Algebras of diffeomorphisms of the N-torus

E. Ramos
Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794-3840

C. H. Sah
Department of Mathematics, State University of New York, Stony Brook, New York 11794-3651

R. E. Shrock
Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794-3840

(Received 12 December 1989; accepted for publication 4 April 1990)

The general algebra $\mathscr{D}(N,V)$ of infinitesimal diffeomorphisms of the N-torus $(S^1)^N$ involving generators $L_{v,m}$ depending on a structure vector $v\in\mathbb{C}^N$ and a vector-valued index $m\in\mathbb{Z}^N$ is constructed. Several results are proved for this algebra. Special cases of the algebra were previously presented in Ramos–Shrock [E. Ramos and R. E. Shrock, Int. J. Mod. Phys. A **4**, 4279 (1989).] The concept of the $\mathbb{Q}$ rank of the space V of structure vectors, denoted $r_{\mathbb{Q}}(V)$, is defined and certain related "Cartan matrices" are introduced. It is shown that $\mathscr{D}(N,V)$, as an ungraded algebra, is simple if and only if $r_{\mathbb{Q}}(V)=N$, i.e., the $\mathbb{Q}$ rank of V is maximal. A classification under isomorphisms is given for the algebra and is shown to reduce to a classification of the Cartan matrices. The space of structure vectors is isomorphic to the unique "Cartan subalgebra." A number of properties concerning the central extensions of these algebras are then proved. For the case dim $V=1$, $r_{\mathbb{Q}}(V)=N$, it is shown that the central extension is unique. For the case dim $V=1, r_{\mathbb{Q}}(V)<N$, a new and greatly enlarged central extension is constructed involving a central charge *function* from $\mathbb{Z}^{N-r}$ to $\mathbb{C}$ (essentially equivalent to an infinite number of central charge parameters), rather than a single central charge parameter. Finally, it is proved that for dim $V\geqslant 2$, this algebra has no nontrivial central extension.

I. INTRODUCTION

The Virasoro algebra is a complex Lie algebra Vir with basis c and L_m, $m\in\mathbb{Z}$, which satisfy the commutation relations

$$[L_m,L_n]=(m-n)L_{m+n}+(c/12)\times(m^3-m)\delta_{m+n,0},[L_m,c]=0, \tag{1.1}$$

where $m,n\in\mathbb{Z}$ and c is called the central charge. According to a theorem of Gel'fand and Fuchs,[1] it is isomorphic to the universal central extension of the complexified algebra Diff(S^1) of infinitesimal diffeomorphisms (or polynomial vector fields) of the circle. (For reviews, see Refs. 2 and 3.) By viewing Vir as a real Lie algebra and tensoring it with $\mathbb{C}$ (over $\mathbb{R}$), we get a direct sum of Vir and its "complex conjugate" $\overline{\text{Vir}}$. This yields the algebra of infinitesimal conformal transformations in two dimensions. In this way, it plays a fundamental role in conformal field theory and has proved useful in elucidating the properties of string theories and two-dimensional statistical mechanical models.[3]

Recently, these algebras have been generalized to algebras comprised of certain infinitesimal diffeomorphisms of the N torus, $(S^1)^N$, in Ref. 4. The generalized algebras involve operators L_n wth indices $n=(n_1,...,n_N)^T\in\mathbb{Z}^N$. (Here and henceforth, m, n, etc. will denote N-dimensional integral column vectors; we will not use boldface notation.) Two different types of central extensions were presented in Ref. 4:

$$[L_m,L_n]=\langle v|m-n\rangle L_{m+n}+A(m)\delta_{m+n,0}, \tag{1.2}$$

$$[L_m,L_n]=\langle v|m-n\rangle L_{m+n}+A(m)\delta_{\langle v|m+n\rangle,0}\,. \tag{1.3}$$

The structure constant vector $v=(v_1,...,v_n)$ is one of the many new features of this albegra. One may take $v\in{}^N\mathbb{Z}$, ${}^N\mathbb{Q}$, ${}^N\mathbb{R}$, or ${}^N\mathbb{C}$ [the upper-left index signifies that these are spaces of bra (or row) vectors]. The scalar product $\langle v|n\rangle$ is a bilinear map from ${}^N\mathbb{C}\times\mathbb{Z}^N$ to $\mathbb{C}$ defined by the rule $\langle v|n\rangle=\Sigma_j v_j n_j$, $1\leqslant j\leqslant N$, (i.e., with no complex conjugation). In general, the algebras (1.2) and (1.3) are not isomorphic as central extensions, the first central extension involves the Kronecker delta function;

$$\delta_{m+n,0}=\prod_{1\leqslant i\leqslant N}\delta_{m_i+n_i,0}, \tag{1.4}$$

(where the 0 on the left-hand side is the zero vector in $\mathbb{Z}^n$), while the second involves the weaker condition that $\langle v|m+n\rangle=0$ so that it is one of a larger family of central extensions. The function $A(m)$ appearing in both (1.2) and (1.3) is given by

$$A(m)=a_1\langle v|m\rangle+a_3\langle v|m\rangle^3. \tag{1.5}$$

Equations (1.2) and (1.3) are identical if and only if the components of the structure vector v are rationally linear-independent complex numbers. If one wants to introduce a particular definition of adjoint:

$$L_n^\dagger=L_{-n}, \tag{1.6}$$

it is necessary to select $v\in{}^N\mathbb{R}$, and take $a_1,a_3\in\mathbb{R}$. Possible generalizations of such adjoints will be considered elsewhere. From the viewpoint of a general complex Lie algebra, we can deal with the case of $c\in\mathbb{C}$ and not impose (1.6). Since shifts of the form $L_0\to L_0+a$ leave the commutation relations invariant but cause the shift $a_1\to a_1+2a\langle v|m\rangle$, one can, without

 0022-2488/90/081805-12$03.00

loss of generality, take $a_1 = -a_3$. Moreover, motivated by the case of Virasoro algebra, one may use the notation

$$a_3 = -a_1 \equiv c/12, \tag{1.7}$$

which defines the central charge c. In Ref. 4 it was shown that the algebras (1.2) and (1.3) with $c = 0$ can be realized in terms of certain infinitesimal generators of diffeomorphisms of $(S^1)^N$. We recall this realization. Let

$$L_{e(s),n} = i \cdot e^{i\langle \vartheta | n \rangle} \frac{\partial}{\partial \vartheta_s}, \tag{1.8}$$

where $e(s) = \langle e_s |$ is the unit row vector in the sth direction, $i = (-1)^{1/2}$, and $\vartheta = (\vartheta_1,...,\vartheta_N)$ with $\vartheta_j \in \mathbb{R} \bmod 2\pi$. This is a generator of an infinitesimal diffeomorphism of $(S^1)^N$. The representation of L_n is given by

$$L_n = \sum_{1 \le j \le N} v_j L_{e(j),n} . \tag{1.9}$$

We recall that the algebras (1.2) and (1.3) retain the same form in terms of generators that differ from the original ones by a global rescaling, $L'_n = aL_n$, provided that one defines a rescaled structure vector and central charge according to $v' = av$ and $c' = a^2 c$. It also retains the same form for an index-dependent rescaling of the form

$$L'_n = \left(\prod_j a_j^{n_j} \right) L_n, \quad \text{with } v' = v,$$

and

$$c' = \left(\prod_j a_j^{m_j + n_j} \right)^{-1} c. \tag{1.10}$$

This is, in fact, the most general automorphism that preserves the grading of the algebra. In addition, the algebra also retains the same form if one transforms the indices by elements $\mathscr{A} \in \mathrm{GL}(N,\mathbb{Z})$. Specifically, let $v' = v \cdot \mathscr{A}^{-1}$ and $m' = \mathscr{A} \cdot m$. This then defines an isomorphism of the algebras (1.2) and (1.3) that does not preserve the grading.

One can also make index-dependent shifts of the generators of the form $L'_n = L_n + a_n$. The resulting algebra has defining relations,

$$[L'_m, L'_n] = \langle v | m - n \rangle L'_{m+n} - \langle v | m - n \rangle \cdot a_{m+n} + A(m) \delta_{\langle v | m+n \rangle, 0}, \tag{1.11}$$

and is equivalent to the original algebra, differing only in a redefinition of the central charge.

The preceding discussion can be generalized easily from $(S^1)^N$ to diffeomorphisms of $(\mathbb{C}^\times)^N$, where $\mathbb{C}^\times$ denotes the multiplicative group of nonzero complex numbers. The generalization of the generators takes on the following form. We take basic generators

$$L_{e(s),n} = -\left(\prod_{1 \le t \le N} z_t^{n_t} \right) \cdot z_s \cdot \frac{\partial}{\partial z_s}, \tag{1.12}$$

where $z_t \in \mathbb{C}^\times$. The generators L_n are defined as before in (1.9). The restrictions to $(S^1)^N$ is obtained by choosing $\rho_s = 1$ in $z_s = \rho_s e^{i\vartheta_s}$.

One can construct a fundamental and far-reaching generalization of Ref. 4 by letting the structure vectors vary for each generator, over some finite or even infinite set. Interestingly, we shall find that once one lets the structure vectors vary, then the properties of the resulting algebra are quite different from those with only one structure vector. In the next section we will begin our discussion of this algebra.

Parenthetically, we note that a subalgebra of the algebra of diffeomorphisms of the two-torus has been studied by Arnold[5] and later by other authors.[6] This subalgebra is based on the very different defining relation

$$[L_m, L_n] = (m_1 n_2 - m_2 n_1) L_{m+n} + \langle a | m \rangle \delta_{m+n,0},$$

where $m,n \in \mathbb{Z}^2$, and consequently has a structure quite different from the $N = 2$ special case of the general algebra of diffeomorphisms of the N-torus considered here. In another different direction, an algebra of diffeomorphisms of the N-torus has been studied in Ref. 7 defined by

$$[L_m, L_n] = r \cdot \sin(\det\{a_1, a_2, ..., m, n\}) L_{m+n} + \langle b | m \rangle \delta_{m+n,0},$$

where $m,n,a_1,...,a_{N-2},b \in \mathbb{Z}^N$. In both of these other algebras, the coefficient of L_{m+n} is a multilinear, rather than linear, function of the indices m and n, in contrast to the Virasoro algebra, to its geneneralization presented in Ref. 4, and to the general algebra of diffeomorphisms of the N-torus studied here, all of which have a coefficient L_{m+n} that is a linear function of $m - n$. Related to this is the fact that in both of the other algebras, $[L_m, L_0] = 0$, again in contrast to the Virasoro algebra, to the algebra of Ref. 4, and to the more general algebra discussed here, where L_0 and $L_{v,0}$, respectively, do not commute with the rest of the algebra and, indeed, serve to define an important Cauchy subalgebra. Finally, yet another fundamental difference is the fact, already indicated, that in the general algebra of diffeomorphisms of the N-torus studied here, each generator $L_{v,n}$ depends not just on an index vector $n \in \mathbb{Z}^N$, but also on a structural vector $v \in \mathbb{C}^N$; it is this property that enables our algebra to be the general, rather than a special case of the, algebra of diffeomorphisms of the N-torus and that adds a great amount of mathematical richness to this algebra, going beyond both the Virasoro algebra and the generalization to a particular algebra of diffeomorphisms of $(S^1)^N$ given in Ref. 4, in which all generators depended on a single structure vector.

II. GENERAL ALGEBRA OF DIFFEOMORPHISMS OF THE N-TORUS

A. Definition of the class

Here we shall define a class of algebras of (infinitesimal) diffeomorphisms of the N-torus. Since the structure vector associated with each generator can be different, we label the generators as $L_{u,m}$. The vector $m = |m\rangle = (m_1,...,m_N)^T \in \mathbb{Z}^N$ will be a ket (column) vector with integer entries. $u = \langle u| = (u_1,...,u_N) \in {}^N\mathbb{C}$ will be a bra (row) vector with complex entries. The space ${}^N\mathbb{C}$ will be viewed as a *right* $\mathbb{C}$-vector space in the sense that a linear operator will be written to the *right* of the vector. The scalar product $\langle u | m \rangle$ will simply be matrix multiplication without complex conjugation on the components of u. Taking an axiomatic approach, we may define the algebra $\mathscr{D}(N)$ by the linearity condition

$$L_{u\alpha + v\beta,m} = \alpha L_{u,m} + \beta L_{v,m}, \quad \alpha,\beta \in \mathbb{C},\ u,v \in {}^N\mathbb{C}, \tag{2.1}$$

together with the Lie bracket

$$[L_{u,m},L_{v,n}]=L_{w,m+n},$$

where $w=u\cdot\langle v|m\rangle-v\cdot\langle u|n\rangle$. (2.2)

It is clear that $\mathscr{D}(N)$ has a vector space basis over $\mathbb{C}$ consisting of all $L_{e(j),m}$, $1\leqslant j\leqslant N$, $m\in\mathbb{Z}^N$, where $e(j)$ is the unit vector in the jth direction. Moreover, every element of $\mathscr{D}(N)$ can be written as a *finite* sum of the form $\Sigma_{u,m}L_{u,m}$, where u and m range over a finite set with each $u\neq 0$. In this sense, $\mathscr{D}(N)$ is a graded vector space over $\mathbb{C}$ with the grading provided by $\mathbb{Z}^N$. However, many of our assertions ignore this grading and are therefore stronger. The bracket operation preserves the grading by the index set $\mathbb{Z}^N$ but does not involve any grading in the structure vector space ${}^N\mathbb{C}$. It is for this reason that we treat $\mathbb{Z}^N$ and ${}^N\mathbb{C}$ as distinct spaces. A specific faithful representation of $\mathscr{D}(N)$ is provided by the general infinitesimal diffeomorphisms of the N-torus, Eq. (1.9). Equivalently, if we define $\mathscr{L}(N)$ to be the algebra $\mathbb{C}[X_j^{\pm 1}]$, $1\leqslant j\leqslant N$, of all Laurent polynomials in N variables over $\mathbb{C}$, then $\mathscr{D}(N)$, the Lie algebra of all $\mathbb{C}$ derivations of $\mathscr{L}(N)$, serves as a concrete faithful model of (2.1) and (2.2). We note that each $\mathscr{A}\in\mathrm{GL}(N,\mathbb{Z})$ defines a Lie algebra automorphism of $\mathscr{D}(N)$ according to the rule

$$\mathscr{A}(L_{u,m})=L_{u\cdot\mathscr{A}^{-1},\mathscr{A}\cdot m}. \tag{2.3}$$

It is worthwhile to point out a basic difference between the present algebra for $N>1$ and the case of $N=1$. We recall that a loop algebra extension of a Lie algebra $\mathfrak{g}$ over $\mathbb{C}$ is $\mathbb{C}[t^{\pm 1}]\otimes\mathfrak{g}$ (where the tensor product is over $\mathbb{C}$ unless noted to the contrary). The Lie bracket is defined by the rule

$$[f\otimes x,g\otimes y]=fg\otimes[x,y]. \tag{2.4}$$

Normally, $\mathfrak{g}$ is taken to be a finite-dimensional Lie algebra over $\mathbb{C}$ (see Refs. 3 and 8 among many others). The fact that $\mathfrak{g}$ is finite dimensional plays a crucial role in the study of loop algebras and the related loop groups as well as the Kac–Moody algebras. Our case is quite different; $\mathfrak{g}$ will turn out to be infinite dimensional. In particular, $\mathfrak{g}$ will be Diff(S^1) or its generalizations. It is this fact that leads us to the occurrence of an infinite number of central charges. We note that loop algebras are usually defined by using (2.4) while Kac–Moody algebras are originally defined by using Cartan matrices (see the reviews given by Garland[8] and Kac[3]). Our abstract approach formally resembles the Kac–Moody approach, but a number of our results use the loop algebra structure. In either case, the infinite dimensionality of $\mathfrak{g}$ forces us to give different arguments.

We next consider a $\mathbb{C}$-vector subspace V of ${}^N\mathbb{C}$. We then define $\mathscr{D}(N,V)$ to be the $\mathbb{C}$-vector subspace of $\mathscr{D}(N)$ spanned by all $L_{v,n}$, $v\in V$, $n\in\mathbb{Z}^N$. It is immediate that $\mathscr{D}(N,V)$ is closed with respect to (2.1) and (2.2) so that it is a Lie subalgebra. The space V will be called the space of structure vectors for $\mathscr{D}(N,V)$. Strictly speaking, we must include as part of the definition the specific choice of $\mathbb{Z}^N$ in $\mathbb{C}^N$ as well as the identification of ${}^N\mathbb{C}$ as the dual space of $\mathbb{C}^N$ through matrix multiplication. It is immediate that the automorphism of $\mathscr{D}(N)$ described in (2.3) carries $\mathscr{D}(N,V)$ isomorphically onto $\mathscr{D}(N,V\cdot\mathscr{A}^{-1})$, $\mathscr{A}\in\mathrm{GL}(N,\mathbb{Z})$.

It is not too difficult to deduce from later results that every automorphism σ of $\mathscr{D}(N,V)$ has the following form (see discussions in Sec. I):

$$\sigma(L_{u,m})=\chi(m)\cdot L_{u\cdot\mathscr{A}^{-1},\mathscr{A}\cdot m},$$

where $\mathscr{A}\in\mathrm{GL}(N,\mathbb{Z})$ with $V\cdot\mathscr{A}=V$, and $\chi:\mathbb{Z}^N\to\mathbb{C}^\times$ is a homomorphism. The function $\chi(m)$ denotes the product factor that multiplies the generator in (1.10). Since we do not need this result, we omit the proof (it can be extracted from the proof of Theorem 3.2). We note that the condition that $\mathscr{A}\in\mathrm{GL}(N,\mathbb{Z})$ should map V onto V is not easy to describe in terms of a basis of V when V has dimension over $\mathbb{C}$ distinct from 1 or N.

B. The $\mathbb{Q}$ rank of the space V of structure vectors. Cartan matrices

The nature of the space V of structure vectors is of fundamental importance for the algebra of infinitesimal diffeomorphisms. We will define two numerical invariants to measure its size. The first invariant of V is $\dim_{\mathbb{C}}V=\ell$ (we often write dim V for $\dim_{\mathbb{C}}V$ when there is no chance of confusion). The orthogonal complement $V^\perp$ is defined as usual as $\{w\in\mathbb{C}^N\,|\,\langle v|w\rangle=0\}$. This is a $\mathbb{C}$ subspace of the space $\mathbb{C}^N$ of ket (column) vectors. It follows that $\mathbb{Z}^N/(V^\perp\cap\mathbb{Z}^N)$ is a free Abelian group with rank denoted by $r_{\mathbb{Q}}(V)$. This number will be called the (relative) $\mathbb{Q}$ *rank* of V (alternatively, one might call it the relative $\mathbb{Z}$ rank). It is immediate that $\dim V\leqslant r_{\mathbb{Q}}(V)\leqslant N$. We note that this concept can be defined in terms of $\mathbb{Q}^N$ rather than $\mathbb{Z}^N$ because of the form invariance of $\mathscr{D}(N,V)$ under rescaling of the generators in addition to the fact that $\mathbb{Q}$ is the quotient field of the principal ideal domain $\mathbb{Z}$. The word "relative" is used in order to emphasize the fact that this concept refers to a choice of $\mathbb{Z}^N$ (or $\mathbb{Q}^N$) in $\mathbb{C}^N$. In mathematical terminology, one would say that $\mathbb{C}^N$ has been given a (particular) $\mathbb{Z}$ structure. Since the scalar product $\langle u|m\rangle$ identifies ${}^N\mathbb{C}$ with the dual space of $\mathbb{C}^N$, it also has a preferred $\mathbb{Z}$ structure with the given selection of $\mathbb{Z}^N$. If we take a free (i.e., indepenent, unconstrained) basis $\ell(j)$, $r_{\mathbb{Q}}(V)<j\leqslant N$, of $V^\perp\cap\mathbb{Z}^N$, then the existence of a complementary basis to make up a basis for $\mathbb{Z}^N$ shows that we can find $\mathscr{A}\in\mathrm{GL}(N,\mathbb{Z})$ so that $\mathscr{A}\cdot\ell(j)=e(j)$ is the standard unit vector in the jth direction. This means that $V\cdot\mathscr{A}^{-1}$ is contained in the subspace of ${}^N\mathbb{C}$ spanned by the standard unit row vectors $e(i)$, $1\leqslant i\leqslant r_{\mathbb{Q}}(V)$, where $r_{\mathbb{Q}}(V)$ is the smallest integer with this property. The preceding assertions are straightforward consequences of the theory of finitely generated modules over principal ideal domains (see Ref. 9, Sec. V.2). In general $\dim V=r_{\mathbb{Q}}(V)$ holds if and only if V is the space of all complex solutions of a system of linear homogeneous equations with integer (or rational) coefficients in N unknowns.

A more concrete, but equivalent, way of looking at the space V of structure vectors is to introduce the concept of a generalized *Cartan matrix* M. By definition, M is an $\ell\times N$ complex matrix of row rank ℓ. Two such matrices M, M' are said to be *equivalent* if there exists $\mathscr{B}$ in $\mathrm{GL}(\ell,\mathbb{C})$ and $\mathscr{A}$ in $\mathrm{GL}(N,\mathbb{Z})$ so that $M'=\mathscr{B}\cdot M\cdot\mathscr{A}^{-1}$. If we select a $\mathbb{C}$-vector space basis $\{v(i)\,|\,1\leqslant i\leqslant\ell\}$ for V consisting of row vectors in ${}^N\mathbb{C}$. We can then arrange these vectors in the form of an $\ell\times N$ matrix M. The (i,j) entry of M is just the scalar product $\langle v(i)|e(j)\rangle$. M is then a Cartan matrix. An examination

of (2.1) and (2.2) shows that $\mathscr{D}(N,V)$ can be defined by using M and generators $L_{v(i),m}$, $1\leqslant i\leqslant \ell$, $m\in\mathbb{Z}^N$. By using (2.3), equivalent Cartan matrices define isomorphic algebras [the $\mathbb{Z}^N$ index group has been changed by $\mathscr{A}\in GL(N,\mathbb{Z})$]. It follows that the isomorphism class of the algebra $\mathscr{D}(N,V)$ comes from at least one equivalence class of Cartan matrices of size $\ell\times N$. The Cartan matrix M has column rank $r_{\mathbb{Q}}(V)$ over $\mathbb{Q}$. Namely, we view the columns as lying in the $\mathbb{Q}$-vector space $\mathbb{C}^{\ell}$ (i.e., we forget the structure of $\mathbb{C}^{\ell}$ as a $\mathbb{C}$-vector space), then $r_{\mathbb{Q}}(V)$ is the maximum number of rationally independent columns of M and coincides with the rank of the free Abelian subgroup of $\mathbb{C}^{\ell}$ consisting of all integral linear combinations of the columns of M. Also, $r_{\mathbb{Q}}(V)$ is an invariant of the equivalence class of M.

In the classical theory, Cartan matrices have integer entries and satisfy suitable restrictions on their entries. The restrictions reflect the fact that the Killing–Cartan form can be used to define a suitable positive-definite scalar product so that the Cartan matrix encodes information about the root vectors. In particular, the root vectors generate a lattice (a discrete free Abelian subgroup of rank ℓ in a real Euclidean space $\mathbb{R}^{\ell}$). Such a picture is not available in our setup. Nevertheless, the equivalence classes of the generalized Cartan matrices so defined will be seen (in Theorem 3.2) to be in one-to-one correspondence with the isomorphism class of our Lie algebras (as ungraded Lie algebras). In this sense, they generalize the classical concept.

Before going further, we look at a few examples. In all these examples, we assume $N=3$.

Example 1: Let dim $V=1$ so V is spanned by a single nonzero vector $v\in^3\mathbb{C}$. (i) Let $v=(2i,6i,-4i)$. It is clear that V is also spanned by $(1,3,-2)=w$. Thus $w\cdot\mathscr{A}^{-1}=(1,0,0)$ holds for

$$\mathscr{A}=\begin{pmatrix}1&3&-2\\0&1&0\\0&0&1\end{pmatrix}.$$

It follows that $r_{\mathbb{Q}}(V)=1$. The sets $(2i,6i,-4i)$, $(1,3,2)$, and $(1,0,0)$ are equivalent Cartan matrices. In contrast, $(1,\pi,0)$ and $(1,2\pi,0)$ are not equivalent Cartan matrices. In general, if dim $V=1$ and v spans V, then $r_{\mathbb{Q}}(V)=1$ if and only if $v_i/v_j\in\mathbb{Q}$ holds for all i and j when $v_j\neq 0$. We will often write $r_{\mathbb{Q}}(v)$ in place of $r_{\mathbb{Q}}(V)$ when v spans V.

(ii) Let $v=(1,1+\pi,\pi)$. The components of v have pairwise irrational ratios. Nevertheless, $r_{\mathbb{Q}}(V)=2$ is not maximal. It is easy to see that $v\cdot\mathscr{A}^{-1}=(1,\pi,0)$ holds for the following element $\mathscr{A}\in GL(3,\mathbb{Z})$:

$$\mathscr{A}=\begin{pmatrix}1&1&0\\0&1&1\\0&0&1\end{pmatrix}.$$

It should be noted that the structure vector v is not unique even if we assume that the first $r_{\mathbb{Q}}(V)$ components are nonzero and begin with 1. For example, we could also use $(1,\pi^{-1},0)$ as a structure vector. This is a consequence of the fact that we are allowed to change v to $\mathscr{B}\cdot v\cdot\mathscr{A}^{-1}$ ($\mathscr{B}\in\mathbb{C}^{\times}$ in the present case since $\ell=1$). This freedom is used in the description of the basic invariants: the dimension and the relative $\mathbb{Q}$ rank of V.

Example 2: dim $V=2$. (i) Let V be spanned by $(1,2,1)$ and $(1,1+\pi,\pi)$ over $\mathbb{C}$. Let $V^1\cap\mathbb{Z}^3$ be spanned by $(1,-1,2)^T$ so that $r_{\mathbb{Q}}(V)=2$. V can also be spanned by $(1,2,1)$ and $(0,1-\pi,1-\pi)$ or by $(1,2,1)$ and $(0,1,1)$. This is seen by using $\mathbb{C}$-linear combination of these row vectors. Thus, $V\cdot\mathscr{A}^{-1}$ is spanned by $(1,0,0)$ and $(0,1,0)$, where $\mathscr{A}\in GL(3,\mathbb{Z})$ is

$$\mathscr{A}=\begin{pmatrix}1&2&1\\0&1&1\\0&0&1\end{pmatrix}.$$

(ii) Let V be spanned by $(0,1,0)$ and $(1,1+\pi,\pi)$ over $\mathbb{C}$, and also spanned by $(1,0,\pi)$ and $(0,1,0)$. These vectors have $\mathbb{Q}$ ranks 2 and 1, respectively. We can also span V by $(1,\sqrt{2},\pi)$ and $(1,\sqrt{3},\pi)$. Each of these latter vectors has $\mathbb{Q}$ rank 3. The space V has $r_{\mathbb{Q}}(V)=3$. Thus the individual structure vectors may or may not determine the $\mathbb{Q}$ rank of V. The following are Cartan matrices:

$$\begin{pmatrix}1&0&\pi\\0&1&0\end{pmatrix},\quad\begin{pmatrix}1&\sqrt{2}&\pi\\1&\sqrt{3}&\pi\end{pmatrix}.$$

The last observation in (ii) of example 2 is quite general. It will be used many times. We will therefore state the general result.

Propositon 2.5: (a) Let v be a nonzero vector in $^N\mathbb{C}$. The one-dimensional subspace of $^N\mathbb{C}$ spanned by v has relative $\mathbb{Q}$ rank r if and only if the components v_i, $1\leqslant i\leqslant N$, of v spans a $\mathbb{Q}$-vector subspace of dimension r in $\mathbb{C}$ (considered as an infinite-dimensional vector space over $\mathbb{Q}$). (b) Let V be a subspace of $^N\mathbb{C}$ with dim $V=\ell$ and relative $\mathbb{Q}$ rank $r_{\mathbb{Q}}(V)=r$. Then V has a basis $\{v(s)\,|\,1\leqslant s\leqslant\ell\}$ over $\mathbb{C}$ such that each $v(s)$ spans a one-dimensional subspace of relative $\mathbb{Q}$ rank r. Conversely, any subspace W of $^N\mathbb{C}$ containing one of these vectors must have relative $\mathbb{Q}$ rank at least r. (c) Every Cartan matrix M associated to V in (b) is equivalent to one where the first ℓ columns form the identity matrix while the last $N-r$ columns are 0.

The preceding assertion can be verified by proving (c) first. We begin with any Cartan matrix M. Since M has only a finite number of columns, the Abelian subgroup of $\mathbb{C}^{\ell}$ generated by the columns is necessarily free. We can therefore use column operations over $\mathbb{Z}$ on M [they correspond to right multiplication by elements of $GL(N,\mathbb{Z})$] to normalize M so that its last $N-r$ columns are 0. Since row operations over $\mathbb{C}$ on M correspond to left multiplication by elements of $GL(\ell,\mathbb{C})$, we may then put M in row echelon normal form. Since column permutations of M corresponds to right multiplication by elements of $GL(N,\mathbb{Z})$, we have proved (c). Since we could have carried out column operations over $\mathbb{Z}$ on the first $N-r$ columns before performing the row operations over $\mathbb{C}$, we can not make general statements about normalizing the columns numbered from $\ell+1$ to $N-r-1$.

We next note that the finite number of entries of M generate a subfield K of $\mathbb{C}$. Since $\mathbb{C}$ is not finitely generated as a field extension of $\mathbb{Q}$, $\mathbb{C}$ must be infinitely generated as a field extension of K. We can now use elements of $\mathbb{C}$ that are linearly independent over K to form linear combination of the rows of M in order to construct a basis of V over $\mathbb{C}$ as indicated in (b). (This requires us to undo the column operations over $\mathbb{Z}$ as needed.) The rest of the assertions are clear.

Remark: There is no uniform procedure to determine the relative $\mathbb{Q}$ rank of V even when $\dim V = 1$. Such a procedure would require the solution to a number of outstanding problems in algebraic and transcendental number theory. For example, $\zeta(3)$ has only recently been proved to be irrational. Thus, $(1,\zeta(3))$ has relative $\mathbb{Q}$ rank 2. We still do not know whether $\zeta(5)$ or γ, the Euler constant, is irrational, and hence we cannot say what the relative $\mathbb{Q}$ rank is for either $(1,\zeta(5))$ or $(1,\gamma)$.

Associated to the concept of Cartan matrices, we can also introduce the notion of *Cartan subalgebra, Cartan decomposition, root spaces, root lattices,* etc. Namely, in $\mathscr{D}(N,V) = \mathfrak{g}$, the Cartan subalgebra $\mathfrak{g}_0$ is defined to be the subspace consisting of $L_{v,0}$, $v \in V$. Each $n \in \mathbb{Z}^N$ will be called a root and the root space $\mathfrak{g}_n$ is defined to be the subspace consisting of $L_{v,n}$, $v \in V$. Except when $r_{\mathbb{Q}}(V) = N$ is maximal, the root lattice $\mathbb{Z}^N$ cannot be identified with a free Abelian subgroup of the dual space of V. We can define an extended Cartan subalgebra $\mathfrak{g}_{\bar{0}}$ as the sum of the subspaces $\mathfrak{g}_n$ with $n \in V^{\perp} \cap \mathbb{Z}^N$. For each $m \in \mathbb{Z}^N$, we can define the extended root space $\mathfrak{g}_{\bar{m}}$ to be the sum of the subspaces $\mathfrak{g}_{n+m}$, $n \in V^{\perp} \cap \mathbb{Z}^N$. These extended root spaces are therefore indexed by elements of the free Abelian group $\mathbb{Z}^r = \mathbb{Z}^N/(V^{\perp} \cap \mathbb{Z}^N)$, $r = r_{\mathbb{Q}}(V)$. We may identify $\mathbb{Z}^r$ with a free Abelian subgroup of the dual space of V. However, there is no natural way of viewing $\mathbb{Z}^r$ as a subgroup of $\mathbb{Z}^N$ and there is also no natural choice of a $\mathbb{Z}$ basis for $V^{\perp} \cap \mathbb{Z}^N$. These concepts will be used shortly.

C. Nonsimplicity of $\mathscr{D}(N,V)$ when $r_{\mathbb{Q}}(V) < N$

The following assertion is straightforward.

Proposition 2.6: If $r_{\mathbb{Q}}(V) = r < N$, then $\mathscr{D}(N,V)$ is not a simply Lie algebra. In fact, we can find a subspace W of $^{r}\mathbb{C}$ with $r_{\mathbb{Q}}(W) = r$ such that $\mathscr{D}(N,V)$ is isomorphic to the $(N-r)$-fold iterated loop extension $\mathbb{C}[X_{r+1}^{\pm 1},\ldots,X_N^{\pm 1}] \otimes \mathscr{D}(r,W)$.

The isomorphism is clear from the discussion on row and column operations on the Cartan matrix. To show the nonsimplicity, we just have to exhibit a single nontrivial ideal in $\mathscr{D}(N,V)$. This is most easily accomplished by using the indicated isomorphism. Namely, $\mathscr{D}(N,V)$ can be mapped onto $\mathscr{D}(r,W)$ by sending each X_j to 1, $r+1 \leqslant j \leqslant N$. If we examine (2.4), it amounts to evaluating the Laurent polynomials at $X_j = 1$, $r+1 \leqslant j \leqslant N$. It is clearly a Lie algebra homomorphism. The kernel is then an ideal $\mathscr{T}$ spanned by the following elements over $\mathbb{C}$:

$$\left\{(M_p - 1) \otimes L_{u,m} | L_{u,m} \in \mathscr{D}(r,W); M_p = \prod_{j=r+1}^{N} X_j^{p_j}\right\}. \tag{2.7}$$

Cleary, $\mathscr{T}$ is a non-Abelian ideal, namely, the Lie bracket of any element of $\mathscr{T}$ with any element of the entire algebra remains in $\mathscr{T}$ and the Lie bracket is nontrivial on $\mathscr{T}$.

Remark: The evaluation at $X_j = 1$ can be modified. Indeed, we can independently evaluate X_j at $\alpha_j \in \mathbb{C}^{\times}$, $r < j \leqslant N$. The generators of the kernel would have to be changed with $M_p - 1$ replaced by $M_p - M_p(\alpha)$ where $M_p(\alpha)$ is the value of the monomial M_p under the evaluation of X_j at α_j. Of course, the ideal $\mathscr{T}$ is not graded. It is much easier to prove the weaker assertion that every proper graded ideal must be 0.

D. Simplicity of $\mathscr{D}(N,V)$ when $r_{\mathbb{Q}}(V) = N$

We now abbreviate $\mathscr{D}(N,V)$ to $\mathfrak{g}$ and use the Cartan decomposition. Equations (2.1) and (2.2) show that $\mathfrak{g}_0$ is an Abelian subalgebra and that the adjoint action of $\mathfrak{g}_0$ on $\mathfrak{g}$ decomposes $\mathfrak{g}$ into the direct sum of the root spaces $\mathfrak{g}_m, m \in \mathbb{Z}^N$. In the classical theory, each root space has dimension 1 when the root is not zero. In our case, all root spaces have the same dimension ℓ. We now have the following theorem.

Theorem 2.8: If $r_{\mathbb{Q}}(V) = N$, then $\mathscr{D}(N,V)$ is simple as an ungraded Lie algebra.

Proof: Let $\mathscr{J}$ be a nonzero ideal of $\mathfrak{g}$ and let $x = \Sigma_{u,m} L_{u,m}$ be a nonzero element of $\mathscr{J}$. The strategy of the proof is to define $|x|$ to be the number of $m \in \mathbb{Z}^N$ in the preceding sum representation with the corresponding u nonzero. In other words, $|x|$ is the number of nonzero components of x with respect to the grading of the algebra. By using the property $[\mathscr{J}, y] \subset \mathscr{J}$ holds for all $y \in \mathfrak{g}$ in conjunction with a minimization of $|x|$, we will then show that $\mathscr{J}$ must contain all the elements of $\mathfrak{g}$. In fact, the argument will show that $x \in \mathscr{J}$ will force $\mathfrak{g}_m \subset \mathscr{J}$ for each $m \in \mathbb{Z}^N$ with $L_{u,m} \neq 0$. Thus, $\mathscr{J}$ is a graded ideal and coincides with $\mathfrak{g}$.

To begin our argument, we may choose x in $\mathscr{J} - \{0\}$ to minimize $|x|$. We assert that $|x| = 1$. Suppose that $|x| > 1$ so that $L_{u,m} + L_{v,n}$ appears as part of the sum representation of $x, m \neq n, u \neq 0 \neq v$. Since $r_{\mathbb{Q}}(V) = N$ and $m - n \notin V^{\perp}$, we can find $w \in V$ with $\langle w | m - n \rangle \neq 0$. In a similar sum representation, the nonzero components of $[L_{w,0}, x]$ must be indexed by a suitable subset of the indices associated to x. For the indices m, n, $[L_{w,0}, x]$ has components $\langle w|m\rangle L_{u,m}$ and $\langle w|n\rangle L_{v,n}$. Since $\langle w|m-n\rangle \neq 0$, one of these two components is definitely not zero. Thus $|[L_{w,0}, x]| = |x|$ by minimality and both components are nonzero. Since $\langle w|m-n\rangle \neq 0$, the following matrix is invertible:

$$\begin{pmatrix} 1 & 1 \\ \langle w|m\rangle & \langle w|n\rangle \end{pmatrix}.$$

Thus $y = [L_{w,0} x] - \langle w|m\rangle x \in \mathscr{J}$ and satisfies $0 < |y| < |x|$. This contradicts the minimality of $|x|$ so that we can assume $x = L_{u,m}$. We have the following cases.

Case 1. $m \neq 0$: Then $[L_{u,m}, L_{v,-m}] = L_{w,0} \in \mathscr{J}$, $w = u \cdot \langle v|m\rangle + v \cdot \langle u|m\rangle$. If $\langle u|m\rangle \neq 0$, we take $v = u$. If $\langle u|m\rangle = 0$, then we use $r_{\mathbb{Q}}(V) = N$ to find v so that $\langle v|m\rangle \neq 0$. Thus, we can find $L_{w,0} \in \mathscr{J}$ with $w \neq 0$. This puts us in case 2.

Case 2. $m = 0$: Suppose that $u_i \neq 0$, then $[x, L_{v,e(i)}] = u_i L_{v,e(i)} \in \mathscr{J}$ for each $v \in V$. Thus $\mathfrak{g}_{e(i)} \subset \mathscr{J}$.

We can repeat case 1 to conclude that $\mathfrak{g}_0 \subset \mathscr{J}$. Using Proposition 2.5, we can find $L_{w,0} \in \mathscr{J}$ where $r_{\mathbb{Q}}(w) = N$. Thus, $[L_{v,n}, L_{w,0}] = \langle w|n\rangle L_{v,n} \in \mathscr{J}$ and $\langle w|n\rangle \neq 0$ if $n \neq 0$. This shows that $\mathfrak{g}_n \subset \mathscr{J}$ holds for every $n \in \mathbb{Z}^n$ so that $\mathscr{J} = \mathfrak{g}$ and $\mathfrak{g}$ is simple as claimed. □

Remark: The preceding argument can be adapted to show that $\mathscr{D}(N,V)$ is simple as a *graded* Lie algebra. For example, the two cases would correspond to $m \notin V^{\perp}$ and

$m \in V^{\wedge}$. Since the other modifications are equally straightforward, we omit further details.

