to Particle Dynamics," in *Lecture Notes in Physics*, Vol. 88 (Springer-Verlag, Berlin, 1983).
[34]A. Lichnerowicz, *Théories Relativistes de la Gravitation et de l'Electromagnètisme* (Masson, Paris, 1955).
[35]H. Kunzle, J. Math. Phys. **13**, 739 (1972); S. Sternberg, Proc. Natl. Acad. Sci. (USA) **74**, 5253 (1977); J. Sniatycki and W. M. Tulczyjew, Ann. Inst. Henri Poincarè, **15**, 177 (1971); A. Weinstein, Lett. Math. Phys. **2**, 417 (1978); J. M. Souriau, Ann. Inst. Henri Poincarè A **20**, 315 (1974); Ch. Duval, Ann. Inst. Henri Poincarè, **25**, 345 (1976); Ch. Duval and P. Horvathy, Ann. Phys. **142**, 10 (1982).
[36]M. Crampin, J. Phys. A **16**, 3755 (1983).
[37]K. Yano and S. Ishihara, *Tangent and Cotangent Bundles* (Dekker, New York, 1973).
[38]V. I. Arnold, Ann. Inst. Fourier (Grenoble) **16**, 319 (1966); J. Marsden and R. Abraham, in *Proceedings on Pure Mathematics* (Am. Math. Soc., Providence, RI), Vol. 16, pp. 237–243; R. Hermann, J. Math. Phys. **13**, 460 (1972); J. Math. Phys. **13**, 833, 878 (1972).
[39]E. Lacomba, Trans. Am. Math. Soc. **185**, 474 (1973).
[40]J. Marsden, Can. Math. Bull. **40**, 247 (1967).
[41]There is a slight abuse of notation here for $Y_a \in \mathfrak{X}(G)$ while $\theta_{\mathscr{L}} \in \mathfrak{X}^*(TG)$. To be more precise we should set

$$i_{\bar{Y}_a} \pi^* df = \pi^* i_{Y_a} df,$$

where π: $TG \to G$ and $f \epsilon F(G)$. This implies $i_{[\bar{Y}_a, \bar{Y}_b]} \pi^* df = \pi^* i_{[Y_a, Y_b]} df$. Moreover (4.7) should be replaced by

$$i_{\bar{Y}_a} d\theta_{\mathscr{L}} = -dP_a = -d i_{\bar{Y}_a} \theta_{\mathscr{L}}.$$

Now $i_{\bar{Y}_a} \theta_{\mathscr{L}}$ depends only on Y_a and $\theta_{\mathscr{L}}$.
[42]C. M. Marle, in *Differential Geometry and Relativity*, edited by M. Cahen and M. Flato (Reidel, Dordrecht, Netherlands, 1976).
[43]J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[44]A. Lichnerowicz, *Geometry of Groups of Transformations* (Noordhoff, Leyden, 1977).
[45]A. Kirillov, *Eléments de la Theorie des Représentations* (Mir, Moscow, 1974).
[46]S. J. Wong, Nuovo Cimento A **65**, 689 (1970).
[47]A. P. Balachandran, "Nonabelian Monopoles Break Colour, I. Classical Mechanics," to appear in Phys. Rev.

Strong amplification of sidebands in a strongly driven three-level atomic system[a)]

Constantine Mavroyannis
Division of Chemistry, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

(Received 25 May 1984; accepted for publication 21 December 1984)

We have considered the excitation spectra for a three-level atom interacting simultaneously with a strong pump field and a weak signal field. The atom consists of an upper excited state $|2\rangle$ and two lower states $|1\rangle$ and $|3\rangle$ where the transition $|1\rangle \leftrightarrow |3\rangle$ is electric-dipole forbidden from parity considerations. The pump field, whose frequency mode ω_b is initially populated, operates near resonance between the states $|2\rangle$ and $|3\rangle$, $|2\rangle \leftrightarrow |3\rangle$, while the signal field describes the radiative decays $|2\rangle \rightarrow |3\rangle$ and $|2\rangle \rightarrow |1\rangle$, respectively. Using the Green function method, the spectral functions for the signal field have been calculated in the limit of high photon densities of the pump field describing the following processes: (i) one-photon process; the spectra consist of a main peak at the signal frequency $\omega = \omega_{21}$ and a pair of sidebands at $\omega = \omega_{21} \pm \Omega - (\omega_{23} - \omega_b)/2$, where $\Omega^2 = \Omega_b^2/2 + (\omega_{23} - \omega_b)^2/4$ with Ω_b and $(\omega_{23} - \omega_b)$ being the Rabi frequency and the detuning of the pump field, respectively. The intensity of the main peak at $\omega = \omega_{21}$ is positive indicating signal-field absorption while those of the sidebands are always negative implying amplification (stimulated emission) of the signal field. The sum of the intensities (in absolute value) of the sidebands is twice that of the main peak at resonance while in the off-resonance case it depends on the value of the ratio $(\omega_{23} - \omega_b)/2\Omega$. (ii) Stimulated three-photon process; the spectra consists of a pair of sidebands peaked at $\omega = \omega_{21} - 2\omega_b \pm \Omega - (\omega_{23} - \omega_b)/2$ whose intensity is negative indicating amplification. At resonance the sidebands have equal intensities but at finite detuning asymmetry arises enhancing one peak while diminishing the intensity of the other. (iii) Stimulated two-photon process; the stimulated Raman spectra consist of a peak at $\omega = \omega_{31}$, which has a delta function distribution, and a pair of strongly amplified sidebands at $\omega = \omega_{31} \pm \Omega + (\omega_{23} - \omega_b)/2$. The computed spectra for the three processes are presented graphically for different values of Rabi frequencies and detunings.

I. INTRODUCTION

Mollow[1] was first to consider the optical amplification arising in a two-level atomic system interacting with a strong pump field and a weak signal field simultaneously. Mollow's approach is semiclassical where the pump field and the signal field are both treated classically and the electron states are quantized. Mollow[1] has shown that at high pump intensities the spectral function takes negative values indicating amplification (stimulated emission) of the signal field. It has been shown[1] that the amplification of the signal field occurs at the expense of the pump field whose rate of attenuation increases as the rate of amplification of the signal field increases. Mollow's theoretical predictions have been confirmed experimentally by Wu *et al.*[2] in a Na atom beam experiment. Mollow's treatment has been recently extended by Galbraith *et al.*[3] in order to calculate double-resonance line shapes for arbitrary angular momentum states of molecules.

The amplification of weak sidebands by a strongly driven two-level vapor system without population inversion has been observed in microwave,[4] rf,[5] and optical[6] transitions. Strong sideband amplification at high pump field intensities has been observed by Tam[7] in self-focused light beam experiments in atomic Na vapors. The observed strong sideband amplification has been explained[7] as due to an off-resonant stimulated Raman scattering process.

The purpose of the present study is to calculate the excitation spectra for the three-level atom shown in Fig. 1, where a strong (pump) field operates between the states $|2\rangle$ and $|3\rangle$, $|2\rangle \leftrightarrow |3\rangle$, while a signal (vacuum) field describes the radiative transitions $|2\rangle \rightarrow |1\rangle$ and $|2\rangle \rightarrow |3\rangle$, respectively; the signal field is taken as a weak perturbing field. We are interested in calculating the spectral functions or equivalently the absorption coefficient for the $|1\rangle \leftrightarrow |2\rangle$ transition of the signal field describing the following processes: (i) one-photon process occurring in the neighborhood of the signal frequency $\omega \approx \omega_{21}$; (ii) stimulated three-photon process near the frequency $\omega \approx \omega_{21} - 2\omega_b$, where one photon of the signal field is absorbed while two photons of the laser field are emitted simultaneously; and (iii) stimulated two-photon or Raman process near the frequency $\omega \approx \omega_{31} \approx \omega_{21} - \omega_b$, where one photon of the signal field is absorbed while a photon of the laser field is emitted simultaneously. It will be shown that the sidebands, which are induced by the laser field near the frequencies of the three-processes under investigation, are strongly amplified. The effects due to the detuning of the laser field are taken into account as well.

The model Hamiltonian is developed in Sec. II which describes the atomic system depicted in Fig. 1. This Hamiltonian is used in Sec. III to derive the equation of motion for

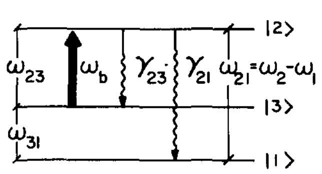


FIG. 1. Energy-level diagram of a three-level atom. The solid line indicates the laser field operating between the states $|2\rangle$ and $|3\rangle$, $|2\rangle \leftrightarrow |3\rangle$. Wiggly lines describe the radiative decays $|2\rangle \rightarrow |3\rangle$ and $|2\rangle \rightarrow |1\rangle$, respectively.

[a)] Issued as NRCC No. 24003.

the Green function of the signal field describing the $|1\rangle \leftrightarrow |2\rangle$ transition. Then expressions for the resulting Green functions are derived representing the one-photon and the three-photon processes, respectively. The excitation spectra for the one-photon and three-photon processes are considered in Sec. IV, where the spectral functions for the processes in question are derived and discussed while the computed spectra are graphically presented. The spectral function for the stimulated Raman spectra is calculated and the computed spectra are graphically presented in Sec. V. The results for the three-processes under investigation are summarized and discussed in Sec. VI.

II. THE MODEL HAMILTONIAN

We consider a three-level atom whose nondegenerate energy levels are depicted in Fig. 1. The energies of the lower ground state $|1\rangle$, the upper ground state $|3\rangle$, and the excited state $|2\rangle$ are denoted by ω_1, ω_3, and ω_2, respectively, and the transition frequencies $\omega_{ij} = \omega_i - \omega_j$ with $i,j = 1,2$, and 3, where units in which $\hbar = 1$ are used throughout. The parity of the state $|2\rangle$ is assumed to be different from those of $|1\rangle$ and $|3\rangle$, hence, the electronic transitions $|1\rangle \leftrightarrow |2\rangle$ and $|2\rangle \leftrightarrow |3\rangle$ are electric dipole allowed while the transition $|1\rangle \leftrightarrow |3\rangle$ is electric dipole forbidden. The atom is pumped between the levels $|3\rangle$ and $|2\rangle$, $|2\rangle \leftrightarrow |3\rangle$, by a strong laser field whose frequency mode ω_b is initially populated and is near resonance with the transition frequency $\omega_{23} = \omega_2 - \omega_3$. The atomic states are simultaneously coupled to the remaining modes of the electromagnetic field (signal or vacuum field), those being initially empty. Since the transition $|1\rangle \leftrightarrow |3\rangle$ is electric dipole forbidden for parity considerations, the electronic state $|3\rangle$ may be considered as a metastable state and, therefore, it is initially populated while the states $|1\rangle$ and $|2\rangle$ are initially empty. The Hamiltonian describing the atomic system shown in Fig. 1, in the electric dipole approximation, may be taken as

$$\begin{aligned} H &= \omega_{21} n_2 + \omega_{31} n_3 + \omega_b n_b \\ &\quad + \tfrac{1}{2} i\omega_p \sqrt{f_b}\,(\alpha_3^+ \alpha_2 - \alpha_2^+ \alpha_3) \\ &\quad \times (\beta_b + \beta_b^+) + \sum_{\mathbf{k},\lambda} ck \beta_{\mathbf{k}\lambda}^+ \beta_{\mathbf{k}\lambda} \\ &\quad + \frac{1}{2} i\omega_p \sum_{\mathbf{k},\lambda} \left[\frac{f_{12}(\mathbf{k},\lambda)\omega_{21}}{ck}\right]^{1/2} \\ &\quad \times (\alpha_1^+ \alpha_2 \beta_{\mathbf{k}\lambda}^+ - \alpha_2^+ \alpha_1 \beta_{\mathbf{k}\lambda}) \\ &\quad + \frac{1}{2} i\omega_p \sum_{\mathbf{k},\lambda} \left[\frac{f_{32}(\mathbf{k},\lambda)\omega_{23}}{ck}\right]^{1/2} \\ &\quad \times (\alpha_3^+ \alpha_2 \beta_{\mathbf{k}\lambda}^+ - \alpha_2^+ \alpha_3 \beta_{\mathbf{k}\lambda}), \end{aligned} \tag{1}$$

where α_i^+ and α_i are the Fermi creation and annihilation operators describing the electron states $|i\rangle$ while $n_i = \alpha_i^+ \alpha_i$ is the corresponding number operator. Also, β_b^+ and β_b are the boson creation and annihilation operators for the laser field with frequency ω_b and number operator $n_b = \beta_b^+ \beta_b$ while the operators $\beta_{\mathbf{k}\lambda}^+$ and $\beta_{\mathbf{k}\lambda}$ are the corresponding ones for the signal field with wave vector $\mathbf{k}$, frequency ck, and transverse polarization λ. The coupling function $f_b = f_{23}(\omega_{23}/\omega_b)$ with f_{23} being the oscillator strength for the transition $|2\rangle \leftrightarrow |3\rangle$ and $\omega_p^2 = 4\pi e^2/mV$ is the plasma frequency, where m is the electron mass, $-e$ is the electron charge, and V is the volume of the sample container. The functions $f_{12}(\mathbf{k},\lambda)$ and $f_{32}(\mathbf{k},\lambda)$ represent the oscillator strengths for the transitions $|1\rangle \leftrightarrow |2\rangle$ and $|2\rangle \leftrightarrow |3\rangle$, respectively. For the three-level system under consideration the occupation number operator $n_i = \alpha_i^+ \alpha_i$ satisfy the supplementary condition

$$n_1 + n_2 + n_3 = 1\,. \tag{2}$$

The first three terms in Eq. (1) describe the free atomic and laser fields, respectively, while the fourth term represents the interaction between them. The last three terms in Eq. (1) describe the free signal field and its interaction with the atomic states, respectively. The laser field with frequency ω_b is initially populated and, hence, it is to be treated as a strong field while the signal field, which describes the decay processes $|2\rangle \to |1\rangle$ and $|2\rangle \to |3\rangle$, is considered as the weak perturbing field since it is initially unpopulated. The Hamiltonian (1) can be derived from the corresponding one, Eq. (1) in Ref. 8, which describes the physical process of third-order mixing of frequencies of two laser fields, when the laser field a operating between the states $|1\rangle$ and $|2\rangle$, $|1\rangle \leftrightarrow |2\rangle$ is discarded. In Eq. (1), the fourth term describing the laser–atom interaction is taken beyond the rotating wave approximation (RWA) because, as it will be discussed later, we will be concerned with physical processes near the frequencies $\omega_{31} \pm \omega_b$ and ω_{31} while the signal field, being the weak perturbing field representing radiative decay processes, is sufficiently described in the RWA.

The dynamics of the system is developed by making use of the double-time retarded Green functions defined as

$$\langle\langle A(t);B(t')\rangle\rangle = -i\theta(t-t')\langle[A(t),B(t')]\rangle\,, \tag{3}$$

where $\theta(t-t')$ is the unit step function, the operators $A(t)$ and $B(t')$ are in the Heisenberg representation, and the angular brackets denote the average over the canonical ensemble appropriate to the total Hamiltonian H. When Eq. (3) is differentiated with respect to t, the Fourier transform $\langle\langle A;B\rangle\rangle$ of $\langle\langle A(t);B(t')\rangle\rangle$ satisfies the equation

$$\omega\langle\langle A;B\rangle\rangle = (1/2\pi)\langle[A,B]\rangle + \langle\langle[A,H];B\rangle\rangle\,. \tag{4}$$

The corresponding result for differentiation with respect to t' is given by

$$\omega\langle\langle A;B\rangle\rangle = (1/2\pi)\langle[A,B]\rangle - \langle\langle A;[B,H]\rangle\rangle\,. \tag{5}$$

These and many other properties of Green's functions as well as the Green function formalism have been described in detail by many authors.[9–12]