III. CLASSIFICATION OF $\mathscr{D}(N,V)$

In this section, we shall give a classification of $\mathscr{D}(N,V)$ as an ungraded Lie algebra. As in the classical theory of finite-dimensional Lie algebras, the Cartan subalgebra will play a crucial role. However, one should note a number of fundamental differences between the present case and the classical case. We have a Cartan subalgebra as well as an extended Cartan subalgebra [when $r_{\mathbb{Q}}(V) < N$]. We can use $\ell = \dim V$ as the analog of the rank in the classical case. As mentioned before, our root spaces all have the same dimension ℓ. This happens in the classical case only when $\ell = 1$. If we select a basis $\{v(i) \mid 1 \leqslant i \leqslant \ell\}$ for V over $\mathbb{C}$, then we have a basis $\{L_{v(i),n} \mid 1 \leqslant i \leqslant \ell\}$ for each root space $\mathfrak{g}_n$, $n \in \mathbb{Z}^N$. Under the right adjoint action, each $L_{v(i),n}$ is then a root vector for the root n. Unless $r_{\mathbb{Q}}(V) = N$, distinct roots can determine the same element of the dual space of V. This led to the notion of extended Cartan subalgebra, extended root space, and a corresponding reduction of the root lattice. In the classical case, judicious choices of normalization of the generators H_i, $1 \leqslant i \leqslant \ell$, of the Cartan subalgebra and basis E_α for the one-dimensional root spaces then led to the familiar Cartan matrices. As indicated, these normalizations use the Killing–Cartan form (it becomes a positive-definite symmetric bilinear form on the real vector space spanned by H_i, $1 \leqslant i \leqslant \ell$). When $\ell = 1$ and $r_{\mathbb{Q}}(V) = N$, our algebras come closer to the classical case. The only difference is that the root lattice $\mathbb{Z}^N$ may not be identifiable with a discrete subgroup of the dual space of V. To get the classical analog so that $\mathbb{Z}^N$ is a discrete subgroup of $\mathbb{R}' = \mathbb{R}$, it is necessary and sufficient that $N = 1$. We are back to the case of $\mathrm{Diff}(S^1)$. In general, when $\ell = N$, we have the full algebra $\mathscr{D}(N)$ and the Cartan matrix can be taken to be the identity matrix of size $\ell \times \ell$. Nevertheless, our algebra $\mathscr{D}(N)$ remains simple rather than breaking up into a direct sum of N simple algebras. These are but some of many striking features of the general algebras of diffeomorphisms when $\ell = \dim V > 1$.

We next prove the following theorem.

Theorem 3.1: The Cartan subalgebra $\mathfrak{g}_0$ of $\mathfrak{g} = \mathscr{D}(N,V)$ is unique and we have (a) The algebra $\mathfrak{g}_0$ is an Abelian subalgebra isomorphic to V. It is maximal Abelian if and only if $r_{\mathbb{Q}}(V) = N$; equivalently, if and only if $\mathscr{D}(N,V)$ is simple. (b) Also, $\mathfrak{g}$ is the direct sum of eigenspaces with respect to the adjoint action of $\mathfrak{g}_0$. (c) the element x of $\mathfrak{g}$ lies in $\mathfrak{g}_0$ if and only if the following condition holds: (*) each y of $\mathfrak{g}$ is contained in a finite-dimensional $\mathrm{ad}\, x$ invariant $\mathbb{C}$ subspace $W(y)$ of $\mathfrak{g}$.

Proof: (a) The map sending $v \in V$ onto $L_{v,0}$ is an isomorphism between V and $\mathfrak{g}_0$. By giving V the zero Lie bracket, we may think of this as an isomorphism of Abelian Lie algebras over $\mathbb{C}$. The remaining assertion follows from the definition and the discussions under parts C and D of section II. (b) This follows from the definition. (c) Since the adjoint action of $\mathfrak{g}_0$ decomposes $\mathfrak{g} = \mathscr{D}(N,V)$ into a direct sum of root spaces $\mathfrak{g}_m, m \in \mathbb{Z}^N$, and each $\mathfrak{g}_m$ has dimension equal to $\ell = \dim V$, it follows that each $x \in \mathfrak{g}_0$ has the property (*). Conversely, assume that $z \notin \mathfrak{g}_0$. We can introduce a lexicographic ordering on $\mathbb{Z}^N$ so that $\mathbb{Z}^N$ becomes a totally ordered Abelian group. For example, given $m \neq n \in \mathbb{Z}^N$, we use $m < n$ to mean that there is an index j so that $m_i = n_i$ holds for $i < j$ and $m_j < n_j$. Since $z = \Sigma L_{u,m} \notin \mathfrak{g}_0$, we can find a nonzero component $L_{u,m}$ with m as large as possible. If $m \leqslant 0$ in $\mathbb{Z}^N$, then we can find a nonzero component $L_{v,n}$ with n as small as possible and be sure that $n < 0$. Without loss of generality, we assume $m > 0$ in our lexicographic ordering. It is enough to find an element $y = L_{w,p}$ so that $(\mathrm{ad}\, z)^q(y), q > 0$, span an infinite-dimensional subspace of $\mathfrak{g}$. There are two cases.

Case 1. $\langle u|m\rangle \neq 0$: Setting $y = L_{u,2m}$ would do the job because $(\mathrm{ad}\, z)^q(y)$ has nonzero component $L_{v,(q+2)m}$ in the largest index $(q+2)m$ with $v = u \cdot \alpha, \alpha = q! \cdot \langle u|m\rangle^q \in \mathbb{C}$.

Case 2. $\langle u|m\rangle = 0$: We select $n \in \mathbb{Z}^N - \{0\}$ with $\langle u|n\rangle \neq 0$. The case $y = L_{u,n}$ would do the job because $(\mathrm{ad}\, z)^q(y)$ has nonzero component $L_{v,qm+n}$ in the largest index $qm + n$ with $v = u \cdot \langle u|n\rangle^q$.

In both cases, the total ordering of the indices shows that $(\mathrm{ad}\, z)^q(y), q \geqslant 0$, cannot lie in a finite-dimensional subspace of $\mathfrak{g}$. □

We now proceed to the classification of $\mathscr{D}(N,V)$.

Theorem 3.2: Fix N and $\ell = \dim V$. The isomorphism classes of the ungraded Lie algebras $\mathscr{D}(N,V)$ stand in one-to-one correspondence with the defining equivalence classes of Cartan matrices, $r_{\mathbb{Q}}(V)$ is an invariant of the isomorphism class of $\mathscr{D}(N,V)$, and $\mathscr{D}(N,V)$ is simple if and only if $r_{\mathbb{Q}}(V) = N$.

Proof: We know from the discussion on Cartan matrices that each equivalence class of such matrices yields an isomorphism class of algebras of the form $\mathscr{D}(N,V)$ through the use of (2.3). We only need to show that distinct equivalence classes do not lead to isomorphic Lie algebras of the form $\mathscr{D}(N,V)$ because the other assertions have been proved already.

Let us assume σ: $\mathscr{D}(N,V) \to \mathscr{D}(N',V')$ is an isomorphism of Lie algebras. We can now assume that they are defined by Cartan matrices M and M'. In view of Theorem 3.1, σ must map the Cartan subalgebra of $\mathscr{D}(N,V)$ isomorphically onto the Cartan subalgebra of $\mathscr{D}(N',V')$. Thus $\ell = \ell'$ and M and M' have the same size. We need to show that M and M' are equivalent. We first treat the case where $r_{\mathbb{Q}}(V) = N = r_{\mathbb{Q}}(V')$. In view of parts C and D in Sec. II, either of the equalities implies the other because σ is an isomorphism. In the root space decomposition of $\mathscr{D}(N,V)$, the maximality of $r_{\mathbb{Q}}(V)$ implies that the *set* $\mathbb{Z}^N$ of roots n may be viewed as subset of the dual space of V. The definition of the Lie bracket in $\mathscr{D}(N,V)$ shows that $\mathbb{Z}^N$ is a free Abelian subgroup of rank $N = r_{\mathbb{Q}}(V)$ of the dual space of V. If we let $v(i)$, $1 \leqslant i \leqslant \ell$, denote a basis of the Cartan subalgebra $\mathfrak{g}_0$ and select root spaces $\mathfrak{g}_{e(j)}$, so that $\{e(j) \mid 1 \leqslant j \leqslant N\}$ is a free basis of the root lattice $\mathbb{Z}^N$, then $[L_{v(i),0}, L_{u,e(j)}] = \langle v(i)|e(j)\rangle L_{u,e(j)}$. Since $L_{u,e(j)} \neq 0$ when $u \neq 0$, the coefficient $\langle v(i)|e(j)\rangle$ is unambiguously defined and we have determined a Cartan matrix M of $\mathscr{D}(N,V)$. Since the only freedom we have comes from basis change of the Cartan subalgebra and the grading, and since both are determined by the Cartan subalgebra and the Lie bracket in $\mathscr{D}(N,V)$, we conclude that distinct equivalence classes of Cartan ma-

trices must correspond to distinct isomorphism classes of $\mathscr{D}(N,V)$ when V has maximal $\mathbb{Q}$ rank N.

We next consider $r_{\mathbb{Q}}(V) = r < N$. Thus

$$\mathscr{D}(N,V) \cong \mathbb{C}[X_{r+1}^{\pm 1},\ldots,X_{N}^{\pm 1}] \otimes \mathscr{D}(r,W)$$

holds for a vector subspace $W \subset {}^{r}\mathbb{C}$ with $W = P \cdot V$ for some $P \in \mathrm{GL}(N,\mathbb{C})$. In particular, we have $W \cong V$ and $r_{\mathbb{Q}}(W) = r$. There is a similar result for $\mathscr{D}(N',V')$. It is enough to show that $r = r'$ and W is equivalent to W' because a Cartan matrix associated to V can be obtained from one for W by adding $N - r$ columns of zero to the right. The equivalence between W and W' can then be extended to an equivalence between M and M' as desired. To get the equivalence between W and W', we not that the extended Cartan subalgebra $\mathfrak{g}_{\bar{0}}$ is the space $\{x \in \mathscr{D}(N,V) \mid [x,y] = 0 \text{ holds for every } y \in \mathfrak{g}_0\}$. It follows that σ must map this extended Cartan subalgebra isomorphically onto the corresponding extended Cartan subalgebra of $\mathscr{D}(N',V')$. By the same kind of reasoning σ must map the extended Cartan decomposition onto the corresponding one. This means that σ induces an isomorphism between the reduced grading groups $\mathbb{Z}^N/(V^{\perp} \cap \mathbb{Z}^N) \cong \mathbb{Z}^r$ and $\mathbb{Z}^{r'}$. Thus we have $r = r'$ and the same kind of argument as used in the case of maximal rank shows that we can extract a Cartan matrix associated to W by choosing a $\mathbb{C}$ basis for W and $\mathbb{Z}$ basis for $\mathbb{Z}^r$. Since these data depend only on V and the Cartan subalgebra, we have the desired conclusion.

Remark: In the general case, we do not know if the isomorphism σ must map root spaces to root spaces. But we do know that it must map extended root spaces to extended root spaces. This is enough to finish the proof. The stronger assertion about preservation of root spaces is an open problem. It is related to the fact that we do not know the exact nature of the group of all automorphisms of the Lie algebra when $r_{\mathbb{Q}}(V) < N$. It is easy to see that there are an uncountable number of such isomorphism classes when $r_{\mathbb{Q}}(V) > \ell$ and just one when $r_{\mathbb{Q}}(V) = \ell$.

IV. UNIVERSAL CENTRAL EXTENSIONS. DIM $V=1$

We begin with a brief review of central extensions. From a mathematical point of view, the basic problem is one of deciding if a particular projective representation ρ of a Lie algebra $\mathfrak{g}$ comes from an ordinary representation of $\mathfrak{g}$. This is not always the case. We therefore try the next best thing and ask if there is a central extension $\mathfrak{g}'$ of $\mathfrak{g}$ [given in the form of a Lie algebra homomorphism pr: $\mathfrak{g}' \to \mathfrak{g}$ with $\mathfrak{g} = \rho(\mathfrak{g}')$ and with $\mathrm{pr}^{-1}(0) = \mathfrak{c}'$ contained in the center of $\mathfrak{g}'$] so that the projective representation ρ of $\mathfrak{g}$ can be lifted to an ordinary representation $\rho' = \rho \circ \mathrm{pr}$ of $\mathfrak{g}'$. We then face the question of the uniqueness of $(\mathfrak{g}',\rho')$ as a function of $(\mathfrak{g},\rho)$. This question was first studied by Schur in a sequence of papers for finite groups; see Ref. 10. His ideas have become a part of homological algebra in terms of the (co)homology theory of groups and algebras; see Hochschild and Serre.[11] For our purposes, Schur's original treatment is adequate. Since the description of all representations ρ of $\mathfrak{g}$ is likely to be quite difficult, we first ask for the existence of a single central extension $\mathfrak{g}'$ that could be used to lift all possible projective representations ρ of $\mathfrak{g}$. Such an extension is called a *versal central extension* of $\mathfrak{g}$. (The work "versal" is artificially obtained by using the second half of the word universal.) It turns out that versal extensions always exist but may not be unique. Namely, we can always form the direct sum of a versal extension with an arbitrary Abelian Lie algebra to get another versal central extension. A versal central extension is called a *universal central extension* when it enjoys the uniqueness property. Namely, for each representation ρ of $\mathfrak{g}$, the lift ρ' to $\mathfrak{g}'$ is unique. By looking at commutators, we can mimic Schur's observation that $\mathfrak{g}$ has a universal central extension $\mathfrak{g}'$ (that is then unique up to uniquely determined isomorphism) if and only if $\mathfrak{g}$ coincides with its commutator subalgebra $[\mathfrak{g},\mathfrak{g}]$. When this happens, $\mathfrak{g}'$ also coincides with its own commutator subalgebra $[\mathfrak{g}',\mathfrak{g}']$. Lie algebras that coincide with their commutator subalgebras are often called *perfect*. In particular, every simple Lie algebra is perfect. In view of (2.4) and Sec. II, $\mathscr{D}(V,N)$ is always perfect even though it is simple only when $r_{\mathbb{Q}}(V) = N$. (It should be noted that an automorphism of $\mathfrak{g}$ does not have to yield the identity on the "central charges" in $\mathfrak{g}'$. Thus the uniqueness of $\mathfrak{g}'$ refers to the identity automorphism of $\mathfrak{g}$).

The description of the universal central extension $\mathfrak{g}'$ of a perfect Lie algebra $\mathfrak{g}$ is usually accomplished through the use of two-cocycles. From an elementary physicist's point of view, this reduces to the requirement that a candidate central extension terms must satisfy the Jacobi identity. A more mathematical approach is as follows. If we only want to have a central extension that is tailor made for a particular irreducible projective representation ρ of $\mathfrak{g}$, then the two-cocycle can be taken to have value in $\mathbb{C}$. Similarly, if the universal central extension $\mathfrak{g}'$ happens to have $\mathfrak{c}'$ of dimension 1 over $\mathbb{C}$, then we can again use $\mathbb{C}$-valued two-cocycles to describe $\mathfrak{g}$. For example, the case of Virasoro algebra leads to the notion of a central charge in the description of the universal central extension of $\mathrm{Diff}(S^1)$. As indicated earlier (1.3) belongs to a larger family of central extensions. This signifies the fact that $\mathfrak{c}'$ may have dimension greater than 1. As a result, we need to consider general two-cocycles with values in a $\mathbb{C}$-vector space. The idea of Schur can be described quite simply. Namely, we first describe a versal central extension with an associated two-cocycle with values in a $\mathbb{C}$-vector space D. Its commutator subalgebra is then the desired universal central extension. This reduces C down to $\mathfrak{c}'$. We recall that a two-cocycle g of $\mathfrak{g}$ with values in a $\mathbb{C}$-vector space C is a map g: $\mathfrak{g} \times \mathfrak{g} \to C$ with the following properties;

g is $\mathbb{C}$ bilinear,

g is skew symmetric, $g(x;y) = -g(y;x)$, (4.1)

g satisfies the Jacobi identity, $g(x;[y,z]) + \text{cyclic} = 0$.

The associated central extension $E(g) = \mathfrak{g}'$ can then be identified with the $\mathbb{C}$-vector space direct sum $\mathfrak{g} \oplus C$ and the Lie bracket is then defined by the rule:

$$[j(x),j(y)] = j([x,y]) + g(x;y), \quad x,y \in \mathfrak{g}. \tag{4.2}$$

Here $j: \mathfrak{g} \to E(g) = \mathfrak{g}'$ so that $\mathrm{pr} \circ j$ is the identity map on $\mathfrak{g}$. It goes without saying that $[j(x),c] = 0$ holds for $x \in \mathfrak{g}$ and $c \in C$. It should be noted that the map j is essential in this description. Namely, j can only be altered to j' so that $j'(x) = j(x) + h(x)$. The $\mathbb{C}$-vector space homomorphism

$h:\mathfrak{g}\to C$ is arbitrary and is called a 1 cochain. The two-cocycle g is then changed to the new two-cocycle g' with $g'(x;y)=g(x;y)-h([x,y])$. The cocycle g' is said to differ from g by the two-coboundary δh. The strategy in describing the universal central extension is to modify g by two-coboundaries δh until we get to a manageable two-cocycle g'. In general, the final g' is not unique. The choice is typically made on the basis of its utility. This is the place where confusion of notations unavoidably arises. Once the universal central extension $\mathfrak{g}'$ is found, the $\mathbb{C}$-valued two-cocycles have the form of $f\circ g$ with f ranging over the $\mathbb{C}$-linear maps from $\mathfrak{c}'$ to $\mathbb{C}$. In view of (4.1), it is then apparent that $\mathfrak{c}'$ can be identified as the quotient of the $\mathbb{C}$-vector space $\Lambda^2_{\mathbb{C}}(\mathfrak{g})$ of skew symmetric two-tensors by the subspace spanned by $x\wedge[y,z]+$ cyclic as x, y, and z range over $\mathfrak{g}$. It is important to note that the collection of all two-cocycles with values in any $\mathbb{C}$-vector space is again a $\mathbb{C}$-vector space. In particular, the two-coboundaries form a vector subspace of the space of all two-cocycles. Usually, two-coboundaries are called "trivial two-cocyles." This does not mean that they are zero. It does mean that the associated central extention $E(\delta h)$ is a direct sum of $\mathfrak{g}$ and an Abelian Lie algebra. To be precise, j' is a Lie algebra homomorphism so that $\mathfrak{g}'$ is the Lie algebra direct sum of $\mathfrak{c}'$ and $j'(\mathfrak{g})\cong\mathfrak{g}$.

There is one other important property of a perfect Lie algebra $\mathfrak{g}$. If $\mathfrak{g}'$ is the direct sum of $\mathfrak{g}$ and an Abelian Lie algebra $\mathfrak{c}'$, then $\mathfrak{g}$ is uniquely determined as the commutator subalgebra $[\mathfrak{g}',\mathfrak{g}']$. In terms of two-cocycle g of $\mathfrak{g}$ with values in $\mathfrak{c}'$, this means that, once g is trivial, the lift j' from $\mathfrak{g}$ to $\mathfrak{g}'$ is uniquely determined (as a Lie algebra homomorphism) rather than unique up to an additive modification by a Lie algebra homomorphism from $\mathfrak{g}$ to $\mathfrak{c}'$. This is a direct consequence of the fact that a Lie algebra homomorphism from a perfect Lie algebra to an Abelian Lie algebra is necessarily the zero map. We will use this fact later.

When $\mathfrak{g}$ is perfect and has a Cartan decomposition, the description of the universal central extension $\mathfrak{g}'$ by means of a two-cocycle is somewhat simpler. To be concrete, we specialize to $\mathfrak{g}=\mathscr{D}(N,V)$. We write $g(u,m;v,n)$ for $g(u;v)$ when $u\in\mathfrak{g}_m$ and $v\in\mathfrak{g}_n$. For any root space $\mathfrak{g}_n$, $n\neq 0$, each nonzero element x of $\mathfrak{g}_n$ is a simultaneous eigenvector for the adjoint action of $\mathfrak{g}_0$. Since $n\neq 0$ and $\mathfrak{c}'+\mathbb{C}\cdot j(x)$ is closed for the adjoint action in $\mathfrak{g}'$ of the Lie subalgebra $\mathfrak{c}'+j(\mathfrak{g}_0)$, it becomes clear that there is a unique lift $j(x)$ so that $j(x)$ remains a simultaneous eigenvector for this lifted adjoint action. Moreover, j can be taken to be $\mathbb{C}$ linear on $\mathfrak{g}_n$. From the Jacobi identity, we get $[\mathfrak{g}_m,\mathfrak{g}_n]\subset\mathfrak{g}_{m+n}$. It then follows without problem that the $\mathbb{C}$ linear map j defined on the root spaces $\mathfrak{g}_n$ for $n\neq 0$ can be assumed to respect the Lie bracket in the sense that

$$j([x,y])=[j(x),j(y)],\quad \text{if } x\in\mathfrak{g}_m, y\in\mathfrak{g}_n, \text{ and } m+n=0.$$

In other words, the two-cocycle g may be assumed to have the property,

$$g(u,m;v,n)=0,\quad \text{if } m+n\neq 0. \tag{4.3}$$

This means that we only have to determine the lift $j(x)$ for $x\in\mathfrak{g}_0$. Since j is assumed to be $\mathbb{C}$ linear and $\mathfrak{g}_0$ has a distinguished $\mathbb{C}$ basis, we are reduced to the description of j on these distinguished basis elements. This will be accomplished by shifting to an examination of g on pairs of distinguished basis elements of $\mathfrak{g}$ through the Jacobi identity.

We will now proceed to the case of $\mathscr{D}(N,V)$ with $\dim V=\ell=1$. In order to have the goal in front of us, we will exhibit the final result immediately. For the sake of concreteness, we will take a single structure vector $v=(v_1,\ldots,v_r,0,\ldots,0)\in{}^N\mathbb{C}$ where $r=r_{\mathbb{Q}}(V)$ so that $v_1,\cdots,v_r$ are rationally-independent complex numbers. We may define $w=(v_1,\ldots,v_r)\in{}^r\mathbb{C}$ so that $\mathscr{D}(N,V)\cong\mathscr{L}(N-r)\otimes\mathscr{D}(r,W)$. It will be more convenient to write $\mathbb{Z}^N$ as the direct sum $\mathbb{Z}^r\oplus\mathbb{Z}^{N-r}$. We note that $\mathbb{Z}^{N-r}=V^{\perp}\cap\mathbb{Z}^N$ is uniquely determined by V (or v) while $\mathbb{Z}^r$ is one of many complements of $\mathbb{Z}^{N-r}$ in $\mathbb{Z}^N$ when $r<N$. We will now let $p,q,\ldots$ vary over $\mathbb{Z}^{N-r}$ and $m,n\ldots$ vary over $\mathbb{Z}^r$. With these notations, we will next define $\tilde{\mathscr{D}}(N,v)$ by generators M_pL_m and c_q together with the commutators:

$$[M_pL_m,M_qL_n]=\langle w|m-n\rangle M_{p+q}L_{m+n}+\delta_{m+n,0}\cdot A(\langle w|m\rangle)\cdot c_{p+q},$$
$$A(t)=(t^3-t)/12,\quad t\in\mathbb{C},$$

and

$$[M_pL_m,c_q]=0, p,q\in\mathbb{Z}^{N-r},\quad m,n\in\mathbb{Z}^r. \tag{4.4}$$

In the above definition, the parameters c_q, $q\in\mathbb{Z}^{N-r}$, are to be viewed as independent basis elements of a $\mathbb{C}$-vector space C, the center of $\tilde{\mathscr{D}}(N,v)$. By setting all c_q equal to 0, we may identify M_pL_m with $L_{v,(m,p)}\in\mathscr{D}(N,V)$, in other words, $M_pL_m=j(L_{v,(m,p)})$. It is then clear that $\tilde{\mathscr{D}}(N,v)$ is a central extension of $\mathscr{D}(N,V)$. If we calculate $[M_pL_{2m},M_qL_{-2m}]-2[M_pL_m,M_qL_{-m}]$, then $t=\langle w|m\rangle\neq 0$ implies that c_s lies in the commutator subalgebra for each $s\in\mathbb{Z}^{N-r}$. This shows that $\tilde{\mathscr{D}}(N,v)$ is equal to its own commutator subalgebra. If ρ denotes an irreducible representation of $\tilde{\mathscr{D}}(N,v)$, then each c_p must be represented by a scalar multiple $c_\rho(p)$ times the identity operator. This then defines a complex-valued function c_ρ on the set $\mathbb{Z}^{N-r}$. It should be noted that one could start with the apparently more general central extension:

$$[M_pL_m,M_qL_n]=\langle w|m-n\rangle M_{p+q}L_{m+n}+\tfrac{1}{12}\cdot\delta_{m+n,0}[c^{(3)}_{p+q}t^3+c^{(1)}_{p+q}t],$$

depending on two different arbitrary functions $c^{(3)}_{p+q}$ and $c^{(1)}_{p+q}$ from $\mathbb{Z}^{N-r}$ to $\mathbb{C}$, multiplying the t^3 and t terms. This form certainly satisfies the Jacobi identity. However, one can redefine the generators M_pL_0 for each $p\in\mathbb{Z}^{N-r}$ in such a way as to render $c^{(1)}_{p+q}=-c^{(3)}_{p+q}$. In general, any complex-valued function on $\mathbb{Z}^{N-r}$ will be called a *central charge function*. These are exactly the equivalence classes of $\mathbb{C}$-valued two-cocycles on $\mathscr{D}(N,v)$. It is not known which of these central charge functions can arise from the irreducible representations of $\tilde{\mathscr{D}}(N,v)$. Such representations would lead to irreducible projective representations of $\mathscr{D}(N,v)$. We now state the principal result in this section.

Theorem 4.5: Let $V=\mathbb{C}v$. Then $\tilde{\mathscr{D}}(N,v)$ as defined in (4.4) is the universal central extension of $\mathscr{D}(N,V)$. In other words, every perfect central extension of $\mathscr{D}(N,V)$ is obtained as $\tilde{\mathscr{D}}(N,v)/J$ for a suitable $\mathbb{C}$-vector subspace J of the center C of $\tilde{\mathscr{D}}(N,v)$.

To begin our proof, we will take C to be the quotient of $\Lambda_{\mathbb{C}}^2(\mathfrak{g})$ as indicated earlier. Thus, we have a definite vector space for the values of the two-cocycle. The defining relations for the universal central extension $E = E(g)$ can now be written in the form;

$$[M_pL_m, M_qL_n] = \langle w|m-n\rangle \cdot M_{p+q}L_{m+n} + g(p,m;q,n). \tag{4.6}$$

Our goal is to identify the form of $g(p,m;q,n)$ as displayed in (4.4). We note that the assumption $\ell = 1$ has the immediate consequence that $C \otimes \mathbb{C} \cdot M_0L_{w,0}$ is automatically an Abelian subalgebra of E. However, this is true but certainly not yet obvious for the subalgebra $C \otimes j(\mathfrak{g}_0)$. Under the adjoint action of $C \otimes j(\mathfrak{g}_0)$ on $E, C \otimes \mathbb{C} \cdot M_pL_m$ is an eigenspace with C corresponds to the zero eigenvalue. Equation (4.3) now translates to the normalization assumption

$$g(p,m;q,n) = 0, \quad m+n \neq 0. \tag{4.7}$$

For the remaining arguments in the proof of Theorem 4.5, we assume (4.7). We divide our task into two cases according to $r = N$ and $r < N$.

Case 1. $r=N$: This case was dealt with by Ramos and Shrock in Ref. 4. We reorganize the proof so as to set the tone for generalization. We can simplify our notation by setting $g(m;n) = g(0,m;0,n)$. As indicated, our problem reduces to the determination of $g(m;-m)$ for $m \neq 0$ in $\mathbb{Z}^N$. The Jacobi identity in E for M_0L_k, M_0L_m, and M_0L_n with $k+m+n=0$ is equivalent to the condition

If $k+m+n=0$ in $\mathbb{Z}^N$, then

$$\langle v|m-n\rangle g(k;-k) + \langle v|n-k\rangle g(m;-m) + \langle v|k-m\rangle g(n;-n) = 0. \tag{4.8}$$

We observe that the collection $\{L_{tk} | t \in \mathbb{Z}\}$, $k \neq 0$, spans a subalgebra of $\mathscr{D}(N,v)$ that is isomorphic to $\mathrm{Diff}(S^1)$ through a rescaling process. This uses the fact that $r_{\mathbb{Q}}(v) = N$. Thus, the two-cocycle has a good description as indicated in the definition of the Virasoro algebra. Our task is to show that these descriptions fit together. At this juncture, we know $\dim C \geqslant 1$ because of the central extension displayed in (4.4). If we can show that $\dim C \leqslant 1$, then we are done. Of course, when $N=1$, we are done by means of rescaling in conjunction with the known result of Gel'fand and Fuchs. Our strategy is to reduce the general case down to the case of $N=2$ by an induction process. Namely, we will show that $\dim C \leqslant 1$. To be precise, we will show that the following assertion holds.

Lemma 4.9: If two-cocycle g with property (4.7) is 0 on a subalgebra of $\mathscr{D}(N,v)$ spanned by all L_{tk}, $t \in \mathbb{Z}$, where $\mathbb{Z}k$ is a particular nonzero direct summand of $\mathbb{Z}^N$, then g is identically 0.

Remark: We observe that Lemma 4.9 is formulated in such a way that the values of the two-cocycle g can lie in any $\mathbb{C}$-vector space. If we wish, we can take C to be $\mathbb{C}$. Furthermore, the proof of Lemma 4.9 can be reduced to the case $N=2$. Namely, to show that g vanishes identically, we only have to show that $g(n;-n)=0$ for all $n \in \mathbb{Z}^N$. Since $\mathbb{Z}^N$ is a free Abelian group, any subgroup generated by M elements is contained in a direct summand with at most M generators. This shows that we can find $\mathscr{A} \in \mathrm{GL}(N,\mathbb{Z})$ $\mathscr{A} \cdot k = e(1)$ and $\mathscr{A} \cdot n \in \mathbb{Z}e(1) + \mathbb{Z}e(2)$. Since $v' = \mathscr{A} \cdot v$ has $\mathbb{Q}$ rank N, we see that L_k and L_n lie in a subalgebra of $\mathscr{D}(N,V)$ that is isomorphic to $\mathscr{D}(2,v'')$ where $v'' = (v_1', v_2')$ has $\mathbb{Q}$ rank 2. This reduces the general case of Lemma 4.9 from N to 2.

We now prove Lemma 4.9 for the case of $N=2$. We may take $k = e(1)$. Let us take $m = ae(1)$, $a \in \mathbb{Z}$, in (4.7). Since $g(m;-m)=0$, we obtain

$$\langle v|n+2ae(1)\rangle \cdot g(n;-n) = \langle v|n-ae(1)\rangle \cdot g(n+ae(1);-n-ae(1)). \tag{4.10}$$

Since we can assume $n \notin \mathbb{Z}e(1)$, the two coefficients in (4.10) are nonzero. This yields the recursion rule

$$g(n+ae(1);-n-ae(1)) = \frac{\langle v|n\rangle + 2a\langle v|e(1)\rangle}{\langle v|n\rangle - a\langle v|e(1)\rangle} \cdot g(n;-n). \tag{4.11}$$

Equation (4.11) permits us to compute $g(n+(a+b)e(1);-n-(a+b)e(1))$ in two different ways. One way involves the direct replacement of a by $a+b$ in (4.11) and the other way involves using the associative law and iterate (4.11) twice. We obtain from this calculation,

$$(3abA^2B - 2ab(a+b)A^3) \cdot g(n;-n) = 0, \quad a,b \in \mathbb{Z} \tag{4.12}$$

where $A = \langle v|e(1)\rangle, B = \langle v|n\rangle$. Since $A \neq 0 \neq B$ and $a,b \in \mathbb{Z}$ are arbitrary, we must have $g(n;-n)=0$. Since $n \notin \mathbb{Z}e(1)$ is arbitrary and $g(ae(1);-ae(1))=0$ is assumed, we see that g is identically 0. This prove Lemma 4.9. We note that A and B are in fact rationally independent so that it is only necessary to take a, b in $\mathbb{Z}$ with $a \neq 0 \neq b$ to deduce the vanishing of $g(n;-n)$.

To deduce $\dim C \leqslant 1$ from Lemma 4.9, we take any subspace J of C that contains the subspace spanned by $g(e(1);-e(1))$. Then, $E' = E/J$ is a central extension with a companion two-cocycle g' satisfying the hypothesis of Lemma 4.9. Since the universal central extension E is perfect, the vanishing of g' shows that E' must be $\mathfrak{g}$ so that $\dim C \leqslant 1$ as desired. We have concluded that proof in case 1.

Case 2. $r<N$: Our two-cocycle g is assumed to satisfy (4.7). Now $\mathscr{D}(N,v)$ is isomorphic to $\mathscr{L}(N-r) \otimes \mathscr{D}(r,w)$ and we can identify $\mathscr{D}(r,w)$ with a subalgebra of $\mathscr{D}(N,V)$ by means of the generators $L_{v,(m,0)}$, $m \in \mathbb{Z}^r$. It follows that $g(0,m;0,n)$, $m,n \in \mathbb{Z}^r$, can be viewed as a two-cocycle on $\mathscr{D}(r,w)$. Since modification of a two-cocycle satisfying (4.7) on $\mathscr{D}(r,w)$ amounts to adjustment of the generator M_0L_0, we can use the result in case 1 to make the adjustment and conclude that $g(0,m;0,n) = \delta_{m+n,0} A(\langle w|m\rangle) \cdot c_0$. This illustrates the general strategy of our proof. We will show that for each $t \in \mathbb{Z}^{N-r}$, $g(p,m;q,n)$ depends only on $t = p+q$ and $m,n \in \mathbb{Z}^r$. Moreover, with t fixed, the resulting function satisfies (4.8) so that it can be viewed as a two-cocycle on $\mathscr{D}(r,w)$. Since modifications of such two-cocycles for fixed $t \in \mathbb{Z}^{N-r}$ can be accomplished by suitable modification of M_tL_0, it means that $g(p,m;q,n) = \delta_{m+n,0} A(\langle w|m\rangle) \cdot c_t$ holds because we already have a central extension with independent c_t, $t \in \mathbb{Z}^{N-r}$.

We begin with the reminder that (4.7) may be assumed

in general. By looking at the Jacobi identity for elements M_sL_k, M_pL_m, and M_qL_n in E with $k+m+n\neq 0$, we obtain the generalization of (4.8),

$$\langle w|m-n\rangle\cdot g(s,k;p+q,-k) + \langle w|n-k\rangle\cdot g(p,m;q+s,-m) + \langle w|k-m\rangle\cdot g(q,n;s+p,-n)=0. \tag{4.13}$$

If we take $k=n\neq 0\neq m$ and $q=0$ in (4.13), we then have

$$g(s,k;p,-k)\equiv g(0,k;s+p,-k),\quad k\neq 0. \tag{4.14}$$

If we take $k=-m\neq 0=n$ in (4.13) and combine the skew symmetry of g with (4.14), then $g(s,0;p,0)=0$ holds for all s, $p\in\mathbb{Z}^{N-r}$. Thus, for each $k\in\mathbb{Z}^r$, $g(s,k;p,-k)=g(0,k;s+p,-k)$ is a function of $s+p\in\mathbb{Z}^{N-r}$. If we denote $g(p,k;q,-k)$ by $g_{p+q}(k;-k), k\in\mathbb{Z}^r$, then (4.13) takes over the role of (4.8) and the skew symmetry follows from the symmetry in p, q and (4.1). As indicated, this concludes the proof of Theorem 4.5.

V. UNIVERSAL CENTRAL EXTENSIONS. DIM $V>1$

The main result in this section is the following theorem.

Theorem 5.1: If $\dim(V)=\ell>1$, then $\mathcal{D}(N,V)$ has no nontrivial central extension.

In order to avoid confusion, we begin with a central extension $\mathfrak{g}'$ of $\mathcal{D}(N,V)$ with an associated two-cocycle g. We will perform modifications on g until it becomes identically 0. Since this is the goal, we may assume that g takes value in $C=\mathbb{C}$. To carry out modification on g means that we modify $j(L_{u,n})=L'_{u,n}$ in $\mathfrak{g}'$ by adding on suitable elements in C. This modification will be done on suitably selected basis elements and extended through linearity over $\mathbb{C}$. Throughout the rest of the section, (4.7) will be assumed so that the modifications are restricted to $n=0$. In analogy with Sec. IV, we divide the proof into various cases. We begin with some notations. The standard basis of $\mathbb{Z}^N$ over $\mathbb{Z}$ will be denoted by $e(j)$, $1\leqslant j\leqslant N$. The standard dual basis of ${}^N\mathbb{C}$ over $\mathbb{C}$ will also be denoted by $e(j)$, $1\leqslant j\leqslant N$, so that $\langle e(i)|e(j)\rangle=\delta_{ij}$. In $\mathcal{D}(N)$, the generator $L_{e(i),e(j)}$ will be abbreviated to $L_{i,e(j)}$, where $1\leqslant i,j\leqslant N$. We consider $\mathcal{D}(N,V)$ as a subalgebra of $\mathcal{D}(N)$.

Case 1. dim $V=\ell>1$ and $r_{\mathbb{Q}}(V)=N$: As indicated above, we assume (4.7). With appropriate modifications, we will show that we can also achieve (4.7) for $m+n=0$. Since g is $\mathbb{C}$ bilinear, we may assume that u and v range over suitable $\mathbb{C}$-vector space basis of V. It is then clear that the case $\ell=2$ holds the key.

We begin with the special case of $\ell=2=r_{\mathbb{Q}}(V)=N$, namely with $\mathcal{D}(2)$. Since $L_{i,ae(i)}$, $a=0,\ \pm 1$, $i=1,2$, span the direct sum of two commuting copies of $\mathfrak{sl}(2,\mathbb{C})$, we may use the semisimplicity of *finite-dimensional* Lie algebras over $\mathbb{C}$ to make the following assumption:

$$g(i,ae(i);j,be(j))=0,\quad \text{for } |ab|\leqslant 1, a,b\in\mathbb{Z},\ i,j=1,2. \tag{5.2}$$

We note that (5.2) refers to a $\mathbb{C}$ basis of V that is dual to the $\mathbb{Z}$ basis of $\mathbb{Z}^2$. When we come to the general case of $N>2$, we need to consider various direct summand of rank 2 in $\mathbb{Z}^N$. This will force us to change the $\mathbb{C}$ basis of V. In the present case, (5.2) together with linearity fixes $L'_{u,0}$ so that we have no more freedom left. It is only natural for us to show the intermediate result.

Lemma 5.3: With the preceding notation, if g is a $\mathbb{C}$-valued 2 cocycle on $\mathcal{D}(2)$ so that (4.7) and (5.2) hold, then g is identically 0. In particular, Theorem 5.1 holds for $D(2)$.