The expression for the absorption coefficient describing the electronic transitions $|1\rangle \leftrightarrow |2\rangle$ and $|2\rangle \leftrightarrow |3\rangle$ is determined by the spectral functions which are obtained from $-2\,\mathrm{Im}\langle\langle \alpha_1^+ \alpha_2; \alpha_2^+ \alpha_1\rangle\rangle$ and $-2\,\mathrm{Im}\langle\langle \alpha_3^+ \alpha_2; \alpha_3^+ \alpha_2\rangle\rangle$, where the symbol Im stands for the imaginary part of the expression in question. The Green functions $G_{12,21}(\omega) = \langle\langle \alpha_1^+ \alpha_2; \alpha_2^+ \alpha_1\rangle\rangle$ and $G_{21,12}(\omega) = \langle\langle \alpha_2^+ \alpha_1; \alpha_1^+ \alpha_2\rangle\rangle$ describe the absorption $|1\rangle \to |2\rangle$ and emission $|2\rangle \to |1\rangle$ processes, respectively, which are due to the weak signal field, while the Green functions $G_{32,23}(\omega) = \langle\langle \alpha_3^+ \alpha_2; \alpha_2^+ \alpha_3\rangle\rangle$ and $G_{23,32}(\omega) = \langle\langle \alpha_2^+ \alpha_3; \alpha_3^+ \alpha_2\rangle\rangle$ are the corresponding ones for the processes $|3\rangle \rightarrow |2\rangle$ and $|2\rangle \rightarrow |3\rangle$, respectively, which describe the strong laser field b operating between $|2\rangle \leftrightarrow |3\rangle$ as well as the decay process $|2\rangle \rightarrow |3\rangle$ occurring through the signal field (Fig. 1). Since our concern is the interference effect arising from the strong laser field b into the signal field describing the transition $|1\rangle \rightarrow |2\rangle$ and $|2\rangle \rightarrow |1\rangle$, which denote the signal field absorption and emission processes, respectively, and thus we have to calculate the expressions for $G_{12,21}(\omega)$ and $G_{21,12}(\omega)$ by means of the Hamiltonian (1). Using the symmetry of the Hamiltonian (1), it can be easily shown that in the complex ω plane we have the following relations:

$$G_{12,21}(\omega) = G_{21,12}(-\omega), \quad G_{21,12}(\omega) = G_{12,21}(-\omega). \tag{6}$$

These indicate that the casuality principle is satisfied. Thus, it is necessary only to calculate the expression for the Green function $G_{12,21}(\omega)$ or that for $G_{21,12}(\omega)$ by means of the Hamiltonian (1). Then the imaginary parts of the derived expressions for $G_{12,21}(\omega)$ and $G_{21,12}(\omega)$ determine the spectral functions which describe the excitation spectra for the radiative processes $|1\rangle \rightarrow |2\rangle$ and $|2\rangle \rightarrow |1\rangle$, respectively; this is described in the next and subsequent sections.

III. THE EQUATION OF MOTION FOR $G_{12,21}(\omega)$

Using the Hamiltonian (1), equations of motion (4) and (5), and the initial condition that at times $t = t' = 0$ only for state $|3\rangle$ is occupied while the states $|1\rangle$ and $|2\rangle$ are empty, namely,

$$\begin{aligned} \bar{n}_1 &= \langle \alpha_1^+ \alpha_1 \rangle = 0\,, \\ \bar{n}_2 &= \langle \alpha_2^+ \alpha_2 \rangle = 0\,, \\ \bar{n}_3 &= \langle \alpha_3^+ \alpha_3 \rangle = 1\,. \end{aligned} \tag{7}$$

We derive the equation of motion for the Green function $G_{12,21}(\omega)$ as

$$G_{12,21}(\omega) = [\omega_p^2 f_b / 4d_{12}^2(\omega)] \times \langle\langle \alpha_1^+ \alpha_3 (\beta_b + \beta_b^+); \alpha_3^+ \alpha_1 (\beta_b + \beta_b^+) \rangle\rangle\,, \tag{8}$$

where the propagator $d_{12}(\omega)$ is

$$d_{12}(\omega) = \omega - \omega_{21} - \gamma_+(\omega)\,, \tag{9}$$

$$2\gamma_+(\omega) = \gamma_{21}(\omega) + \gamma_{23}(\omega)\,. \tag{10}$$

The damping function $\gamma_{21}(\omega)$ is defined as

$$\gamma_{21}(\omega) = \frac{1}{2}\,\omega_p^2 \sum_{\mathbf{k},\lambda} \frac{f_{12}(\mathbf{k},\lambda)(\omega_{21}/ck)}{\omega - ck + i\epsilon}\,. \tag{11}$$

The function $\gamma_{21}(\omega)$ is a complex quantity, $\gamma_{21}(\omega) = \gamma'_{21} - i\gamma^0_{21}$, where γ'_{21} gives a small energy shift (Lamb shift), which will be discarded, while γ^0_{21} represents the spontaneous radiative decay for the transition $|2\rangle \rightarrow |1\rangle$ and is given by

$$\gamma^0_{21} = 4/3(\omega_{21}/c)^3 |\boldsymbol{\mu}_{21}|^2\,, \tag{12}$$

where $\boldsymbol{\mu}_{21}$ is the matrix element of the electric dipole moment operator for the transition $|1\rangle \leftrightarrow |2\rangle$. The damping function $\gamma_{23}(\omega) = \gamma'_{23} - i\gamma^0_{23}$ and γ^0_{23} can be obtained from Eqs. (11) and (12), respectively, after making the replacement of 1 by 3 $(1 \leftrightarrow 3)$. In this approximation, the propagator $d_{12}(\omega)$ becomes equal to

$$d_{12}(\omega) = \omega - \omega_{21} + i\gamma^0_+\,, \tag{13a}$$

$$\gamma^0_+ = (\gamma^0_{21} + \gamma^0_{23})/2\,. \tag{13b}$$

It should be pointed out that Eq. (8) is exact provided that the conditions (7) are satisfied and the signal field is a weak perturbing field. We may rewrite Eq. (8) as

$$G_{12,21}(\omega) = (\omega_p^2 f_b / 4d_{12}^2(\omega)) \times [G_{13b,31(b+b^+)}(\omega) + G_{13b^+,31(b+b^+)}(\omega)]\,, \tag{14}$$

where

$$G_{13b,31(b+b^+)}(\omega) = \langle\langle \alpha_1^+ \alpha_3 \beta_b; \alpha_3^+ \alpha_1 (\beta_b + \beta_b^+) \rangle\rangle\,, \tag{15}$$

$$G_{13b^+,31(b+b^+)}(\omega) = \langle\langle \alpha_1^+ \alpha_3 \beta_b^+; \alpha_3^+ \alpha_1 (\beta_b + \beta_b^+) \rangle\rangle\,. \tag{16}$$

The Green function $G_{13b,31(b+b^+)}(\omega)$ describes the physical process where the state $|3\rangle$ of the atom absorbs one photon from the laser field to generate an excitation near the frequency $\omega_{31} + \omega_b$ while the Green function $G_{13b^+,31(b+b^+)}(\omega)$ describes the corresponding process where the state $|3\rangle$ emits one photon to produce the frequency in the neighborhood $\omega_{31} - \omega_b$. In the first order of perturbation approximation, the coupling function between excitations occurring at the frequencies $\omega = \omega_{31} + \omega_b$ and $\omega = \omega_{31} - \omega_b$ is of the order of magnitude $(\Omega_b / 2\omega_b)^2$, where

$$\Omega_b^2 = \omega_p^2 f_b (\tfrac{1}{2} + \bar{n}_b) \approx \omega_p^2 f_b \bar{n}_b\,, \quad \bar{n}_b = \langle \beta_b^+ \beta_b \rangle\,, \tag{17}$$

Ω_b and $\bar{n}_b$ being the Rabi frequency and the average number of photons of the laser field, respectively, with $\bar{n}_b \gg 1$. This coupling between the excitations in question occurs because of the expression

$$\tfrac{1}{2} i\omega_p \sqrt{f_b}\,(\alpha_3^+ \alpha_2 \beta_b - \alpha_2^+ \alpha_3 \beta_b^+)\,, \tag{18}$$

which is part of the fourth term in the Hamiltonian (1) and describes the laser–atom interaction beyond the RWA. In general, the coupling between different types of excitations due to expression (18) is of the order of $(\Omega_b / s\omega_b)^s$, where s is the number of the participating photons in the excitations in question. This is due to the fact that the operators in the expression (18) generate s number of coupled differential equations.

In the optical, ultraviolet (UV) and vacuum-ultraviolet (VUV) region of frequencies, Ω_b is much less than ω_b, $\Omega_b \ll \omega_b$, while in the infrared, radio-, and microwave region of frequencies the relations $\Omega_b < \omega_b$ and $\Omega_b \lessdot \omega_b$ hold, respectively. In the present study, we are concerned with the optical, UV, and VUV region of frequencies for which $\Omega_b \ll \omega_b$ and, hence, the Green functions $G_{13b,31(b+b^+)}(\omega)$ and $G_{13b^+,31(b+b^+)}(\omega)$ will be calculated in the RWA, where the excitations near the frequencies $\omega_{31} + \omega_b$ and $\omega_{31} - \omega_b$ are independent of each other. In this approximation, Eq. (15) may be written as

$$G_{12,21}(\omega) \approx (\omega_p^2 f_b / 4d_{12}^2(\omega)) \times [G^{\mathrm{RWA}}_{13b,31(b+b^+)}(\omega) + G^{\mathrm{RWA}}_{13b^+,31(b+b^+)}(\omega)]\,. \tag{19}$$

Notice that the relation

$$G_{12,21}(\omega) = G^{\rm RWA}_{12,21}(\omega) + G^{\rm RWA}_{21,12}(\omega) \tag{20}$$

is also satisfied.

Using the Hamiltonian (1) in the RWA for the laser–atom interaction and Eq. (4), we derive the following equations of motion:

$$G^{\rm RWA}_{13b,31(b+b^+)}(\omega) = -\frac{(\bar n_3 - \bar n_1)\bar n_b d_{12}(\omega)}{2\pi[d_{12}(\omega)d_{13b}(\omega) - \Omega_b^2/2]}, \tag{21}$$

$$G^{\rm RWA}_{13b^+,31(b+b^+)}(\omega) = -\frac{(\bar n_3 - \bar n_1)\bar n_b d_{12b^+b^+}(\omega)}{2\pi[d_{13b^+}(\omega)d_{12b^+b^+}(\omega) - \Omega_b^2/2]}, \tag{22}$$

where the propagators in Eqs. (21) and (22) are defined as

$$d_{13b}(\omega) = \omega - \omega_{31} - \omega_b, \tag{23}$$

$$d_{13b^+}(\omega) = \omega - \omega_{31} + \omega_b, \tag{24}$$

$$d_{12b^+b^+}(\omega) = \omega - \omega_{21} + 2\omega_b + i\gamma^0_+ . \tag{25}$$

In deriving Eqs. (21) and (22), we have made use of the decoupling approximation[11–13]

$$\langle\langle \alpha_1^+\alpha_3\beta_b^+\beta_b\beta_b;\alpha_3^+\alpha_1(\beta_b+\beta_b^+)\rangle\rangle \approx 2\bar n_b G_{13b,31(b+b^+)}(\omega), \tag{26}$$

which permits photon–photon correlations of the laser field to be taken into account and it is valid in the limit of high photon densities, namely, when $\bar n_b \gg 1$. For a three-level atom,

$$\bar n_1 + \bar n_2 + \bar n_3 = 1, \tag{27a}$$

and when the initial conditions (7) are applicable at $t = t' = 0$, then the factor $(\bar n_3 - \bar n_1)$ in Eqs. (21) and (22) is of the order of unity. At times different than zero, $t = t' \neq 0$, $(\bar n_3 - \bar n_1)$ varies between 0 and 1, $0 < \bar n_3 - \bar n_1 < 1$, and its deviation from unity depends on the depletion of the electron population of the metastable state $|3\rangle$ or, equivalently, on the increase of the electron population of the states $|1\rangle$ and $|2\rangle$, namely,

$$\bar n_3 - \bar n_1 = 1 - (2\bar n_1 + \bar n_2). \tag{27b}$$

If N is the total density of the atoms then

$$\bar n_1 + \bar n_2 + \bar n_3 = N, \tag{27c}$$

and, hence

$$\bar n_3 - \bar n_1 = N - (2\bar n_2 + \bar n_1). \tag{27d}$$

Thus the expressions (21) and (22) are valid in the RWA and in the limit of high photon densities, $\bar n_b \gg 1$. Substitution of Eqs. (21) and (22) into Eq. (19) gives

$$G_{12,21}(\omega) = \Pi_{12,21}(\omega) + S_{12,21}(\omega), \tag{28}$$

where

$$\Pi_{12,21}(\omega) = \frac{(\bar n_3 - \bar n_1)}{4\pi}\left[\frac{1}{d_{12}(\omega)} - \frac{d_{13b}(\omega)}{d_{12}(\omega)d_{13b}(\omega) - \Omega_b^2/2}\right], \tag{29}$$

$$S_{12,21}(\omega) = -\frac{(\bar n_3 - \bar n_1)\Omega_b^2 d_{12b^+b^+}(\omega)}{8\pi d_{12}^2(\omega)[d_{13b^+}(\omega)b_{12b^+b^+}(\omega) - \Omega^2_{b/2}]}. \tag{30}$$

Equations (28)–(30) will be used in the next section to discuss the excitation spectra for the problem under investigation. In the limit when $\Omega_b \to 0$ the functions $\Pi_{12,21}(\omega) \mapsto 0$ and $S_{12,21}(\omega) \mapsto 0$ and, hence the functions $\Pi_{12,21}(\omega)$ and $S_{12,21}(\omega)$, describe stimulated one-photon and three-photon spectra, respectively.

IV. EXCITATION SPECTRA

To facilitate the study of the excitation spectra, we introduce the dimensionless variables

$$X = (\omega - \omega_{21})/\gamma^0_+, \quad Y = (\omega - \omega_{21} + 2\omega_b)/\gamma^0_+, \tag{31}$$

$$\nu_b = (\omega_{23} - \omega_b)/\gamma^0_+, \quad \eta_b = \Omega_b/\gamma^0_+, \quad \eta = \frac{\Omega}{\gamma^0_+} = \left(\frac{\nu_b^2}{4} + \frac{\eta_b^2}{2}\right)^{1/2}, \tag{32}$$

where

$$\Omega^2 = (\omega_{23} - \omega_b)^2/4 + \Omega_b^2/2. \tag{33}$$

In Eqs. (31)–(33), X and Y are the reduced frequencies for the signal field and for the three-photon process, respectively; ν_b and η_b represent the relative detuning and the Rabi frequency of the laser field, respectively; while η denotes the total relative frequency shift arising from the Rabi frequency and the detuning of the laser field. Within this convention the radiative width appears in units of γ^0_+. Substituting Eqs. (31)–(33) into Eqs. (29) and (30), we obtain

$$\Pi_{12,21}(\omega) = \frac{(\bar n_3 - \bar n_1)}{4\pi\gamma^0_+}\left[\frac{1}{X+i} - \frac{1}{2}\left(\frac{1+\nu_b/2\eta - i/2\eta}{X+\frac12\nu_b - \eta + i/2} + \frac{1-\nu_b/2\eta + i/2\eta}{X+\frac12\nu_b+\eta+i/2}\right)\right], \tag{34}$$

$$S_{12,21}(\omega) = -\frac{(\bar n_3 - \bar n_1)\Omega_b^2}{16\pi\gamma^0_+ d_{12}^2(\omega)}\left[\frac{1+\nu_b/2\eta - i/2\eta}{Y+\frac12\nu_b - \eta + i/2} + \frac{1-\nu_b/2\eta + i/2\eta}{Y+\frac12\nu_b+\eta+i/2}\right]. \tag{35}$$

The excitation spectra are described by the spectral function which is determined by the imaginary part of the Green function $G_{12,21}(\omega)$, namely, $-2\,\mathrm{Im}\,G_{12,21}(\omega)$, and it is derived from Eqs. (28), (34), and (35) in the form

$$\begin{aligned} P_{12}(\omega) &\equiv -2\,\mathrm{Im}\,G_{12,21}(\omega) \\ &= -2\,\mathrm{Im}\,[\Pi_{12,21}(\omega) + S_{12,21}(\omega)] \\ &= \frac{(\bar n_3 - \bar n_1)}{2\pi\gamma^0_+}\left[F_{12}(X) + \frac{\Omega_b^2}{4d_{12}^2(\omega)}J_{12}(Y)\right], \end{aligned} \tag{36}$$

where the shape functions $F_{12}(X)$ and $J_{12}(Y)$ are given by

$$F_{12}(X) = 1/(X^2+1) - \tfrac12 L(X,\nu_b,\eta) - \tfrac12 L(X,\nu_b,-\eta), \tag{37}$$

$$J_{12}(Y) = -L(Y,\nu_b,\eta) - L(Y,\nu_b,-\eta), \tag{38}$$

while the functions $L(X,\nu_b,\mp\eta)$ and $L(Y,\nu_b,\pm\eta)$ describing the shape of the sidebands at the relative frequencies

$X = -\frac{1}{2}\nu_b \pm \eta$ and $Y = -\frac{1}{2}\nu_b \pm \eta$, respectively, are defined as

$$L(X,\nu_b, \pm\eta) = \frac{\frac{1}{2}(1 \pm \nu_b/2\eta) \pm (X + \frac{1}{2}\nu_b \mp \eta)/2\eta}{(X + \frac{1}{2}\nu_b \mp \eta)^2 + \frac{1}{4}}, \tag{39}$$

and the function $L(Y,\nu_b, \pm\eta)$ is derived from Eq. (39) when X is replaced everywhere by $Y, X \leftrightarrow Y$.