Proof: The Jacobi identity for $L'_{1,e(1)}, L'_{2,p}, L'_{2,q}$ with $e(1)+p+q=0$ yields

$$p_1\cdot g(2,p+e(1);2,-p-e(1))=(p_1+1)\cdot g(2,p;2,-p). \tag{5.4}$$

Here $p=(p_1,p_2)\in\mathbb{Z}^2$ and (5.3) is independent of p_2. Similarly, let $q=(q_1,q_2)\in\mathbb{Z}^2$. If we take $q_1=0$, then

$$g(2,be(2);2,-be(2))=0 \quad \text{holds for all } b\in\mathbb{Z}. \tag{5.5}$$

By symmetry, we obtain from (5.5) also the result

$$g(1,ae(1);1,-ae(1))=0 \quad \text{holds for all } a\in\mathbb{Z}. \tag{5.6}$$

We repeat with $L'_{1,ae(1)}, L'_{2,p}, L'_{2,q}$, $ae(1)+p+q=0$, $a\in\mathbb{Z}$, to get

$$q_1\cdot g(2,q+ae(1);2,-q-ae(1)) = (q_1+a)\cdot g(2,q;2,-q). \tag{5.7}$$

Computing $p_1\cdot g(2,p+(a+b)e(1);2,-p-(a+b)e(1))$ in two different ways by (5.7), we get

$$(p_1+a+b)\cdot g(2,p;2,-p) = (p_1+a)\cdot(p_1+b)\cdot g(2,p;2,-p). \tag{5.8}$$

Since $a,b\in\mathbb{Z}$ are arbitrary, we conclude from symmetry that

$$g(i,q;i-q)=0 \quad \text{holds for any } q\in\mathbb{Z}^2,\quad i=1,2. \tag{5.9}$$

We next examine the Jacobi identity for $L'_{1,p}, L'_{2,q}, L'_{2,s}$ with $p+q+s=0$. This gives

$$\{s_2-q_2\}\cdot g(1,p;2,-p) = p_2\cdot\{g(1,-q;2,q)-g(1,-s;2,s)\}. \tag{5.10}$$

With $p_2=0\neq q_2=-s_2$, we see that $g(1,p;2,-p)=0$ holds whenever $p_2=0$. By symmetry, $g(1,p;2,-p)=0$ holds if either p_1 or p_2 is 0. With $r_2=q_2\neq 0$ so that $p_2=-2q_2\neq 0$, we conclude from (5.10) that $g(1,s;2,-s)$ depends only on s_2. Since $g(1,s;2,-s)=0$ holds when $s_1=0$, we can conclude that $g(1,s;2,-s)=0$ holds for all $s\in\mathbb{Z}^2$. Since (4.7) is assumed, g is identically 0. □

We now upgrade our special case to the case where $\ell=2$ but $r_{\mathbb{Q}}(V)=N$ is arbitrary. We can assume $N\geqslant 3$. For any subset J of $\{1,\ldots,N\}$, let $\mathcal{D}_J$ be the subalgebra spanned by $L_{u,m}$ where $u\in V$ and m ranges over the integral linear combinations of $e(j)$ in $\mathbb{Z}^N$, $j\in J$. Thus $\mathcal{D}_{\{1,2\}}\cong\mathcal{D}(2)$ under an isomorphism of the type described by (2.3). We can therefore carry out the modification and use Lemma 5.3 to assume that $g(u,m;v,n)$ is identically 0 on $\mathcal{D}_{\{1,2\}}$ in addition to (4.3). We must now show that $g(u,m;v,-m)$ is 0 for m in $\mathbb{Z}^N$. Of course, we may assume $m\notin\mathbb{Z}e(1)+\mathbb{Z}e(2)$. By using an isomorphism of the form (2.3), we may assume that $m=s\cdot e(3)$ for some nonzero integer s. We can choose a $\mathbb{C}$-vector basis $v(1)$ and $v(2)$ for V so that a Cartan matrix M for $\mathcal{D}(N,V)$ has the following form:

$$\begin{pmatrix} 1 & 0 & a & \cdots \\ 0 & 1 & b & \cdots \end{pmatrix},\quad a \text{ or } b \text{ is irrational.}$$

Without loss of generality, we can assume $a\notin\mathbb{Q}$ so that $\mathcal{D}_{\{2,3\}}\cong\mathcal{D}(2)$ by an isomorphism of type (2.3). We note that such

isomorphisms exist because of the assumption $r_{\mathbb{Q}}(V)=N$, the display of the Cartan matrix merely exhibits this assumption in a more precise way. We can therefore conclude that the restriction of the two-cocycle to $\mathscr{D}_{\{2,3\}}$ is a trivial two-cocycle. By modifying g to g', we can conclude that g' is identically zero on $\mathscr{D}_{\{2,3\}}$. We note that the modifications involve only basis elements of the form $L'_{u,0}, u\in V$. It is easy to see that $\mathscr{D}_{\{2\}}=\mathscr{D}_{\{1,2\}}\cap\mathscr{D}_{\{2,3\}}\cong\mathscr{L}(1)\otimes\mathscr{D}(1)$ is a perfect subalgebra. By choice, the restriction g_1 of g to $\mathscr{D}_{\{2\}}$ through $\mathscr{D}_{\{1,2\}}$ is identically 0. Similarly, the restriction g_3 of g' to $\mathscr{D}_{\{2\}}$ through $\mathscr{D}_{\{2,3\}}$ is also identically 0. We know that g and g' differ by a 2 coboundary δh. It follows that g_1 and g_3 differ by the two-coboundary $\delta h'$, where h' is the restriction of h to $\mathscr{D}_{\{2\}}$. From the observation made at the end of the review of central extensions in Sec. IV, we conclude that $h'=0$. Since $\mathscr{D}_{\{2\}}$ contains the Cartan subalgebra $\mathfrak{g}_0$, the vanishing of h' is the same as the vanishing of h. Namely, no modification of g is needed at all. This is the same as saying that $g(u,m;v,-m)=0$ already holds. We have upgraded our assertion to the case of $\ell=2$ and $r_{\mathbb{Q}}(V)=N$ is arbitrary.

We next upgrade to the general case of $2\leqslant\ell\leqslant r_{\mathbb{Q}}(V)=N$. We use a similar idea. By Proposition 2.5, we can find a $\mathbb{C}$-vector space basis $\{v(i)\,|\,1\leqslant i\leqslant\ell\}$ with $r_{\mathbb{Q}}(v(i))=N$ for $1\leqslant i\leqslant N$. Our goal is to carry out modifications on $L'_{u,0}$ in order to get $g(u,m;v,-m)$ to be 0. Since g is $\mathbb{C}$ bilinear, it is enough to achieve this for u and v independently ranging over $\{v(i)\,|\,1\leqslant i\leqslant\ell\}$. For each nonempty subset J of $\{1,\ldots,\ell\}$, let $\mathscr{D}^J$ be the subalgebra spanned by all $L_{u,m}$, $m\in\mathbb{Z}^N$, and u ranging over the $\mathbb{C}$ subspace V^J of V spanned by $v(j)$ with $j\in J$. It is then clear that $\mathscr{D}^J\cong\mathscr{D}(N,V^J)$ with $\dim V^J=\#(J)$ and $r_{\mathbb{Q}}(V)=N$. If $J=\{i,j\}$ with $i\neq j$, then the previous upgrade shows that the restriction g^J of g to $\mathscr{D}^J$ is a trivial two-cocycle. If i,j,k are distinct, then $\mathscr{D}^{\{i,j\}}\cap\mathscr{D}^{\{j,k\}}=\mathscr{D}^{\{j\}}$ is perfect. We can use the same kind of reasoning to show that the independently carried out modifications to get g to be identically zero on each $\mathscr{D}^{\{i,j\}}$, $i\neq j$, in fact fit together without further modifications. This completes the argument in case 1.

Case 2. Let dim $V=\ell>1$ and $r_{\mathbb{Q}}(V)=r<N$: We have $\mathscr{D}(N,V)\cong\mathscr{L}(N-r)\otimes\mathscr{D}(r,W)$. We mimic case 2 of Sec. IV. The Jacobi identity for $M_sL_{u,k}$, $M_pL_{v,m}$, $M_qL_{w,n}$ in E with $k+m+n=0$ yields the following analog of (4.13):

$$\begin{aligned}&g(s,u,k;p+q,v\cdot\langle w|m\rangle-w\cdot\langle v|n\rangle,-k)\\&\quad+g(p,v,m;q+s,w\cdot\langle u|n\rangle-u\cdot\langle w|k\rangle,-m)\\&\quad+g(q,w,n;s+p,u\cdot\langle v|k\rangle-v\cdot\langle u|m\rangle,-n)=0.\end{aligned}\tag{5.11}$$

Since g is $\mathbb{C}$ bilinear in the vectors varying over W, we can take u,v,w to range over a $\mathbb{C}$ basis of W consisting of vectors with $\mathbb{Q}$ rank $r=r_{\mathbb{Q}}(W)$. If we take $k=n\neq0\neq m$, $q=0$, and $u=w$ in (5.11), then we obtain

$$\begin{aligned}&g(s,u,k;p,v\cdot\langle u|2k\rangle+u\cdot\langle v|k\rangle,-k)\\&\quad=g(0,u,k;s+p,v\cdot\langle u|2k\rangle+u\cdot\langle v|k\rangle,-k),\quad k\neq0.\end{aligned}\tag{5.12}$$

If we first take $u=v$ in (5.12) to be a vector in W with $r_{\mathbb{Q}}(u)=r$ so that $\langle u|k\rangle\neq0$, we can then combine the intermediate result with the $\mathbb{C}$ bilinearity of g to deduce from (5.12)

$$g(s,u,k;p,v,-k)=g(0,u,k;s+p,v,-k)\,,\tag{5.13}$$

where $k\neq0$, $r_{\mathbb{Q}}(u)=r_{\mathbb{Q}}(v)=r$. By Proposition 2.5 and the $\mathbb{C}$ bilinearity of g, (5.13) then holds for all pairs $u,v\in W$. If we take $k=-m\neq0=n$, $u=v$, we obtain from skew symmetry, $\mathbb{C}$ bilinearity and (5.13) to conclude that

$$g(p,w,0;q,u,0)\equiv0\,.$$

The argument is now similar to case 2 of Sec. IV. The only difference is that for each $t\in\mathbb{Z}^{N-r}$, we can modify $M_tL_{u,0}$ in order to reach the conclusion that g can be made identically 0 by way of case 1. This concludes the proof of Theorem 5.1 □

VI. CONCLUSIONS

In this paper we have elucidated the properties of the algebra $\mathscr{D}(N,V)$ of diffeomorphisms of the N-torus. From a physics point of view, one is also motivated to study such algebras as a natural generalization of the study of theories invariant under conformal transformations, corresponding to (two commuting copies of) the algebra of diffeomorphisms of the circle.

The present work concentrates on the mathematical aspects of this algebra. We find that our general algebra exhibits a very interesting and rich set of mathematical properties. We have defined the notion of a $\mathbb{Q}$ rank, $r_{\mathbb{Q}}(V)$, of the space V spanned by the structure vectors. We have shown that $\mathscr{D}(N,V)$ is simple if and only if $r_{\mathbb{Q}}(V)$ is maximal. We have then carried out a classification of $\mathscr{D}(N,V)$. Finally, we have established a number of properties concerning the central extensions of this algebra. In the case dim $V=1$, $r_{\mathbb{Q}}(V)=N$, we have shown that the central extension given in Ref. 4 is unique. In this sense, the universal central extension $\tilde{\mathscr{D}}(N,v)$ can be viewed as generalizations of the Virasoro algebra Vir. When dim $V=1$ and $r_{\mathbb{Q}}(V)<N$, we have constructed a much larger central extension involving a central charge *function* from $\mathbb{Z}^{N-r}$ to $\mathbb{C}$, rather than a single central charge parameter. We note that Diff(S^1) is somewhat like a relative of the loop algebra over $\mathfrak{sl}(2,\mathbb{C})$. Both contain $\mathfrak{sl}(2,\mathbb{C})$ as a maximal finite-dimensional simple subalgebra. Although the reasons are quite different, their universal central extensions have one dimensional centers. From a mathematical viewpoint, the iterated loop algebra extensions suggests that there could be some interesting mathematics related to the iterated loop algebra extensions of classical finite dimensional simple Lie algebras. Of course, such n-fold iterated loop algebra extensions naturally admit the action of $\mathscr{D}(n)$. In general, we do not know the precise structure of the universal central extensions of the n-fold iterated loop algebra extension of a general Lie algebra $\mathfrak{g}$. For example, in contrast to the loop algebra extension of a finite-dimensional simple Lie algebra $\mathfrak{g}$, in the case dim $V\geqslant2$, we have proved that $\mathscr{D}(N,V)$ has no nontrivial central extension. Currently, work is in progress on a number of further topics associated with the algebra $\mathscr{D}(N,V)$. One especially challenging area is that of representation theory.

Note: After this work was completed and the preprint circulated in June 1989, a paper by A. A. Balinskiĭ and S. P.

Novikov, "Poisson brackets of hydrodynamic type, Frobenius algebras and Lie algebras, " Sov. Math. Dokl. **32**, 228–231 (1985), was brought to our attention. Their specific example involved an Abelian (noncentral) ideal that led to an infinite number of central charges. In contrast, our algebras are either simple or are iterated loop algebras over simple algebras and do not have Abelian ideals.

ACKNOWLEDGMENTS

R. E. S. and E. R. are partially supported by the National Science Foundation Grant NSF-PHY-85-07627. C. H. S. is partially supported by the National Science Foundation Grant DMS-8901905 and by a grant from the Paul and Gabriella Rosenbaum Foundation.

[1]B. L. Feigin and D. B. Fuchs, Funct. Anal. Appl. **16**, 114 (1982); **17**, 241 (1983).
[2]P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 304 (1986).
[3]V. Kac, *Infinite-Dimensional Lie Algebras* (Birkhauser, Boston, 1983).
[4]E. Ramos and R. E. shrock, Int. J. Mod. Phys. A **4**, 4279 (1989).
[5]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[6]B. de Wit, U. Marquard, and H. Nicolai, "Area-preserving Diffeomorphisms and Supermembrane Lorentz Invariance," THU-89-5; E. G. Floratos and J. Iliopoulos, Phys. Lett. B **201**, 237 (1988); J. Hoppe, *ibid.* **215**, 706 (1988).
[7]D. B. Fairlie, P. Fletcher, and C. K. Zachos, Phys. Lett. B **218**, 203 (1989); D. B. Fairlie and C. K. Zachos, *ibid.* **224**, 101 (1989).
[8]H. Garland, J. Algebra **53**, 480 (1978); A. Pressley and G. Segal, *Loop Groups* (Clarendon, London, 1986).
[9]C. H. Sah, *Abstract Algebra* (Academic, New York, 1967).
[10]I. Schur and J. Reine, Angew. Math. **127**, 20 (1904); **132**, 85 (1907); **139**, 155 (1911).
[11]G. P. Hochschild and J. P. Serre, Ann. Math. **57**, 591 (1953).

The complete root systems of the affine Kac–Moody superalgebras

I. Tsohantjis and J. F. Cornwell
Department of Physics and Astronomy, University of St. Andrews, St. Andrews, Fife KY16 9SS, Scotland

(Received 26 October 1989; accepted for publication 21 March 1990)

By using certain special automorphisms, the complete root structures of all the affine Kac–Moody superalgebras, both twisted and untwisted, are derived in detail.

I. INTRODUCTION

In recent years, affine Kac–Moody algebras have found a number of significant applications in theoretical physics, probably the best known being its appearance in the theory of the heterotic string.[1] (For comprehensive reviews of the usage of Kac–Moody algebras in quantum physics see Goddard and Olive.[2]) Various "super" extensions of Kac–Moody algebras have also been considered in physical contexts.[3] The object of the present paper is to facilitate this development by giving a complete description of the root structure of the affine Kac–Moody superalgebras that were introduced by Kac.[4]

The structure of this paper is as follows. In Sec. II the basic structure of Kac–Moody superalgebras is briefly summarized, the aims being to establish notations, conventions, and objectives. Section II A describes the "abstract" formulation in terms of generators, while Sec. II B gives the more practical explicit construction. In Sec. III the roots of the untwisted affine Kac–Moody superalgebras are briefly discussed, the treatment being brief because the procedures are very similar to those for the untwisted affine Kac–Moody algebras. This brings us to the construction of the twisted Kac–Moody superalgebras, which is the main substance of this paper. The problems with the canonical automorphisms are described in general terms in Sec. IV. By using certain, more convenient choices of the automorphisms the complete roots structures of $A^{(2)}(2l-1/0)$, $A^{(4)}(2l/0)$, and $C^{(2)}(l+1)$ are investigated in Secs. V–VII. Because in every case no less than three superalgebras are involved, each with their own sets of simple roots, invariant supersymmetric bilinear forms, and so on, it is necessary to distinguish clearly between these similar but different quantities. The arguments are presented in sufficient detail to accomplish this.

II. BASIC STRUCTURE OF KAC–MOODY SUPERALGEBRAS

A. Introduction

The original investigation of Kac–Moody superalgebras (under the name of "contragredient Lie superalgebras") was carried out by Kac,[4,5] the underlying idea being to generalize the concepts of Kac–Moody algebras[6–8] from the Lie algebra to the Lie superalgebra situation. Although the general definition of Kac–Moody superalgebras was deliberately set up in such a way that the Kac–Moody algebras appear as a special case, for ease of exposition it will henceforth be assumed here that (unless otherwise stated) all the Kac–Moody superalgebras which will be discussed have *nontrivial odd* parts. With this assumption it was shown by Kac[5] that the set of *finite*-dimensional Kac–Moody superalgebras consists only of the basic classical simple Lie superalgebras $A(r/s)$ (for $r>s\geqslant 0$), $B(r/s)$ (for $r\geqslant 0$ and $s\geqslant 1$), $C(s)$ (for $s\geqslant 2$), $D(r/s)$ (for $r\geqslant 2$ and $s\geqslant 1$), $D(2/1;\alpha)$ (for any complex α except 0, -1, or ∞), $F(4)$, and $G(3)$, together with $\mathrm{sl}(r+1/r+1;\mathrm{C})$ (for $r\geqslant 1$), whose factor algebra with its invariant Abelian subalgebra is $A(r/r)$. The new and more significant feature was the appearance of sets of *infinite*-dimensional Lie superalgebras, of which the most interesting (and tractable) are the "affine Kac–Moody superalgebras" that form the subject of this paper.

One consequence of the explicit construction of affine Kac–Moody superalgebras that is outlined below is that it is convenient to denote such superalgebra by two quantities, namely an integer m, which may take the values 1, 2, or 4, and which appears as a superscript, and a set of symbols labeling a corresponding basic classical simple complex Lie superalgebra $\mathscr{L}_s^0$. The whole set of affine Kac–Moody superalgebras is then:

$$B^{(1)}(0/l), \quad \text{for } l=1,2,3,...;$$

$$A^{(2)}(2l-1/0), \quad \text{for } l=2,3,4,...;$$

$$C^{(2)}(l), \quad \text{for } l=2,3,4,...;$$

and

$$A^{(4)}(2l/0), \text{ for } l=1,2,3,... .$$

In the general discussion, an affine Kac–Moody superalgebra will be denoted by $\mathscr{L}_s$.

For a complex affine Kac–Moody superalgebra $\mathscr{L}_s$ the generalized Cartan matrix $\mathbf{A}$ is defined to be a $(l+1)\times(l+1)$ matrix of rank l with the index set $I=\{0,1,2,...,l\}$ labeling the rows and columns, with the diagonal entries all taking the value 2, and with off-diagonal entries being nonpositive integers such that $A_{jk}=0$ if and only if $A_{kj}=0$. [The quantity l here is the same as the the integer l appearing in the list of affine Kac–Moody superalgebras listed above, so that, for example, $A^{(2)}(2l-1/0)$ corresponds to a generalized Cartan matrix of dimension $(l+1)\times(l+1)$.] It will also be assumed that $\mathbf{A}$ is indecomposable in the sense that it does not have the block form

$$\begin{pmatrix} \mathbf{A}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{22} \end{pmatrix},$$

where $\mathbf{A}^{11}$ and $\mathbf{A}^{22}$ are nontrivial submatrices, nor can it be put in this form by any reordering of the index set I. Each column of $\mathbf{A}$ is associated with a simple root, which can be either even or odd.

 0022-2488/90/081817-22$03.00

The Cartan subalgebra $\mathscr{H}$ of $\mathscr{L}_s$ is introduced as a complex vector space of dimension $l+2$, and is assumed to be part of the even subspace $\mathscr{L}_0$ of $\mathscr{L}_s$. Let $h_j\,(j\in I)$ be defined to be any $l+1$ linearly independent elements of $\mathscr{H}$. As always, the dual space $\mathscr{H}^*$ of $\mathscr{H}$ is the set of linear functionals defined on $\mathscr{H}$ and has the same dimension $l+2$. Let $\alpha_j\,(j\in I)$ be $l+1$ linear functionals on $\mathscr{H}$ which are both linearly independent of each other and are such that

$$\alpha_j(h_k)=A_{kj},$$

for $j,k\in I$, so that α_j is the simple root associated with the jth column of **A**. The notation for these basis elements can be refined by writing $H_{\alpha_k}=h_k$ for $k\in I$, so that

$$\alpha_j(H_{\alpha_k})=A_{kj}, \tag{2.1}$$

for $j,k\in I$, thereby indicating the link that these basis elements have acquired with the linear functionals α_j. Let τ denote the set of simple odd roots.

The last stage in the abstract construction is to set up the complex Lie superalgebra $\mathscr{L}_s$ whose set of generators consists of the basis elements of $\mathscr{H}$ together with $2(l+1)$ elements E_{α_k} and $E_{-\alpha_k}$ defined for each $k\in I$, with the whole set being assumed to satisfy following relations:

$$[h,h']=0, \quad \text{for all } h,h'\in\mathscr{H}, \tag{2.2}$$

$$[E_{\alpha_j},E_{-\alpha_k}]=\delta_{jk}H_{\alpha_j}, \quad \text{for } j,k\in I, \tag{2.3}$$

$$[h,E_{\alpha_k}]=\alpha_k(h)E_{\alpha_k}, \quad \text{for all } h\in\mathscr{H} \text{ and } k\in I, \tag{2.4}$$

$$[h,E_{-\alpha_k}]=-\alpha_k(h)E_{-\alpha_k}, \quad \text{for all } h\in\mathscr{H} \text{ and } k\in I. \tag{2.5}$$

Here it is assumed that E_{α_k} and $E_{-\alpha_k}$ are both even if α_k is even and are both odd if α_k is odd. This superalgebra $\mathscr{L}_s$ is defined so that it is spanned not only these generators but also *all* the generalized Lie products of the form:

$$[E_{\alpha_k},E_{\alpha_{k'}}],[E_{\alpha_k},[E_{\alpha_{k'}},E_{\alpha_{k''}}]], \text{ and so on,}$$

together with those of the form:

$$[E_{-\alpha_k},E_{-\alpha_{k'}}],[E_{-\alpha_k},[E_{-\alpha_{k'}},E_{-\alpha_{k''}}]], \text{ etc.,}$$

subject only to the constraint that

$$(\mathrm{ad}\,E_{\alpha_j})(1-A_{jk})E_{\alpha_k}=0 \tag{2.6}$$

and

$$(\mathrm{ad}\,E_{-\alpha_j})(1-A_{jk})E_{-\alpha_k}=0, \tag{2.7}$$

for $j\neq k$ $(j,k\in I)$. The resulting complex Lie algebra $\mathscr{L}_s$ may be referred to as "the Kac–Moody superalgebra based on the generalized Cartan matrix **A** that has the set of odd simple roots τ."

It should be noted that with $h=H_{\alpha_j}$ (2.4) and (2.5) reduce to

$$[H_{\alpha_j},E_{\alpha_k}]=A_{jk}E_{\alpha_k} \tag{2.8}$$

and

$$[H_{\alpha_j},E_{-\alpha_k}]=A_{jk}E_{-\alpha_k}, \tag{2.9}$$

respectively, by virtue of (2.1).

The basic ideas and terminology of roots and root subspaces for a complex Kac–Moody superalgebra $\mathscr{L}_s$ are very similar to those for a simple complex Lie algebras and superalgebras and for Kac–Moody algebras. In particular, the commutative subalgebra $\mathscr{H}$ of $\mathscr{L}_s$ is still referred to as its Cartan subalgebra and the set of elements a_α of $\mathscr{L}_s$ that have the property that

$$[h,a_\alpha]=\alpha(h)a_\alpha, \tag{2.10}$$

for all $h\in\mathscr{H}$ is again said to form the root subspace $\mathscr{L}_{s\alpha}$ corresponding to the root α. Because of (2.4) and (2.5) the generators E_{α_k} and $E_{-\alpha_k}$ are members of $\mathscr{L}_{\alpha_k}$ and $\mathscr{L}_{-\alpha_k}$, respectively, for all $k\in I$. Also nonzero generalized Lie products of the form $[E_{\alpha_k},E_{\alpha_{k'}}]$, $[E_{\alpha_k},[E_{\alpha_{k'}},E_{\alpha_{k''}}]]$, and so on, all belong to root subspaces for which the corresponding root α has the form

$$\alpha=\sum_{k\in I}\kappa_k^\alpha\alpha_k \tag{2.11}$$

where each κ_k^α $(k\in I)$ is a non-negative integer. Here, κ_k^α is the number of times the generator E_{α_k} appears in the corresponding generalized Lie product. Naturally, such a root is called a positive root, and in particular the simple roots are positive roots. Similarly, generalized Lie products of the form:

$$[E_{-\alpha_k},E_{-\alpha_{k'}}],[E_{-\alpha_k},[E_{-\alpha_{k'}},E_{-\alpha_{k''}}]], \text{ and so on,}$$

all belong to root subspaces for which the root α is given by (2.11) but with each κ_k^α being a nonpositive integer, such a root being called a negative root. The set of all nonzero roots of $\mathscr{L}_s$ will be denoted by Δ, the set of positive roots by Δ_+, and the set of negative roots by Δ_-. If $\alpha\in\Delta_+$ then $-\alpha\in\Delta_-$ and vice versa. The direct sum of the positive and negative root subspaces will be denoted by $\mathscr{L}_{s^+}$ and $\mathscr{L}_{s^-}$, respectively, so

$$\mathscr{L}_s=\mathscr{L}_{s^+}\oplus\mathscr{H}\oplus\mathscr{L}_{s^-}, \tag{2.12}$$

where the $\oplus$ symbol indicates only a vector space direct sum, and does not imply that separate parts mutually commute.

The Kac–Moody superalgebra $\mathscr{L}_s$ possesses a nondegenerate bilinear supersymmetric form $B(\ ,\ )$, which is unique up to a constant multiplicative factor, and which is also nondegenerate on $\mathscr{H}$, so that for every linear functional α defined on $\mathscr{H}$ there exists a unique element h_α of $\mathscr{H}$ that is defined by

$$B(h_\alpha,h)=\alpha(h), \quad \text{for all } h\in\mathscr{H}. \tag{2.13}$$

Then $\langle\alpha,\beta\rangle$ may be defined by

$$\langle\alpha,\beta\rangle=B(h_\alpha,h_\beta) \tag{2.14}$$

for any pair of linear functionals α and β defined on $\mathscr{H}$. It then follows that the elements of the generalized Cartan matrix are given by the usual expressions:

$$A_{jk}=2\langle\alpha_j,\alpha_k\rangle/\langle\alpha_j,\alpha_j\rangle, \quad \text{for } j,k\in I. \tag{2.15}$$

The fundamental Weyl reflections S_α may be defined in the usual way by

$$(S_\alpha\beta)(h)=\beta(h)-\{2\langle\beta,\alpha\rangle/\langle\alpha,\alpha\rangle\}\alpha(h), \tag{2.16}$$

the set of fundamental Weyl reflections $S\alpha_k$ generating the Weyl group $\mathscr{W}$ of the Kac–Moody superalgebra $\mathscr{L}_s$. If $\alpha\in\Delta$ and if there exists an element $S\in\mathscr{W}$ such that $\alpha=S\alpha_k$ for some simple root α_k then α is said to be a "real" root of $\mathscr{L}_s$, the other roots of $\mathscr{L}_s$ being described as being "imaginary."

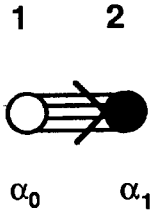


FIG. 1. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $B^{(1)}(0/1)$.

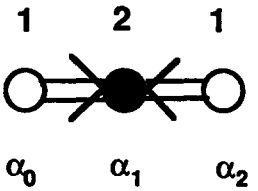


FIG. 3. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $A^{(2)}(3/0)$.

The root $\alpha \in \Delta$ is real if and only if $\langle \alpha,\alpha \rangle > 0$.

Every generalized Cartan matrix can be associated with a generalized Dynkin diagram by first assigning to each simple root α_k ($k \in I$) a vertex, which is drawn as a small "white" circle (○) if α_k is even and as a small "black" circle (●) if α_k is odd, then drawing L_{jk} lines from the α_j vertex to the α_k vertex, where $L_{jk} = \max\{|A_{jk}|,|A_{jk}|\}$, and adding an arrow pointing from the α_j vertex to the α_k vertex if $|A_{kj}| > 1$. Finally, a set of numerical marks $\{N_0,N_1,...,N_l\}$ may be attached to the vertices. These are defined by the relations

$$\sum_{j=0}^{l} N_j A_{kj} = 0, \quad \text{for } k = 0,1,...,l$$

and the requirement that the lowest value of the set $\{N_0,N_1,...,N_l\}$ has value 1. The set of generalized Dynkin diagrams corresponding to the affine Kac–Moody superalgebras is exhibited in Figs. 1–8.

B. Explicit construction of affine Kac–Moody superalgebras

The above procedure for setting up an affine Kac–Moody superalgebra was rather abstract but fortunately Kac[4] has shown that there exists a more *explicit* construction that is related to the above notation. Suppose that the simple Lie superalgebra $\mathcal{L}_s^0$ has Killing form $B^0(\,,\,)$ (which for every simple Lie superalgebra mentioned above is nondegenerate), that $\mathcal{L}_s^0$ has rank l^0, that $\mathcal{H}^0$ is its Cartan subalgebra, that α_k^0, for $k = 1,2,...,l^0$, are its distinguished simple roots, and that Δ^0, Δ_+^0, and Δ_-^0 are its nonzero, positive, and negative root systems, respectively. Suppose also that $h_{\alpha^0}^0$ is the element of $\mathcal{H}^0$ that is defined by analogy with (2.13) for each linear functional α^0 on $\mathcal{H}^0$ by

$$B^0(h_{\alpha^0}^0,h^0) = \alpha^0(h^0), \quad \text{for all } h^0 \in \mathcal{H}^0. \tag{2.17}$$

Then $\langle \alpha^0,\beta^0 \rangle^0$ may be defined by

$$\langle \alpha^0,\beta^0 \rangle^0 = B^0(h_{\alpha^0}^0,h_{\beta^0}^0), \tag{2.18}$$

for any pair of linear functionals α^0 and β^0 on $\mathcal{H}^0$ [cf. (2.14)]. A Weyl-type canonical basis may be chosen for $\mathcal{L}_s^0$. It consists of $h_{\alpha_k^0}^0$, for $k = 1,2,...,l^0$, together with $e_{\alpha^0}^0$, for all $\alpha^0 \in \Delta^0$, and these are assumed to satisfy the usual commutation and anticommutation relations. In a realization of $\mathcal{L}_s^0$ in which the elements of $\mathcal{L}_s^0$ are represented by supermatrices, with $e_{\alpha^0}^0$ being represented by $\mathbf{e}_{\alpha^0}^0$, the convention will be adopted that

$$\mathbf{e}_{-\alpha^0}^0 = -(\mathbf{e}_{\alpha^0}^0)^{\mathrm{st}}, \tag{2.19}$$

for all α^0 of Δ_+^0, the superscripts st indicating that the supertranspose must be taken.

The first stage in the construction is to consider the loop superalgebra that consists of all complex linear combinations of the products $t^j \otimes a_r^0$, where j takes any integer value, a_r^0 are the basis elements of $\mathcal{L}_s^0$, and t is a real number. The generalized Lie product of this loop algebra may be defined by

$$[t^j \otimes a^0, t^k \otimes b^0] = t^{j+k} \otimes [a^0,b^0], \tag{2.20}$$

for all integers j and k and all $a^0,b^0 \in \mathcal{L}_s^0$, where the generalized Lie product of the right-hand side of (2.20) is that of $\mathcal{L}_s^0$. Here, it is assumed that $t^j \otimes a_r^0$ is even if a_r^0 is even and that $t^j \otimes a_r^0$ is odd if a_r^0 is odd. This superalgebra may be extended by introducing an additional even element c (so that it then consists of all complex linear combinations of the $t^j \otimes a_r^0$ and of c), with the generalized Lie product (2.20) being modified to become

$$[t^j \otimes a^0, t^k \otimes b^0] = t^{j+k} \otimes [a^0,b^0] + j\delta^{j+k,0} B^0(a^0,b^0)c, \tag{2.21}$$

for all integers j and k and all $a^0,b^0 \in \mathcal{L}_s^0$, and where it is assumed that

$$[t^j \otimes a^0, c] = 0, \tag{2.22}$$

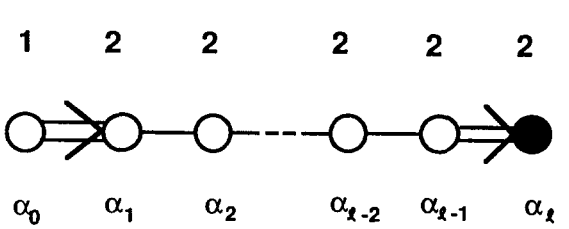


FIG. 2. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $B^{(1)}(0/l)$ (for $l > 2$).

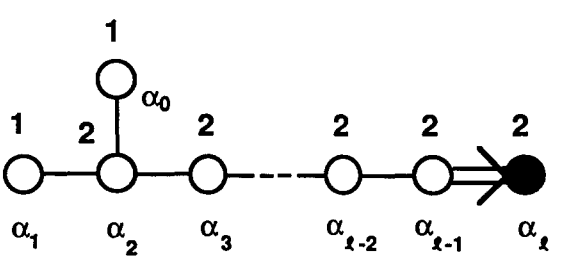


FIG. 4. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $A^{(2)}(2l-1/0)$ (for $l > 3$).

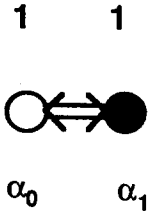


FIG. 5. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $A^{(4)}(2/0)$.

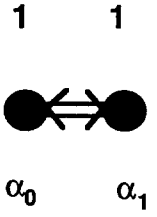


FIG. 7. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $C^{(2)}(2)$.

for all integers j and all $a^0 \in \mathscr{L}_s^0$. Again the generalized Lie product on the right-hand side of (2.21) is that of $\mathscr{L}_s^0$. This Lie superalgebra may be enlarged by adding a further even element d, for which it is assumed that

$$[d, t^j \otimes a^0] = j t^j \otimes a^0, \tag{2.23}$$

for all integers j and all $a^0 \in \mathscr{L}_s^0$, and that

$$[d, c] = 0. \tag{2.24}$$

The resulting Lie superalgebra will be denoted by $\mathscr{L}_s^{(1)}$. Its composition may be summarized by the statement:

$$\mathscr{L}_s^{(1)} = (\mathbb{C}c) \oplus (\mathbb{C}d) \oplus \sum_{j=-\infty}^{\infty} (t^j \otimes \mathscr{L}_s^0). \tag{2.25}$$

Equation (2.21) shows that the set of elements $t^0 \otimes a^0$, where $a^0 \in \mathscr{L}_s^0$, form a subalgebra of $\mathscr{L}_s^{(1)}$ that is isomorphic to $\mathscr{L}_s^0$.

A bilinear supersymmetric invariant form $B^{(1)}(\ ,\ )$ can be set up on $\mathscr{L}_s^{(1)}$ by the definitions

$$B^{(1)}(t^j \otimes a^0, t^k \otimes b^0) = \delta^{j+k,0} B^0(a^0, b^0), \tag{2.26}$$

$$B^{(1)}(t^j \otimes a^0, c) = 0, \tag{2.27}$$

$$B^{(1)}(t^j \otimes a^0, d) = 0, \tag{2.28}$$

$$B^{(1)}(c, c) = 0, \tag{2.29}$$

$$B^{(1)}(c, d) = 1, \tag{2.30}$$

$$B^{(1)}(d, d) = 0, \tag{2.31}$$

for all $a^0, b^0 \in \mathscr{L}^0$ and all integers j and k. Clearly, $B^{(1)}(\ ,\ )$ coincides with $B^0(\ ,\ )$ on the subalgebra of $\mathscr{L}_s^{(1)}$ that is isomorphic to $\mathscr{L}_s^0$.

Now consider the generalized Dynkin diagram of the Kac–Moody superalgebra $\mathscr{L}_s$ and choose the node corresponding to *any* simple root α_k of $\mathscr{L}_s$. Suppose that the corresponding numerical mark is N_k. Let q be the integer defined by

$$q = mN_k, \tag{2.32}$$

and let ϕ be an automorphism of $\mathscr{L}_s^0$ of order q. Inspection of Figs. 1–8 shows that the only possible values of q are 1, 2, and

4. Let $\mathscr{L}_{sp}^{0(q)}$ be the subspace of $\mathscr{L}_s^0$ that consists of all the elements a^0 of $\mathscr{L}_s^0$ that are such that

$$\phi(a^0) = e^{2\pi p i/q} a^0, \tag{2.33}$$

where $p = 0, 1, \ldots, q-1$. It follows that $\mathscr{L}_{s0}^{0(q)}$ is a Lie superalgebra, and that for each p taking the value 1,2,..., or $q-1$ the subspace $\mathscr{L}_{sp}^{0(q)}$ provides a carrier space for a representation Γ^p of $\mathscr{L}_{s0}^{0(q)}$ by the prescription

$$[a_{0r}^0, a_{pr'}^0] = \sum_{r''=1}^{n_p} \Gamma^p(a_{0r}^0)_{r''r'} a_{pr''}^0, \tag{2.34}$$

for all a_{0r}^0 of $\mathscr{L}_{s0}^{0(q)}$, where n_p is the dimension of $\mathscr{L}_{sp}^{0(q)}$ and a_{pr}^0 (for $r = 1,2,\ldots,n_p$) are the basis elements of $\mathscr{L}_{sp}^{0(q)}$. Also

$$\mathscr{L}_s^0 = \oplus \sum_{p=0}^{q-1} \mathscr{L}_{sp}^{0(q)}.$$

The key result established by Kac[4] is that $\mathscr{L}_s$ may be taken to be the subalgebra of $\mathscr{L}_s^{(1)}$ whose set of basis elements consists of c, d, and, for $p = 0,1,\ldots,$ and $q-1$, of all $t^j \otimes a^0$ for every integer j that is such that $j \bmod q = p$ and every basis element $a_{pr}^0 \in \mathscr{L}_{sp}^{0(q)}$. This may be summarized by the statement that

$$\mathscr{L}_s = (\mathbb{C}c) \oplus (\mathbb{C}d) \oplus \sum_{p=0}^{q-1} \sum_{j=-\infty,\, j \bmod q = p}^{\infty} (t^j \otimes \mathscr{L}_{sp}^{0(q)}). \tag{2.35}$$

The generalized Lie products of $\mathscr{L}_s$ are then those inherited from $\mathscr{L}_s^{(1)}$ and so are given by (2.21), (2.22), (2.23), and (2.24). Moreover the supersymmetric bilinear invariant form $B(\ ,\ )$ may be taken to be such that

$$B(a,b) = \mu B^{(1)}(a,b), \tag{2.36}$$

for all $a, b \in \mathscr{L}_s$, μ being an arbitrary constant that may be chosen in any way. As the subset of elements of $\mathscr{L}_s$ of the form $t^0 \otimes a^0$ (for all the elements a^0 of $\mathscr{L}_s^0$) form a subalgebra that is isomorphic to $\mathscr{L}_s^0$, a particularly convenient choice to let μ be such that $B(\ ,\ )$ coincides with the Killing form $B_{s0}^{0(q)}(\ ,\ )$ of $\mathscr{L}_{s0}^{0(q)}$, that is, so that

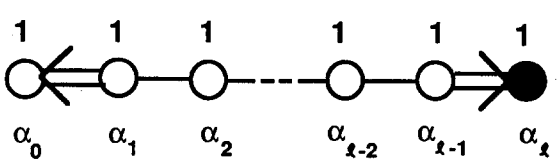


FIG. 6. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $A^{(4)}(2l/0)$ (for $l \geq 2$).

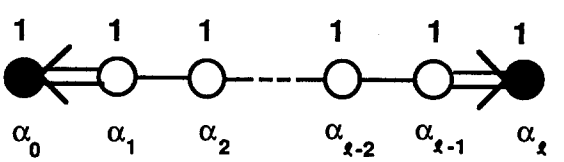


FIG. 8. Generalized Dynkin diagram of the affine Kac–Moody superalgebra $C^{(2)}(l+1)$ (for $l \geq 2$).

$$B(t^0\otimes a^0,t^0\otimes b^0)=B^{0(q)}_{s0}(a^0,b^0)\text{for all } a^0,b^0 \text{ of } \mathscr{L}^{0(q)}_{s0}. \tag{2.37}$$

Kac[4] has also shown that if the node corresponding to the simple root α_k of $\mathscr{L}_s$ associated with (2.32) is removed, together with all the lines attached to it, then the resulting generalized Dynkin diagram corresponds to the Lie superalgebra $\mathscr{L}^{0(q)}_{s0}$, which is either simple or is the direct sum of simple Lie superalgebras. Clearly, if $\mathscr{L}_s$ has only one odd simple root and if the chosen node corresponds to this odd simple root, then $\mathscr{L}^{0(q)}_{s0}$ will contain no nontrivial odd part, and so, in this case, $\mathscr{L}^{0(q)}_{s0}$ will be a semisimple Lie algebra.

III. ROOTS OF THE UNTWISTED AFFINE KAC–MOODY SUPERALGEBRAS

The simplest situation is that in which $m=1$ and the numerical mark N_k corresponding to the chosen simple root also has value 1, implying [by (2.32)] that $q=1$. Then the automorphism ϕ is the identity mapping of $\mathscr{L}^0_s$ onto itself, there is only one subspace $\mathscr{L}^{0(1)}_{s0}$, and this coincides with $\mathscr{L}^0_s$, and the resulting affine Kac–Moody superalgebra $\mathscr{L}_s$ is the superalgebra $\mathscr{L}^{(1)}_s$ introduced above. Such a superalgebra may be described as being "untwisted." The only affine Kac–Moody superalgebras for which $m=1$ are the $B^{(1)}(0/l)$ (for $l=1,2,\ldots$), and as $N_0=1$ (with the labeling of simple roots in Figs. 1 and 2), they all have this form.

With this choice, the structure of these untwisted affine Kac–Moody superalgebras $B^{(1)}(0/l)$ is easily determined in complete detail. Clearly, for $B^{(1)}(0/l)$ the Lie superalgebra $\mathscr{L}^0_s(=\mathscr{L}^{0(1)}_{s0})$ is $B(0/l)$, which, with the labeling conventions of the generalized Dynkin diagram of Fig. 9, has a distinguished set of simple roots consisting of $l-1$ even simple roots $\alpha^0_1,\alpha^0_2,\ldots,\alpha^0_{l-1}$ and one odd simple root α^0_l. Then, with

$$l=l^0, \tag{3.1}$$

(2.21) to (2.24) together indicate that a linearly independent and mutually commuting set of $l+2$ even elements is formed by c, d and $t^0\otimes h^0_{\alpha^0_k}$ (for $k=1,2,\ldots,l$).

Let $\mathscr{H}^{(1)}$ be the (l^0+2)-dimensional complex vector space that has these elements as its basis, so that

$$\mathscr{H}^{(1)}=(\mathbb{C}c)\oplus(\mathbb{C}d)\oplus\sum_{k=1}^{l^0}(t^0\otimes h^0_{\alpha^0_k}). \tag{3.2}$$

Every linear functional α^0 that is defined on $\mathscr{H}^0$ can be extended to become a linear functional on $\mathscr{H}^{(1)}$ by the definitions:

$$\alpha^0(t^0\otimes h^0_{\alpha^0_k})=\alpha^0(h^0_{\alpha^0_k}), \tag{3.3}$$

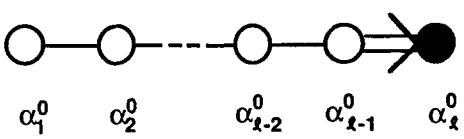


FIG. 9. Generalized Dynkin diagram of the basic simple superalgebra $B(0/l)$ (for $l\geq 1$).

$$\alpha^0(c)=0, \tag{3.4}$$

$$\alpha^0(d)=0. \tag{3.5}$$

No confusion will be caused by denoting (as here) both α^0 and its extension by the same symbol.

Let δ be the linear functional on $\mathscr{H}^{(1)}$ defined by

$$\delta(t^0\otimes h^0_{\alpha^0_k})=0,\quad \text{for } k=1,2,\ldots,l^0, \tag{3.6}$$

$$\delta(c)=0, \tag{3.7}$$

$$\delta(d)=1. \tag{3.8}$$

Then, as (2.21), (2.22), and (2.23) give,

$$[t^0\otimes h^0_{\alpha^0_k},t^j\otimes e^0_{\alpha^0}]=\alpha(t^0\otimes h^0_{\alpha^0_k})(t^j\otimes e^0_{\alpha^0}), \tag{3.9}$$

$$[c,t^j\otimes e^0_{\alpha^0}]=0, \tag{3.10}$$

$$[d,t^j\otimes e^0_{\alpha^0}]=j(t^j\otimes e^0_{\alpha^0}), \tag{3.11}$$

for any $\alpha^0\in\Delta^0$ and for any integer j (and for $k=1,\ldots,l^0$), it follows from (3.6)–(3.11) that

$$[h,t^j\otimes e^0_{\alpha^0}]=\{j\delta(h)+\alpha^0(h)\}(t^j\otimes e^0_{\alpha^0}), \tag{3.12}$$

for *all* $h\in\mathscr{H}^{(1)}$. Thus $t^j\otimes e^0_{\alpha^0}$ corresponds to a root $j\delta+\alpha^0$ of $\mathscr{L}_s$. Similarly, for any $\beta^0\in\Delta^0$ and any nonzero integer j

$$[t^0\otimes h^0_{\alpha^0_k},t^j\otimes h^0_{\beta^0}]=0, \tag{3.13}$$

$$[c,t^j\otimes h^0_{\beta^0}]=0, \tag{3.14}$$

$$[d,t^j\otimes h^0_{\beta^0}]=j(t^j\otimes h^0_{\beta^0}), \tag{3.15}$$

so, by (3.6) to (3.8),

$$[h,t^j\otimes h^0_{\beta^0}]=j\delta(h)(t^j\otimes h^0_{\beta^0}) \tag{3.16}$$

for all $h\in\mathscr{H}^{(1)}$. Thus $t^j\otimes h^0_{\beta^0}$ corresponds to a root $j\delta$ of $\mathscr{L}_s$. Moreover, there are l linearly independent elements with this property, namely, $h^0_{\alpha^0_k}$ (for $k=1,2,\ldots,l$), and, as there are no further elements of $\mathscr{L}_s$ to consider, the root subspace of $j\delta$ must have dimension l (for $j\neq 0$).

Taking $\mu=1$ in (2.36), it follows from (2.13), (2.17), and (2.26) to (2.28) that

$$h_{\alpha^0}=t^0\otimes h^0_{\alpha^0}, \tag{3.17}$$

for each $\alpha^0\in\Delta^0$ (and its extension), and hence by (2.14), (2.18), (2.26), and (3.3) to (3.5) that

$$\langle\alpha^0,\beta^0\rangle=\langle\alpha^0,\beta^0\rangle^0, \tag{3.18}$$

for every pair $\alpha^0,\beta^0\in\Delta^0$ (and their extensions). Also (2.13) and (2.27), (2.29), (2.30), and (3.6) to (3.8) imply that

$$h_\delta=c. \tag{3.19}$$

Thus if α^0_k is the extension of any simple root of $\mathscr{L}^0_s$ then

$$\langle\delta,\alpha^0_k\rangle=0 \tag{3.20}$$

and

$$\langle j\delta,j\delta\rangle=0. \tag{3.21}$$

Thus $\langle j\delta,j\delta\rangle=0$ for every integer j, so every nonzero root of the form $j\delta$ is "imaginary." Moreover, as $\langle j\delta+\alpha^0,j\delta+\alpha^0\rangle=\langle\alpha^0,\alpha^0\rangle^0$ and as $\mathscr{L}^0_s=B(0/l)$ has the exceptional property that $\langle\alpha^0,\alpha^0\rangle^0>0$ for *every* root α^0 (cf. Cornwell[9]), it follows that every root of the form $j\delta+\alpha^0$ is "real."

Every root can be put in the form (2.11) if the simple

roots $\alpha_0,\alpha_1,...,\alpha_l$ of the Kac–Moody superalgebra $B^{(1)}(0/l)$ are taken to be

$$\alpha_0 = \delta - \alpha_H^0, \tag{3.22}$$

where

$$\alpha_H^0 = 2\sum_{r=1}^{l} \alpha_r^0 \tag{3.23}$$

is the highest root of $\tilde{\mathscr{L}}_s^0$, and

$$\alpha_k = \alpha_k^0 \text{ for } k = 1,2,...,l, \tag{3.24}$$

[where the α_k^0 of (3.23) and (3.24) are the extensions of the simple roots of $\tilde{\mathscr{L}}_s^0$]. Then the simple roots $\alpha_0,\alpha_1,...,\alpha_{l-1}$ are all even and α_l is odd. It is easily checked that the Cartan matrices of $B^{(1)}(0/1)$ and $B^{(1)}(0/l)$ (for $l\geqslant 2$) evaluated using (2.15) correspond to the generalized Dynkin diagrams given in Figs. 1 and 2, respectively.

IV. TWISTED AFFINE KAC–MOODY SUPERALGEBRAS CORRESPONDING TO AUTOMORPHISMS OF ORDER 4: CHOICE OF THE AUTOMORPHISM

Turning to the opposite extreme case in which the q of (2.32) has the value 4, the obvious choice of the automorphism ϕ of the simple Lie superalgebra $\tilde{\mathscr{L}}_s^0$ is the "canonical" fourfold automorphism ψ that is defined by

$$\psi(h^0) = -h^0 \text{(for all } h^0 \text{ of } \mathscr{H}^0), \tag{4.1}$$

$$\psi(e^0_{\alpha^0}) = e^0_{-\alpha^0}$$

$$[\text{if } \alpha^0(\in\Delta^0) \text{ is even or is odd and negative}], \tag{4.2}$$

and

$$\psi(e^0_{\alpha^0}) = -e^0_{-\alpha^0} [\text{if } \alpha^0(\in\Delta^0) \text{ is odd and positive}] \tag{4.3}$$

(cf. Scheunert[10]). With this choice (2.33) implies that (i) the basis elements of $\tilde{\mathscr{L}}_{s0}^{0(q)}$ may be taken to be $e^0_{\alpha^0} + e^0_{-\alpha^0}$, for all even positive roots $\alpha^0\in\Delta^0$; (ii) the basis elements of $\tilde{\mathscr{L}}_{s1}^{0(q)}$ may be taken to be $e^0_{\alpha^0} + ie^0_{-\alpha^0}$, for all odd positive roots $\alpha^0\in\Delta^0$; (iii) the basis elements of $\tilde{\mathscr{L}}_{s2}^{0(q)}$ may be taken to be $e^0_{\alpha^0} - e^0_{-\alpha^0}$, for all even positive roots $\alpha^0\in\Delta^0$ and $h^0_{\alpha_k^0}$, for $k = 1,2,...,l^0$; (iv) the basis elements of $\tilde{\mathscr{L}}_{s3}^{0(q)}$ may be taken to be $e^0_{\alpha^0} - ie^0_{-\alpha^0}$, for all odd positive roots $\alpha^0\in\Delta^0$.

This automorphism has been used by Golitzin[11] to find the simple roots and generators of $A^{(2)}(2l-1/0)$ and $A^{(4)}(2l/0)$.

Although these basis elements are very straightforward, the difficulties start arising with this choice of automorphism when one tries to determine explicitly the complete root structure of the Kac–Moody superalgebra $\tilde{\mathscr{L}}_s$. The problem is that if the Cartan subalgebra of $\tilde{\mathscr{L}}_s$ is chosen to be in $\tilde{\mathscr{L}}_{s0}^{0(q)}$ (as in the case $q = 1$), it *cannot* consist of c, d, and elements of the form $t^0\otimes h^0_{\alpha_k^0}$ (for $k = 1,2,...,l^0$), for the elements $h^0_{\alpha_k^0}$ are not members of $\tilde{\mathscr{L}}_{s0}^{0(q)}$. Instead, the simplest choice is c, d, and certain linear combinations of $t^0\otimes(e^0_{\alpha^0} + e^0_{-\alpha^0})$ (for the even positive roots $\alpha^0\in\Delta^0$). To find the roots it is then necessary to evaluate the generalized Lie products of these with all the elements of the sets (i) to (iv) above, taking appropriate linear combinations of the latter in order satisfy the root equation (2.10). Not only is this messy, it also makes no direct use of the known root structure of the simple Lie superalgebra $\tilde{\mathscr{L}}_s^0$. Indeed the situation here is very similar to the one that occurs in the standard method of determination of the Iwasawa and Langlands decompositions of the simple Lie algebras, and the resolution of the problem is based on essentially the same idea as that of the "direct" determination of these decompositions that was given by Cornwell.[12,13]

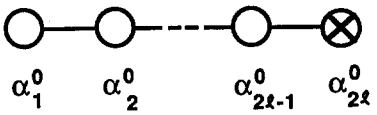


FIG. 10. Generalized Dynkin diagram of the basic simple superalgebra $A(2l-1/0)$ (for $l\geqslant 1$).

Incidentally, it is clear that the canonical fourfold automorphism ψ of $\tilde{\mathscr{L}}_s^0$ defined in (4.1) is not associated with any rotation of the usual generalized Dynkin diagram of $\tilde{\mathscr{L}}_s^0$ based on the distinguished simple roots, because for $A(2l-1/0)$, $A(2l/0)$, and $C(l+1)$ the generalized Dynkin diagrams exhibited in Figs. 10–12 possess no symmetries. Nevertheless it is convenient to describe the corresponding Kac–Moody superalgebras as being "twisted." In this connection it may be noted that Frappat *et al.*[14] have shown that it is sometimes possible by using *nondistinguished* sets of simple roots to construct generalized Dynkin diagrams for the basic simple Lie superalgebras that possess rotational symmetries that do correspond to outer automorphisms of these superalgebras. However, this is not possible in every case that is relevant here, the simplest example where it cannot be done is $A(2/0)$.

The most general fourfold automorphism ϕ of $\tilde{\mathscr{L}}_s^0$ has the form

$$\phi = \theta^{-1}\psi\theta, \tag{4.4}$$

where ψ is the canonical fourfold automorphism of $\tilde{\mathscr{L}}_s^0$ defined in (4.1) to (4.3) and θ is any automorphism of $\tilde{\mathscr{L}}_s^0$. If θ can be chosen so that enough elements of the form $t^0\otimes h^0_{\alpha_k^0}$ lie in $\tilde{\mathscr{L}}_{s0}^{0(q)}$ then the roots of the Kac–Moody superalgebra will be very easy to obtain. In investigating this condition it is useful to note that if the simple Lie superalgebra $\tilde{\mathscr{L}}_s^0$ is expressed in terms of supermatrices with the graded partitioning

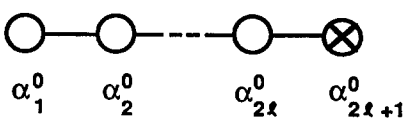


FIG. 11. Generalized Dynkin diagram of the basic simple superalgebra $A(2l/0)$ (for $l\geqslant 1$).

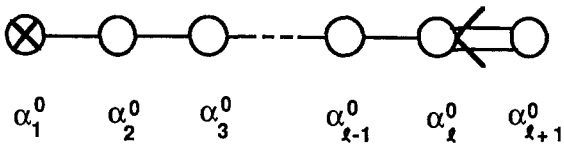


FIG. 12. Generalized Dynkin diagram of the basic simple superalgebra $C(l+1)$ (for $l \geqslant 1$).

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{4.5}$$

then

$$\psi(\mathbf{M}) = -\mathbf{M}^{\mathrm{st}} = \begin{pmatrix} -\tilde{\mathbf{A}} & \tilde{\mathbf{C}} \\ -\tilde{\mathbf{B}} & -\tilde{\mathbf{D}} \end{pmatrix}, \tag{4.6}$$

where $\tilde{\mathbf{A}}$ denotes the ordinary transpose of $\mathbf{A}$.

The choice of θ will first be investigated for the Kac–Moody superalgebras of the form $A^{(2)}(2l-1/0)$ (for $l = 2,3,4...$).

V. TWISTED AFFINE KAC–MOODY SUPERALGEBRAS CORRESPONDING TO AUTOMORPHISMS OF ORDER 4: ROOTS OF $A^{(2)}(2l-1/0)$ (FOR $l=2,3,4,...$)

A. The fourfold automorphisms

An explicit realization of the simple Lie superalgebra $A(2l-1/0)$ is provided by $\mathrm{sl}(2l/1)$, considered as a complex superalgebra, $\mathrm{sl}(2l/1)$ being defined as the set of $(2l+1)\times(2l+1)$ complex supermatrices with the grading partitioning

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{5.1}$$

that are subject to the supertrace condition that

$$\mathrm{str}\,\mathbf{M} = \mathbf{0}. \tag{5.2}$$

(Here, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are of dimensions $2l\times 2l$, $2l\times 1$, $1\times 2l$, and 1×1, respectively.) The rank l^0 of $A(2l-1/0)$ is given by

$$l^0 = 2l.$$

The generalized Dynkin diagram of $A(2l-1/0)$ is shown in Fig. 10, which indicates that its distinguished simple roots α_k^0 are even for $k = 1,2,...,2l-1$, but that α_{2l}^0 is odd. [The rules for interpreting the generalized Dynkin diagrams for basic simple Lie superalgebras are as given for the affine Kac–Moody superalgebras in Sec. II A, except that now two types of distinguished odd simple roots may be distinguished, α_k^0 being associated with a small closed circle (●) if $\langle\alpha_k^0,\alpha_k^0\rangle^0 \neq 0$ and with a small grey circle ⊗ if $\langle\alpha_k^0,\alpha_k^0\rangle^0 = 0$.] With the bilinear form $B^0(\ ,\ )$ being defined by

$$B^0(\mathbf{M},\mathbf{N}) = 2(2l-1)\ \mathrm{str}\ (\mathbf{MN}), \tag{5.3}$$

the basis elements of its Cartan subalgebra $\mathscr{H}^0$ may be taken to be

$$\mathbf{h}^0_{\alpha_k^0} = \{\tfrac{1}{2}(2l-1)\}\{\mathbf{e}_{k,k} - \mathbf{e}_{k+1,k+1}\} \quad (\text{for } k = 1,2,...,2l-1) \tag{5.4a}$$

and

$$\mathbf{h}^0_{\alpha_{2l}^0} = \{\tfrac{1}{2}(2l-1)\}\{\mathbf{e}_{2l,2l} + \mathbf{e}_{2l+1,2l+1}\}. \tag{5.4b}$$

Here, $\mathbf{e}_{r,s}$ is the matrix of dimension $(2l+1)\times(2l+1)$ that is defined by

$$(\mathbf{e}_{r,s})_{jk} = \delta_{rj}\delta_{sk} \quad (\text{for } j,k = 1,2,...,2l+1), \tag{5.5}$$

so that with this choice all the matrices of $\mathscr{H}^0$ are *diagonal*. The positive even roots $\beta^0_{(j,k)}$ and positive odd roots $\delta^0_{(j)}$ of $A(2l-1/0)$ are given in terms of the distinguished set of simple roots $\alpha_1^0,\alpha_2^0,...,\alpha_{2l}^0$ of $A(2l-1/0)$ by

$$\beta^0_{(j,k)} = \sum_{r=j}^{k-1} \alpha_r^0 \quad (\text{for } j,k = 1,2,...,2l, \text{ with } j<k), \tag{5.6a}$$

and

$$\delta^0_{(j)} = \sum_{r=j}^{2l} \alpha_r^0 \quad (\text{for } j = 1,2,...,2l), \tag{5.6b}$$

for which the corresponding basis elements of $A(2l-1/0)$ may be taken to be

$$\mathbf{e}^0_{\beta^0_{(j,k)}} = \mathbf{e}^0_{\alpha_j^0+\alpha_{j+1}^0+\cdots+\alpha_{k-1}^0} = \mathbf{e}_{j,k} \quad (\text{for } j,k = 1,2,...,2l; j<k) \tag{5.7a}$$

and

$$\mathbf{e}^0_{\delta^0_{(j)}} = \mathbf{e}^0_{\alpha_j^0+\alpha_{j+1}^0+\cdots+\alpha_{2l}^0} = \mathbf{e}_{j,2l+1} \quad (\text{for } j = 1,2,...,2l). \tag{5.7b}$$

The basis elements belonging to the corresponding negative roots may be chosen in accordance with (2.19). [For further information on $A(2l-1/0)$ see Cornwell.[9]]

Taking the node corresponding to the odd simple root α_l of $A^{(2)}(2l-1/0)$ for $l \geqslant 3$, and to the odd simple root α_1 of $A^{(2)}(2l-1/0)(=A^{(2)}(3/0))$ for $l=2$, as the corresponding numerical mark has value 2 (cf. Figs. 3 and 4) (2.32) shows that $q=4$. It follows from (2.33) that if the automorphism (4.6) is employed then the subalgebra $\mathscr{L}_{s0}^{0(4)}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\tilde{\mathbf{A}} = \mathbf{A},\ -\tilde{\mathbf{B}} = \mathbf{C},\ \tilde{\mathbf{C}} = \mathbf{B},\ \text{and}\ -\tilde{\mathbf{D}} = \mathbf{D},$$

which when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$-\tilde{\mathbf{A}} = \mathbf{A},\ \mathbf{B} = \mathbf{0},\ \mathbf{C} = \mathbf{0},\ \text{and}\ \mathbf{D} = \mathbf{0}. \tag{5.8}$$

Thus subalgebra $\mathscr{L}_{s0}^{0(q)}$ is isomorphic to the set of $2l\times 2l$ complex antisymmetric matrices and hence is isomorphic to the simple complex Lie algebra D_l, which is simple if $l>2$ but is only semisimple if $l=2$, for then $D_2 = A_1 \oplus A_1$. (Inspection of Figs. 3 and 4 shows that the generalized Dynkin diagram with the chosen node and attached lines removed does indeed correspond to D_l.) As expected from the comments at the end of the previous section, none of the basis elements of the Cartan subalgebra $\mathscr{H}^0$ of $A(2l-1/0)$ are members of this $\mathscr{L}_{s0}^{0(q)}$ (because all the members of this $\mathscr{L}_{s0}^{0(q)}$ are *non*diagonal matrices).

A realization of D_l in which the basis elements of the Cartan subalgebra of D_l are given by diagonal matrices is given by the $2l \times 2l$ complex matrices $\mathbf{A}'$ that satisfy the condition

$$\tilde{\mathbf{A}}'\mathbf{G} + \mathbf{G}\mathbf{A}' = \mathbf{0}, \tag{5.9}$$

where

$$\mathbf{G} = \begin{pmatrix} \mathbf{0} & \mathbf{1}_l \\ \mathbf{1}_l & \mathbf{0} \end{pmatrix}. \tag{5.10}$$

This realization will be referred to as the "canonical" form of D_l. These matrices $\mathbf{A}'$ are related to the $2l \times 2l$ antisymmetric matrices $\mathbf{A}$ by

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{A}', \tag{5.11}$$

where $\mathbf{T}$ is a certain $2l \times 2l$ complex matrix that satisfies the condition

$$\tilde{\mathbf{T}}\mathbf{T} = \mathbf{G} \tag{5.12}$$

(cf. Cornwell[15]). This mapping can be extended to an automorphism of θ of $\mathscr{L}_s^0 (= sl(2l/1))$ by the definition

$$\theta(\mathbf{M}) = \begin{pmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_1 \end{pmatrix} \mathbf{M} \begin{pmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_1 \end{pmatrix}, \tag{5.13}$$

for all $\mathbf{M}$ of $\mathrm{sl}(2l/1)$. Then, by (4.4), (5.12), and (5.13),

$$\phi\left(\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}\right) = \begin{pmatrix} -\mathbf{G}\tilde{\mathbf{A}}\mathbf{G} & \mathbf{G}\tilde{\mathbf{C}} \\ -\tilde{\mathbf{B}}\mathbf{G} & -\tilde{\mathbf{D}} \end{pmatrix}. \tag{5.14}$$

B. The subspaces $\mathscr{L}_{sp}^{0(4)}$ (for $p=0,1,2,3$)

The four subspaces $\mathscr{L}_{sp}^{0(4)}$ (for $p = 0,1,2,3$) corresponding to the automorphism ϕ of (5.14) will now be considered in turn.

1. $\mathscr{L}_{s0}^{0(4)}$

By (2.33) the subalgebra $\mathscr{L}_{s0}^{0(4)}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G} = \mathbf{A}, \quad -\tilde{\mathbf{B}}\mathbf{G} = \mathbf{C}, \quad \mathbf{G}\tilde{\mathbf{C}} = \mathbf{B}, \quad \text{and} \quad -\tilde{\mathbf{D}} = \mathbf{D},$$

which when taken together, along with the fact that D is 1×1, imply that

$$\tilde{\mathbf{A}}\mathbf{G} + \mathbf{G}\mathbf{A} = \mathbf{0}, \quad \mathbf{B} = \mathbf{0}, \quad \mathbf{C} = \mathbf{0}, \quad \text{and} \quad \mathbf{D} = \mathbf{0}, \tag{5.15}$$

and so is isomorphic to the canonical form of D_l.

Before proceeding it will be useful to recall some properties of the canonical form of D_l (cf. Konuma *et al.*[16] and Cornwell[15]). Its Killing form $B^{D_l}(\ ,\ )$ is given by

$$B^{D_l}(\mathbf{A},\mathbf{A}') = 2(l-1)\mathrm{tr}(\mathbf{A}\mathbf{A}') \tag{5.16}$$

(for all $\mathbf{A}$ and $\mathbf{A}'$ of the canonical form). Thus, by (5.3),

$$B^0\left(\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right) = \frac{(2l-1)}{(l-1)} B^{D_l}(\mathbf{A},\mathbf{A}'), \tag{5.17}$$

for all $\mathbf{A}$ and $\mathbf{A}'$ of the canonical form. This implies that (2.37) is satisfied if

$$\mu = (l-1)/(2l-1). \tag{5.18}$$

Denoting the simple roots of D_l by $\alpha_k^{D_l}$ (for $k = 1,2,...,l$), the corresponding basis elements of the Cartan subalgebra $\mathscr{H}^{D_l}$ of D_l defined by

$$B^{D_l}(\mathbf{h}_{\alpha_k^{D_l}}^{D_l}, \mathbf{h}) = \alpha_k^{D_l}(\mathbf{h}) \quad \text{for all } \mathbf{h} \in \mathscr{H}^{D_l} \tag{5.19}$$

[cf. (2.13) and (2.17)] are

$$\mathbf{h}_{\alpha_k^{D_l}}^{D_l} = \{\tfrac{1}{4}(l-1)\}\{\mathbf{e}_{k,k} - \mathbf{e}_{k+l,k+l} - \mathbf{e}_{k+1,k+1} + \mathbf{e}_{k+l+1,k+l+1}\} \quad (\text{for } k = 1,2,...,l-1) \tag{5.20}$$

and

$$\mathbf{h}_{\alpha_l^{D_l}}^{D_l} = \{\tfrac{1}{4}(l-1)\}\{\mathbf{e}_{l-1,l-1} - \mathbf{e}_{2l-1,2l-1} + \mathbf{e}_{l,l} - \mathbf{e}_{2l,2l}\}. \tag{5.21}$$

The associated root subspace basis elements are

$$\mathbf{e}_{\alpha_k^{D_l}}^{D_l} = \{\tfrac{1}{2}(l-1)\}\{\mathbf{e}_{k,k+1} - \mathbf{e}_{k+l+1,k+l}\} \quad (\text{for } k = 1,2,...,l-1) \tag{5.22}$$

and

$$\mathbf{e}_{\alpha_l^{D_l}}^{D_l} = \{\tfrac{1}{2}(l-1)\}\{\mathbf{e}_{l-1,2l} - \mathbf{e}_{l,2l-1}\}, \tag{5.23}$$

the normalization factors being chosen so that

$$B^{D_l}(\mathbf{e}_{\alpha_k^{D_l}}^{D_l}, \mathbf{e}_{-\alpha_k^{D_l}}^{D_l}) = -1, \tag{5.24}$$

where, as usual,

$$\mathbf{e}_{-\alpha_k^{D_l}}^{D_l} = -\tilde{\mathbf{e}}_{\alpha_k^{D_l}}^{D_l}. \tag{5.25}$$

The *diagonal* basis elements of $\mathscr{L}_{s0}^{0(4)}$ will be considered first. As they may be taken to consist of the set $\{\mathbf{e}_{k,k} - \mathbf{e}_{k+l,k+l} | \text{ for } k = 1,2,...,l\}$, it follows that they are all members of the Cartan subalgebra $\mathscr{H}^{0(4)} (= \mathscr{H}^{D_l})$ of D_l (as expected). As

$$\mathbf{e}_{k,k} - \mathbf{e}_{k+l,k+l} = 2(2l-1) \sum_{r=k}^{k+l-1} \mathbf{h}_{\alpha_r^0}^0 \tag{5.26}$$

[for $k = 1,2,...,l$, by (5.4a)], the most general element of $\mathscr{H}^{0(4)}$ has the form

$$\sum_{k=1}^{l} \kappa_k (\mathbf{e}_{k,k} - \mathbf{e}_{k+l,k+l}) = 2(2l-1) \sum_{k=1}^{2l-1} \rho k \, \mathbf{h}_{\alpha_k^0}^0, \tag{5.27}$$

where $\kappa_1, \kappa_2, ..., \kappa_l$ are any complex numbers, and where

$$\rho_k = \sum_{r=1}^{k} \kappa_r \quad (\text{for } k = 1,2,...,l)$$

and

$$\rho_{k+l} = \sum_{r=k+1}^{l} \kappa_r \quad (\text{for } k = 1,2,...,l-1).$$

Thus on $\mathscr{H}^{0(4)}$ the simple roots of $A(2l-1/0)$ are given by

$$\alpha_k^0(h) = \kappa_k - \kappa_{k+1} \quad (\text{for } k = 1,2,...,l-1), \tag{5.28}$$

$$\alpha_l^0(h) = \kappa_1 + k_l, \tag{5.29}$$

$$\alpha_{k+l}^0(h) = -(\kappa_k - \kappa_{k+1}) \quad (\text{for } k = 1,2,...,l-1), \tag{5.30}$$

and

$$\alpha_{2l}^0(h) = -\kappa_l. \tag{5.31}$$

However, from (5.20) and (5.21)

$$\mathbf{e}_{k,k} - \mathbf{e}_{k+l,k+l}$$
$$= 4(l-1)\left\{\sum_{r=k}^{k+l-1} \mathbf{h}^{D_l}_{\alpha_r^{D_l}} - \frac{1}{2}\mathbf{h}^{D_l}_{\alpha_{l-1}^{D_l}} - \frac{1}{2}\mathbf{h}^{D_l}_{\alpha_l^{D_l}}\right\}$$
(for $k = 1,2...,l-2$),

$$\mathbf{e}_{l-1,l-1} - \mathbf{e}_{2l-1,2l-1} = 2(l-1)\left\{\mathbf{h}^{D_l}_{\alpha_{l-1}^{D_l}} + \mathbf{h}^{D_l}_{\alpha_l^{D_l}}\right\},$$

and

$$\mathbf{e}_{l,l} - \mathbf{e}_{2l,2l} = 2(l-1)\left\{-\mathbf{h}^{D_l}_{\alpha_{l-1}^{D_l}} + \mathbf{h}^{D_l}_{\alpha_l^{D_l}}\right\},$$

so

$$\sum_{k=1}^{l} \kappa_k(\mathbf{e}_{k,k} - \mathbf{e}_{k+l,k+l}) = 4(l-1)\sum_{k=1}^{l} \mu_k \mathbf{h}^{D_l}_{\alpha_k^{D_l}} \tag{5.32}$$

where

$$\mu_k = \sum_{r=1}^{k} \kappa_r \quad (\text{for } k = 1,2,...,l-2),$$

$$\mu_{l-1} = \frac{1}{2}\sum_{r=1}^{l-1} \kappa_r - \frac{1}{2}\kappa_l,$$

and

$$\mu_l = \frac{1}{2}\sum_{r=1}^{l} \kappa_r.$$

Thus for $h \in \mathscr{H}^{0(4)}(=\mathscr{H}^{D_l})$

$$\alpha_k^{D_l}(h) = \kappa_k - \kappa_{k+1} \quad (\text{for } k = 1,2,...,l-1) \tag{5.33}$$

and

$$\alpha_l^{D_l}(h) = \kappa_{l-1} + \kappa_l. \tag{5.34}$$

Comparison of (5.28)–(5.31) with (5.33) and (5.34) then shows that for $h \in \mathscr{H}^{0(4)}(=\mathscr{H}^{D_l})$ of D_l the simple roots $\alpha_k^{D_l}$ of D_l and α_k^0 of $A(2l-1/0)$ are related by

$$\alpha_k^0(h) = -\alpha_{k+l}^0(h) = \alpha_k^{D_l}(h) \quad \text{for } k = 1,2,...,l-1, \tag{5.35}$$

$$\alpha_l^0(h) = \sum_{r=1}^{l-2} \alpha_r^{D_l}(h) + \alpha_l^{D_l}(h), \tag{5.36}$$

and

$$\alpha_{2l}^0(h) = \tfrac{1}{2}\alpha_{l-1}^{D_l}(h) - \tfrac{1}{2}\alpha_l^{D_l}(h). \tag{5.37}$$

[when $l = 2$ the first term of (5.36) does not appear].

Finally it follows from (2.13), (2.26)–(2.28), (2.36), (2.37), (5.20), (5.21), and (5.26) that corresponding elements of the Cartan subalgebra of the Kac–Moody superalgebra are

$$h_{\alpha_k^{D_l}} = t^0 \otimes h^{D_l}_{\alpha_k^{D_l}} = \{(2l-1)/2(l-1)\}t^0 \otimes \{\mathbf{h}^0_{\alpha_k^0} - \mathbf{h}^0_{\alpha_{k+l}^0}\}$$
(for $k = 1,2,...,l-1$), (5.38)

and

$$h_{\alpha_l^{D_l}} = t^0 \otimes h^{D_l}_{\alpha_l^{D_l}}$$
$$= \frac{(2l-1)}{2(l-1)}\, t^0 \otimes \left\{\mathbf{h}^0_{\alpha_{l-1}^0} + 2\sum_{r=l}^{2l-2} \mathbf{h}^0_{\alpha_r^0} + \mathbf{h}^0_{\alpha_{2l-1}^0}\right\}. \tag{5.39}$$

The *nondiagonal* basis elements of $\mathscr{L}^{0(4)}_{s0}$ will now be examined. They fall into four sets.

(i) For $j,k = 1,2,...,l$, with $j<k$:

$$\mathbf{e}_{j,k} - \mathbf{e}_{k+l,j+l} = \mathbf{e}^0_{\beta^0_{(j,k)}} + \mathbf{e}^0_{-\beta^0_{(j+l,k+l)}}, \tag{5.40}$$

where $\mathbf{e}^0_{\beta^0_{(j,k)}}$ and $\mathbf{e}^0_{-\beta^0_{(j+l,k+l)}}$ are given by (2.19) and (5.6a). As (5.35) implies that

$$\beta^0_{(j,k)}(h) = -\beta^0_{(j+l,k+l)}(h) = \sum_{r=j}^{k-1} \alpha_r^{D_l}(h) \tag{5.41}$$

[for $j,k = 1,2,...,l$, with $j<k$, and for all $h \in \mathscr{H}^{0(4)}(=\mathscr{H}^{D_l})$], the basis element (5.40) corresponds to the root $\beta^0_{(j,k)}(h)$ of D_l.

(ii) For $j,k = 1,2,...,l$, with $j<k$:

$$-\mathbf{e}_{k,j} + \mathbf{e}_{j+l,k+l} = \mathbf{e}^0_{-\beta^0_{(j,k)}} + \mathbf{e}^0_{\beta^0_{(j+l,k+l)}}, \tag{5.42}$$

which corresponds to the root $-\beta^0_{(j,k)}(h)$ of D_l, where $\beta^0_{(j,k)}(h)$ is given by (5.41).

(iii) For $j,k = 1,2,...,l$, with $j<k$:

$$\mathbf{e}_{j,k+l} - \mathbf{e}_{k,j+l} = \mathbf{e}^0_{\beta^0_{(j,k+l)}} - \mathbf{e}^0_{\beta^0_{(k,j+l)}}, \tag{5.43}$$

where $\mathbf{e}^0_{\beta^0_{(j,k+l)}}$ and $\mathbf{e}^0_{\beta^0_{(k,j+l)}}$ are again given by (5.6a). As (5.35) and (5.36) imply that for all $h \in \mathscr{H}^{0(4)}(=\mathscr{H}^{D_l})$

$$\beta^0_{(j,k+l)}(h) = \beta^0_{(k,j+l)}(h)$$
$$= \begin{cases} \displaystyle\sum_{r=j}^{k-1} \alpha_r^{D_l}(h) + 2\sum_{r=k}^{l-2} \alpha_r^{D_l}(h) + \alpha_{l-1}^{D_l}(h) + \alpha_l^{D_l}(h) & (\text{for } j,k = 1,2,...,l-2, \text{ with } j<k), \\ \displaystyle\sum_{r=j}^{l-2} \alpha_r^{D_l}(h) + \alpha_{l-1}^{D_l}(h) + \alpha_l^{D_l}(h) & (\text{for } j = 1,2,...,l-2, \text{ and } k = l-1), \\ \displaystyle\sum_{r=j}^{l-2} \alpha_r^{D_l}(h) + \alpha_l^{D_l}(h) & (\text{for } j = 1,2,...,l-2, \text{ and } k = l), \\ \text{and} & \\ \alpha_l^{D_l}(h) & (\text{for } j = l-1 \text{ and } k = l), \end{cases} \tag{5.44}$$

the basis element (5.43) corresponds to the root $\beta^0_{(j,k+l)}(h)$ of D_l.

(iv) For $j,k = 1,2,...,l$, with $j<k$:

$$\mathbf{e}_{j+l,k} - \mathbf{e}_{k+l,j} = -\mathbf{e}^0_{-\beta^0_{(k,j+l)}} + \mathbf{e}^0_{-\beta^0_{(j,k+l)}}, \tag{5.45}$$

which corresponds to the root $-\beta^0_{(j,k+l)}(h)$ of D_l, where $\beta^0_{(j,k+l)}(h)$ is given by (5.44).

As expected, the elements of (5.40), (5.42), (5.43), and (5.45) are *even* members of $A(2l-1/0)$. It is easily checked that the set of $2l(l-1)$ nonzero roots of (i)–(iv) above,

together with the l zero roots, are all weights of the adjoint representation of D_l. For $l \geqslant 4$ the highest weight is

$$\Lambda = \Lambda_2^{D_l} = \alpha_1^{D_l} + 2\sum_{k=2}^{l-2} \alpha_k^{D_l} + \alpha_{l-1}^{D_l} + \alpha_l^{D_l}, \tag{5.46}$$

while for $l = 2$ and 3 the second term on the right-hand side of (5.46) does not appear and

$$\Lambda = \Lambda_{l-1}^{D_l} + \Lambda_l^{D_l} = \sum_{k=1}^{l} \alpha_k^{D_l} \tag{5.47}$$

as expected [cf. Cornwell[17] (Appendix F)].