The expression (36) for $P_{12}(\omega)$ defines the absorption coefficient of the signal field describing the physical processes near the frequencies $\omega = \omega_{21}$ $(X = 0)$ and $\omega = \omega_{21} - 2\omega_b$ $(Y = 0)$, respectively. Hence, positive and negative values of the function $P_{12}(\omega)$ indicate that the physical processes of absorption and amplification (stimulated emission) of the signal field takes place, respectively. If ω in (36) is replaced everywhere by $-\omega$ then the expression $P_{12}(-\omega)$ describes the emission spectra in the neighborhood of the frequencies $\omega = \omega_{21}$ and $\omega_b = 2\omega_b - \omega_{21}$, respectively.

A. Stimulated one-photon process

The spectral function $F_{12}(X)$ given by Eq. (37) describes the excitation spectra near the reduced frequency X. The first term on the right-hand side (rhs) describes the central peak of the signal field, which is a Lorentzian line with maximum relative intensity equal to unity $(h_c = 1)$ at $X = 0$ and half-width equal to γ^0_+. The second and third terms, respectively, describe a pair of sidebands with maximum relative intensities at zero detunings $(\nu_b = 0)$ equal to $-1(h_\pm = -1)$ which are peaked at $X = \pm\eta$; the half-width of each sideband is $\frac{1}{2}\gamma^0_+$. The shape of the sidebands is determined by Eq. (39), which are asymmetric Lorentzian lines peaked at $X = \pm\eta - \frac{1}{2}\nu_b$ with negative relative intensities and the shape of the lines depends strongly on the value of the detuning ν_b. The ratio of the maximum height h_c of the central peak at $X = 0$ to those of the sidebands $h_\pm$ at $X = \pm\eta - \frac{1}{2}\nu_b$, respectively, is given by

$$h_c/h_\pm = -1/(1 \pm \nu_b/2\eta). \tag{40}$$

Since the intensity of the central peak is positive while those of the sidebands are negative, the spectral function $F_{12}(X)$ given by Eq. (37) describes the physical process of absorption of $X = 0$ and amplification (stimulated emission) at $X = \pm\eta - \frac{1}{2}\nu_b$ and $\nu_b = 0$ as well as with $\nu_b \neq 0$, respectively. As an illustration, the function $F_{12}(X)$ denoted as the relative intensity $= F_{12}(X)$ is plotted versus the reduced relative frequency X in Figs. 2(a)–2(d) for a constant value of $\eta_b = 10$ and various values of detunings ν_b. Figure 2(a) illustrates the spectra at resonance $(\nu_b = 0)$ and for $\eta_b = 10$, where the relative intensity of the central peak at $X = 0$ and those of the sidebands at $X = \pm\eta_b/\sqrt{2}$ are equal but opposite in sign indicating that signal field absorption and amplification will take place at the corresponding frequencies. Figures 2(b)–2(d) depict the spectra for values of ν_b equal to 5, 10, and 20, respectively. It is shown from Figs. 2(b)–2(d) and Eq. (39) that as the value of the detuning increases the intensities of the sidebands at $X = \eta - \frac{1}{2}\nu_b$ and $X = -\eta - \frac{1}{2}\nu_b$ increase and decrease, respectively, provided that $\nu_b/2\eta < 1$, which is always true for $\eta_b > 1$. In the

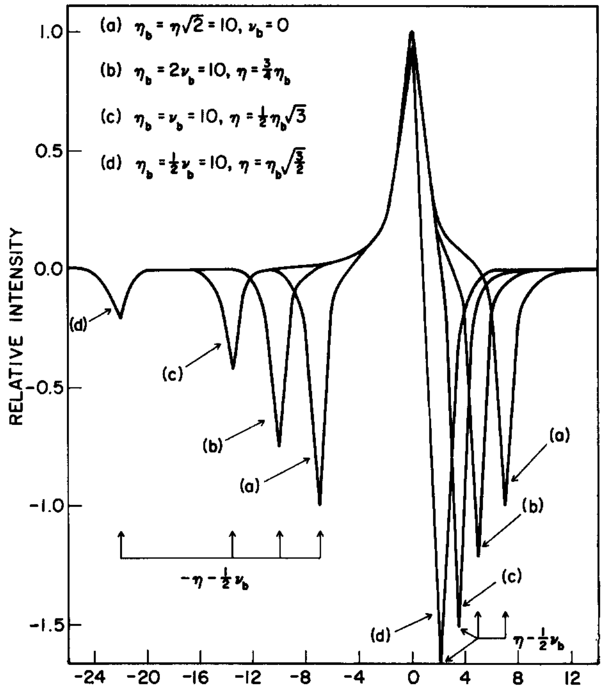


FIG. 2. One-photon spectra. The relative intensity $= F_{12}(X)$ is computed from the rhs of Eq. (37) and is plotted versus the relative frequency $X = (\omega - \omega_{21})/\gamma^0_+$ for the relative Rabi frequency $\eta_b = 10$ and various detunings. (a) $\nu_b = 0$, (b) $\nu_b = 5$, (c) $\nu_b = 10$, and (d) $\nu_b = 20$.

limit for values of $\nu_b > \eta_b$ for $\nu_b/2\eta < 1$, $h_c \to 1$, $h_- \to 0$ while h_+ takes its maximum negative value $h_+ \to -2$ at $X = \eta - \frac{1}{2}\nu_b$. A schematic representation of the splitting of the excitate state $|2\rangle$ is shown in Fig. 3.

B. Stimulated three-photon processes

The second term on the rhs of Eq. (36) represents stimulated three-photon absorption spectra, which are generated near the frequency $\omega = \omega_{21} - 2\omega_b > 0$, and describe the physical process where one photon of the signal field with frequency ω_{21} is absorbed while two photons of the laser field with frequency $2\omega_b$ are emitted. Correspondingly, the emission process generates the spectra near the frequency $\omega = 2\omega_b - \omega_{21} > 0$, where two photons of the laser field with frequency $2\omega_b$ are absorbed while a photon of the signal field with frequency ω_{21} is emitted. In Eq. (36), the factor $\Omega^2_b/4d^2_{12}(\omega)$ may be taken as

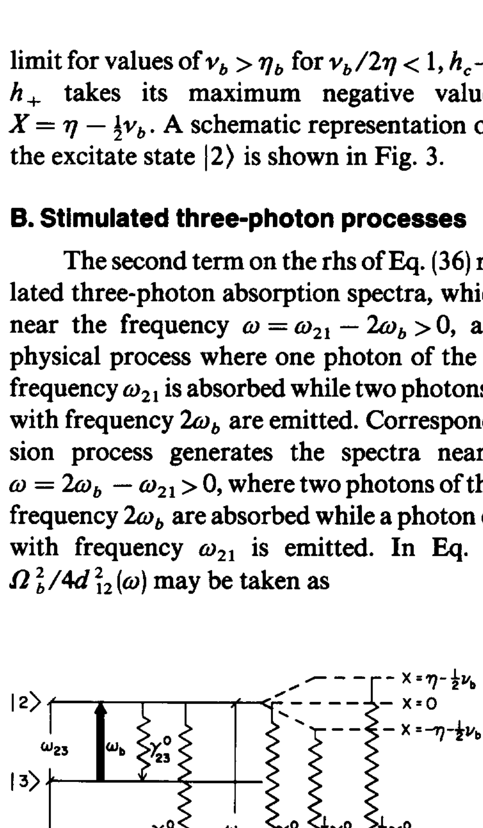


FIG. 3. Schematic representation of the energy splitting at the relative frequency $X = (\omega - \omega_{21})/\gamma^0_+$.

$$\frac{\Omega_b^2}{4d_{12}^2(\omega)} = \frac{\Omega_b^2}{4(\omega - \omega_{21} + i\gamma_+^0)^2} \approx \frac{\Omega_b^2}{4(-2\omega_b + i\gamma_+^0)^2} \approx \left(\frac{\Omega_b}{4\omega_b}\right)^2, \tag{41}$$

where ω has been replaced by its approximate value at $\omega \approx \omega_{21} - 2\omega_b$. The factor $\Omega_b^2/4d_{12}^2(\omega)$ implies that the spectral function $J_{12}(Y)$ in Eq. (36) describes stimulated third-order nonlinear spectra near the frequency Y. At resonance, $\nu_b = 0$, and when Eq. (27b) is satisfied then for $\omega_{21} = \omega_a$, the second term on the rhs of Eq. (36) is identical to the corresponding one given by Eqs. (11) and (12) in Ref. 8, which describe the spectra near the frequency $\omega = \omega_a - 2\omega_b$.

The spectral function $J_{12}(Y)$ given by Eqs. (38) describes a doublet with maximum intensities $i_\pm$ at the frequencies $Y = \pm\eta - \frac{1}{2}\nu_b$ equal to

$$i_\pm = -2(1 \pm \nu_b/2\eta), \tag{42}$$

respectively, and half-widths equal to $\frac{1}{2}\gamma_+^0$. The intensities of the doublet are always negative and, hence, the physical process of amplification is expected to take place at the frequencies $Y = \pm\eta - \frac{1}{2}\nu_b$ and the shape of the doublet depends on the value of the ratio $\nu_b/2\eta$. The spectra are illustrated in Figs. 4(a)–4(d), where the function $J_{12}(Y)$ is plotted as $J_{12}(Y)$ = relative intensity versus the relative frequency Y for the Rabi frequency $\eta_b = 10$ and various detunings ν_b. The resonance case, $\nu_b = 0$, is depicted in Fig. 4(a), which is identical to Fig. 2(a) in Ref. 8 when $\omega_{21} = \omega_a$ and Eq. (27b) is satisfied. In this case the maximum relative intensity of the doublet remains constant and equal to -2. Figures 4(b)–4(d) illustrate the spectra for values of the detunings $\nu_b = 5$, 10, and 20, respectively. It is shown that the shape of the doublet in Figs. 4(b)–4(d) depends strongly on the value of the ratio $\nu_b/2\eta$. A schematic representation of the energy splitting which occurs at the relative frequency Y is depicted in Fig. 5.

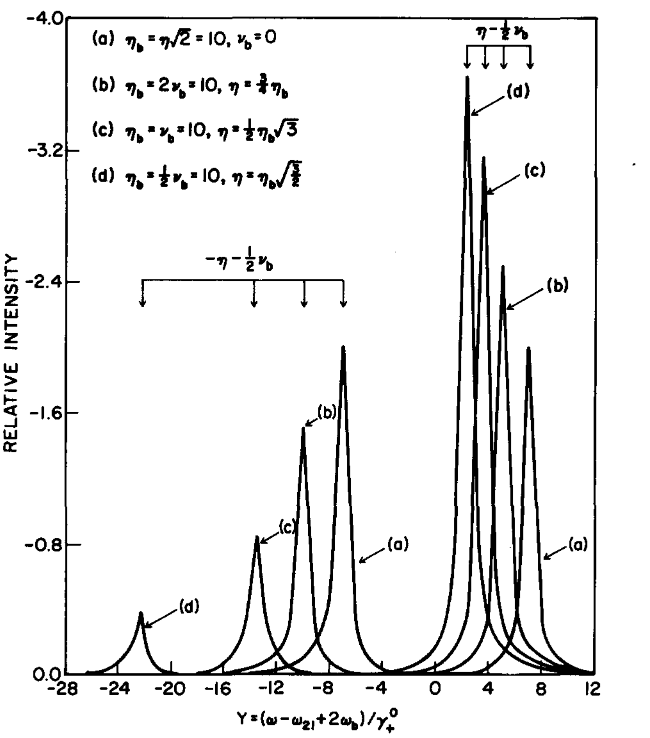


FIG. 4. Three-photon spectra. The relative intensity $= J_{12}(Y)$ is computed from the rhs of Eq. (38) and is plotted versus the relative frequency $Y = (\omega - \omega_{21} + 2\omega_b)/\gamma_+^0$ for the relative Rabi frequency $\eta_b = 10$ and various detunings. (a) $\nu_b = 0$, (b) $\nu_b = 5$, (c) $\nu_b = 10$, and (d) $\nu_b = 20$.

V. STIMULATED RAMAN SPECTRA

In order to calculate the excitation spectra near the frequency $\omega = \omega_{31} \approx \omega_{21} - \omega_b$, we decouple the Green function appearing on the rhs of Eq. (8) as follows:

$$\begin{aligned}\langle\langle \alpha_1^+\alpha_3(\beta_b + \beta_b^+); \alpha_3^+\alpha_1(\beta_b + \beta_b^+)\rangle\rangle &\approx \langle(\beta_b + \beta_b^+)(\beta_b + \beta_b^+)\rangle G_{13,31}(\omega) \\ &\quad - \langle n_3(1 - n_1)\rangle G_{b,b}(\omega),\end{aligned} \tag{43}$$

where

$$G_{13,31}(\omega) = \langle\langle \alpha_1^+\alpha_3; \alpha_3^+\alpha_1\rangle\rangle,$$
$$G_{b,b}(\omega) = \langle\langle \beta_b + \beta_b^+; \beta_b + \beta_b^+\rangle\rangle \tag{44}$$

are the Green functions describing excitations near the frequencies $\omega \approx \omega_{31} \approx \omega_{21} - \omega_b$ and $\omega = \pm\omega_b$, respectively, while $\langle(\beta_b^+ + \beta_b)(\beta_b + \beta_b^+)\rangle \approx 1 + 2\bar{n}_b \approx 2\bar{n}_b$, for $\bar{n}_b \gg 1$. The decoupling approximation (43) indicates that the photon mode ω_b does not interact with the atomic excitation operator $\alpha_1^+\alpha_3$. This is a reasonable assumption since the electronic transition $|1\rangle \leftrightarrow |3\rangle$ is electric dipole forbidden from parity considerations. Such decouplings have been used in the literature for physical systems describing electron–phonon[14] and electron–electron[15–17] interactions in solids as well as for electron–photon processes[18–20] in atomic systems. Substituting Eq. (43) into Eq. (8) we have

$$G_{12,21}(\omega) = \widetilde{G}_{12,21}(\omega) + R_{12}(\omega), \tag{45}$$

where

$$R_{12}(\omega) = (\Omega_b^2/2d_{12}^2(\omega))G_{13,31}(\omega). \tag{46}$$

The Green function $\widetilde{G}_{12,21}(\omega)$ describes excitations in the neighborhood of the frequencies $\omega = \pm\omega_b$ and will not be relevant to our investigation while the Green function $G_{13,31}(\omega)$ describes excitations near $\omega = \omega_{31} \approx \omega_{21} - \omega_b$, which are of interest to us. These excitations in the neighborhood of the frequency $\omega = \omega_{31} \approx \omega_{21} - \omega_b$ are generated by interference effects arising between the signal field and the laser field and represent the physical process, where, simultaneously, a photon of the signal field with frequency ω_{21} is absorbed while a photon of the laser field with frequency ω_b is emitted. Thus the spectra described by the function $R_{12}(\omega)$ given by Eq. (46) may be attributed as due to the absorption

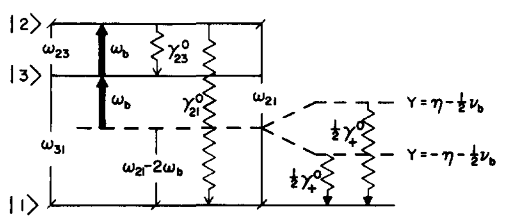


FIG. 5. Schematic representation of the energy splitting at the relative frequency $Y = (\omega - \omega_{21} + 2\omega_b)/\gamma_+^0$.

of the Raman frequency $\omega = \omega_{31} \approx \omega_{21} - \omega_b$. Then $d_{12}(\omega)$ in Eq. (46) may be approximated by

$$d_{12}^2(\omega) = (\omega - \omega_{21} + i\gamma_+^0)^2 \approx d_{12}^2(\omega_{31})$$
$$= (\omega_{31} - \omega_{21} + i\gamma_+^0)^2 \approx (-\omega_b + i\gamma_+^0)^2 \approx \omega_b^2 , \quad (47)$$

where ω has been replaced by its approximate value at $\omega \approx \omega_{31} \approx \omega_{21} - \omega_b$ and Eq. (47) may be written as

$$R_{12}(\omega) \approx (\Omega_b^2/2\omega_b^2) G_{13,31}(\omega) . \quad (48)$$

It should be pointed out that, apart from a factor of 2, Eq. (46) could have been derived if the original Hamiltonian had been taken in the RWA. The factor of 2, which is derived from the first term on the rhs of Eq. (43), is due to the fact that the Hamiltonian (1) is beyond the RWA. The expressions (46) and (48) are analogous to the corresponding Eqs. (8) and (11) in Ref. (19), which describe the Raman spectra at the frequencies $\omega = \pm(\omega_a - \omega_b)$ arising from the interference between two laser fields with frequencies ω_a and ω_b, respectively, interacting with a three-level atom in the "V" configuration.[19,20] The factors $\Omega_b^2/2d_{12}^2(\omega)$ and $\Omega_b^2/2\omega_b^2$ in Eqs. (46) and (48), respectively, imply that $R_{12}(\omega)$ describes stimulated third-order nonlinear spectra near the frequency $\omega = \omega_{31} = \omega_{21} - \omega_b$.