2. $\mathcal{L}_{s1}^{0(4)}$

By (2.33) the subspace $\mathcal{L}_{s1}^{0(4)}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G} = i\mathbf{A}, \quad -\tilde{\mathbf{B}}\mathbf{G} = i\mathbf{C}, \quad \mathbf{G}\tilde{\mathbf{C}} = i\mathbf{B}, \text{ and } -\tilde{\mathbf{D}} = i\mathbf{D},$$

which when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$\mathbf{A} = \mathbf{0}, \quad \mathbf{D} = \mathbf{0}, \quad \text{and } \mathbf{C} = i\tilde{\mathbf{B}}\mathbf{G}. \tag{5.48}$$

The basis elements of $\mathcal{L}_{s1}^{0(4)}$ fall into two sets:
(i) For $j = 1,2,\ldots,l$:

$$\mathbf{e}_{j,2l+1} + i\mathbf{e}_{2l+1,j+l} = \mathbf{e}^0_{\delta^0_{(j)}} - i\mathbf{e}^0_{-\delta^0_{(j+l)}}, \tag{5.49}$$

where $\mathbf{e}^0_{\delta^0_{(j)}}$ and $i\mathbf{e}^0_{-\delta^0_{(j+l)}}$ are given by (2.19) and (5.6b). For all $h \in \mathcal{H}^{0(4)} (= \mathcal{H}^{D_l})$ (5.35), (5.36), and (5.37) imply that

$$\delta^0_{(j)}(h) = -\delta^0_{(j+l)}(h) = \begin{cases} \sum_{r=j}^{l-2} \alpha_r^{D_l}(h) + \frac{1}{2}\alpha_{l-1}^{D_l}(h) + \frac{1}{2}\alpha_l^{D_l}(h), & \text{for } j \leqslant l-2, \\ \frac{1}{2}\alpha_{l-1}^{D_l}(h) + \frac{1}{2}\alpha_l^{D_l}(h), & \text{for } j = l-1, \\ -\frac{1}{2}\alpha_{l-1}^{D_l}(h) + \frac{1}{2}\alpha_l^{D_l}(h), & \text{for } j = l. \end{cases} \tag{5.50}$$

In all cases the basis element (5.49) corresponds to the root $\delta^0_{(j)}(h)$ of D_l.

(ii) For $j = 1,2,\ldots,l$:

$$-\mathbf{e}_{2l+1,j} + i\mathbf{e}_{j+l,2l+1} = \mathbf{e}^0_{-\delta^0_{(j)}} + i\mathbf{e}^0_{\delta^0_{(j+l)}}, \tag{5.51}$$

which corresponds to the weight $-\delta^0_{(j)}(h)$ of D_l [$\delta^0_{(j)}(h)$ being as in (5.50)].

These weights all belong to a $2l$-dimensional irreducible representation of D_l with highest weight:

$$\Lambda = \Lambda_1^{D_l} = \sum_{k=1}^{l-2} \alpha_k^{D_l} + \frac{1}{2}\alpha_{l-1}^{D_l} + \frac{1}{2}\alpha_l^{D_l}, \tag{5.52}$$

[where for $l = 2$ the first term on the right-hand side of (5.52) does not appear]. It should be noted that all the elements of (5.49) and (5.51) are odd members of $A(2l-1/0)$, so all the elements of $\mathcal{L}_{s2}^{0(4)}$ are odd.

3. $\mathcal{L}_{s2}^{0(4)}$

By (2.33) the subalgebra $\mathcal{L}_{s2}^{0(4)}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G} = -\mathbf{A}, -\tilde{\mathbf{B}}\mathbf{G} = -\mathbf{C}, \mathbf{G}\tilde{\mathbf{C}} = -\mathbf{B}, \quad \text{and } -\tilde{\mathbf{D}} = -\tilde{\mathbf{D}},$$

which when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$\tilde{\mathbf{A}}\mathbf{G} - \mathbf{G}\mathbf{A} = \mathbf{0}, \quad \mathbf{B} = \mathbf{0}, \mathbf{C} = \mathbf{0}, \tag{5.53}$$

with $\mathbf{D}$ being determined only by the supertrace condition $\mathrm{tr}\,\mathbf{A} = \mathrm{tr}\,\mathbf{D}$. On using (5.4a) and (5.4b), the diagonal basis elements of $\mathcal{L}_{s2}^{0(4)}$ may be taken to consist of the set

$$\{\tfrac{1}{2}(2l-1)\}\{\mathbf{e}_{k,k} + \mathbf{e}_{k+l,k+l} + 2\mathbf{e}_{2l+1,2l+1}\} = \sum_{r=k}^{k+l-1} \mathbf{h}^0_{\alpha^0_r} + 2\sum_{r=k+l}^{2l} \mathbf{h}^0_{\alpha^0_r} \tag{5.54}$$

(for $k = 1,2,\ldots,l$), which each corresponds to zero weight of D_l.

The nondiagonal basis elements of $\mathcal{L}_{s2}^{0(4)}$ fall into six sets.

(i) For $j,k = 1,2,\ldots l$, with $j<k$:

$$\mathbf{e}_{j,k} + \mathbf{e}_{k+l,j+l} = \mathbf{e}^0_{\beta^0_{(j,k)}} - \mathbf{e}^0_{-\beta^0_{(j+l,k+l)}}, \tag{5.55}$$

where $\mathbf{e}^0_{\beta^0_{(j,k)}}$ and $\mathbf{e}^0_{-\beta^0_{(j+l,k+l)}}$ are given by (2.19) and (5.6a), and $\beta^0_{(j,k)}(h) (= -\beta^0_{(j+l,k+l)}(h))$ is given for all $h \in \mathcal{H}^{0(4)} (= \mathcal{H}^{D_l})$ by (5.41), so this basis element (5.55) again corresponds to the root $\beta^0_{(j,k)}(h)$ of D_l.

(ii) For $j,k = 1,2,\ldots,l$, with $j<k$:

$$-\mathbf{e}_{k,j} - \mathbf{e}_{j+l,k+l} = \mathbf{e}^0_{-\beta^0_{(j,k)}} - \mathbf{e}^0_{\beta^0_{(j+l,k+l)}}, \tag{5.56}$$

which corresponds to the root $-\beta^0_{(j,k)}(h)$ of D_l, where $\beta^0_{(j,k)}(h)$ is given by (5.41).

(iii) For $j = 1,2,\ldots l$:

$$\mathbf{e}_{j,j+l} = \mathbf{e}^0_{\beta^0_{(j,j+l)}}, \tag{5.57}$$

where $\mathbf{e}^0_{\beta^0_{(j,j+l)}}$ is given by (5.6a), which corresponds to the weight $\beta^0_{(j,j+l)}$ of D_l. By a further application of (5.35) and (5.36) $\beta^0_{(j,j+l)}(h)$ can be rewritten for $h \in \mathcal{H}^{0(4)} (= \mathcal{H}^{D_l})$ as

$$\beta^0_{(j,j+l)}(h)=\begin{cases}2\sum_{r=j}^{l-2}\alpha_r^{D_l}(h)+\alpha_{l-1}^{D_l}(h)+\alpha_l^{D_l}(h) & (\text{for } j=1,2,\ldots,l-2),\\ \alpha_{l-1}^{D_l}(h)+\alpha_l^{D_l}(h) & (\text{for } j=l-1),\\ -\alpha_{l-1}^{D_l}(h)+\alpha_l^{D_l}(h) & (\text{for } j=l).\end{cases} \tag{5.58}$$

(iv) For $j=1,2,\ldots,l$:

$$-\mathbf{e}_{j+l,j}=\mathbf{e}^0_{-\beta^0_{(j,j+l)}}, \tag{5.59}$$

which corresponds to the weight $-\beta^0_{(j,j+l)}(h)$ of D_l, where $\beta^0_{(j,j+l)}(h)$ is given by (5.58).

(v) For $j,k=1,2,\ldots,l$, with $j<k$:

$$\mathbf{e}_{j,k+l}+\mathbf{e}_{kj+l}=\mathbf{e}^0_{\beta^0_{(j,k+l)}}+\mathbf{e}^0_{\beta^0_{(kj+l)}}, \tag{5.60}$$

where $\mathbf{e}^0_{\beta^0_{(j,k+l)}}$ and $\mathbf{e}^0_{\beta^0_{(kj+l)}}$ are again given by (5.6a). As $\beta^0_{(j,k+l)}(h)(=\beta^0_{(kj+l)}(h))$ is given for all $h\in\mathcal{H}^{O(4)}(=\mathcal{H}^{D_l})$ by (5.44), so this basis element (5.60) again corresponds to the root $\beta^0_{(j,k+l)}(h)$ of D_l.

(vi) For $j,k=1,2,\ldots,l$, with $j<k$:

$$-\mathbf{e}_{k+l,j}-\mathbf{e}_{j+l,k}=\mathbf{e}^0_{-\beta^0_{(j,k+l)}}+\mathbf{e}^0_{-\beta^0_{(kj+l)}}, \tag{5.61}$$

which corresponds to the root $-\beta^0_{(j,k+l)}(h)$ of D_l, where $\beta^0_{(j,k+l)}(h)$ is given by (5.44).

These $2l^2+l$ weights belong to a representation of D_l which is the direct sum of the trivial 1-dimensional irreducible representation with highest weight $\Lambda=0$ and the $(2l^2+l-1)$-dimensional irreducible representation with highest weight

$$\Lambda=2\Lambda_1^{D_l}=2\sum_{k=1}^{l-2}\alpha_k^{D_l}+\alpha_{l-1}^{D_l}+\alpha_l^{D_l}, \tag{5.62}$$

[where for $l=2$ the first term on the right-hand side of (5.62) does not appear]. It should be noted that all the elements of $\mathcal{L}^{O(4)}_{s2}$ are even members of $A(2l-1/0)$.

4. $\mathcal{L}^{O(4)}_{s3}$

By (2.33) the subspace $\mathcal{L}^{O(4)}_{s3}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G}=-i\mathbf{A},\quad -\tilde{\mathbf{B}}\mathbf{G}=-i\mathbf{C},\quad \mathbf{G}\tilde{\mathbf{C}}=-i\mathbf{B},$$
$$\text{and}\quad -\tilde{\mathbf{D}}=-i\mathbf{D},$$

which when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$\mathbf{A}=0,\quad \mathbf{D}=0,\quad \text{and}\quad \mathbf{C}=-i\tilde{\mathbf{B}}\mathbf{G}. \tag{5.63}$$

The basis elements of $\mathcal{L}^{O(4)}_{s3}$ fall into two sets.

(i) For $j=1,2,\ldots,l$:

$$\mathbf{e}_{j,2l+1}-i\mathbf{e}_{2l+1,j+l}=\mathbf{e}^0_{\delta^0_{(j)}}+i\mathbf{e}^0_{-\delta^0_{(j+l)}}, \tag{5.64}$$

where $\mathbf{e}^0_{\delta^0_{(j)}}$ and $i\mathbf{e}^0_{-\delta^0_{(j+l)}}$ are given by (2.19) and (5.6b). As $\delta^0_{(j)}(h)(=-\delta^0_{(j+l)}(h))$ is given for all $h\in\mathcal{H}^{O(4)}(=\mathcal{H}^{D_l})$ by (5.50) so the basis element (5.64) again corresponds to the root $\delta^0_{(j)}(h)$ of D_l.

(ii) For $j=1,2,\ldots,l$:

$$-\mathbf{e}_{2l+1,j}-i\mathbf{e}_{j+l,2l+1}=\mathbf{e}^0_{-\delta^0_{(j)}}-i\mathbf{e}^0_{\delta^0_{(j+l)}}, \tag{5.65}$$

which corresponds to the weight $-\delta^0_{(j)}(h)$ of D_l, $\delta^0_{(j)}(h)$ being as in (5.50).

These two sets of weights are exactly the same as for $\mathcal{L}^{O(4)}_{s1}$, so they all belong to a $2l$-dimensional irreducible representation of D_l with highest weight Λ is given by (5.58) and (5.52). All the elements of $\mathcal{L}^{O(4)}_{s3}$ are odd.

C. The roots of $A^{(2)}(2l-1/0)$

Defining $\delta(h)$ as in (3.6) to (3.8), it follows that the roots $\alpha(h)$ and the corresponding basis elements e_α of $A^{(2)}(2l-1/0)$ are as follows.

(i) $\alpha(h)=4J\delta(h)$, (for $J=0,\pm1,\pm2,\ldots$). There are l linearly independent basis elements $e^{(k)}_\alpha$ corresponding to this root which may be labeled by an additional superscript, so that

$$e^{(k)}_\alpha=\{(2l-1)/2(l-1)\}t^{4J}\otimes\{\mathbf{h}^0_{\alpha^0_k}-\mathbf{h}^0_{\alpha^0_{k+l}}\}$$
$$(\text{for } k=1,2,\ldots,l-1)$$

and

$$e^{(l)}_\alpha=\{(2l-1)/2(l-1)\}t^{4J}$$
$$\otimes\left\{\mathbf{h}^0_{\alpha^0_{l-1}}+2\sum_{r=l}^{2l-2}\mathbf{h}^0_{\alpha^0_r}+\mathbf{h}^0_{\alpha^0_{2l-1}}\right\},$$

[which reduce to (5.38) and (5.39) in the special case $J=0$].

(ii) $\alpha(h)=4J\delta(h)\pm\beta^0_{(j,k)}(h)$ (for $j,k=1,2,\ldots,l$, with $j<k$, and for $J=0,\pm1,\pm2,\ldots$), where $\beta^0_{(j,k)}(h)$ is the extension of the weight of D_l that is given by (5.41) and

$$e_\alpha=t^{4J}\otimes\{\mathbf{e}^0_{\pm\beta^0_{(j,k)}}+\mathbf{e}^0_{\mp\beta^0_{(j+l,k+l)}}\}.$$

(iii) $\alpha(h)=4J\delta(h)\pm\beta^0_{(j,k+l)}(h)$, (for $j,k=1,2,\ldots,l$, with $j<k$, and for $J=0,\pm1,\pm2,\ldots$), where $\beta^0_{(j,k+l)}(h)$ is the extension of the weight of D_l that is given by (5.44) and

$$e_\alpha=t^{4J}\otimes\{\mathbf{e}^0_{\pm\beta^0_{(j,k+l)}}-\mathbf{e}^0_{\pm\beta^0_{(kj+l)}}\}.$$

(iv) $\alpha(h)=(4J+1)\delta(h)\pm\delta^0_{(j)}(h)$, (for $j=1,2,\ldots,l$, and for $J=0,\pm1,\pm2,\ldots$), where $\delta^0_{(j)}(h)$ is the extension of the weight of D_l that is given by (5.50) and

$$e_\alpha=t^{4J+1}\otimes\{\mathbf{e}^0_{\pm\delta^0_{(j)}}\mp i\mathbf{e}^0_{\mp\delta^0_{(j+l)}}\}.$$

(v) $\alpha(h)=(4J+2)\delta(h)$, (for $J=0,\pm1,\pm2,\ldots$). There are l linearly independent basis elements $e^{(k)}_\alpha$ corresponding to this root which may be labeled by an additional superscript, so that

$$e^{(k)}_\alpha=t^{4J+2}\otimes\left\{\sum_{r=k}^{k+l-1}\mathbf{h}^0_{\alpha^0_r}+2\sum_{r=k+l}^{2l}\mathbf{h}^0_{\alpha^0_r}\right\}$$
$$(\text{for } k=1,2,\ldots,l);$$

(vi) $\alpha(h)=(4J+2)\delta(h)\pm\beta^0_{(j,k)}(h)$, (for j

$= 1,2,...,l$, with $j<k$, and for $J=0,\pm 1,\pm 2,...$), where $\beta^0_{(j,k)}(h)$ is the extension of the weight of D_l that is given by (5.41) and

$$e_\alpha = t^{4J+2}\otimes\{\mathbf{e}^0_{\pm\beta^0_{(j,k)}} - \mathbf{e}^0_{\mp\beta^0_{(j+l,k+l)}}\}.$$

(vii) $\alpha(h) = (4J+2)\delta(h) \pm \beta^0_{(jj+l)}(h)$, (for $j=1,2,...,l$, and for $J=0,\pm 1,\pm 2,...$), where $\beta^0_{(jj+l)}(h)$ is the extension of the weight of D_l that is given by (5.58) and $e_\alpha = t^{4J+2}\otimes \mathbf{e}^0_{\pm\beta^0_{(jj+l)}}$.

(viii) $\alpha(h) = (4J+2)\delta(h) \pm \beta^0_{(j,k+l)}(h)$, (for $j,k=1,2,...,l$, with $j<k$, and for $J=0,\pm 1,\pm 2,...$), where $\beta^0_{(j,k+l)}(h)$ is the extension of the weight of D_l that is given by (5.44) and

$$e_\alpha = t^{4J+2}\otimes\{\mathbf{e}^0_{\pm\beta^0_{(j,k+l)}} + \mathbf{e}^0_{\pm\beta^0_{(k,j+l)}}\}.$$

(ix) $\alpha = (4J+3)\delta(h) \pm \delta^0_{(j)}(h)$, (for $j=1,2,...,l$, and for $J=0,\pm 1,\pm 2,...$), where $\delta^0_{(j)}(h)$ is the extension of the weight of D_l that is given by (5.50) and

$$e_\alpha = t^{4J+3}\otimes\{\mathbf{e}^0_{\pm\delta^0_{(j)}} \pm i\mathbf{e}^0_{\mp\delta^0_{(j+l)}}\}.$$

(x) $\alpha(h)=0$, with c and d as basis elements.

With μ chosen as in (5.18) [so that (2.37) is valid], it follows that

$$\langle\alpha^0,\beta^0\rangle = \langle\alpha^0,\beta^0\rangle^{D_l}, \tag{5.66}$$

where on the right-hand side of (5.66) α^0 and β^0 are any pair of linear functionals defined on $\mathcal{H}^{0(4)}(=\mathcal{H}^{D_l})$, the evaluation being performed with respect to the Killing form of D_l, and where on the left-hand side of (5.66) α^0 and β^0 denote the corresponding extensions to the Cartan subalgebra of the Kac–Moody superalgebra $A^{(2)}(2l-1/0)$, the evaluation being performed with respect to its supersymmetric bilinear invariant form $B(\ ,\ )$. As D_l is a semisimple Lie algebra, $\langle\alpha^0,\alpha^0\rangle^{D_l}>0$ for every nonzero linear functional α^0 defined on $\mathcal{H}^{D_l}$, so $\langle\alpha^0,\alpha^0\rangle>0$ for the corresponding extension. Moreover (2.13), (2.28), (2.30), (2.31), (2.36), and (5.18) imply that

$$h_\delta = \{(2l-1)/(l-1)\}c. \tag{5.67}$$

Thus, if α^0_k is the extension of any simple root of $\mathcal{L}^0_s$, then

$$\langle\delta,\alpha^0_k\rangle = 0 \tag{5.68}$$

and

$$\langle j\delta, j\delta\rangle = 0. \tag{5.69}$$

Thus $\langle j\delta, j\delta\rangle = 0$ for integer j, so every nonzero root of $A^{(2)}(2l-1/0)$ belonging to the sets (i) and (v) is "imaginary." Moreover, because $\langle j\delta+\alpha^0, j\delta+\alpha^0\rangle = \langle\alpha^0,\alpha^0\rangle^{D_l}$ and because $\langle\alpha^0,\alpha^0\rangle^{D_l}>0$ for linear functional α^0 and its corresponding extension (as has just been noted), it follows that every root of $A^{(2)}(2l-1/0)$ belonging to the sets (ii), (iii), (iv), (vi), (vii), (viii), and (ix) is "real." All the elements mentioned in the above sets are even, except for those in the sets (iv) and (ix), which are odd.

In relating these roots to the simple roots of the Kac–Moody superalgebra $A^{(2)}(2l-1/0)$ it is necessary to consider the cases $l=2$ and $l>2$ separately because the labeling of the generalized Dynkin diagrams of $A^{(2)}(2l-1/0)$ is different in the two cases.

For $A^{(2)}(3/0)$ (i.e., for $l=2$) the simple roots may be taken to be

$$\alpha_0 = \alpha_1^{D_2}, \quad \alpha_1 = \delta - \alpha^0_H, \quad \alpha_2 = \alpha_2^{D_2},$$

where

$$\alpha^0_H = \Lambda_1^{D_2} = \frac{1}{2}\sum_{k=1}^{2}\alpha_k^{D_2} \tag{5.70}$$

is the highest weight of the representation of $\mathcal{L}^{0(4)}_{s0}$ for which $\mathcal{L}^{0(4)}_{s1}$ is the carrier space [cf. (5.58)] and $\alpha_1^{D_2}$ and $\alpha_2^{D_2}$ are the extensions of the simple roots α^0_1 and α^0_2 of D_2. As e_{α_1} appears in the set (iv) it follows that e_{α_1} is odd, so α_1 is an odd root of the Kac–Moody superalgebra $A^{(2)}(3/0)$. All the other simple roots of $A^{(2)}(3/0)$ are even.

For $A^{(2)}(2l-1/0)$ for $l>2$ the simple roots may be taken to be

$$\alpha_l = \delta - \alpha^0_H \tag{5.71}$$

and

$$\alpha_k = \alpha^{D_l}_{l-k} \quad (\text{for } k=0,1,...,l-1), \tag{5.72}$$

where

$$\alpha^0_H = \Lambda_1^{D_l} = \sum_{k=1}^{l-2}\alpha_k^{D_l} + \frac{1}{2}\alpha^{D_l}_{l-1} + \frac{1}{2}\alpha_l^{D_l} \tag{5.73}$$

is the highest weight of the representation of $\mathcal{L}^{0(4)}_{s0}$ for which $\mathcal{L}^{0(4)}_{s1}$ is the carrier space [cf. (5.52)] and the $\alpha_k^{D_l}$ are the extensions of the simple roots of D_l. As e_{α_l} appears in the set (iv) it follows that e_{α_l} is odd, so α_l is an odd root of the Kac–Moody superalgebra $A^{(2)}(2l-1/0)$ (for $l>2$). All the other simple roots of $A^{(2)}(2l-1/0)$ are even (for $l>2$). It is then easily checked that the Cartan matrices of $A^{(2)}(3/0)$ and $A^{(2)}(2l-1/0)$ (for $l\geqslant 3$) evaluated using (2.15) correspond to the generalized Dynkin diagrams given in Figs. 3 and 4.

VI. TWISTED AFFINE KAC–MOODY SUPERALGEBRAS CORRESPONDING TO AUTOMORPHISMS OF ORDER 4: ROOTS OF $A^{(4)}(2l/0)$ (FOR $l=1,2,3,...$)

A. The fourfold automorphisms

The general line of argument for $A^{(4)}(2l/0)$ is very similar to that given for $A^{(2)}(2l-1/0)$ in the previous section, so its presentation can be given more briefly. An explicit realization of the simple Lie superalgebra $A(2l/0)$ is provided by $\mathrm{sl}(2l+1/1)$, considered as a complex superalgebra, where $\mathrm{sl}(2l+1/1)$ is defined as the set of $(2l+2)\times(2l+2)$ complex supermatrices that satisfy the condition (5.2). The grading partitioning may be taken to be as in (5.1), but now **A**, **B**, **C**, and **D** are of dimensions $(2l+1)\times(2l+1)$, $(2l+1)\times 1$, $1\times(2l+1)$, and 1×1, respectively. The rank l^0 of $A(2l/0)$ is given by

$$l^0 = 2l+1. \tag{6.1}$$

The generalized Dynkin diagram of $A(2l/0)$ is shown in Fig. 11, which indicates that its distinguished simple roots α^0_k are even for $k=1,2,...,2l$, but that α^0_{2l+1} is odd. With the bilinear form $B^0(\ ,\ )$ being defined by

$$B^0(\mathbf{M},\mathbf{N}) = 4l\,\mathrm{str}\,(\mathbf{MN}), \tag{6.2}$$

the basis elements of its Cartan subalgebra $\mathscr{H}^0$ may be taken to be

$$\mathbf{h}^0_{\alpha^0_k} = \{1/4l\}\{\mathbf{e}_{k,k} - \mathbf{e}_{k+1,k+1}\} \quad (\text{for } k = 1,2,...,2l) \tag{6.3}$$

and

$$\mathbf{h}^0_{\alpha^0_{2l+1}} = \{1/4l\}\{\mathbf{e}_{2l+1,2l+1} + \mathbf{e}_{2l+2,2l+2}\}. \tag{6.4}$$

Now $\mathbf{e}_{r,s}$ is the matrix of dimension $(2l+2)\times(2l+2)$ that is defined by

$$(\mathbf{e}_{r,s})_{jk} = \delta_{rj}\delta_{sk} \quad (\text{for } j,k = 1,2,...,2l+2), \tag{6.5}$$

so that with this choice all the matrices of $\mathscr{H}^0$ are again *diagonal*. The positive even roots $\beta^0_{(j,k)}$ and positive odd roots $\delta^0_{(j)}$ of $A(2l/0)$ are given in terms of the distinguished set of simple roots $\alpha^0_1, \alpha^0_2, ..., \alpha^0_{2l+1}$ of $A(2l/0)$ by

$$\beta^0_{(j,k)} = \sum_{r=j}^{k-1} \alpha^0_r \quad (\text{for } j,k = 1,2,...,2l+1; j<k) \tag{6.6a}$$

and

$$\delta^0_{(j)} = \sum_{r=j}^{2l+1} \alpha^0_r \quad (\text{for } j = 1,2,...,2l+1), \tag{6.6b}$$

for which the corresponding basis elements of $A(2l/0)$ may be taken to be

$$\mathbf{e}^0_{\beta^0_{(j,k)}} = \mathbf{e}^0_{\alpha^0_j + \alpha^0_{j+1} + \cdots + \alpha^0_{k-1}} = \mathbf{e}_{j,k} \quad (\text{for } j,k = 1,2,...,2l+1; j<k) \tag{6.7a}$$

and

$$\mathbf{e}^0_{\delta^0_{(j)}} = \mathbf{e}^0_{\alpha^0_j + \alpha^0_{j+1} + \cdots + \alpha^0_{2l+1}} = \mathbf{e}_{j,2l+2} \quad (\text{for } j = 1,2,...,2l+1). \tag{6.7b}$$

The basis elements corresponding to the corresponding negative roots may be chosen in accordance with (2.19). [For further information on $A(2l/0)$ see Cornwell.[9]]

Taking the node corresponding to the odd simple root α_l of $A^{(4)}(2l/0)$ for $l \geqslant 1$, as the corresponding numerical mark has value 1 (cf. Figs. 5 and 6), (2.32) shows that $q = 4$ again. It follows from (2.33) that if the automorphism (4.6) is employed then the subalgebra $\mathscr{T}^{0(4)}_{s0}$ consists of the supermatrices whose submatrices satisfy the conditions (5.8), so that the subalgebra $\mathscr{T}^{0(q)}_{s0}$ is isomorphic to the set of $(2l+1)\times(2l+1)$ complex antisymmetric matrices and hence is isomorphic to the simple complex Lie algebra B_l. (Inspection of Figs. 5 and 6 shows that the generalized Dynkin diagram with the chosen node and attached lines removed does indeed correspond to B_l.) As expected, none of the basis elements of the Cartan subalgebra $\mathscr{H}^0$ of $A(2l/0)$ are members of this $\mathscr{T}^{0(q)}_{s0}$ (because all the members of this $\mathscr{T}^{0(q)}_{s0}$ are nondiagonal matrices).

A realization of B_l in which the basis elements of the Cartan subalgebra of B_l are given by diagonal matrices is given by the $(2l+1)\times(2l+1)$ complex matrices $\mathbf{A}'$ that satisfy the condition (5.9), but where now

$$\mathbf{G} = \begin{pmatrix} \mathbf{1}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_l \\ \mathbf{0} & \mathbf{1}_l & \mathbf{0} \end{pmatrix}. \tag{6.8}$$

This realization will be referred to as the "canonical" form of B_l. These matrices $\mathbf{A}'$ are related to the $(2l+1)\times(2l+1)$ antisymmetric matrices $\mathbf{A}$ by

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{A}', \tag{6.9}$$

where $\mathbf{T}$ is a certain $(2l+1)\times(2l+1)$ complex matrix that satisfies the condition

$$\tilde{\mathbf{T}}\mathbf{T} = \mathbf{G}, \tag{6.10}$$

$\mathbf{G}$ being as defined in (6.8) (cf. Cornwell[15]). This mapping can be extended to an automorphism of θ of $\mathscr{T}^0_s[= sl(2l+1/1)]$ by using the definition (5.13), so that ϕ is again given by (5.14). [Of course in (5.13) and (5.14) the dimensions of all the submatrices must be modified in the obvious way.]

B. The subspaces $\mathscr{T}^{0(4)}_{sp}$ (for $p=0,1,2,3$)

The four subspaces $\mathscr{T}^{0(4)}_{sp}$ (for $p = 0,1,2,3$) corresponding to the automorphism ϕ of (5.14) will now be considered in turn.

1. $\mathscr{T}^{0(4)}_{s0}$

By (2.33) the subalgebra $\mathscr{T}^{0(4)}_{s0}$ consists of the supermatrices whose submatrices satisfy the conditions (5.15) [with $\mathbf{G}$ given by (6.8)], and so is isomorphic to the canonical form of B_l.

Some properties of the canonical form of B_l (cf. Konuma *et al.*[16] and Cornwell[15]) will first be summarized. Its killing form $B^{B_l}(,)$ is given by

$$B^{B_l}(\mathbf{A},\mathbf{A}') = (2l-1)\mathrm{tr}(\mathbf{A}\mathbf{A}'), \tag{6.11}$$

(for all $\mathbf{A}$ and $\mathbf{A}'$ of the canonical form). Thus, by (6.2),

$$B^0\left(\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right) = \frac{4l}{(2l-1)} B^{B_l}(\mathbf{A},\mathbf{A}'), \tag{6.12}$$

for all $\mathbf{A}$ and $\mathbf{A}'$ of the canonical form. This implies that (2.37) is satisfied if

$$\mu = (2l-1)/(4l). \tag{6.13}$$

Denoting the simple roots of B_l by $\alpha^{B_l}_k$ (for $k = 1,2,...,l$), the corresponding basis elements of the Cartan subalgebra $\mathscr{H}^{B_l}$ of B_l defined by

$$B^{B_l}(\mathbf{h}^{B_l}_{\alpha^{B_l}_k}, \mathbf{h}) = \alpha^{B_l}_k(\mathbf{h}) \text{ for all } \mathbf{h} \in \mathscr{H}^{B_l} \tag{6.14}$$

[cf. (2.13) and (2.17)] are

$$\mathbf{h}^{B_l}_{\alpha^{B_l}_k} = \{1/(2l-1)\}\{\mathbf{e}_{k+1,k+1} - \mathbf{e}_{k+l+1,k+l+1} - \mathbf{e}_{k+2,k+2} + \mathbf{e}_{k+l+2,k+l+2}\}$$
$$(\text{for } k = 1,2,...,l-1) \tag{6.15}$$

and

$$\mathbf{h}^{B_l}_{\alpha^{B_l}_l} = \{1/(2l-1)\}\{\mathbf{e}_{l+1,l+1} - \mathbf{e}_{2l+1,2l+1}\}. \tag{6.16}$$

The associated root subspace basis elements are

$$\mathbf{e}^{B_l}_{\alpha^{B_l}_k} = \{1/2(2l-1)\}\{\mathbf{e}_{k+1,k+2} - \mathbf{e}_{k+l+2,k+l+1}\} \quad (\text{for } k = 1,2,...,l-1) \tag{6.17}$$

and

$$\mathbf{e}^{B_l}_{\alpha^{B_l}_l} = \{1/2(2l-1)\}\{\mathbf{e}_{1,2l+1} - \mathbf{e}_{l+1,1}\}, \tag{6.18}$$

the normalization factors being chosen so that

$$B^{B_l}(\mathbf{e}^{B_l}_{\alpha^{B_l}_k}, \mathbf{e}^{B_l}_{-\alpha^{B_l}_k}) = -1, \tag{6.19}$$

where, as usual,

$$\mathbf{e}^{B_l}_{-\alpha^{B_l}_k} = -\tilde{\mathbf{e}}^{B_l}_{\alpha^{B_l}_k}. \tag{6.20}$$

The *diagonal* basis elements of $\mathscr{T}^{0(4)}_{s0}$ will be considered first. As they may be taken to consist of the set $\{\mathbf{e}_{k+1,k+1} - \mathbf{e}_{k+l+1,k+l+1} \mid \text{for } k = 1,2,...,l\}$, it follows that they are all members of the Cartan subalgebra $\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$ of B_l (as expected). Thus the most general element of $\mathscr{H}^{0(4)}$ is of the form

$$\sum_{k}^{l} = 1\ \kappa_k(\mathbf{e}_{k+1,k+1} - \mathbf{e}_{k+l+1,k+l+1}), \tag{6.21}$$

where $\kappa_1, \kappa_2, ..., \kappa_l$ are any complex numbers, which can be rewritten, by (6.3), as

$$4l\sum_{k}^{2l} = 2\rho_k \mathbf{h}^0_{\alpha^0_k}, \tag{6.22}$$

where

$$\rho_k = \sum_{r=1}^{k-1} \kappa_r \quad (\text{for } k = 2,...,l+1) \tag{6.23}$$

and

$$\rho_{k+l} = \sum_{r=k}^{l} \kappa_r \quad (\text{for } k = 2,...,l). \tag{6.24}$$

Thus on $\mathscr{H}^{0(4)}$ the simple roots of $A(2l/0)$ are given by

$$\alpha^0_1(h) = -\kappa_1, \tag{6.25}$$

$$\alpha^0_k(h) = \kappa_{k-1} - \kappa_k \quad (\text{for } k = 2,3,...,l), \tag{6.26}$$

$$\alpha^0_{l+1}(h) = \kappa_1 + \kappa_l, \tag{6.27}$$

$$\alpha^0_{k+l}(h) = -(\kappa_{k-1} - \kappa_k) \quad (\text{for } k = 2,3,...,l), \tag{6.28}$$

and

$$\alpha^0_{2l+1}(h) = -\kappa_l,$$

which implies that on $\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$

$$\alpha^0_k(h) = -\alpha^0_{k+l}(h) \quad (\text{for } k = 2,3,...,l), \tag{6.29}$$

$$\alpha^0_1(h) = -\frac{1}{2}\sum_{k=2}^{l+1} \alpha^0_k(h), \tag{6.30}$$

and

$$\alpha^0_{2l+1}(h) = -\frac{1}{2}\alpha^0_{l+1}(h) + \frac{1}{2}\sum_{k=2}^{l} \alpha^0_k(h). \tag{6.31}$$

Consideration of a similar argument for the simple roots $\alpha^{B_l}_k$ of B_l then shows that on the Cartan subalgebra $\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$ of B_l the simple roots $\alpha^{B_l}_k$ of B_l and α^0_k of $A(2l/0)$ are related by

$$\alpha^0_1(h) = -\sum_{r=1}^{l} \alpha^{B_l}_r(h), \tag{6.32}$$

$$\alpha^0_k(h) = -\alpha^0_{k+l}(h) = \alpha^{B_l}_{k-1}(h) \quad (\text{for } k = 2,3,...,l), \tag{6.33}$$

$$\alpha^0_{l+1}(h) = \sum_{r=1}^{l-1} \alpha^{B_l}_r(h) + 2\alpha^{B_l}_l(h), \tag{6.34}$$

and

$$\alpha^0_{2l+1}(h) = -\alpha^{B_l}_l(h), \tag{6.35}$$

Finally it follows from (2.13), (2.17), (2.26)–(2.28), (2.36), (2.37), (6.3), (6.4), (6.15), and (6.16) that corresponding elements of the Cartan subalgebra of the Kac–Moody superalgebra are

$$h_{\alpha^{B_l}_k} = t^0 \otimes h^{B_l}_{\alpha^{B_l}_k} = \{2l/(2l-1)\}t^0 \otimes \{\mathbf{h}^0_{\alpha^0_{k+1}} - \mathbf{h}^0_{\alpha^0_{k+l+1}}\}$$
$$(\text{for } k = 1,2,...,l-1) \tag{6.36}$$

and

$$h_{\alpha^{B_l}_l} = t^0 \otimes h^{B_l}_{\alpha^{B_l}_l} = \{2l/(2l-1)\}t^0 \otimes \left\{\sum_{r=l+1}^{2l} \mathbf{h}^0_{\alpha^0_r}\right\}. \tag{6.37}$$

The *nondiagonal* basis elements of $\mathscr{T}^{0(4)}_{s0}$ will now be examined. They fall into six sets.

(i) For $j = 1,2,...,l$:

$$\mathbf{e}_{1,j+1} - \mathbf{e}_{j+l+1,1} = \mathbf{e}^0_{\beta^0_{(1,j+1)}} + \mathbf{e}^0_{-\beta^0_{(1,j+l+1)}}, \tag{6.38}$$

where $\mathbf{e}^0_{\beta^0_{(1,j+1)}}$ and $\mathbf{e}^0_{-\beta^0_{(1,j+l+1)}}$ are given by (2.19) and (6.7a). As (6.32) and (6.33) imply that

$$\beta^0_{(1,j+1)}(h) = -\beta^0_{(1,j+l+1)}(h) = -\sum_{r=j}^{l} \alpha^{B_l}_r(h) \tag{6.39}$$

[for $j = 1,2,...,l$, and for all $h \in \mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$], the basis element (6.41) corresponds to the root $\beta^0_{(1,j+1)}(h)$ of B_l.

(ii) For $j = 1,2,...,l$:

$$-\mathbf{e}_{j+1,1} + \mathbf{e}_{1,j+l+1} = \mathbf{e}^0_{-\beta^0_{(1,j+1)}} + \mathbf{e}^0_{\beta^0_{(1,j+l+1)}}, \tag{6.40}$$

which corresponds to the root $-\beta^0_{(1,j+1)}(h)$ of B_l, where $\beta^0_{(1,j+1)}(h)$ is given by (6.39).

(iii) For $j,k = 1,2,...,l$, with $j<k$:

$$\mathbf{e}_{j+1,k+1} - \mathbf{e}_{k+l+1,j+l+1} = \mathbf{e}^0_{\beta^0_{(j+1,k+1)}} + \mathbf{e}^0_{-\beta^0_{(j+l+1,k+l+1)}}. \tag{6.41}$$

where $\mathbf{e}^0_{\beta^0_{(j+1,k+1)}}$ and $\mathbf{e}^0_{-\beta^0_{(j+l+1,k+l+1)}}$ are given by (2.19) and (6.7a). As (6.33) implies that

$$\beta^0_{(j+1,k+1)}(h) = -\beta^0_{(j+l+1,k+l+1)}(h) = \sum_{r=j}^{k-1} \alpha^{B_l}_r(h) \tag{6.42}$$

[for $j,k = 1,2,...,l$, with $j<k$, and for all $h \in \mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$], the basis element (6.41) corresponds to the root $\beta^0_{(j+1,k+1)}(h)$ of B_l. (This set does not appear when $l = 1$).

(iv) For $j,k = 1,2,...,l$, with $j<k$:

$$-\mathbf{e}_{k+1,j+1} + \mathbf{e}_{j+l+1,k+l+1}$$
$$= \mathbf{e}^0_{-\beta^0_{(j+1,k+1)}} + \mathbf{e}^0_{\beta^0_{(j+l+1,k+l+1)}}, \tag{6.43}$$

which corresponds to the root $-\beta^0_{(j+1,k+1)}(h)$ of B_l, where $\beta^0_{(j+1,k+1)}(h)$ is as in (6.42). (This set does not appear when $l = 1$).

(v) For $j,k = 1,2,...,l$, with $j<k$:

$$\mathbf{e}_{j+1,k+l+1} - \mathbf{e}_{k+1,j+l+1} = \mathbf{e}^0_{\beta^0_{(j+1,k+l+1)}} - \mathbf{e}^0_{\beta^0_{(k+1,j+l+1)}}. \tag{6.44}$$

where $\mathbf{e}^0_{\beta^0_{(j+1,k+l+1)}}$ and $\mathbf{e}^0_{\beta^0_{(k+1,j+l+1)}}$ are given by (2.19) and (6.7a). As (6.33) and (6.34) imply that

$$\beta^0_{(j+1,k+l+1)}(h) = \beta^0_{(k+1,j+l+1)}(h) = \sum_{r=j}^{k-1} \alpha_r^{B_l}(h) + 2\sum_{r=k}^{l} \alpha_r^{B_l}(h) \tag{6.45}$$

[for $j,k = 1,2,\ldots,l$, with $j<k$, and for all $h \in \mathscr{H}^{0(4)} \times (=\mathscr{H}^{B_l})$], the basis element (6.41) corresponds to the root $\beta^0_{(j+1,k+l+1)}(h)$ of B_l. (This set does not appear when $l=1$).

(vi) For $j,k = 1,2,\ldots,l$, with $j<k$:

$$-\mathbf{e}_{k+l+1,j+1} + \mathbf{e}_{j+l+1,k+1} = \mathbf{e}^0_{-\beta^0_{(j+1,k+l+1)}} - \mathbf{e}^0_{-\beta^0_{(k+1,j+l+1)}}. \tag{6.46}$$

This corresponds to the root $-\beta^0_{(j+1,k+l+1)}(h)$ of B_l, where $\beta^0_{(j+1,k+1)}(h)$ is given by (6.45). (This set does not appear when $l=1$).

As expected the elements of (6.38), (6.40), (6.41), (6.43), (6.44), and (6.46) are *even* members of $A(2l/0)$.

It is easily checked that the set of $2l^2$ nonzero roots of (i) to (vi) above, together with the l zero roots, are all weights of the adjoint representation of B_l. For $l \geqslant 2$ its highest weight is

$$\Lambda = \Lambda_2^{B_l} = \alpha_1^{B_l} + 2\sum_{k=2}^{l} \alpha_k^{B_l}, \tag{6.47}$$

while for $l=1$ it is

$$\Lambda = 2\Lambda_1^{B_l} = \alpha_1^{B_l} \tag{6.48}$$

as expected [cf. Cornwell[17] (Appendix F)].

2. $\mathscr{T}^{0(4)}_{s1}$

By (2.33) the subspace $\mathscr{T}^{0(4)}_{s1}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G} = i\mathbf{A}, \quad -\tilde{\mathbf{B}}\mathbf{G} = i\mathbf{C}, \quad \mathbf{G}\tilde{\mathbf{C}} = i\mathbf{B}, \text{ and } -\tilde{\mathbf{D}} = i\mathbf{D},$$

which when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$\mathbf{A} = \mathbf{O}, \quad \mathbf{D} = \mathbf{O}, \text{ and } \mathbf{C} = i\tilde{\mathbf{B}}\mathbf{G}. \tag{6.49}$$

The basis elements of $\mathscr{T}^{0(4)}_{s1}$ fall into three sets.

(i) For $j = 1,2,\ldots,l$:

$$\mathbf{e}_{j+1,2l+2} + i\mathbf{e}_{2l+2,j+l+1} = \mathbf{e}^0_{\delta^0_{(j+1)}} - i\mathbf{e}^0_{-\delta^0_{(j+l+1)}}, \tag{6.50}$$

where $\mathbf{e}^0_{\delta^0_{(j+1)}}$ and $\mathbf{e}^0_{-\delta^0_{(j+l+1)}}$ are given by (2.19) and (6.7b). As (6.33), (6.34), and (6.35) imply that for $j = 1,2,\ldots,l$ and for all $h \in \mathscr{H}^{0(4)} (= \mathscr{H}^{B_l})$

$$\delta^0_{(j+1)}(h) = -\delta^0_{(j+l+1)}(h) = \sum_{r=j}^{l} \alpha_r^{B_l}(h), \tag{6.51}$$

the basis element (6.50) corresponds to the weight $\delta^0_{(j+1)}(h)$ of B_l.

(ii) For $j = 1,2,\ldots,l$:

$$-\mathbf{e}_{2l+2,j+1} + i\mathbf{e}_{j+l+1,2l+2} = \mathbf{e}^0_{-\delta^0_{(j+1)}} + i\mathbf{e}^0_{\delta^0_{(j+l+1)}}, \tag{6.52}$$

which corresponds to the weight $-\delta^0_{(j+1)}(h)$ of B_l, where $\delta^0_{(j+1)}(h)$ is given by (6.51).

(iii) The single basis element:

$$\mathbf{e}_{1,2l+2} + i\mathbf{e}_{2l+2,1} = \mathbf{e}^0_{\delta^0_{(1)}} - i\mathbf{e}^0_{-\delta^0_{(1)}}, \tag{6.53}$$

where $\mathbf{e}^0_{\delta^0_{(1)}}$ and $i\mathbf{e}^0_{-\delta^0_{(1)}}$ are given by (2.19). However, by (6.29)–(6.31) $\delta^0_{(1)}(h) = 0$ for all $h \in \mathscr{H}^{0(4)} (= \mathscr{H}^{B_l})$, so (6.53) corresponds to a zero weight of B_l.

These weights all belong to a $(2l+1)$-dimensional irreducible representation of B_l with highest weight

$$\Lambda = \Lambda_1^{B_l} = \sum_{k=1}^{l} \alpha_k^{B_l}. \tag{6.54}$$

It should be noted that all the elements of (6.50), (6.52), and (6.53) are odd members of $A(2l/0)$, so all the elements of $\mathscr{T}^{0(4)}_{s1}$ are odd.

3. $\mathscr{T}^{0(4)}_{s2}$

By (2.33) the subalgebra $\mathscr{T}^{0(4)}_{s2}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G} = -\mathbf{A}, \quad -\tilde{\mathbf{B}}\mathbf{G} = -\mathbf{C},$$
$$\mathbf{G}\tilde{\mathbf{C}} = -\mathbf{B}, \text{ and } -\tilde{\mathbf{D}} = -\mathbf{D},$$

which when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$\tilde{\mathbf{A}}\mathbf{G} - \mathbf{G}\mathbf{A} = \mathbf{0}, \quad \mathbf{B} = \mathbf{0}, \quad \mathbf{C} = \mathbf{0}, \tag{6.55}$$

with $\mathbf{D}$ being determined only by the supertrace condition $\operatorname{tr}\mathbf{A} = \operatorname{tr}\mathbf{D}$. On using (6.3) and (6.4), the diagonal basis elements of $\mathscr{T}^{0(4)}_{s2}$ may be taken to consist of two sets.

(i) The single basis element

$$(1/4l)\{\mathbf{e}_{1,1} + \mathbf{e}_{2l+2,2l+2}\} = \sum_{r=1}^{2l+1} \mathbf{h}^0_{\alpha^0_r}. \tag{6.56}$$

(ii) For $k = 1,2,\ldots,l$:

$$(1/4l)\{\mathbf{e}_{k+1,k+1} + \mathbf{e}_{k+l+1,k+l+1} + 2\mathbf{e}_{2l+2,2l+2}\} = \sum_{r=k+1}^{k+l} \mathbf{h}^0_{\alpha^0_r} + 2\sum_{r=k+l+1}^{2l+1} \mathbf{h}^0_{\alpha^0_r}. \tag{6.57}$$

Each of these corresponds to zero weight of B_l, so that the zero weight has multiplicity $l+1$.

The nondiagonal basis elements of $\mathscr{T}^{0(4)}_{s2}$ fall into eight sets.

(i) For $j = 1,2,\ldots,l$:

$$\mathbf{e}_{1,j+1} + \mathbf{e}_{j+l+1,1} = \mathbf{e}^0_{\beta^0_{(1,j+1)}} - \mathbf{e}^0_{-\beta^0_{(1,j+l+1)}}, \tag{6.58}$$

where $\mathbf{e}^0_{\beta^0_{(1,j+1)}}$ and $\mathbf{e}^0_{-\beta^0_{(1,j+l+1)}}$ are given by (2.19) and (6.7a), and $\beta^0_{(1,j+1)}(h)\,[=-\beta^0_{(1,j+l+1)}(h)]$ is given for all $h \in \mathscr{H}^{0(4)} (= \mathscr{H}^{B_l})$ by (6.39), so this basis element (6.58) again corresponds to the weight $\beta^0_{(1,j+1)}(h)$ of B_l.

(ii) For $j = 1,2,\ldots,l$:

$$-\mathbf{e}_{j+1,1} - \mathbf{e}_{1,j+l+1} = \mathbf{e}^0_{-\beta^0_{(1,j+1)}} - \mathbf{e}^0_{\beta^0_{(1,j+l+1)}}, \tag{6.59}$$

which corresponds to the weight $-\beta^0_{(1,j+1)}(h)$ of B_l, where $\beta^0_{(1,j+1)}(h)$ is given by (6.39).

(iii) For $j,k = 1,2,\ldots,l$, with $j<k$:

$$\mathbf{e}_{j+1,k+1} + \mathbf{e}_{k+l+1,j+l+1} = \mathbf{e}^0_{\beta^0_{(j+1,k+1)}} - \mathbf{e}^0_{-\beta^0_{(j+l+1,k+l+1)}}, \tag{6.60}$$

where $\mathbf{e}^0_{\beta^0_{(j+1,k+1)}}$ and $\mathbf{e}^0_{-\beta^0_{(j+l+1,k+l+1)}}$ are given by

(2.19) and (6.7a), and $\beta^0_{(j+1,k+1)}(h)$ $[=-\beta^0_{(j+l+1,k+l+1)}(h)]$ is given for all $h\in\mathscr{H}^{0(4)}[=\mathscr{H}^{B_l}]$ by (6.42), so this basis element (6.60) again corresponds to the weight $\beta^0_{(j+1,k+1)}(h)$ of B_l. (This set does not appear when $l=1$.)

(iv) For $j,k=1,2,\ldots,l$, with $j<k$:

$$-\mathbf{e}_{k+1,j+1}-\mathbf{e}_{j+l+1,k+l+1}=\mathbf{e}^0_{-\beta^0_{(j+1,k+1)}}-\mathbf{e}^0_{\beta^0_{(j+l+1,k+l+1)}}. \tag{6.61}$$

This corresponds to the weight $-\beta^0_{(j+1,k+1)}(h)$ of B_l, where $\beta^0_{(j+1,k+1)}(h)$ is given by (6.42). (This set does not appear when $l=1$.)

(v) For $j,k=1,2,\ldots,l$, with $j<k$:

$$\mathbf{e}_{j+1,k+l+1}+\mathbf{e}_{k+1,j+l+1}=\mathbf{e}^0_{\beta^0_{(j+1,k+l+1)}}+\mathbf{e}^0_{\beta^0_{(k+1,j+l+1)}}. \tag{6.62}$$

where $\mathbf{e}^0_{\beta^0_{(j+1,k+l+1)}}$ and $\mathbf{e}^0_{\beta^0_{(k+1,j+l+1)}}$ are given by (2.19) and (6.7a), and $\beta^0_{(j+1,k+l+1)}(h)[=\beta^0_{(k+1,j+l+1)}(h)]$ is given for all $h\in\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$ by (6.45), so this basis element (6.62) again corresponds to the weight $\beta^0_{(j+1,k+l+1)}(h)$ of B_l. (This set does not appear when $l=1$.)

(vi) For $j,k=1,2,\ldots,l$, with $j<k$:

$$-\mathbf{e}_{k+l+1,j+1}-\mathbf{e}_{j+l+1,k+1}=\mathbf{e}^0_{-\beta^0_{(j+1,k+l+1)}}+\mathbf{e}^0_{-\beta^0_{(k+1,j+l+1)}}. \tag{6.63}$$

This corresponds to the weight $-\beta^0_{(j+1,k+l+1)}(h)$ of B_l, where $\beta^0_{(j+1,k+l+1)}$ is given by (6.45). (This set does not appear when $l=1$.)

(vii) For $j=1,2,\ldots,l$:

$$\mathbf{e}_{j+1,j+l+1}=\mathbf{e}^0_{\beta^0_{(j+1,j+l+1)}}, \tag{6.64}$$

where $\mathbf{e}^0_{\beta^0_{(j+1,j+l+1)}}$ is given by (2.19) and (6.7a). Thus the basis element (6.64) corresponds to the weight $\beta^0_{(j+1,j+l+1)}(h)$ of B_l, where (6.33) and (6.34) imply that

$$\beta^0_{(j+1,j+l+1)}(h)=2\sum_{r=j}^{l}\alpha_r^{B_l}(h), \tag{6.65}$$

[for $j=1,2,\ldots,l$, and for all $h\in\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$].

(viii) For $j=1,2,\ldots,l$:

$$-\mathbf{e}_{j+l+1,j+1}=\mathbf{e}^0_{-\beta^0_{(j+1,j+l+1)}}, \tag{6.66}$$

which corresponds to the weight $-\beta^0_{(j+1,j+l+1)}(h)$ of B_l, where $\beta^0_{(j+1,j+l+1)}(h)$ is given by (6.65).

These $2l^2+3l+1$ weights belong to a representation of B_l which is the direct sum of the trivial 1-dimensional irreducible representation with highest weight $\Lambda=0$ and the $(2l^2+3l)$-dimensional irreducible representation with highest weight

$$\Lambda=2\Lambda_1^{B_l}=2\sum_{k=1}^{l}\alpha_l^{B_l}. \tag{6.67}$$

It should be noted that all the elements of $\mathscr{L}^{0(4)}_{s2}$ are even members of $A(2l/0)$.

4. $\mathscr{L}^{0(4)}_{s3}$

By (2.33) the subspace $\mathscr{L}^{0(4)}_{s3}$ consists of the supermatrices whose submatrices satisfy the conditions

$$-\mathbf{G}\tilde{\mathbf{A}}\mathbf{G}=-i\mathbf{A},\quad -\tilde{\mathbf{B}}\mathbf{G}=-i\mathbf{C},$$

$$\mathbf{G}\tilde{\mathbf{C}}=-i\mathbf{B},\text{ and }-\tilde{\mathbf{D}}=-i\mathbf{D},$$

which, when taken together, along with the fact that $\mathbf{D}$ is 1×1, imply that

$$\mathbf{A}=0,\quad \mathbf{D}=0,\text{ and }\mathbf{C}=-i\tilde{\mathbf{B}}\mathbf{G}. \tag{6.68}$$

The basis elements of $\mathscr{L}^{0(4)}_{s3}$ fall into three sets.

(i) For $j=1,2,\ldots,l$:

$$\mathbf{e}_{j+1,2l+2}-i\mathbf{e}_{2l+2,j+l+1}=\mathbf{e}^0_{\delta^0_{(j+1)}}+i\mathbf{e}^0_{-\delta^0_{(j+l+1)}}, \tag{6.69}$$

where $\mathbf{e}^0_{\delta^0_{(j+1)}}$ and $\mathbf{e}^0_{-\delta^0_{(j+l+1)}}$ are given by (2.19) and (6.7b), and $\delta^0_{(j+1)}(h)[=-\delta^0_{(j+l+1)}(h)]$ is given for all $h\in\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$ by (6.51), so this basis element (6.69) again corresponds to the weight $\delta^0_{(j+1)}(h)$ of B_l.

(ii) For $j=1,2,\ldots,l$:

$$-\mathbf{e}_{2l+2,j+1}-i\mathbf{e}_{j+l+1,2l+2}=\mathbf{e}^0_{-\delta^0_{(j+1)}}-i\mathbf{e}^0_{\delta^0_{(j+l+1)}}, \tag{6.70}$$

which corresponds to the root $-\delta^0_{(j+1)}(h)$ of B_l, where $\delta^0_{(j+1)}(h)$ is given by (6.51).

(iii) The single basis element:

$$\mathbf{e}_{1,2l+2}-i\mathbf{e}_{2l+2,1}=\mathbf{e}^0_{\delta^0_{(1)}}+i\mathbf{e}^0_{-\delta^0_{(1)}}, \tag{6.71}$$

where $\mathbf{e}^0_{\delta^0_{(1)}}$ is given by (2.19) and (6.6b). As $\delta^0_{(1)}(h)=0$ for all $h\in\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$, (6.71) corresponds to a zero weight of B_l.

These weights all belong to a $(2l+1)$-dimensional irreducible representation of B_l whose highest weight is given by (6.54). All the elements of $\mathscr{L}^{0(4)}_{s3}$ are odd.

C. The roots of $A^{(4)}(2l/0)$

Defining $\delta(h)$ as in (3.6) to (3.8), it follows that the roots $\alpha(h)$ and the corresponding basis elements e_α of $A^{(4)}(2l/0)$ are as follows.

(i) $\alpha=4J\delta(h)$, (for $J=0,\pm1,\pm2,\ldots$). There are l linearly independent basis elements $e^{(k)}_\alpha$ corresponding to this root that may be labeled by an additional superscript, so that

$$e^{(k)}_\alpha=\{2l/(2l-1)\}t^{4J}\otimes\{\mathbf{h}^0_{\alpha^0_{k+1}}-\mathbf{h}^0_{\alpha^0_{k+l+1}}\}$$

(for $k=1,2,\ldots,l-1$),

and

$$e^{(l)}_\alpha=\frac{2l}{(2l-1)}t^{4J}\otimes\left\{\sum_{r=l+1}^{2l}\mathbf{h}^0_{\alpha^0_r}\right\}$$

[which reduce to (6.36) and (6.37) in the special case $J=0$].

(ii) $\alpha(h)=4J\delta(h)\pm\beta^0_{(1,j+1)}(h)$ (for $j=1,2,\ldots,l$, and for $J=0,\pm1,\pm2,\ldots$), where $\beta^0_{(1,j+1)}(h)$ is the extension of the weight of B_l that is given by (6.39) and

$$e_\alpha=t^{4J}\otimes\{\mathbf{e}^0_{\pm\beta^0_{(1,j+1)}}+\mathbf{e}^0_{\mp\beta^0_{(1,j+l+1)}}\}.$$

(iii) $\alpha(h)=4J\delta(h)\pm\beta^0_{(j+1,k+1)}(h)$, (for $j,k=1,2,\ldots,l$, with $j<k$, and for $J=0,\pm1,\pm2,\ldots$), where $\beta^0_{(j+1,k+1)}(h)$ is the extension of the weight of B_l that is given by (6.42) and

$$e_\alpha = t^{4J} \otimes \{\mathbf{e}^0_{\pm\beta^0_{(j+1,k+1)}} + \mathbf{e}^0_{\mp\beta^0_{(j+l+1,k+l+1)}}\}.$$

(iv) $\alpha(h) = 4J\delta(h) \pm \beta^0_{(j+1,k+l+1)}(h)$, (for $j,k = 1,2,...,l$, with $j<k$, and for $J = 0, \pm 1, \pm 2,...$), where $\beta^0_{(j+1,k+l+1)}(h)$ is the extension of the weight of B_l that is given by (6.45) and

$$e_\alpha = t^{4J} \otimes \{\mathbf{e}^0_{\pm\beta^0_{(j+1,k+l+1)}} - \mathbf{e}^0_{\pm\beta^0_{(k+1,j+l+1)}}\}.$$

(v) $\alpha(h) = (4J+1)\delta(h) \pm \delta^0_{(j+1)}(h)$ (for $j = 1,2,...,l$, and for $J = 0, \pm 1, \pm 2,...$), where $\delta^0_{(j+1)}(h)$ is the extension of the weight of B_l that is given by (6.51) and

$$e_\alpha = t^{4J+1} \otimes \{\mathbf{e}^0_{\pm\delta^0_{(j+1)}} \mp i\mathbf{e}^0_{\mp\delta^0_{(j+l+1)}}\}.$$

(vi) $\alpha(h) = (4J+1)\delta(h)$ (for $J = 0, \pm 1, \pm 2,...$), with

$$e_\alpha = t^{4J+1} \otimes \{\mathbf{e}^0_{\delta^0_{(1)}} - i\mathbf{e}^0_{-\delta^0_{(1)}}\}.$$

(vii) $\alpha(h) = (4J+2)\delta(h)$ (for $J = 0, \pm 1, \pm 2,...$). There are $l+1$ linearly independent basis elements $e^{(k)}_\alpha$ corresponding to this root which may be labeled by an additional superscript, so that

$$e^{(k)}_\alpha = t^{4J+2} \otimes \left\{ \sum_{r=k+1}^{k+l} \mathbf{h}^0_{\alpha^0_r} + 2 \sum_{r=k+l+1}^{2l+1} \mathbf{h}^0_{\alpha^0_r} \right\}$$

(for $k = 1,2,...,l$)

and

$$e^{(l+1)}_\alpha = t^{4J+2} \otimes \sum_{r=1}^{2l+1} \mathbf{h}^0_{\alpha^0_r}.$$

(viii) $\alpha(h) = (4J+2)\delta(h) \pm \beta^0_{(1,j+1)}(h)$ (for $j = 1,2,...,l$, and for $J = 0, \pm 1, \pm 2,...$), where $\beta^0_{(1,j+1)}(h)$ is the extension of the weight of B_l that is given by (6.39) and

$$e_\alpha = t^{4J+2} \otimes \{\mathbf{e}^0_{\pm\beta^0_{(1,j+1)}} - \mathbf{e}^0_{\mp\beta^0_{(1,j+l+1)}}\}.$$

(ix) $\alpha(h) = (4J+2)\delta(h) \pm \beta^0_{(j+1,k+1)}(h)$ (for $j,k = 1,2,...,l$, with $j<k$, and for $J = 0, \pm 1, \pm 2,...$), where $\beta^0_{(j+1,k+1)}(h)$ is the extension of the weight of B_l that is given by (6.42) and

$$e_\alpha = t^{4J+2} \otimes \{\mathbf{e}^0_{\pm\beta^0_{(j+1,k+1)}} - \mathbf{e}^0_{\mp\beta^0_{(j+l+1,k+l+1)}}\}.$$

(x) $\alpha(h) = (4J+2)\delta(h) \pm \beta^0_{(j+1,k+l+1)}(h)$ (for $j,k = 1,2,...,l$, with $j<k$, and for $J = 0, \pm 1, \pm 2,...$), where $\beta^0_{(j+1,k+l+1)}(h)$ is the extension of the weight of B_l that is given by (6.45) and

$$e_\alpha = t^{4J+2} \otimes \{\mathbf{e}^0_{\pm\beta^0_{(j+1,k+l+1)}} + \mathbf{e}^0_{\pm\beta^0_{(k+1,j+l+1)}}\}.$$

(xi) $\alpha(h) = (4J+2)\delta(h) \pm \beta^0_{(j+1,j+l+1)}(h)$ (for $j = 1,2,...,l$, and for $J = 0, \pm 1, \pm 2,...$), where $\beta^0_{(j+1,j+l+1)}(h)$ is the extension of the weight of B_l that is given by (6.65) and

$$e_\alpha = t^{4J+2} \otimes \mathbf{e}^0_{\pm\beta^0_{(j+1,j+l+1)}}.$$

(xii) $\alpha(h) = (4J+3)\delta(h) \pm \delta^0_{(j+1)}(h)$ (for $j = 1,2,...,l$, and for $J = 0, \pm 1, \pm 2,...$), where $\delta^0_{(j+1)}(h)$ is the extension of the weight of B_l that is given by (6.51) and

$$e_\alpha = t^{4J+3} \otimes \{\mathbf{e}^0_{\pm\delta^0_{(j+1)}} \pm i\mathbf{e}^0_{\mp\delta^0_{(j+l+1)}}\}.$$

(xiii) $\alpha(h) = (4J+3)\delta(h)$ (for $J = 0, \pm 1, \pm 2,...$), with

$$e_\alpha = t^{4J+3} \otimes \{\mathbf{e}^0_{\delta^0_{(1)}} + i\mathbf{e}^0_{-\delta^0_{(1)}}\}.$$

(xiv) $\alpha(h) = 0$, with c and d as basis elements.

With μ chosen as in (6.13) [so that (2.37) is valid], it follows that

$$\langle \alpha^0, \beta^0 \rangle = \langle \alpha^0, \beta^0 \rangle^{B_l}, \tag{6.72}$$

where on the right-hand side of (6.72) α^0 and β^0 are any pair of linear functionals defined on $\mathscr{H}^{0(4)}(=\mathscr{H}^{B_l})$, the evaluation being performed with respect to the Killing form of B_l, and where on the left-hand side of (6.72) α^0 and β^0 denote the corresponding extensions to the Cartan subalgebra of the Kac–Moody superalgebra $A^{(4)}(2l/0)$, the evaluation being performed with respect to its supersymmetric bilinear invariant form $B(,)$. As B_l is a simple Lie algebra, $\langle \alpha^0, \alpha^0 \rangle^{B_l} > 0$ for every nonzero linear functional α^0 defined on $\mathscr{H}^{B_l}$, so $\langle \alpha^0, \alpha^0 \rangle > 0$ for the corresponding extension. Moreover (2.13), (2.28), (2.30), (2.31), (2.36), and (6.13) imply that

$$h_\delta = \{4l/(2l-1)\}c. \tag{6.73}$$

Thus if α^0_k is the extension of any simple root of $\mathscr{L}^0_s$, then

$$\langle \delta, \alpha^0_k \rangle = 0 \tag{6.74}$$

and

$$\langle j\delta, j\delta \rangle = 0. \tag{6.75}$$

Thus $\langle j\delta, j\delta \rangle = 0$ for integer j, so every nonzero root of $A^{(4)}(2l/0)$ belonging to the sets (i), (vi), (vii), and (xiii) is "imaginary." Moreover, because $\langle j\delta + \alpha^0, j\delta + \alpha^0 \rangle = \langle \alpha^0, \alpha^0 \rangle^{B_l}$ and because $\langle \alpha^0, \alpha^0 \rangle^{B_l} > 0$ for linear functional α^0 and its corresponding extension (as has just been noted), it follows that every root of $A^{(4)}(2l/0)$ belonging to the sets (ii), (iii), (iv), (v), (viii), (ix), (x), (xi), and (xii) is "real." All the elements mentioned in the above sets are even, except for those in the sets (v), (vi), (xii), and (xiii), which are odd.

For $A^{(4)}(2l/0)$ for $l \geqslant 1$ the simple roots may be taken to be

$$\alpha_l = \delta - \alpha^0_H, \tag{6.76}$$

and

$$\alpha_k = \alpha^{B_l}_{l-k} \quad (\text{for } k = 0,1,...,l-1), \tag{6.77}$$

where

$$\alpha^0_H = \Lambda_1^{B_l} = \sum_{k=1}^{l} \alpha_k^{B_l} \tag{6.78}$$

is the highest weight of the representation of $\mathscr{L}^{0(4)}_{s0}$ for which $\mathscr{L}^{0(4)}_{s1}$ is the carrier space [cf. (6.54)] and the $\alpha_k^{B_l}$ are the extensions of the simple roots of B_l. As e_{α_l} appears in the set (v) it follows that e_{α_l} is odd, so α_l is an odd root of the Kac–Moody superalgebra $A^{(4)}(2l/0)$. All the other simple roots of $A^{(4)}(2l/0)$ are even.

It is then easily checked that the Cartan matrices of $A^{(4)}(2/0)$ and $A^{(4)}(2l/0)$ (for $l \geqslant 2$) evaluated using (2.15) correspond to the generalized Dynkin diagrams given in Figs. 5 and 6, respectively.

VII. TWISTED AFFINE KAC–MOODY SUPERALGEBRAS CORRESPONDING TO AUTOMORPHISMS OF ORDER 2: ROOTS OF $C^{(2)}(l+1)$ (FOR $l=1,2,3,...$)

A. The two-fold automorphisms

An explicit realization of the simple Lie superalgebra $C(l+1)$ is provided by the orthosymplectic algebra $\mathrm{osp}(2/2l;\mathbb{C})$, considered as a complex superalgebra, where $\mathrm{osp}(2/2l;\mathbb{C})$ is defined as the set of $(2l+2)\times(2l+2)$ complex supermatrices with the grading partitioning

$$\mathbf{M}=\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}, \tag{7.1}$$

that are subject to the condition that

$$\mathbf{M}^{st}\mathbf{K}+(-1)^{\deg \mathbf{M}}\mathbf{K}\mathbf{M}=\mathbf{0}, \tag{7.2}$$

where

$$\mathbf{K}=\begin{pmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{J}\end{pmatrix}, \tag{7.3}$$

with

$$\mathbf{G}=\begin{pmatrix}\mathbf{0} & \mathbf{1}_l\\ \mathbf{1}_l & \mathbf{0}\end{pmatrix}, \tag{7.4}$$

and

$$\mathbf{J}=\begin{pmatrix}\mathbf{0} & \mathbf{1}_l\\ -\mathbf{1}_l & \mathbf{0}\end{pmatrix}, \tag{7.5}$$

[Here **A**, **B**, **C**, **D**, **K**, **G**, and **J** are of dimensions 2×2, $2\times2l$, $2l\times2$, $2l\times2l$, $(2l+2)\times(2l+2)$, 2×2, and $2l\times2l$, respectively.] The condition (7.2) implies that

$$\tilde{\mathbf{A}}\mathbf{G}+\mathbf{G}\mathbf{A}=\mathbf{0}, \tag{7.6}$$

$$\tilde{\mathbf{D}}\mathbf{J}+\mathbf{J}\mathbf{D}=\mathbf{0}, \tag{7.7}$$

and

$$\tilde{\mathbf{B}}\mathbf{G}-\mathbf{J}\mathbf{C}=\mathbf{0}. \tag{7.8}$$

The rank l^0 of $C(l+1)$ is given by

$$l^0=l+1.$$

The generalized Dynkin diagram of $C(l+1)$ is shown in Fig. 12, which indicates that its distinguished simple roots α_k^0 are even for $k=2,3,...,l+1$, but that α_1^0 is odd. With the bilinear form $B^0(,)$ being defined by

$$B^0(\mathbf{M},\mathbf{N})=-2l\,\mathrm{str}\,(\mathbf{MN}), \tag{7.9}$$

the basis elements of its Cartan subalgebra $\mathscr{H}^0$ may be taken to be

$$\mathbf{h}^0_{\alpha_1^0}=-\{1/4l\}\{\mathbf{e}_{1,1}-\mathbf{e}_{2,2}+\mathbf{e}_{3,3}-\mathbf{e}_{l+3,l+3}\}, \tag{7.10}$$

$$\mathbf{h}^0_{\alpha_k^0}=\{1/4l\}\{\mathbf{e}_{k+1,k+1}-\mathbf{e}_{k+l+1,k+l+1}-\mathbf{e}_{k+2,k+2}+\mathbf{e}_{k+l+2,k+l+2}\} \quad (\text{for } k=2,3,...,l) \tag{7.11}$$

and

$$\mathbf{h}^0_{\alpha_{l+1}^0}=\{1/2l\}\{\mathbf{e}_{l+2,l+2}-\mathbf{e}_{2l+2,2l+2}\}. \tag{7.12}$$

Again $\mathbf{e}_{r,s}$ is the matrix of dimension $(2l+2)\times(2l+2)$ that is defined by (6.5), so that with this choice all the matrices of $\mathscr{H}^0$ are again *diagonal*. The positive even roots $\beta^{0+}_{(j,k)}$ and $\beta^{0-}_{(j,k)}$ and positive odd roots $\delta^{0+}_{(j)}$ and $\delta^{0-}_{(j)}$ of $C(l+1)$ are given in terms of the distinguished set of simple roots α_1^0, $\alpha_2^0,...,\alpha_{l+1}^0$ of $C(l+1)$ by

$$\beta^{0-}_{(j,k)}=\sum_{r=j+1}^{k}\alpha_r^0 \quad (\text{for } j,k=1,2,...,l, j<k), \tag{7.13}$$

$$\beta^{0+}_{(j,k)}=\sum_{r=j+1}^{k}\alpha_r^0+2\sum_{r=k+1}^{l}\alpha_r^0+\alpha_{l+1}^0 \quad (\text{for } j,k=1,2,...,l-1, j<k), \tag{7.14}$$

$$\beta^{0+}_{(j,l)}=\sum_{r=j+1}^{l}\alpha_r^0+\alpha_{l+1}^0 \quad (\text{for } j=1,2,...,l-1), \tag{7.15}$$

$$\beta^{0+}_{(jj)}=2\sum_{r=j+1}^{l}\alpha_r^0+\alpha_{l+1}^0 \quad (\text{for } j=1,2,...,l-1), \tag{7.16}$$

$$\beta^{0+}_{(l,l)}=\alpha_{l+1}^0, \tag{7.17}$$

$$\delta^{0-}_{(j)}=\sum_{r=1}^{j}\alpha_r^0 \quad (\text{for } j=1,2,...,l), \tag{7.18}$$

$$\delta^{0+}_{(j)}=\sum_{r=1}^{j}\alpha_r^0+2\sum_{r=j+1}^{l}\alpha_r^0+\alpha_{l+1}^0 \quad (\text{for } j=1,2,...,l-1), \tag{7.19}$$

and

$$\delta^{0+}_{(l)}=\sum_{r=1}^{l+1}\alpha_r^0. \tag{7.20}$$

The corresponding basis elements of $C(l+1)$ may be taken to be

$$\mathbf{e}^0_{\beta^{0-}_{(j,k)}}=\mathbf{e}_{j+2,k+2}-\mathbf{e}_{k+l+2,j+l+2} \quad (\text{for } j,k=1,2,...,l; j<k), \tag{7.21}$$

$$\mathbf{e}^0_{\beta^{0+}_{(j,k)}}=\mathbf{e}_{j+2,k+l+2}+\mathbf{e}_{k+2,j+l+2} \quad (\text{for } j,k=1,2,...,l; j\leqslant k), \tag{7.22}$$

$$\mathbf{e}^0_{\delta^{0-}_{(j)}}=\mathbf{e}_{1,j+2}+\mathbf{e}_{j+l+2,2} \quad (\text{for } j=1,2,...,l), \tag{7.23}$$

and

$$\mathbf{e}^0_{\delta^{0+}_{(j)}}=\mathbf{e}_{1,j+l+2}-\mathbf{e}_{j+2,2} \quad (\text{for } j=1,2,...,l). \tag{7.24}$$

The basis elements corresponding to the corresponding negative roots may be chosen in accordance with (2.19). [For further information on $C(l+1)$ see Cornwell.[9]]

Taking the node corresponding to the odd simple root α_0 of $C^{(2)}(l+1)$, as the corresponding numerical mark has value 1 (cf. Figs. 7 and 8], (2.32) shows that $q=2$. Moreover, inspection of Figs. 7 and 8 shows that the generalized Dynkin diagram with the chosen node and attached lines removed corresponds to $B(0/l)$, the subalgebra $\mathscr{L}^{0(2)}_{s0}$ has to be isomorphic to $B(0/l)$.

The complex simple Lie superalgebra $B(0/l)$ may be realized as $\mathrm{osp}(1/2l;\mathbb{C})$, which is the set of $(2l+1)\times(2l+1)$ supermatrices $\mathbf{m}$ of the form

$$\mathbf{m}=\begin{pmatrix}\mathbf{0} & \mathbf{b}\\ \mathbf{c} & \mathbf{D}\end{pmatrix}, \tag{7.25}$$

where $\mathbf{b}$ and $\mathbf{c}$ are submatrices of dimensions $1\times2l$ and $2l\times1$, respectively, that experience the constraint

$$\tilde{\mathbf{b}} - \mathbf{J}\mathbf{c} = \mathbf{0}, \tag{7.26}$$

and $\mathbf{D}$ is a $2l \times 2l$ submatrix such that

$$\tilde{\mathbf{D}}\mathbf{J} + \mathbf{J}\mathbf{D} = \mathbf{0} \tag{7.27}$$

[$\mathbf{J}$ being defined in (7.5)]. This will be called the canonical form of $B(0/l)$.

One possible twofold automorphism of $C(l+1)$ is provided by ψ^2, where ψ is the automorphism of (4.6). However, as

$$\psi^2(\mathbf{M}) = -(-\mathbf{M}^{st})^{st} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ -\mathbf{C} & \mathbf{D} \end{pmatrix},$$

it follows from (2.33) that if this automorphism is employed then the subalgebra $\mathscr{L}_{s0}^{0(2)}$ would consist of the supermatrices with $\mathbf{B} = \mathbf{C} = \mathbf{0}$, and with $\mathbf{A}$ and $\mathbf{D}$ satisfying (7.6) and (7.7), respectively, so that the subalgebra $\mathscr{L}_{s0}^{0(2)}$ would be isomorphic to the even part of $C(l+1)$, and not to the superalgebra $B(0/l)$. Consequently, ψ^2 is *not* an appropriate choice of automorphism.

As will be demonstrated explicitly in the next subsection the correct choice is actually given by

$$\phi(\mathbf{M}) = \mathbf{L}^{-1}(-\mathbf{M}^{st})\mathbf{L}, \tag{7.28}$$

where

$$\mathbf{L} = \begin{pmatrix} \mathbf{1}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{pmatrix},$$

$\mathbf{J}$ being as in (7.5), so that

$$\phi(\mathbf{M}) = \begin{pmatrix} -\tilde{\mathbf{A}} & \tilde{\mathbf{C}}\mathbf{J} \\ -\mathbf{J}^{-1}\tilde{\mathbf{B}} & -\mathbf{J}^{-1}\tilde{\mathbf{D}}\mathbf{J} \end{pmatrix}. \tag{7.29}$$

It is easily checked that this provides a twofold automorphism of $C(l+1)$.

B. The subspaces $\mathscr{L}_{sp}^{0(2)}$ (for $p=0,1$)

The two subspaces $\mathscr{L}_{sp}^{0(2)}$ (for $p = 0,1$) corresponding to the automorphism ϕ of (7.28) will now be considered in turn.

1. $\mathscr{L}_{s0}^{0(2)}$

By (2.33) the subalgebra $\mathscr{L}_{s0}^{0(2)}$ consists of the supermatrices whose submatrices satisfy the conditions $-\tilde{\mathbf{A}} = \mathbf{A}$ and $\tilde{\mathbf{C}}\mathbf{J} = \mathbf{B}$ in addition to (7.6), (7.7), and (7.8). Together these imply that $\mathbf{A} = \mathbf{0}$ and that

$$\mathbf{B} = 2^{-1/2}\begin{pmatrix} -\mathbf{b} \\ \mathbf{b} \end{pmatrix} \text{ and } \mathbf{C} = 2^{-1/2}(-\mathbf{c} \quad \mathbf{c}), \tag{7.30}$$

where $\mathbf{b}$ and $\mathbf{c}$ are submatrices of dimensions $1 \times 2l$ and $2l \times 1$, respectively, that experience the constraint (7.26). It is easily checked that subject to these conditions the mapping

$$\Psi\left(\begin{pmatrix} 0 & \mathbf{b} \\ \mathbf{c} & \mathbf{D} \end{pmatrix}\right) = \begin{pmatrix} 0 & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{7.31}$$

is an isomorphic mapping of $B(0/l)$ onto $\mathscr{L}_{s0}^{0(2)}$ [the factors of $2^{-1/2}$ in (7.30) being inserted to help give this result].