To proceed further we have to calculate the Green function $G_{13,31}(\omega)$ by means of the Hamiltonian (1) in the limit of high photon densities of the laser field where $\bar{n}_b \gg 1$. We are interested in calculating the excitation spectra near the frequency $\omega = \omega_{31} \approx \omega_{21} - \omega_b$ including the spectra of the sidebands which are induced by the laser field. However, we are not concerned with the spectra arising from the coupling between the two-photon excitation at the frequencies $\omega = \omega_{21} - \omega_b$ and $\omega = \omega_{21} + \omega_b$ whose coupling constant is of the order $(\Omega_b/2\omega_b)^2$, namely, the coupling between these two excitations describe nonlinear effects induced into the excitation spectra at the frequencies in question, respectively. Hence, the discard of the coupling between the two excitations in question is equivalent of calculating the Green function $G_{13,31}(\omega)$ through the Hamiltonian (1) in the RWA, which is found to be

$$G_{13,31}^{\mathrm{RWA}}(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1)}{4\pi} \times \left[\frac{1}{d_{13}(\omega)} + \frac{d_{12b^+}(\omega)}{d_{13}(\omega) d_{12b^+}(\omega) - \Omega_b^2/2}\right], \quad (49)$$

where the propagators $d_{13}(\omega)$ and $d_{12b^+}(\omega)$ are defined as

$$d_{13}(\omega) = \omega - \omega_{31} , \quad (50)$$

$$d_{12b^+}(\omega) = \omega - \omega_{21} + \omega_b + i\gamma_+^0 . \quad (51)$$

In deriving Eq. (49) we have made use of the decoupling approximation (26) and, therefore, the expression (49) is valid in the limit of high photon densities of the laser field for which $\bar{n}_b \gg 1$.

If we introduced the dimensionless variable

$$Z = (\omega - \omega_{31})/\gamma_+^0 , \quad (52)$$

then using Eqs. (46)–(52) we obtain

$$R_{12}(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1)\Omega_b^2}{8\pi\gamma_+^0 d_{12}^2(\omega)} \left[\frac{1}{Z} + \frac{1}{2}\left(\frac{1 - \nu_b/2\eta + i/2\eta}{Z - \frac{1}{2}\nu_b - \eta + i/2} + \frac{1 + \nu_b/2\eta - i/2\eta}{Z - \frac{1}{2}\nu_b + \eta + i/2}\right)\right] . \quad (53)$$

Taking the imaginary of Eq. (53), we have

$$-2\,\mathrm{Im}\,R_{12}(\omega) = \frac{(\bar{n}_3 - \bar{n}_1)\Omega_b^2}{4\pi\gamma_+^0 d_{12}^2(\omega)} [-\pi\delta(Z) + I_{12}(Z)] , \quad (54)$$

where

$$I_{12}(Z) = -\frac{1}{2}\frac{\frac{1}{2}(1 - \nu_b/2\eta) - (Z - \frac{1}{2}\nu_b - \eta)/2\eta}{(Z - \frac{1}{2}\nu_b - \eta)^2 + \frac{1}{4}} - \frac{1}{2}\frac{\frac{1}{2}(1 + \nu_b/2\eta) + (Z - \frac{1}{2}\nu_b + \eta)/2\eta}{(Z - \frac{1}{2}\nu_b + \eta)^2 + \frac{1}{4}} . \quad (55)$$

The expression (54) describes the stimulated Raman spectra near the relative frequency Z and it is valid in the limit when $\bar{n}_b \gg 1$. The first term on the rhs of Eq. (54) describes the main peak, which has a delta-function distribution at the frequency $Z = 0$ indicating the stability of the mode in question. This is in agreement with Breit's[21] suggestion denoting the absence of spontaneous emission for atoms having a common upper level and two different lower levels.[22,23] The function $I_{12}(Z)$ given by Eq. (55) describes a pair of sidebands, which are peaked at the frequencies $Z = \frac{1}{2}\nu_b \pm \eta$ and have radiative half-widths equal to $\frac{1}{2}\gamma_+^0$, respectively. Thus, although the main peak at $Z = 0$ is stable (nonradiative), the two sidebands decay radiatively with a lifetime of the order $2/\gamma_+^0$. The shape of the sidebands depends strongly on the value of the ratio $\nu_b/2\eta$ while their intensities are always negative, which implies that the physical process of amplification occurs at the frequencies $Z = \frac{1}{2}\nu_b \pm \eta$ and their maximum values $I_\pm$ are given by

$$I_\pm = -(1 \mp \nu_b/2\eta) . \quad (56)$$

The Raman spectra are illustrated in Figs. 6(a)–6(d), where the function $I_{12}(Z)$ is plotted as the relative intensity $= I_{12}(Z)$ versus the relative frequency Z for the Rabi frequency $\eta_b = 10$ and for various values of the detuning $\nu_b = 0$, 5, 10, and 20, respectively. Figure 6(a) implies that at resonance, $\nu_b = 0$, the doublet has equal negative intensities; for finite values of the detuning ν_b, as the value of ν_b increases the negative intensity of the peak at $Z = \frac{1}{2}\nu_b + \eta$ decreases while the corresponding one at $Z = \frac{1}{2}\nu_b - \eta$ increases as is depicted in Figs. 6(b)–6(d). Finally as ν_b takes values greater than η_b, $\nu_b > \eta_b$, but the inequality $\nu_b/2\eta < 1$ must be always satisfied, the intensity of the sideband at $Z = \frac{1}{2}\nu_b + \eta$ becomes negligibly small, $I_+ \to -0$ while the corresponding one at $Z = \frac{1}{2}\nu_b - \eta$ takes its maximum negative value, $I_- \to -2$. The splitting of the state $|3\rangle$ at the relative frequency Z is shown schematically in Fig. 7.

VI. DISCUSSION

We have considered the excitation spectra for a three-level atom as shown in Fig. 1, where a strong laser field operates near resonance between the states $|2\rangle$ and $|3\rangle$, $|2\rangle \leftrightarrow |3\rangle$, while the signal field, which is assumed to be a weak perturbing field, describes the radiative decays

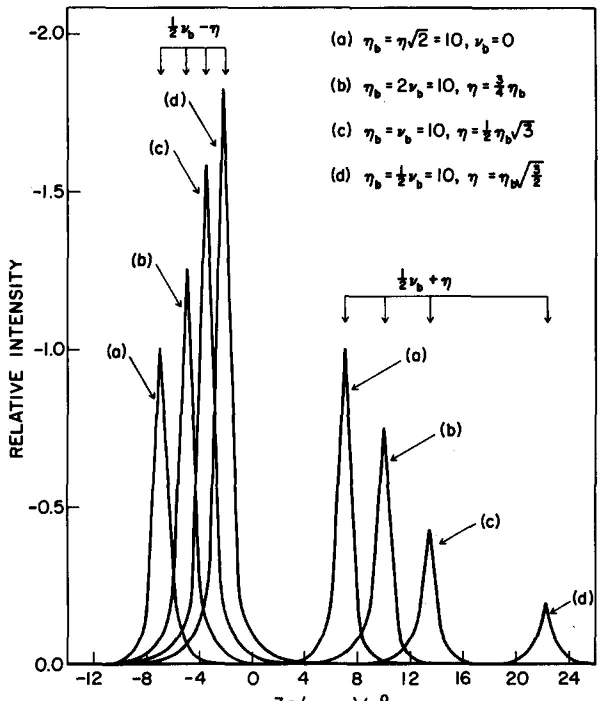


FIG. 6. Two-photon or Raman spectra. The relative intensity $= I_{12}(Z)$ computed from the rhs of Eq. (55) is plotted versus the relative frequency $Z = (\omega - \omega_{31})/\gamma_+^0$ for the relative Rabi frequency $\eta_b = 10$ and various detunings. (a) $\nu_b = 0$, (b) $\nu_b = 5$, (c) $\nu_b = 10$, and (d) $\nu_b = 20$.