Some properties of the canonical form of $B(0/l)$ and its image under the mapping (7.31) will first be summarized (the conventions being those of Cornwell[9]). The Killing form $B^{B(0/l)}(,)$ is given by

$$B^{B(0/l)}(\mathbf{m},\mathbf{m}') = -(2l+1)\mathrm{str}(\mathbf{m}\mathbf{m}') \tag{7.32}$$

[for all $\mathbf{m}$ and $\mathbf{m}'$ of the canonical form of $B(0/l)$]. Then, by (7.31),

$$B^{B(0/l)}(\mathbf{m},\mathbf{m}') = -(2l+1)\mathrm{str}(\Psi(\mathbf{m})\Psi(\mathbf{m}')), \tag{7.33}$$

and so, by (7.9),

$$B^{0}(\Psi(\mathbf{m}),\Psi(\mathbf{m}')) = \{2l/(2l+1)\}B^{B(0/l)}(\mathbf{m},\mathbf{m}') \tag{7.34}$$

for all $\mathbf{m}$ and $\mathbf{m}'$ of the canonical form. This implies that (2.37) is satisfied if

$$\mu = (2l+1)/(2l). \tag{7.35}$$

Denoting the simple even roots of $B(0/l)$ by $\alpha_k^{B(0/l)}$ (for $k = 1,2,...,l-1$) and the simple odd root of $B(0/l)$ by $\alpha_l^{B(0/l)}$, the corresponding basis elements of the Cartan subalgebra $\mathscr{H}^{B(0/l)}$ of $B(0/l)$ defined by

$$B^{B(0/l)}(\mathbf{h}_{\alpha_k^{B(0/l)}}^{B(0/l)},\mathbf{h}) = \alpha_k^{B(0/l)}(\mathbf{h}) \quad \text{for all } \mathbf{h} \in \mathscr{H}^{B(0/l)} \tag{7.36}$$

[cf. (2.13) and (2.17)] are

$$\mathbf{h}_{\alpha_k^{B(0/l)}}^{B(0/l)} = \{1/(2l+1)\}\{\mathbf{e}_{k+1,k+1} - \mathbf{e}_{k+l+1,k+l+1} - \mathbf{e}_{k+2,k+2} + \mathbf{e}_{k+l+2,k+l+2}\} \quad (\text{for } k = 1,2,...,l-1) \tag{7.37}$$

and

$$\mathbf{h}_{\alpha_l^{B(0/l)}}^{B(0/l)} = \{1/(2l+1)\}\{\mathbf{e}_{l+1,l+1} - \mathbf{e}_{2l+1,2l+1}\}. \tag{7.38}$$

Thus

$$\Psi(\mathbf{h}_{\alpha_k^{B(0/l)}}^{B(0/l)}) = \{1/(2l+1)\}\{\mathbf{e}_{k+2,k+2} - \mathbf{e}_{k+l+2,k+l+2} - \mathbf{e}_{k+3,k+3} + \mathbf{e}_{k+l+3,k+l+3}\} \quad (\text{for } k = 1,2,...,l-1) \tag{7.39}$$

and

$$\Psi(\mathbf{h}_{\alpha_l^{B(0/l)}}^{B(0/l)}) = \{1/(2l+1)\}\{\mathbf{e}_{l+2,l+2} - \mathbf{e}_{2l+2,2l+2}\}. \tag{7.40}$$

The *diagonal* basis elements of $\mathscr{L}_{s0}^{0(2)}$ will be considered first. As they may be taken to consist of the set $\{\mathbf{e}_{k+2,k+2} - \mathbf{e}_{k+l+2,k+l+2} \mid \text{for } k = 1,2,...,l\}$, it follows that they are all members of the Cartan subalgebra $\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$ of $B(0/l)$ (as expected). By (7.11), the most general element of $\mathscr{H}^{0(2)}$ has the form

$$\sum_k^l = {}_1\kappa_k(\mathbf{e}_{k+2,k+2} - \mathbf{e}_{k+l+2,k+l+2}) = 4l\sum_{k=2}^{l+1}\rho_k\mathbf{h}_{\alpha_k^0}^0,$$

where $\kappa_1,\kappa_2,...,\kappa_l$ are any complex numbers, and where

$$\rho_k = \sum_{r=1}^{k-1}\kappa_r \quad (\text{for } k = 2,...,l)$$

and

$$\rho_{l+1} = \frac{1}{2}\sum_{r=1}^{l}\kappa_r.$$

Thus on $\mathscr{H}^{0(2)}$ the simple roots of $C(l+1)$ are given by

$$\alpha_1^0(h) = -\kappa_1, \tag{7.41}$$

$$\alpha_k^0(h) = \kappa_{k-1} - \kappa_k \quad (\text{for } k = 2,3,...,l), \tag{7.42}$$

and

$$\alpha_{l+1}^0(h) = 2\kappa_l, \tag{7.43}$$

which implies that on $\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$

$$\alpha_{l+1}^0(h) = -2\sum_{k=1}^{l}\alpha_k^0(h). \tag{7.44}$$

However, as (7.39) and (7.40) imply that

$$\sum_{k}^{l} = {}_1\kappa_k(\mathbf{e}_{k+2,k+2} - \mathbf{e}_{k+l+2,k+l+2})$$

$$= 2(2l+1)\sum_{k=1}^{l}\mu_k\Psi(\mathbf{h}_{\alpha_k^{B(0/l)}}^{B(0/l)}),$$

with

$$\mu_k = \sum_{r=1}^{k}\kappa_r \quad (\text{for } k = 1,2,...,l),$$

it follows that on $\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$

$$\alpha_k^{B(0/l)}(h) = \kappa_k + \kappa_{k-1} \quad (\text{for } k = 1,2,...,l-1), \tag{7.45}$$

and

$$\alpha_l^{B(0/l)}(h) = \kappa_l. \tag{7.46}$$

Comparison of (7.41) to (7.43) with (7.45) and (7.46) then shows that on the Cartan subalgebra on $\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$ of $B(0/l)$ the simple roots $\alpha_k^{B(0/l)}$ of $B(0/l)$ and α_k^0 of $C(l+1)$ are related by

$$\alpha_1^0(h) = -\sum_{r=1}^{l}\alpha_r^{B(0/l)}(h), \tag{7.47}$$

$$\alpha_k^0(h) = \alpha_{k-1}^{B(0/l)}(h) \quad (\text{for } k = 2,3,...,l), \tag{7.48}$$

and

$$\alpha_{l+1}^0(h) = 2\alpha_l^{B(0/l)}(h). \tag{7.49}$$

Finally it follows from (2.13), (2.17), (2.26)–(2.28), (2.36), (2.37), (7.11), (7.12), (7.39), and (7.40) that corresponding elements of the Cartan subalgebra of the Kac–Moody superalgebra are

$$h_{\alpha_k^{B(0/l)}} = t^0\otimes\Psi(\mathbf{h}_{\alpha_k^{B(0/l)}}^{B(0/l)})$$

$$= \{2l/(2l+1)\}t^0\otimes\mathbf{h}_{\alpha_{k+1}}^0 \tag{7.50}$$

$$(\text{for } k = 1,2,...,l).$$

The *nondiagonal* basis elements of $\mathscr{T}_{s0}^{0(2)}$ will now be examined. They fall into six sets.

(i) For $j,k = 1,2,...,l$, with $j<k$: For the basis element $\mathbf{e}_{\beta_{(j,k)}^{0-}}^0$ of (7.21), it is implied by (7.13) and (7.48) that this corresponds to the root

$$\beta_{(j,k)}^{0-}(h) = \sum_{r=j}^{k-1}\alpha_r^{B(0/l)}(h), \tag{7.51}$$

of $B(0/l)$ [for $j,k = 1,2,...,l$, with $j<k$, and for all $h\in\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$].

(ii) For $j,k = 1,2,...,l$, with $j<k$:

$$-\mathbf{e}_{k+2,j+2} + \mathbf{e}_{j+l+2,k+l+2} = \mathbf{e}_{-\beta_{(j,k)}^{0-}}^0, \tag{7.52}$$

which corresponds to the root $-\beta_{(j,k)}^{0-}(h)$ of $B(0/l)$, where $\beta_{(j,k)}^{0-}(h)$ is as in (7.51).

(iii) For $j,k = 1,2,...,l$, with $j\leqslant k$: For the basis element $\mathbf{e}_{\beta_{(j,k)}^{0+}}^0$ of (7.22), it is implied by (7.14), (7.15), (7.48), and (7.49) that this corresponds to the root

$$\beta_{(j,k)}^{0+}(h) = \sum_{r=j}^{k-1}\alpha_r^{B(0/l)}(h) + 2\sum_{r=k}^{l}\alpha_r^{B(0/l)}(h), \tag{7.53}$$

of $B(0/l)$ [for $j,k = 1,2,...,l$, with $j\leqslant k$, and for all $h\in\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$, the first term on the right-hand side of (7.53) not appearing if $j=k$].

(iv) For $j,k = 1,2,...,l$, with $j\leqslant k$:

$$-\mathbf{e}_{k+l+2,j+2} - \mathbf{e}_{j+l+2,k+2} = \mathbf{e}_{-\beta_{(j,k)}^{0+}}^0, \tag{7.54}$$

which corresponds to the root $-\beta_{(j,k)}^{0+}(h)$ of $B(0/l)$, where $\beta_{(j,k)}^{0+}(h)$ is as in (7.53).

(v) For $j = 1,2,...,l$:

$$-\mathbf{e}_{j+2,1} + \mathbf{e}_{j+2,2} + \mathbf{e}_{2,j+l+2} - \mathbf{e}_{1,j+l+2}$$

$$= -\mathbf{e}_{\delta_{(j)}^{0+}}^0 + \mathbf{e}_{-\delta_{(j)}^{0-}}^0, \tag{7.55}$$

where $\mathbf{e}_{\delta_{(j)}^{0-}}^0$ and $\mathbf{e}_{-\delta_{(j)}^{0+}}^0$ are given by (2.17), (7.23), and (7.24). As (7.18), (7.19), (7.20), (7.47), (7.48), and (7.49) imply that

$$\delta_{(j)}^{0+}(h) = -\delta_{(j)}^{0-}(h) = \sum_{r=j}^{l}\alpha_r^{B(0/l)}(h), \tag{7.56}$$

[for $j = 1,2,...,l$, and for all $h\in\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$], the basis element (7.55) corresponds to the root $\delta_{(j)}^{0+}(h)$ of $B(0/l)$.

(vi) For $j = 1,2,...,l$:

$$\mathbf{e}_{1,j+2} - \mathbf{e}_{2,j+2} + \mathbf{e}_{j+l+2,2} - \mathbf{e}_{j+l+2,1} = \mathbf{e}_{\delta_{(j)}^{0-}}^0 + \mathbf{e}_{-\delta_{(j)}^{0+}}^0, \tag{7.57}$$

where $\mathbf{e}_{\delta_{(j)}^{0-}}^0$ and $\mathbf{e}_{-\delta_{(j)}^{0+}}^0$ are given by (2.17), (7.23), and (7.24), which corresponds to the root $-\delta_{(j)}^{0+}(h)$ of $B(0/l)$, where $\delta_{(j)}^{0+}(h)$ is as in (7.56).

All the elements of the above sets are even members of $C(l+1)$, except for those of (7.55) and (7.57), which are odd.

It is easily checked that the set of $2l(l+1)$ nonzero roots of (i) to (vi) above, together with the l zero roots, are all weights of the adjoint representation of $B(0/l)$, whose highest weight is

$$\Lambda = 2\Lambda_1^{B(0/l)} = 2\sum_{k=1}^{l}\alpha_k^{B(0/l)}. \tag{7.58}$$

2. $\mathscr{T}_{s1}^{0(2)}$

By (2.33) the subspace $\mathscr{T}_{s1}^{0(2)}$ consists of the supermatrices whose submatrices satisfy the conditions $\tilde{\mathbf{A}} = \mathbf{A}, \tilde{\mathbf{D}}\mathbf{J} - \mathbf{J}\mathbf{D} = \mathbf{0}$, and $\tilde{\mathbf{C}}\mathbf{J} = -\mathbf{B}$ in addition to (7.6), (7.7), and (7.8). Together these imply that $\mathbf{D} = \mathbf{0}$, that

$$\mathbf{B} = 2^{-1/2}\begin{pmatrix}\mathbf{b}\\ \mathbf{b}\end{pmatrix} \text{ and } \mathbf{C} = 2^{-1/2}(\mathbf{c} \quad \mathbf{c}), \tag{7.59}$$

where $\mathbf{b}$ and $\mathbf{c}$ are submatrices of dimensions $1\times 2l$ and $2l\times 1$, respectively, that experience the constraint (7.26), and that

$$\mathbf{A}=\begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}, \tag{7.60}$$

where a is any complex number. Thus $\mathscr{L}_{s1}^{0(2)}$ possesses a single diagonal basis element

$$\left(\frac{1}{4}l\right)\{\mathbf{e}_{1,1}-\mathbf{e}_{2,2}\}=-\frac{1}{2}\mathbf{h}^0_{\alpha^0_{l+1}}-\sum_{r=1}^{l}\mathbf{h}^0_{\alpha^0_r}, \tag{7.61}$$

which corresponds to a zero weight of $B(0/l)$.

The nondiagonal basis elements of $\mathscr{L}_{s1}^{0(2)}$ fall into two sets.

(i) For $j=1,2,...,l$:

$$-\mathbf{e}_{j+2,1}-\mathbf{e}_{j+2,2}+\mathbf{e}_{2,j+l+2}+\mathbf{e}_{1,j+l+2}=\mathbf{e}^0_{\delta^0_{(j)}}+\mathbf{e}^0_{-\delta^0_{(j)}}, \tag{7.62}$$

where $\mathbf{e}^0_{\delta^0_{(j)}}$ and $\mathbf{e}^0_{-\delta^0_{(j)}}$ are given by (2.17), (7.23), and (7.24), and $\delta^0_{(j)}{}^+(h)[=-\delta^0_{(j)}{}^-(h)]$ is given for $j=1,2,...,l$, and for all $h\in\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$ by (7.56), the basis element (7.62) corresponds to the weight $\delta^0_{(j)}{}^+(h)$ of $B(0/l)$.

(ii) For $j=1,2,...,l$:

$$\mathbf{e}_{1,j+2}+\mathbf{e}_{2,j+2}+\mathbf{e}_{j+l+2,2}+\mathbf{e}_{j+l+2,1}=\mathbf{e}^0_{\delta^0_{(j)}}-\mathbf{e}^0_{-\delta^0_{(j)}}, \tag{7.63}$$

where $\mathbf{e}^0_{\delta^0_{(j)}}$ and $\mathbf{e}^0_{-\delta^0_{(j)}}$ are given by (2.17), (7.23), and (7.24), which corresponds to the root $-\delta^0_{(j)}{}^+(h)$ of $B(0/l)$, where $\delta^0_{(j)}{}^+(h)$ is as in (7.56).

The diagonal basis element (7.61) is an even element of $C(l+1)$, but all the nondiagonal elements of the sets (i) and (ii) are odd members of $C(l+1)$.

They form the carrier space of an irreducible representation of $B(0/l)$ of dimension $2l+1$ whose highest weight is

$$\Lambda=\Lambda_1^{B(0/l)}=\sum_{k=1}^{l}\alpha_k^{B(0/l)}. \tag{7.64}$$

[See Tsohantjis and Cornwell[18] for a discussion of the supercharacters and superdimensions of $B(0/l)$.]

C. The roots of $C^{(2)}(l+1)$

Defining $\delta(h)$ as in (3.6) to (3.8), it follows that the roots $\alpha(h)$ and the corresponding basis elements e_α of $C^{(2)}(l+1)$ are as follows.

(i) $\alpha=2J\delta(h)$ (for $J=0,\pm1,\pm2,...$). There are l linearly independent basis elements $e_\alpha^{(k)}$ corresponding to this root which may be labeled by an additional superscript, so that

$$e_\alpha^{(k)}=\{2l/(2l+1)\}t^{2J}\otimes\mathbf{h}^0_{\alpha^0_{k+1}}\quad(\text{for } k=1,2,...,l)$$

[which reduces to (7.50) in the special case $J=0$].

(ii) $\alpha(h)=2J\delta(h)\pm\beta^0_{(j,k)}{}^-(h)$ (for $j,k=1,2,...,l$, with $j<k$, and for $J=0,\pm1,\pm2,...$), where $\beta^0_{(j,k)}{}^-(h)$ is the extension of the weight of $B(0/l)$ that is given by (7.51) and

$$e_\alpha=t^{2J}\otimes\mathbf{e}^0_{\pm\beta^0_{(j,k)}}.$$

(iii) $\alpha(h)=2J\delta(h)\pm\beta^0_{(j,k)}{}^+(h)$ (for $j,k=1,2,...,l$, with $j\leqslant k$, and for $J=0,\pm1,\pm2,...$), where $\beta^0_{(j,k)}{}^+(h)$ is the extension of the weight of $B(0/l)$ that is given by (7.53) and

$$e_\alpha=t^{2J}\otimes\mathbf{e}^0_{\pm\beta^0_{(j,k)}}.$$

(iv) $\alpha(h)=2J\delta(h)\pm\delta^0_{(j)}{}^+(h)$ (for $j=1,2,...,l$, and for $J=0,\pm1,\pm2,...$), where $\delta^0_{(j)}{}^+(h)$ is the extension of the weight of $B(0/l)$ that is given by (7.56) and

$$e_\alpha=t^{2J}\otimes\{\mathbf{e}^0_{\mp\delta^0_{(j)}}\mp\mathbf{e}^0_{\pm\delta^0_{(j)}}\}.$$

(v) $\alpha(h)=(2J+1)\delta(h)$ (for $J=0,\pm1,\pm2,...$), with

$$e_\alpha=t^{2J+1}\otimes\left\{\frac{1}{2}\mathbf{h}^0_{\alpha^0_{l+1}}+\sum_{r=1}^{l}\mathbf{h}^0_{\alpha^0_r}\right\}.$$

(vi) $\alpha(h)=(2J+1)\delta(h)\pm\delta^0_{(j)}{}^+(h)$ (for $j=1,2,...,l$, and for $J=0,\pm1,\pm2,...$), where $\delta^0_{(j)}{}^+(h)$ is the extension of the weight of $B(0/l)$ that is given by (7.56) and

$$e_\alpha=t^{2J}\otimes\{\mathbf{e}^0_{\mp\delta^0_{(j)}}\pm\mathbf{e}^0_{\pm\delta^0_{(j)}}\}.$$

(vii) $\alpha(h)=0$, with c and d as basis elements.

With μ chosen as in (7.35) [so that (2.37) is valid], it follows that

$$\langle\alpha^0,\beta^0\rangle=\langle\alpha^0,\beta^0\rangle^{B(0/l)}, \tag{7.65}$$

where on the right-hand side of (7.65) α^0 and β^0 are any pair of linear functionals defined on $\mathscr{H}^{0(2)}(=\mathscr{H}^{B(0/l)})$, the evaluation being performed with respect to the Killing form of $B(0/l)$, and where on the left-hand side of (7.65) α^0 and β^0 denote the corresponding extensions to the Cartan subalgebra of the Kac–Moody superalgebra $C^{(2)}(l+1)$, the evaluation being performed with respect to its supersymmetric bilinear invariant form $B(,)$. As $\langle\alpha^0,\alpha^0\rangle^{B(0/l)}>0$ for every nonzero linear functional α^0 defined on $\mathscr{H}^{B(0/l)}$, then $\langle\alpha^0,\alpha^0\rangle>0$ for the corresponding extension. Moreover (2.13), (2.28), (2.30), (2.31), (2.36), and (7.35) imply that

$$h_\delta=\{2l/(2l+1)\}c. \tag{7.66}$$

Thus if α_k^0 is the extension of any simple root of $\mathscr{L}_s^0$, then

$$\langle\delta,\alpha_k^0\rangle=0 \tag{7.67}$$

and

$$\langle j\delta,j\delta\rangle=0. \tag{7.68}$$

Thus $\langle j\delta,j\delta\rangle=0$ for integer j, so every nonzero root of $C^{(2)}(l+1)$ belonging to the sets (i) and (v) is imaginary. Moreover, because $\langle j\delta+\alpha^0,j\delta+\alpha^0\rangle=\langle\alpha^0,\alpha^0\rangle^{B(0/l)}$ and because $\langle\alpha^0,\alpha^0\rangle^{B(0/l)}>0$ for linear functional α^0 and its corresponding extension (as has just been noted), it follows that every nonzero root of $C^{(2)}(l+1)$ belonging to all the above sets except (i) and (v) is real. All the elements mentioned in the above sets are even, except for those in the sets (iv) and (vi), which are odd.

For $C^{(2)}(l+1)$ the simple roots may be taken to be

$$\alpha_0=\delta-\alpha_H^0, \tag{7.69}$$

and

$$\alpha_k=\alpha_k^{B(0/l)}\quad(\text{for } k=1,...,l), \tag{7.70}$$

where

$$\alpha_H^0=\Lambda_1^{B(0/l)}=\sum_{k=1}^{l}\alpha_k^{B(0/l)} \tag{7.71}$$

is the highest weight of the representation of $\mathscr{L}_s^0$ for which $\mathscr{L}_{s1}^{0(2)}$ is the carrier space [cf. (7.64)] and the $\alpha_k^{B(0/l)}$ are the extensions of the simple roots of $B(0/l)$. As e_{α_0} and e_{α_l} appear in the sets (vi) and (iv), respectively, it follows that e_{α_0} and e_{α_l} are odd, so α_0 and α_l are odd roots of the Kac–Moody superalgebra $C^{(2)}(l+1)$. All the other simple roots of $C^{(2)}(l+1)$ are even.

It is then easily checked that the Cartan matrices of $C^{(2)}(2)$ and $C^{(2)}(l+1)$ (for $l \geqslant 2$), evaluated using (2.15), correspond to the generalized Dynkin diagrams given in Figs. 7 and 8.

[1] D. J. Gross, J. A. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **54**, 502 (1985); Nucl. Phys. B **256**, 253 (1985); **267**, 75 (1986).

[2] P. Goddard and D. Olive, Int. J. Mod. Phys. A **4**, 303 (1986); *Kac–Moody and Virasoro Algebras* (World Scientific, Singapore, 1988).

[3] A. Bilal and J. L. Gervais, Phys. Lett. B **177**, 313 (1986); P. Goddard, D. Olive, and G. Waterson, Comm. Math. Phys. **112**, 591 (1986); V. G. Kac and I. T. Todorov, *ibid.* **102**, 337 (1986); **104**, 175 (1986); F. Langouche and T. Schucker, Lett. Math. Phys. **11**, 275 (1986); P. Windey, Comm. Math. Phys. **105**, 511 (1986); E. Kiritsis and G. Siopsis, Phys. Lett. B **184**, 353 (1987); S. Nam, *ibid.* **187**, 340 (1987); F. Englert, A. Sevrin, W. Troost, A. van Proeyen, and Ph. Spindel, J. Math. Phys. **29**, 281 (1988); K. Hamada and M. Takeo, Phys. Lett. B **210**, 120 (1988); E. Abdalla, M. C. B. Abdalla, and A. Zadra, Mod. Phys. Lett. A **4**, 849 (1989); P. Bouwknegt, A. Ceresole, J. McCarthy, and P. van Nieuwenhuizen, Phys. Rev. D **39**, 2971 (1989); P. Bouwknegt, A. Ceresole, P. van Nieuwenhuizen, and J. McCarthy, *ibid.* **40**, 415 (1989); A. C. Cadavid and R. J. Finkelstein, Int. J. Mod. Phys. A **4**, 2967 (1989); S. Olafsson, J. Phys. A **22**, 157 (1989).

[4] G. Kac, Adv. Math. **30**, 85 (1978).

[5] V. G. Kac, Adv. Math. **26**, 8 (1977).

[6] V. G. Kac, Math. USSR Izv. **2**, 1271 (1968). [Original: Izv. Akad. Nauk SSSR Ser. Mat. Tom. **32** (6), 1923 (1968)].

[7] V. G. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U. P., Cambridge, England, 1985).

[8] R. V. Moody, Bull. Am. Math. Soc. **73**, 217 (1967); J. Algebra **10**, 211 (1968); Canad. J. Math. **21**, 1432 (1969).

[9] J. F. Cornwell, *Group Theory in Physics* (Academic, London, 1989), vol. III.

[10] M. Scheunert, *The Theory of Lie Superalgebras*, Lecture Notes in Mathematics, Vol. 716 (Springer-Verlag, Berlin, 1979).

[11] G. Golitzin, J. Algebra **117**, 198 (1988).

[12] J. F. Cornwell, J. Math. Phys. **16**, 992 (1975).

[13] J. F. Cornwell, J. Math. Phys. **20**, 547 (1979).

[14] L. Frappat, A. Sciarrino, and P. Sorba, Comm. Math. Phys. **121**, 457 (1989).

[15] J. F. Cornwell, Int. J. Theor. Phys. **12**, 333 (1975).

[16] M. Konuma, K. Shima, and M. Wada, Progr. Theor. Phys. Suppl. **28**, 1 (1963).

[17] J. F. Cornwell, *Group Theory in Physics* (Academic, London, 1984), Vol. II.

[18] I. Tsohantjis and J. F. Cornwell, Int. J. Theor. Phys. **29**, 351 (1990).

On quantum fields that generate local algebras

Detlev Buchholz
II. Institut für Theoretische Physik, Universität Hamburg, D-2000 Hamburg 50, Federal Republic of Germany

(Received 9 January 1990; accepted for publication 21 March 1990)

The properties of quantum fields generating local nets of von Neumann algebras are discussed. For fields that satisfy certain energy bounds, necessary and sufficient conditions for the existence of such a net and for the existence of product states are given in terms of the vacuum expectation values.

I. INTRODUCTION

The relationship between the two familiar settings of local quantum field theory based on unbounded field operators,[1] respectively, algebras of bounded operators,[2] has long been an interesting problem. (For a recent review and list of references cf. Ref. 3.) Here we are concerned with the characterization of quantum fields which generate local nets of von Neumann algebras. This problem was thoroughly discussed in Ref. 4, where several algebraic conditions involving the field operators have been given that all imply that such a net exists. It is the aim of the present investigation to amend these results by conditions that can be stated directly in terms of the vacuum expectation values (Wightman functions) of the fields.

In the interest of simplicity we treat here only the case of a single Hermitian local scalar field ϕ, as described in Chap. III of the monograph of Streater and Wightman.[1] We adhere to the usual assumptions concerning the locality, the Poincaré covariance, the spectrum condition, and the existence of a unique vacuum vector Ω. By and large our notation is standard, but a few points have to be mentioned here.

For any open subset $\mathscr{R}$ of Minkowski space $\mathscr{M}$ we define $\mathfrak{P}(\mathscr{R})$ as the smallest unital * algebra that contains the averaged field operators $\phi(f)$ for every test function f with compact support in $\mathscr{R}$. (For clarity we consider only open subsets $\mathscr{R}$ of $\mathscr{M}$. Accordingly, the spacelike complement $\mathscr{R}'$ of $\mathscr{R}$ is defined as the set of all points in $\mathscr{M}$ which are spacelike separated from the closure of $\mathscr{R}$. We also use the notation $\mathscr{R}_1 \subset\subset \mathscr{R}_2$ if the closure of $\mathscr{R}_1$ is contained in the interior of $\mathscr{R}_2$.) The elements $X \in \mathfrak{P}(\mathscr{R})$ are regarded as defined on the usual domain $\mathscr{D}$ including the vacuum vector Ω and their closures X^{**} and adjoints X^* are always with respect to this domain. The Hermitian conjugate of X is defined as $X^\dagger = X^* \restriction \mathscr{D}$ and for the field one has, in particular, $\phi(f)^\dagger = \phi(f^*)$. We recall the well-known fact[5] that $\mathfrak{P}(\mathscr{R})\Omega$ is a core for the operators X^{**}, i.e., $(X \restriction \mathfrak{P}(\mathscr{R})\Omega)^- = X^{**}$.

In this paper we employ the notation $\mathfrak{P}^w$ for the weak commutant of any * algebra $\mathfrak{P}$ of operators defined on a common dense domain $\mathscr{D}$: It is the set of all *bounded* operators B such that

$$(B^*\Phi, X\Psi) = (X^\dagger\Phi, B\Psi) \tag{1.1a}$$

for all $X \in \mathfrak{P}$ and $\Phi, \Psi \in \mathscr{D}$. An equivalent statement of condition (1.1a) is that

$$X^{\dagger *} B \supset BX \tag{1.1b}$$

for all $X \in \mathfrak{P}$. We recall that the weak commutant of a * algebra of unbounded operators is a weakly closed linear manifold of operators that is stable under taking adjoints. In general, however, it need not be stable under taking products, i.e., it need not be an algebra.

It is apparent that the structure of the weak commutants of the algebras $\mathfrak{P}(\mathscr{R})$ in quantum field theory is of importance for the question of whether the quantum field ϕ is affiliated with a system of local von Neumann algebras. This question has an affirmative answer if these commutants satisfy the following two clear-cut conditions.

(a) *The weak commutants $\mathfrak{P}(\mathscr{K})^w$ are algebras for each double cone $\mathscr{K}$.*

(b) *The vacuum vector Ω is cyclic for $\mathfrak{P}(\mathscr{K}')^w$ for each double cone $\mathscr{K}$.*

It can be shown that condition (a) implies that the map $K \to \mathfrak{A}(\mathscr{K}) := \mathfrak{P}(\mathscr{K}')^w$ defines a local net of von Neumann algebras on $\mathscr{H}$. Yet these algebras might well be trivial. That they are sufficiently big in order to accommodate bounded functions of the field operators follows from condition (b). In fact, slightly less information is required in order to arrive at the desired conclusion. The following proposition is based on such an apparently weaker version of the above conditions.

Proposition 1.1: Let ϕ be a quantum field such that $\mathfrak{P}(\mathscr{K})^w$ is an algebra for each open double cone $\mathscr{K}$ and Ω is cyclic for $\bigcup_{\mathscr{K}} \mathfrak{P}(\mathscr{K}')^w$. Then the following holds.

(i) The mapping (as usual, the commutant of any set $\mathfrak{C}$ of bounded operators on $\mathscr{H}$ is denoted by $\mathfrak{C}'$)

$$\mathscr{K} \to \mathfrak{A}(\mathscr{K}) := \mathfrak{P}(\mathscr{K})^{w\prime}$$

defines a local, irreducible, Poincaré-covariant net of von Neumann algebras. Moreover, Ω is cyclic for each $\mathfrak{A}(\mathscr{K})$.

(ii) Each operator $X \in \mathfrak{P}(\mathscr{K})$ has some closed extension $X_e \subset X^{\dagger *}$ which is affiliated with $\mathfrak{A}(\mathscr{K})$.

Proof: Proposition 1.1 is a variant of known results; it is, for example, a simple corollary of Theorem 2.7 in Ref. 4. For the convenience of the reader we indicate here the relevant steps in the argument. To begin we note that for any wedge region $\mathscr{W}$ (i.e., any region of Minkowski space that is Poincaré equivalent to $\mathscr{W}_1 = \{x \in \mathscr{M} : x_1 > |x_0|\}$) one has

$$\mathfrak{P}(\mathscr{W})^w = \bigcap_{\mathscr{K} \subset \mathscr{W}} \mathfrak{P}(\mathscr{K})^w \tag{1.2}$$

as a consequence of the definition of the algebras $\mathfrak{P}(\mathscr{R})$. It

 0022-2488/90/081839-08$03.00

thus follows from our assumptions and the Reeh–Schlieder theorem that $\mathfrak{P}(\mathscr{W})^w$ is a von Neumann algebra for which the vacuum vector Ω is cyclic and separating. The results in Ref. 6, Lemma 11 on the relation between the modular conjugation corresponding to the pair $\mathfrak{P}(\mathscr{W})^w$, Ω and the PCT operator can therefore be applied, leading to the crucial equation

$$\mathfrak{P}(\mathscr{W})^{w\prime} = \mathfrak{P}(\mathscr{W}')^w. \tag{1.3}$$

Proceeding from relation (1.2) to the commutants we obtain $\mathfrak{P}(\mathscr{W})^{w\prime} = \bigvee_{\mathscr{K}\subset\mathscr{W}} \mathfrak{A}(\mathscr{K})$ and in view of Eq. (1.3) it is then obvious after some geometrical considerations that the net $\mathscr{K}\to\mathfrak{A}(\mathscr{K})$ is local. The Poincaré covariance of the net is a consequence of the covariance properties of the underlying quantum field ϕ.

Setting $X_e = (X^{\dagger *}\mathscr{D}_1)^-$, where $X\in\mathfrak{P}(\mathscr{K})$ and $\mathscr{D}_1$ is the linear span of $\mathfrak{P}(\mathscr{K})^w\mathscr{D}$, the assertion about the extensions in Proposition (1.1) (ii) follows from the assumption that $\mathfrak{P}(\mathscr{K})^w$ is an algebra. Hence Ω is cyclic for $\mathfrak{A}(\mathscr{K})$ [since it is cyclic for $\mathfrak{P}(\mathscr{K})$ by the Reeh–Schlieder theorem] and, consequently, the net is irreducible. □

If the conditions described in Proposition (1.1) are satisfied, then the theory has many desirable properties.[4] Among the interesting features we mention here only that $\mathfrak{P}(\mathscr{K}')^w\supseteq\mathfrak{A}(\mathscr{K})$, which means that condition (b) holds. Cf. Ref. 4 for further consequences. We also adopt from Ref. 4 the following definition.

Definition: A quantum field theory is said to comply with *scenario G* if statements (i) and (ii) in Proposition 1.1 are satisfied.

Conditions (a) and (b) guarantee that there is a local net of von Neumann algebras that is associated with the quantum field ϕ. Yet not much is known about the necessity of these conditions. Whereas condition (b) must be satisfied if ϕ is to be affiliated with a local net in the sense of Proposition (1.1) (ii), it is conceivable that condition (a) does not need to hold in this case. An apparently weaker (necessary and sufficient) condition is the requirement that the weak commutants $\mathfrak{P}(\mathscr{K})^w$ contain sufficiently large subalgebras for which Ω is cyclic. Thus the somewhat awkward possibility that the weak commutants $\mathfrak{P}(\mathscr{K})^w$ themselves might not be algebras under these circumstances is left open.

However, as far as condition (a) is concerned, the situation is quite simple for the class of fields satisfying so-called energy bounds. We recall that a field ϕ is said to satisfy an H bound, where H is the Hamiltonian, if for each test function f there exists some number n such that $\phi(f)\cdot(1+H)^{-n}$ is a bounded operator,

$$\|\phi(f)\cdot(1+H)^{-n}\| < \infty. \tag{1.4}$$

(For the notion of generalized H bounds cf. Ref. 4.) The specific implication that is of interest here is the following result, based on Lemma 5.4 in Ref. 4.

Proposition 1.2: Let ϕ be a quantum field satisfying a (generalized) H bound. Then the weak commutants $\mathfrak{P}(\mathscr{R})^w$ are algebras for all open regions $\mathscr{R}\subset\mathscr{M}$.

Proof: The regularized field operators $\phi(f)\in\mathfrak{P}(\mathscr{R})$ have the property that they also belong to $\mathfrak{P}(\mathscr{R})$ for small time translations. (Recall that $\operatorname{supp} f$ is compact and $\mathscr{R}$ is open.) We can therefore make use of the assumption that ϕ satisfies a (generalized) H bound and apply the above-mentioned result of Ref. 4, saying that each $B\in\mathfrak{P}(\mathscr{R})^w$ commutes strongly with $\phi(f)$, i.e., $\phi(f)^{**}B\supset B\phi(f)^{**}$. It is apparent from this relation, also, that the product of any two elements of $\mathfrak{P}(\mathscr{R})^w$ commutes strongly with $\phi(f)$. However, the elements of $\mathfrak{P}(\mathscr{R})$ are finite sums and products of field operators $\phi(f)$ with $\operatorname{supp} f\subset\mathscr{R}$, so that Proposition 1.2 follows by repeated application of these facts in the defining relation for $\mathfrak{P}(\mathscr{R})^w$. □

The H-bound condition (1.4) can be stated directly in terms of the system of Wightman functions of the field ϕ and therefore fits naturally into a field-theoretic setting: Moreover, it has a simple intuitive meaning which says that the field strength is bounded in all states of limited energy and does not grow too rapidly with the energy of these states. Energy bounds have been established in various models (cf., for example, Refs. 7 and 8). There also exist counter examples, such as the free massless scalar field in three space-time dimensions, for which states of unlimited field strength and arbitrarily small energy exist because of infrared problems.[9] However, apart from theories with such peculiar properties, the H-bound condition seems to be a reasonable requirement in quantum field theory.

The taming influence of H bounds on the domain problems that plague quantum field theory is well known and has been used in previous work. For example, it has been shown by Driessler and Fröhlich[8] that fields satisfying a linear H bound [where one can put $n=1$ in relation (1.4)] are affiliated with a local net in the sense of Proposition 1.1. Unfortunately, the requirement of a linear H bound seems to be too restrictive for certain composite fields (currents, etc.) appearing in theories of physical interest. Consequently, one must allow for exponents $n>1$ in the general condition (1.4). However, in the latter case it is not known whether such fields ϕ are necessarily affiliated with a local net.

Whereas the H-bound condition provides a useful tool that suffices in many cases for verifying that a given field satisfies condition (a), no such simple test is known in the case of condition (b). (Cf., however, the recent interesting publication in Ref. 10.) As a step toward such a more managable criterion we demonstrate in the present investigation that condition (b) is related to positivity requirements on the Wightman functions, which in principle can be checked directly.

Our results are based on the observation that the size of the weak commutant of a * algebra $\mathfrak{P}$ of unbounded operators can be determined with the help of a suitable seminorm on $\mathfrak{P}$. The same device can also be used to establish the existence of weak intertwiners between different representations of $\mathfrak{P}$. We discuss this more general issue in Sec. II.

In Sec. III we apply our results to the question of when a quantum field admits "well-localized states,"[11,12] respectively, "product states."[13] The existence of such states in sufficient abundance is on one hand closely related to condition (b): On the other hand, it manifests itself in certain specific positivity properties of the above-mentioned seminorms, which can be expressed in terms of the Wightman functions. This fact will lead us to several new criteria which imply the desirable Scenario G.

II. WEAK INTERTWINERS AND SEMINORMS

Before we enter into the discussion of quantum fields we collect here some general facts about representations of algebras of unbounded operators.

Definition: Let $\mathfrak{P}$ be a unital * algebra. A representation π of $\mathfrak{P}$ on a separable Hilbert space $\mathscr{H}$ is said to be *standard* if

(i) there exists a common dense domain $\mathscr{D} \subset \mathscr{H}$ for the elements of $\pi(\mathfrak{P})$;

(ii) $\pi(X^\dagger) \subset \pi(X)^*$ for all $X \in \mathfrak{P}$;

(iii) there is a vector $\Omega \in \mathscr{D}$ that is cyclic for $\pi(\mathfrak{P})$ and the dense set of vectors $\pi(\mathfrak{P})\Omega$ is a core for each $\pi(X)$, $X \in \mathfrak{P}$. Such vectors will be called *standard vectors.*

Standard representations are obtained from any state ω on $\mathfrak{P}$ by the GNS construction. Then

$$\omega(X) = (\Omega, \pi(X)\Omega), \quad X \in \mathfrak{P}, \tag{2.1}$$

where Ω is the GNS vector representing ω. We will only deal with such representations. In order to be able to explore the relation between different representations of $\mathfrak{P}$ we introduce the notion of the "weak intertwiner."