$|2\rangle \to |3\rangle$ and $|2\rangle \to |1\rangle$, respectively. The state $|3\rangle$ has the same parity as the state $|1\rangle$ and, hence, the state $|3\rangle$ as a metastable one is initially populated while the remaining states $|1\rangle$ and $|2\rangle$ are empty. The electron population of the state $|3\rangle$ is depleted by the action of the laser which excites the electrons into the state $|2\rangle$, where the electrons emit photons through the signal field and decay into the states $|3\rangle$ and $|1\rangle$ simultaneously. The energy shifts (Lamb shifts) induced by the signal field have been discarded as being negligibly small in comparison to the corresponding Rabi frequency induced by the laser field.

Using the Green function method, we have calculated the excitation spectra of the signal field in the limit of high photon densities of the laser field for the following processes.

(i) One photon process. The spectral function is given by the first term on the rhs of Eq. (36), where the function $F_{12}(X)$ is defined by Eqs. (37) and (39) and is graphically presented in Figs. 2(a)–2(d); the splitting of the state $|2\rangle$ at the frequency X is schematically depicted in Fig. 3. The intensity of the central peak at $X = 0$ or at $\omega = \omega_{21}$ is positive indicating signal-field absorption while those of the pair of sidebands at $X = \pm\eta - \frac{1}{2}\eta_b$ are negative implying the stimulated emission of the signal field. The one-photon process discussed here is the three-level analog to that for the two-level system predicted by Mollow[1] and confirmed experimentally by Wu *et al.*[2] It is pointed out that for the three-level system under investigation the sum of the intensities is equal to $h_+ + h_- = -2$ for $\nu_b = 0$ and $h_+ + h_- = -2/(1 - \nu_b^2/4\eta^2)$ for $\nu_b \neq 0$ while the intensity of the central peak is $h_c = 1$. For the two-level system,[1,2] stimulated emission takes place for certain values of photon densities but the gain is rather small; for instance, the observed maximum gain[2] was approximately 0.4% with a resonant laser-field intensity of 130 mW/cm^2. Thus the results of the present study are encouraging and suggesting that the amplification of the signal field in a three-level system might occur more easily and greater gains might be achieved than for the corresponding two-level system. Of course, this could only be verified experimentally.

(ii) Stimulated three-photon process. The second term on the rhs of Eq. (36) describes the stimulated three-photon process near the frequency $\omega = \omega_{21} - 2\omega_b$, where one photon of the signal field is absorbed while two photons of the laser field are emitted simultaneously. The shape function $J_{12}(Y)$ is defined by Eq. (38) and describes a doublet whose intensity is always negative implying amplification while the shape of the doublet depends on the value of the ratio $\nu_b/2\eta$. The calculated spectra are shown in Figs. 4(a)–4(d) while the splitting that occurs at the frequency Y is depicted schematically in Fig. 5. The stimulated three-photon process discussed here seems to be analogous to the three-photon process which has been used in Ref. 2 to interpret the observed absorption and amplification peaks in the off-resonance spectra, Fig. 4(c) in Ref. 2. However, the schematic representation given by Fig. 5 may be the appropriate one for our three-photon process.

(iii) Stimulated two-photon process. The spectral function for the stimulated Raman spectra at $\omega = \omega_{31} \approx \omega_{21} - \omega_b$ is determined by the expressions (54) and (55). The spectra consist of a main peak at $\omega = \omega_{31}$ or $Z = 0$, which has a delta-function distribution, and a pair of sidebands peaked at $Z = \frac{1}{2}\nu_b \pm \eta$ which have negative intensities implying amplification at the frequencies in question; the half-widths of the sidebands are $\frac{1}{2}\gamma_+^0$. The spectra of the doublet are shown in Figs. 6(a)–6(d) while the splitting of the state $|3\rangle$ is depicted schematically in Fig. 7. The spectra for the stimulated Raman process discussed here have, qualitatively, the three main characteristic properties that have been observed experimentally by Tam,[7] namely: (1) the sidebands appear at frequencies $Z \neq 0$ and, in particular, at $Z = \frac{1}{2}\nu_b \pm \eta$; (2) the sidebands always have negative intensities indicating amplification; and (3) the factor $(\bar{n}_3 - \bar{n}_1)\Omega_b^2$ in Eq. (54) implies that the relative intensity of the sidebands depends on the photon density $\bar{n}_b$ of the laser field as well as on the atomic density N through Eq. (27d). Thus, at suitable atomic density, photon density of the laser field and detuning will generate strong sideband amplification at the frequencies $Z = \frac{1}{2}\nu_b \pm \eta$. Hence, the stimulated Raman spectra discussed here agree at least qualitatively with the observed spectra by Tam.[7] For a quantitative comparison

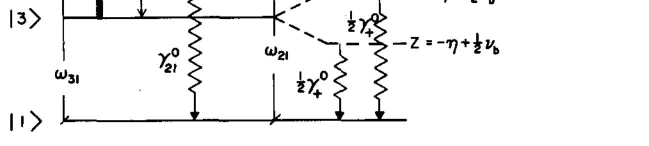


FIG. 7. Schematic representation of the energy splitting at the relative frequency $Z = (\omega - \omega_{31})/\gamma_+^0$.

one has to have accurate experimental data regarding the parameters involved. We hope that the present study will encourage experiments such as to facilitate a quantitative comparison between theory and experiment.

[1] B. R. Mollow, Phys. Rev. A **5**, 2217 (1972).
[2] F. Y. Wu, S. Ezekiel, M. Ducloy, and B. R. Mollow, Phys. Rev. Lett. **38**, 1077 (1977).
[3] H. W. Galbraith, M. Dubs, and J. I. Steinfeld, Phys. Rev. A **26**, 1528 (1982).
[4] B. Senitzky, G. Gould, and S. Cutler, Phys. Rev. **130**, 1460 (1963); B. Senitzky and S. Cutler, Microwave J. **1**, 62 (1964).
[5] A. M. Bonch-Bruevich, V. A. Khodovoi, and N. A. Chigir, Zh. Eksp. Teor. Fiz. **67**, 2069 (1974) [Sov. Phys. JETP **40**, 1027 (1975)].
[6] S. L. McCall, Phys. Rev. A **9**, 1515 (1974); H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, Phys. Rev. Lett. **36**, 1135 (1976).
[7] A. C. Tam, Phys. Rev. A **19**, 1971 (1979).
[8] C. Mavroyannis, Opt. Commun. **46**, 323 (1983).
[9] D. N. Zubarev, Usp. Fiz. Nauk **71**, 71 (1960) [Sov. Phys. Usp. **3**, 320 (1960)]; *Nonequilibrium Statistical Thermodynamics* (Plenum, New York, 1974), Chap. 16.
[10] V. L. Bonch-Bruevich and S. V. Tyablikov, *The Green Function Method in Statistical Mechanics* (North-Holland, Amsterdam, 1962).
[11] S. V. Tyablikov and V. L. Bonch-Bruevich, Adv. Phys. **11**, 317 (1962).
[12] C. Mavroyannis, "The Green Function Method," in *Physical Chemistry*, edited by H. Eyring, W. Jost, and D. Henderson (Academic, New York, 1975), Vol. XI A, Chap. 8.
[13] C. Mavroyannis, Phys. Rev. A **18**, 185 (1978); **22**, 1129 (1980); **27**, 1414 (1983).
[14] V. L. Bonch-Bruevich, "Optical Properties of Solids," in *Proceedings of the International School "Enrico Fermi,"* Varenna, Italy (Academic, New York, 1966), p. 358.
[15] V. L. Bonch-Bruevich and R. Rozmann, Sov. Phys. Solid State **5**, 2117 (1964).
[16] V. L. Bonch-Bruevich, *The Electronic Theory of Heavily Doped Semiconductors* (Elsevier, New York, 1966), Chap. 13.
[17] C. Mavroyannis, Phys. Rev. B **6**, 2463 (1972).
[18] C. Mavroyannis, Can. J. Phys. **58**, 957 (1980).
[19] C. Mavroyannis, Mol. Phys. **48**, 847 (1983).
[20] C. Mavroyannis, J. Phys. Chem. **88**, 4868 (1984).
[21] G. Breit, Rev. Mod. Phys. **5**, 91 (1933); P. W. Milonni, Phys. Rep. **25**, 53 (1976); A. Schenzle and R. G. Brewer, in *Laser Spectroscopy*, edited by S. Haroche, J. C. Pebay-Peyroula, T. W. Hänsch, and S. E. Harris (Springer, New York, 1975), p. 420.
[22] W. W. Chow, M. O. Scully, and J. O. Stoner, Phys. Rev. A **11**, 1380 (1970).
[23] R. M. Herman, H. Grotch, R. Kornblith, and J. H. Eberly, Phys. Rev. A **11**, 1389 (1975).