Definition: Let $(\pi_1, \mathscr{H}_1)$ and $(\pi_2, \mathscr{H}_2)$ be standard representations of a unital * algebra $\mathfrak{P}$. A linear operator T: $\mathscr{H}_1 \to \mathscr{H}_2$ is called a *weak intertwiner* between these representations if

$$T \cdot \pi_1(X) \subset \pi_2(X^\dagger)^* \cdot T \tag{2.2}$$

for all $X \in \mathfrak{P}$. (If one interchanges the order of the two representations an analogous relation involving T^* holds.)

We note that the set of weak intertwiners between two standard representations is a weakly closed linear manifold of bounded operators. We are interested in conditions on the underlying standard vectors that imply that such intertwiners exist in sufficient abundance.

From the definition of weak intertwiners and the fact that $\mathfrak{P}$ is an algebra we see that for any decomposition of a given $Z \in \mathfrak{P}$ of the form

$$Z = \sum_i X_i^\dagger Y_i \text{ with } X_i, Y_i \in \mathfrak{P}, \tag{2.3}$$

the equality for any pair of standard vectors $\Omega_i \in \mathscr{H}_i$, $i = 1,2$ holds:

$$(T\Omega_1, \pi_2(Z)\Omega_2) = \sum_i (T\pi_1(X_i)\Omega_1, \pi_2(Y_i)\Omega_2). \tag{2.4}$$

Hence we obtain the bound

$$|(T\Omega_1, \pi_2(Z)\Omega_2)| \leqslant \|T\| \cdot \sum_i \|\pi_1(X_i)\Omega_1\| \|\pi_2(Y_i)\Omega_2\|. \tag{2.5}$$

Since the lhs of (2.5) does not depend on the particular decomposition of Z, we may replace the sum on the rhs by

$$\tau(Z) = \inf \left\{ \sum_i \|\pi_1(X_i)\Omega_1\| \|\pi_2(Y_i)\Omega_2\| : Z = \sum_i X_i^\dagger Y_i \right\}, \tag{2.6}$$

where the infimum refers to all decompositions of Z in $\mathfrak{P}$. Some elementary properties of this quantity are given in the following lemma.

Lemma 2.1: We have that

(i) $\tau(.)$ is a seminorm on $\mathfrak{P}$. (After completion of this work we became aware of the fact that such seminorms were also introduced in Ref. 14 in a similar context),

(ii) $\tau(X^\dagger Y) \leqslant \|\pi_1(X)\Omega_1\| \cdot \|\pi_2(Y)\Omega_2\|$ for X, $Y \in \mathfrak{P}$.

We omit the simple proof of Lemma 2.1. It is of interest here that on one hand, the seminorm τ can be expressed directly in terms of the standard vectors Ω_1 and Ω_2. We note in this context an alternative expression for τ:

$$\tau(Z) = \inf \left\{ \frac{1}{2} \sum_i (\omega_1(X_i^\dagger X_i) + \omega_2(Y_i^\dagger Y_i)) : Z = \sum_i X_i^\dagger Y_i \right\}. \tag{2.7}$$

Here, again, the infimum refers to all decompositions of Z in $\mathfrak{P}$ and ω_1, ω_2 are the states induced by Ω_1 and Ω_2, respectively. (We omit the straightforward proof of this assertion.) On the other hand, τ contains some information about the size of the set $\mathfrak{T}$ of weak intertwiners. This is made apparent in the following lemma.

Lemma 2.2: Let τ be the seminorm defined in (2.6), respectively, (2.7). Then one has, for any $Z \in \mathfrak{P}$,

$$\tau(Z) = \sup\{|(T\Omega_1, \pi_2(Z)\Omega_2)| : T \in \mathfrak{T}, \|T\| \leqslant 1\},$$

where $\mathfrak{T}$ is the set of weak intertwiners between π_1 and π_2.

Proof: We have seen before that the expression on the rhs of the inequality in Lemma 2.2 is not larger than $\tau(Z)$. To see that equality is obtained, let $Z_0 \in \mathfrak{P}$ be fixed and let φ be any linear functional on $\mathfrak{P}$ such that $\varphi(Z_0) = \tau(Z_0)$ and $|\varphi(Z)| \leqslant \tau(Z)$, $Z \in \mathfrak{P}$. That such a functional exists follows from the Hahn–Banach theorem. From Lemma 2.2 we obtain

$$|\varphi(X^\dagger Y)| \leqslant \tau(X^\dagger Y) \leqslant \|\pi_1(X)\Omega_1\| \cdot \|\pi_2(Y)\Omega_2\| \tag{2.8}$$

for all X, $Y \in \mathfrak{P}$. Bearing in mind that Ω_1 and Ω_2 are cyclic for $\pi_1(\mathfrak{P})$ and $\pi_2(\mathfrak{P})$, respectively, we see that there exists some operator $T_0: \mathscr{H}_1 \to \mathscr{H}_2$ of norm $\|T_0\| \leqslant 1$ such that

$$(T_0\pi_1(X)\Omega_1, \pi_2(Y)\Omega_2) = \varphi(X^\dagger Y) = (T_0\Omega_1, \pi_2(X^\dagger Y)\Omega_2). \tag{2.9}$$

Since the domains $\pi_1(\mathfrak{P})\Omega_1$ and $\pi_2(\mathfrak{P})\Omega_2$ are cores for the respective representations of $\mathfrak{P}$ we infer from Eq. (2.9) that T_0 is a weak intertwiner. On the other hand, we have, by construction, $\tau(Z_0) = \varphi(Z_0) = (T_0\Omega_1, \pi_2(Z_0)\Omega_2)$ and since Z_0 was arbitrary, Lemma 2.2 follows. □

We conclude from the above result that nontrivial weak intertwiners between two given standard representations exist iff $\tau \not\equiv 0$. This raises the question of whether one can also characterize the size of $\mathfrak{T}$ in terms of the seminorm τ. In particular, we are interested in the cases where $\overline{\mathfrak{T}\Omega_1} = \mathscr{H}_2$ (resp., $\overline{\mathfrak{T}^*\Omega_2} = \mathscr{H}_1$). If this is to be the case, then τ must induce a norm on $\pi_2(\mathfrak{P})\Omega_2$ [resp., $\pi_1(\mathfrak{P})\Omega_1$], yet this condition need not be sufficient. Some more stringent requirements are discussed in the following lemma.

Lemma 2.3: Let $(\pi, \mathscr{H})$ be a standard representation of $\mathfrak{P}$ and let σ be a seminorm on $\mathfrak{P}$ such that $\sigma(X) \leqslant \|\pi(X)\Omega\|$, $X \in \mathfrak{P}$, where Ω is a standard vector. Then the following statements are equivalent.

(i) The map

$$X \to \pi(X)\Omega, \quad X \in \mathfrak{P}$$

is σ closable. This means that if $\sigma(X_n)\to 0$ and $\pi(X_n)\Omega\to\Phi$ for some sequence $X_n\in\mathfrak{P}$, then $\Phi=0$.

(ii) There exists an absolutely convex and total set $\mathscr{K}$ of vectors in the unit ball of $\mathscr{H}$ such that

$$\sigma(X)=\sup\{|(\Psi,\pi(X)\Omega)|:\Psi\in\mathscr{K}\}$$

for all $X\in\mathfrak{P}$.

(iii) There exists a bounded invertable operator P on $\mathscr{H}$ such that $\sigma(X)\geqslant\|P\pi(X)\Omega\|$ for all $X\in\mathfrak{P}$.[15]

Proof: Making use of the assumption that $\sigma(X)\leqslant\|\pi(X)\Omega\|$ we conclude by a similar argument as in Lemma 2.2 that there exists an absolutely convex set $\mathscr{K}$ of vectors in the unit ball of $\mathscr{H}$ such that for all $X\in\mathfrak{P}$,

$$\sigma(X)=\sup\{|(\Psi,\pi(X)\Omega)|:\Psi\in\mathscr{K}\}.$$

Now let $\Phi\in\mathscr{H}$ and let $X_n\in\mathfrak{P}$ be any sequence such that $\pi(X_n)\Omega\to\Phi$. [Recall that Ω is cyclic for $\pi(\mathfrak{P})$.] Condition (i) of Lemma 2.3 then says that $\Phi\in\mathscr{K}^{\perp}$ is only possible if $\Phi=0$. Hence $\mathscr{K}$ is total, i.e., condition (ii) holds. Under these circumstances we can select from $\mathscr{K}$ (since $\mathscr{H}$ is separable) a sequence of vectors Ψ_n, $n\in\mathbb{N}$ which is still total in $\mathscr{H}$. The operator P given by

$$P\Phi:=\sum_n 2^{-n}\cdot(\Psi_n,\Phi)\cdot\Psi_n$$

then satisfies condition (iii) in Lemma 2.3. Finally, given any operator P that satisfies condition (iii), it follows from $\pi(X_n)\Omega\to\Phi\neq 0$ that $\liminf_n \sigma(X_n)\geqslant\|P\Phi\|>0$; hence condition (i) holds. □

Definition: A seminorm σ on $\mathfrak{P}$ that satisfies the conditions in Lemma 2.3 relative to some standard representation $(\pi,\mathscr{H})$ is said to be π *compatible*. With the help of Lemmas 2.2 and 2.3 we can now establish the following proposition.

Proposition 2.4: Let $(\pi_1,\mathscr{H}_1)$ and $(\pi_2,\mathscr{H}_2)$ be two standard representations of $\mathfrak{P}$, let $\mathfrak{T}$ be the corresponding set of weak intertwiners defined in (2.2), and let τ be the seminorm defined in (2.6), respectively, (2.7). Then

(i) $\mathfrak{T}\neq\{0\}$ iff $\tau\not\equiv 0$,

(ii) $\overline{\mathfrak{T}\Omega_1}=\mathscr{H}_2$ iff τ is π_2 compatible.

(An analogous statement holds for the adjoint intertwiners in $\mathfrak{T}^*$.)

Proof: Statement (i) of Proposition 2.4 is a direct consequence of Lemma 2.2. If $\overline{\mathfrak{T}\Omega_1}=\mathscr{H}_2$, then Lemmas 2.1 and 2.2 show that τ satisfies the assumptions in part (ii) of Lemma 2.3, i.e., τ is π_2 compatible. Conversely, if $\Phi\in(\mathfrak{T}\Omega_1)^{\perp}$ we pick a sequence $X_n\in\mathfrak{P}$ such that $\pi_2(X_n)\Omega_2\to\Phi$. According to Lemma 2.2 this implies that $\tau(X_n)\to 0$. Thus if τ is π_2 compatible it follows from part (i) of Lemma 2.3 that $\Phi=0$, i.e., $\mathfrak{T}\Omega_1$ is total in $\mathscr{H}_2$. □

It is a nice feature of the τ norm that it allows us to characterize the relation between the representations of $\mathfrak{P}$ by properties of the corresponding states. In order to obtain some further insight into the meaning of τ let us briefly discuss the simple case that $\mathfrak{P}=\mathscr{B}(\mathscr{H})$ and that ω_1, ω_2 in relation (2.7) are two normal states on $\mathscr{B}(\mathscr{H})$. In the universal representation of $\mathscr{B}(\mathscr{H})$ the states ω_1, ω_2 are represented by density matrices ρ_1,ρ_2 and the set $\mathfrak{T}$ of intertwiners is represented by the elements of $\mathscr{B}(\mathscr{H})$ which act on vectors by right multiplication. Applying Lemma 2.2 we therefore have, for $X\in\mathscr{B}(\mathscr{H})$,

$$\tau(X)=\sup\{|\mathrm{Tr}((\rho_1^{1/2}T)^*X\rho_2^{1/2})|:T\in\mathscr{B}(\mathscr{H}),\|T\|\leqslant 1\}$$
$$=\mathrm{Tr}|\rho_1^{1/2}X\rho_2^{1/2}|. \tag{2.10}$$

Expression (2.10) is precisely the transition probability between the (unnormalized) states $X^*\omega_1 X$ and ω_2 invented by Uhlmann.[16] Making use of results of Haagerup[17] one arrives at the same conclusion if $\mathfrak{P}$ is an arbitrary von Neumann algebra.[18] It would be of some interest to also establish this simple interpretation of τ for algebras $\mathfrak{P}$ of unbounded operators.

III. QUANTUM FIELDS AND LOCAL ALGEBRAS

Let us now return to the problem of characterizing quantum fields that generate local algebras. Since the assumption of an energy bound for the fields provides a quite satisfactory substitute for condition (a) we will concentrate on condition (b). We will reformulate condition (b) on one hand in terms of seminorms of the type considered in Sec. II. Condition (b) will then appear as a stringent positivity requirement on the underlying Wightman functions, which fits naturally into the Wightman framework. On the other hand, we will relate condition (b) to properties of the state space: Then it has a more transparent physical interpretation. We begin by considering well-localized states for quantum fields.

A. Well-localized states for quantum fields

If scenario G holds in quantum field theory, then there exists an abundance of states that may be regarded as well-localized excitations of the vacuum: Given any open double cone $\mathscr{K}$ we obtain, for each $A\in\mathfrak{A}(\mathscr{K})$, a positive functional ω_A on $\mathfrak{P}(\mathscr{K}')$ by setting

$$\omega_A(X)=(A\Omega,X^{\dagger *}A\Omega),\quad X\in\mathfrak{P}(\mathscr{K}'). \tag{3.1}$$

[That ω_A is well defined and positive follows from the facts that $\mathfrak{A}(\mathscr{K})\subset\mathfrak{P}(\mathscr{K}')^{w}$ and $\mathfrak{A}(\mathscr{K})$ is an algebra.] Since

$$\omega_A(X^{\dagger}X)=(A\Omega,(X^{\dagger}X)^*A\Omega)$$
$$=(A^*A\Omega,X^{\dagger}X\Omega)$$
$$=(X\Omega,A^*AX\Omega)\leqslant\|A\|^2\cdot(\Omega,X^{\dagger}X\Omega) \tag{3.2}$$

we see that ω_A is dominated by the vacuum state; if A is an isometry the two states even coincide. Hence the functionals ω_A correspond to subensembles of the vacuum state on $\mathfrak{P}(\mathscr{K}')$ and in this sense the operation A produces a well-localized excitation of the vacuum in the region $\mathscr{K}$.

The fact that there exist many such well-localized states can be expressed in terms of a seminorm on $\mathfrak{P}(\mathscr{K}')$ given by

$$\tau_1(X)=\sup\{|\omega(X)|:\omega\leqslant\omega_0\restriction\mathfrak{P}(\mathscr{K}')\}, \tag{3.3}$$

where ω_0 is the vacuum state. It follows from the Cauchy–Schwarz inequality that $\tau_1(X)\leqslant\|X\Omega\|$. On the other hand, we see from (3.2) that

$$\tau_1(X)\geqslant\sup\{|(A^*A\Omega,X\Omega)|:A\in\mathfrak{A}(\mathscr{K}),\|A\|\leqslant 1\}. \tag{3.4}$$

If scenario G holds, then the set of vectors $\mathfrak{A}(\mathscr{K})\Omega$ is dense in $\mathscr{H}$ and, consequently, the set $A^*A\Omega$, $A\in\mathfrak{A}(\mathscr{K})$ is total.

According to Lemma 2.3 this implies that τ_1 is ι compatible, where ι denotes the defining ("identical") representation of $\mathfrak{P}(\mathscr{K}')$ on $\mathscr{H}$, i.e., $\iota(X) = X$. These facts motivate the following definition.

Definition: A quantum field theory is said to have a *full set of well-localized states* for some double cone $\mathscr{K}$ if the seminorm τ_1 defined in (3.3) is ι compatible.

It is of interest that the existence of a full set of well-localized states for some double cone $\mathscr{K}$ in turn implies scenario G if condition (a) holds. For the proof of this assertion we pick any $\omega_1 \leqslant \omega_0 \mid \mathfrak{P}(\mathscr{K}')$ and consider the corresponding GNS representation $(\pi_1, \mathscr{H}_1, \Omega_1)$. Since for $X \in \mathfrak{P}(\mathscr{K}')$,

$$\|\pi_1(X)\Omega_1\|^2 = \omega_1(X^\dagger X) \leqslant \omega_0(X^\dagger X) = \|X\Omega\|^2, \quad (3.5)$$

we can define an operator $S: \mathscr{H} \to \mathscr{H}_1$ by

$$SX\Omega = \pi_1(X)\Omega_1, \quad X \in \mathfrak{P}(\mathscr{K}'). \quad (3.6)$$

Clearly, $\|S\| \leqslant 1$. From

$$\begin{aligned}(ZY\Omega, S^*SX\Omega) &= (\pi_1(ZY)\Omega_1, \pi_1(X)\Omega_1) \\ &= (\pi_1(Y)\Omega_1, \pi_1(Z^\dagger X)\Omega_1) \\ &= (Y\Omega, S^*SZ^\dagger X\Omega) \end{aligned} \quad (3.7)$$

we see that $S^*S \in \mathfrak{P}(\mathscr{K}')^w$. Since ω_1 was arbitrary we conclude that for any $X \in \mathfrak{P}(\mathscr{K}')$ the bound

$$\begin{aligned}&\tau_1(X) \leqslant \tau_0(X) \\ &\quad := \sup\{|(T\Omega, X\Omega)| : T \in \mathfrak{P}(\mathscr{K}')^w, \|T\| \leqslant 1\} \end{aligned} \quad (3.8)$$

holds. Hence if τ_1 is ι compatible it follows from part (iii) of Lemma 2.3 that τ_0 is also ι compatible.

Now the elements of $\mathfrak{P}(\mathscr{K}')^w$ are just the weak intertwiners between ι and ι [cf. relations (1.1b) and (2.1)]; hence we may apply the general results of Sec. II. Making use of Lemma 2.2 we can represent τ_0 in the form

$$\tau_0(Z) = \inf\left\{\frac{1}{2}\sum_i \omega_0(X_i^\dagger X_i + Y_i^\dagger Y_i) : Z = \sum_i X_i^\dagger Y_i\right\}, \quad (3.9)$$

where the infimum refers to all possible decompositons of Z in $\mathfrak{P}(\mathscr{K}')$. The dependence of τ_0 on the underlying system of Wightman functions is made explicit here. The requirement that τ_0 is ι compatible may thus be regarded as a stringent positivity condition on these functions. On the other hand, we see from Proposition 2.4 that $\overline{\mathfrak{P}(\mathscr{K}')^w\Omega} = \mathscr{H}$ iff τ_0 is ι compatible.

However, if condition (a) is given and if $\overline{\mathfrak{P}(\mathscr{K}')^w\Omega} = \mathscr{H}$ for some double cone $\mathscr{K}$, then scenario G holds according to Proposition 1.1, as claimed. We summarize these findings in the following proposition.

Proposition 3.1: Consider the following statements about a quantum field theory:

(i) scenario G holds,

(ii) There exists a full set of well-localized states for some double cone $\mathscr{K}$,

(iii) The seminorm τ_0 is ι compatible for some double cone $\mathscr{K}$,

(iv) $\overline{\mathfrak{P}(\mathscr{K}')^w\Omega} = \mathscr{H}$ for some double cone $\mathscr{K}$.

The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Leftrightarrow$ (iv) hold. The same is true if the term "some" in the above statements is replaced by "every." If condition (a) holds, then all the above statements are equivalent.

B. Product states for quantum fields

The condition that operators that represent physical observables in spacelike separated regions must commute is clearly an indispensible requirement in relativistic quantum field theory. A somewhat stronger formulation of the principle of relativistic causality is the assumption that there exist physical states ω without any correlations between the observables A, B localized in fixed spacelike separated regions. In formula form,

$$\omega(A \cdot B) = \omega(A) \cdot \omega(B). \quad (3.10)$$

We refer to Refs. 19–21 for physical motivations of this postulate. It is our aim to demonstrate that the existence of such product states manifests itself in the structure of the underlying quantum fields.

We begin by recalling some relevant facts. Let us assume first that scenario G holds. Then the map

$$\mathscr{K} \to \mathfrak{A}(\mathscr{K}) := \mathfrak{P}(\mathscr{K})^{w\prime} \quad (3.11)$$

defines a local net of von Neumann algebras on $\mathscr{H}$. We define the algebra $\mathfrak{A}(\mathscr{K}')$ as the smallest von Neumann algebra containing all local algebras $\mathfrak{A}(\mathscr{K}_0)$ with $\mathscr{K}_0 \subset \mathscr{K}'$. By a standard partition-of-unity argument it follows that $\mathfrak{A}(\mathscr{K}') = \mathfrak{P}(\mathscr{K}')^{w\prime}$.

It has been shown by Roos[22] that for any pair of double cones $\mathscr{K}_1 \subset\subset \mathscr{K}_2$ the algebras $\mathfrak{A}(\mathscr{K}_1)$ and $\mathfrak{A}(\mathscr{K}_2')$ are algebraically independent. This means that the * algebra generated by $\mathfrak{A}(\mathscr{K}_1)$ and $\mathfrak{A}(\mathscr{K}_2')$ is isomorphic to the algebraic tensor product $\mathfrak{A}(\mathscr{K}_1) \odot \mathfrak{A}(\mathscr{K}_2')$. It will be advantageous to work with this tensor product.

Systems without any correlations between the spacelike separated regions $\mathscr{K}_1$ and $\mathscr{K}_2'$ are described by vector states in the representation $(\pi_P, \mathscr{H}_P)$ of $\mathfrak{A}(\mathscr{K}_1) \odot \mathfrak{A}(\mathscr{K}_2')$, where $\mathscr{H}_P = \mathscr{H} \otimes \mathscr{H}$ and π_P is given by

$$\pi_P(A \odot B) = A \otimes B \quad (3.12)$$

for $A \in \mathfrak{A}(\mathscr{K}_1)$, $B \in \mathfrak{A}(\mathscr{K}_2')$. Each normalized vector $\Phi \otimes \Psi \in \mathscr{H}_P$ induces a product state on $\mathfrak{A}(\mathscr{K}_1) \odot \mathfrak{A}(\mathscr{K}_2')$.

The original (defining) representation is recovered by setting

$$\pi(A \odot B) = A \cdot B \quad (3.13)$$

on $\mathscr{H}$, where the product on the rhs refers to $\mathscr{B}(\mathscr{H})$. [Note that π is a * representation since $\mathfrak{A}(\mathscr{K}_1) \subset \mathfrak{A}(\mathscr{K}_2')'$.] We call $(\pi, \mathscr{H})$ the vacuum representation.

The assumption that product states exist in the vacuum representation can be expressed in several ways.[13] Here we employ the condition that the representations π_P and π are not disjoint. (This means that there exists some nontrivial intertwiner between π_P and π.) This rather weak requirement implies that there exist product states in the vacuum representation of the algebra $\mathfrak{A}(\mathscr{K}_a) \odot \mathfrak{A}(\mathscr{K}_b')$ for any pair of double cones $\mathscr{K}_a$, $\mathscr{K}_b$ with $\mathscr{K}_a \subset\subset \mathscr{K}_1$, $\mathscr{K}_b \supset\supset \mathscr{K}_2$.[13] We note for later reference the following definition.

Definition: A quantum field theory is said to comply

with *scenario P* for a given pair of double cones $\mathscr{K}_1 \subset\subset \mathscr{K}_2$ if

(i) scenario G holds,

(ii) the representations π_P and π of $\mathfrak{A}(\mathscr{K}_1)\odot\mathfrak{A}(\mathscr{K}_2')$ are not disjoint.

Scenario P has interesting consequences for the structure of the underlying quantum fields that we wish to discuss now. In analogy to the preceding procedure we introduce the algebraic tensor product $\mathfrak{P} = \mathfrak{P}(\mathscr{K}_1)\odot\mathfrak{P}(\mathscr{K}_2')$. The elements of $\mathfrak{P}$ will be denoted by the boldfaced letters $\mathbf{X}$, $\mathbf{Y}$, etc. We recall that the * operation on $\mathfrak{P}$ is defined by $(X\odot Y)^\dagger = X^\dagger\odot Y^\dagger$.

The product state representation $(\pi_P,\mathscr{H})$ of the algebra $\mathfrak{P}$ is defined on the domain $\mathscr{D}_P = \mathrm{span}\{\Phi\otimes\Psi:\Phi,\Psi\in\mathscr{D}\}$ by

$$\pi_P(X\odot Y) = X\otimes Y \tag{3.14}$$

for $X\in\mathfrak{P}(\mathscr{K}_1)$, $Y\in\mathfrak{P}(\mathscr{K}_2')$. We note that π_P is a standard representation since $\Omega_P = \Omega\otimes\Omega\in\mathscr{D}_P$ has all properties of a standard vector; cf. Sec. II. The vacuum representation $(\pi,\mathscr{H})$ of $\mathfrak{P}$ acts on $\mathfrak{D}$ by

$$\pi(X\odot Y) = X\cdot Y. \tag{3.15}$$

Note that the range of π coincides with the algebra $\mathfrak{P}(\mathscr{K}_1\cup\mathscr{K}_2')$.

Now let T: $\mathscr{H}_P\to\mathscr{H}$ be any intertwiner between the representations π_P and π. This means that for any $A\in\mathfrak{A}(\mathscr{K}_1)$ and $B\in\mathfrak{A}(\mathscr{K}_2')$ the equality

$$AB\cdot T = T\cdot(A\otimes B) \tag{3.16}$$

holds. We make use of the fact that in the case of scenario G every operator $X\in\mathfrak{P}(\mathscr{K}_1)$ has some closed extension $X_e \subset X^{\dagger *}$ which is affiliated with $\mathfrak{A}(\mathscr{K}_1)$. Hence each member of the sequence of bounded operators $A_n = X_e\cdot(1+(1/n)X_e^{*}X_e)^{-1}$, $n\in\mathbb{N}$ is an element of $\mathfrak{A}(\mathscr{K}_1)$. Moreover, $A_n\to X$ and $A_n^{*}\to X^\dagger$ in the sense of strong convergence on $\mathscr{D}$. Similarly, we can find for any $Y\in\mathfrak{P}(\mathscr{K}_2')$ some sequence $B_n\in\mathfrak{A}(\mathscr{K}_2')$ such that $B_n\to Y$ and $B_n^{*}\to Y^\dagger$ on $\mathscr{D}$. We can therefore proceed from relation (3.16) to the equality

$$(XY)^{\dagger *}\cdot T = T\cdot(X\otimes Y), \tag{3.17}$$

which holds on the domain $\mathscr{D}_P$. Hence we conclude that T is a weak intertwiner between the representations π_P and π. From relation (3.16) it also follows that $T^*T\in(\mathfrak{A}(\mathscr{K}_1)\bar{\otimes}\mathfrak{A}(\mathscr{K}_2'))' = (\mathfrak{P}(\mathscr{K}_1)^w\bar{\otimes}\mathfrak{P}(\mathscr{K}_2')^w)$ and, consequently, $T^*T\in\pi_P(\mathfrak{P})^w$.

These facts allow us to define a positive functional ω on $\mathfrak{P}$, setting

$$\omega(\mathbf{X}) = (T\Omega_P,\pi(\mathbf{X}^\dagger)^*T\Omega_P). \tag{3.18}$$

Making use of the properties of T we have

$$\begin{aligned}\|\pi(\mathbf{X}^\dagger)^*T\Omega_P\|^2 &= \|T\pi_P(\mathbf{X})\Omega_P\|^2 = (T^*T\Omega_P,\pi_P(\mathbf{X}^\dagger\mathbf{X})\Omega_P)\\ &= (T\Omega_P,\pi(\mathbf{X}^\dagger\mathbf{X})^*T\Omega_P) = \omega(\mathbf{X}^\dagger\mathbf{X}).\end{aligned} \tag{3.19}$$

From Eqs. (3.18) and (3.19) it follows on one hand that ω is positive, as claimed. On the other hand, we see from the second member of (3.19) that $\omega(\mathbf{X}^\dagger\mathbf{X})\leqslant\|T\|^2\cdot\omega_P(\mathbf{X}^\dagger\mathbf{X})$, where ω_P is the product state induced by the vector $\Omega_P = \Omega\otimes\Omega$ in the representation π_P. Hence ω corresponds to a subensemble of the uncorrelated state ω_P. If the intertwiner T happens to be an isometry, then the two states even coincide. We specify the relevant properties of ω in the following definition.

Definition:

(i) A state ω on $\mathfrak{P}$ is said to be a vector state in the representation $(\pi,\mathscr{H})$ if there exists a vector $\Phi\in\mathscr{H}$ that lies in the domain of all operators $\pi(\mathbf{X}^\dagger)^*$, $\mathbf{X}\in\mathfrak{P}$ and for which $\omega(\mathbf{X}^\dagger\mathbf{X}) = \|\pi(\mathbf{X}^\dagger)^*\Phi\|^2$.

(ii) A state ω on $\mathfrak{P}$ is said to be weakly correlated if there exists some number $c>0$ such that $\omega\leqslant c\cdot\omega_P$.

The existence of weakly correlated vector states in the vacuum representation $(\pi,\mathscr{H})$ is thus a necessary condition if scenario P is to be valid. We will now demonstrate that the existence of such a state ω in quantum field theory in turn implies scenario P, provided that the theory satisfies condition (a).

Given ω we can construct a nontrivial weak intertwiner T between the representations π_P and π, setting

$$T\pi_P(\mathbf{X})\Omega_P = \pi(\mathbf{X}^\dagger)^*\Phi,\quad \mathbf{X}\in\mathfrak{P}, \tag{3.20}$$

where $\Phi\in\mathscr{H}$ is the vector that represents ω. Since $\omega\leqslant c\cdot\omega_P$ for some $c>0$ we have $\|T\|\leqslant c^{1/2}$. Making use of the fact that Ω_P is a standard vector in the representation π_P and $\pi((\mathbf{XY})^\dagger)^*\Phi = \pi(\mathbf{X}^\dagger)^*\pi(\mathbf{Y}^\dagger)^*\Phi$ it is apparent from (3.20) that T is a nontrivial weak intertwiner between π_P and π.

In order to proceed we have to analyze the properties of the von Neumann algebras $\mathfrak{A}(\mathscr{K}_1) = \mathfrak{P}(\mathscr{K}_1)^{w'}$ and $\mathfrak{A}(\mathscr{K}_2') = \mathfrak{P}(\mathscr{K}_2')^{w'}$. Denoting the set of weak intertwiners T between the representations π_P and π by $\mathfrak{T}$ we have the following lemma.

Lemma 3.2:

(i) Let $T\in\mathfrak{T}$. Then

$$AB\cdot T = T\cdot(A\otimes B)$$

for all $A\in\mathfrak{A}(\mathscr{K}_1)$ and $B\in\mathfrak{A}(\mathscr{K}_2')$.[15]

(ii) If there exists a family of weak intertwiners $T_\alpha\in\mathfrak{T}$ whose ranges generate $\mathscr{H}$, then $\mathfrak{A}(\mathscr{K}_1)\subset\mathfrak{A}(\mathscr{K}_2')'$ and Ω is separating for $\mathfrak{A}(\mathscr{K}_2')$.

Proof:

(i) Let $T\in\mathfrak{T}$ and $\Phi\in\mathscr{D}$. Then we define an operator T_Φ: $\mathscr{H}\to\mathscr{H}$ by

$$T_\Phi\cdot\Psi = T\cdot(\Psi\otimes\Phi)\quad\text{for}\quad\Psi\in\mathscr{H}.$$

We easily see that $\|T_\Phi\|\leqslant\|T\|\cdot\|\Phi\|$. If $\Psi\in\mathscr{D}$ and $X\in\mathfrak{P}(\mathscr{K}_1)$ we have

$$\begin{aligned}T_\Phi X\Psi &= T\cdot(X\Psi\otimes\Phi) = T\cdot\pi_P(X\odot 1)(\Psi\otimes\Phi)\\ &= \pi(X^\dagger\odot 1)^*T(\Psi\otimes\Phi) = X^{\dagger *}T_\Phi\Psi.\end{aligned}$$

Hence $T_\Phi\in\mathfrak{P}(\mathscr{K}_1)^w$ and if $A\in\mathfrak{A}(\mathscr{K}_1) = \mathfrak{P}(\mathscr{K}_1)^{w'}$ we obtain

$$AT(\Psi\otimes\Phi) = AT_\Phi\Psi = T_\Phi A\Psi = T(A\otimes 1)(\Psi\otimes\Phi).$$

Since the vectors $\Psi\otimes\Phi$ generate $\mathscr{H}_P$ it follows that $A\cdot T = T\cdot(A\otimes 1)$. By a similar reasoning we prove that $B\cdot T = T\cdot(1\otimes B)$ for $B\in\mathfrak{A}(\mathscr{K}_2')$.

(ii) According to step (i) we have for $T_\alpha\in\mathfrak{T}$ that $AB\cdot T_\alpha = T_\alpha\cdot(A\otimes 1)(1\otimes B) = T_\alpha\cdot(1\otimes B)(A\otimes 1) = BA\cdot T_\alpha$ and, consequently, $AB = BA$ if the ranges of T_α generate $\mathscr{H}$. If $B\in\mathfrak{A}(\mathscr{K}_2')$ and $B\Omega = 0$ we obtain, for $\Phi_P\in\mathscr{D}_P$ and $X\in\mathfrak{P}(\mathscr{K}_1)$,

$$(T_a(1\otimes B^*)\Phi_P, X\Omega) = (T_a(X^\dagger\otimes 1)^{**}(1\otimes B^*)\Phi_P,\Omega)$$
$$= (T_a(1\otimes B^*)(X^\dagger\otimes 1)\Phi_P)$$
$$= (T_a(X^\dagger\otimes 1)\Phi_P, B\Omega) = 0.$$

From the above equation we conclude that $T_a(1\otimes B^*) = B^*T_a = 0$ since the sets $\mathfrak{P}(\mathscr{K}_1)\Omega$ and $\mathscr{D}_P$ are dense in $\mathscr{H}$ and $\mathscr{H}_P$, respectively, and, consequently, $B = 0$. □

The second technical result that we need is the following lemma.

Lemma 3.3: Let $0 \neq T \in \mathfrak{T}$, let $\mathscr{N}$ be any open set of space-time translations x, and let $U(x)$ be the corresponding unitary representers on $\mathscr{H}$. Then the ranges of the operators $U(x)T$, $x \in \mathscr{N}$ generate $\mathscr{H}$.

Proof: Let $\Phi \in \mathscr{H}$ be any vector such that $T^*U(x)\Phi = 0$ for $x \in -\mathscr{N}$: It then follows from the spectrum condition, with the help of the edge-of-the-wedge theorem, that this equality holds for all x. On the other hand, we will see below that for any $\Psi \in \mathscr{H}$,

$$\liminf_x \|T^*U(x)\Psi\| \geqslant \|T^*\Omega\|\cdot\|\Psi\| \qquad (*)$$

if x approaches spacelike infinity. Since standard vectors are separating for weak intertwiners we have $T^*\Omega \neq 0$. Hence $\Phi = 0$, as claimed. For the proof of inequality (*) we fix a double cone $\mathscr{K}$ and pick any $Y \in \mathfrak{P}(\mathscr{K})$. If x is a sufficiently large spacelike translation we have $Y(x) := U(x)YU(x)^{-1} \in \mathfrak{P}(\mathscr{K}_2')$ and, consequently,

$$\|T^*U(x)Y\Omega\| = \|T^*\pi(1\odot Y(x))\Omega\|$$
$$= \|\pi_P(1\odot Y(x)^\dagger)^*T^*\Omega\|.$$

Multiplying the above equation by $\|\pi_P(1\odot Y(x))\Psi_P\|$, where $\Psi_P \in \mathscr{D}_P$, we proceed with the help of the Cauchy–Schwarz inequality to the estimate

$$\|\pi_P(1\odot Y(x))\Psi_P\|\cdot\|T^*U(x)Y\Omega\|$$
$$\geqslant |(\pi_P(1\odot(Y^\dagger Y)(x))\Psi_P, T^*\Omega)|.$$

Now, according to the cluster theorem we have

$$w-\lim_x \pi_P(1\odot(Y^\dagger Y)(x))\Psi_P = \|Y\Omega\|^2\cdot\Psi_P$$

if x tends to spacelike infinity; it therefore follows from the above estimate that

$$\|\Psi_P\|\cdot\liminf_x \|T^*U(x)Y\Omega\| \geqslant |(\Psi_P, T^*\Omega)|\cdot\|Y\Omega\|.$$

Since $\mathfrak{P}(\mathscr{K})\Omega$ and $\mathscr{D}_P$ are dense in $\mathscr{H}$ and $\mathscr{H}_P$, respectively, and T is bounded we conclude that inequality (*) holds. □

After these preparations we choose any pair of double cones $\mathscr{K}_a$, $\mathscr{K}_b$ such that $\mathscr{K}_a \subset\subset \mathscr{K}_1$ and $\mathscr{K}_b \supset\supset \mathscr{K}_2$ and consider the subalgebra $\tilde{\mathfrak{P}} = \mathfrak{P}(\mathscr{K}_a)\odot\mathfrak{P}(\mathscr{K}_b')$ of $\mathfrak{P}$. We also introduce the representations $\tilde{\pi} = \pi \restriction \tilde{\mathfrak{P}}$ and $\tilde{\pi}_P = \pi_P \restriction \tilde{\mathfrak{P}}$ and the set $\tilde{\mathfrak{T}}$ of weak intertwiners between $\tilde{\pi}_P$ and $\tilde{\pi}$. With our choice of $\mathscr{K}_a$, $\mathscr{K}_b$ there exists an open set $\mathscr{N}$ of translations such that $\mathscr{K}_a + x \subset \mathscr{K}_1$ and $\mathscr{K}_b' + x \subset \mathscr{K}_2'$ for $x \in \mathscr{N}$ and, consequently, $\mathfrak{P}(\mathscr{K}_a + x)\odot\mathfrak{P}(\mathscr{K}_b' + x) \subset \mathfrak{P}$ if $x \in \mathscr{N}$. Hence if T is any nontrivial weak intertwiner between the representations π_P and π of $\mathfrak{P}$ we have $U(x)T \in \tilde{\mathfrak{T}}$ for $x \in -\mathscr{N}$.

From Lemma 3.3 it follows that the ranges of the operators $U(x)T$, $x \in -\mathscr{N}$ span $\mathscr{H}$ and applying the second part of Lemma 3.2 we conclude that $\mathfrak{A}(\mathscr{K}_a) \subset \mathfrak{A}(\mathscr{K}_b')'$. From the above we see in particular that $\mathfrak{A}(\mathscr{K}_a) \subset \mathfrak{A}(\mathscr{K}_a + x)'$ for large spacelike x, i.e., the operators in $\mathfrak{A}(\mathscr{K}_a)$ are local. However, we cannot exclude the possibility that the algebra $\mathfrak{A}(\mathscr{K}_a)$ is too small in the sense that the elements of $\mathfrak{P}(\mathscr{K}_a)$ do not have any extensions affiliated with it.

It is at this point where condition (a) enters: If $\mathfrak{P}(\mathscr{K})^w$ is an algebra (and hence a von Neumann algebra) for all double cones $\mathscr{K}$, the same is true for $\mathfrak{P}(\mathscr{K}_b')^w = \bigcap_{\mathscr{K}\subset\mathscr{K}_b'} \mathfrak{P}(\mathscr{K})^w$. Making use of part (ii) of Lemma 3.2 we see that the vacuum vector Ω is separating for $\mathfrak{P}(\mathscr{K}_b')^{w\prime} = \mathfrak{A}(\mathscr{K}_b')$ and therefore cyclic for $\mathfrak{P}(\mathscr{K}_b')^{w\prime\prime} = \mathfrak{P}(\mathscr{K}_b')^w$. Hence we infer from Proposition 1.1 that scenario G is obtained. Moreover, applying part (i) of Lemma 3.2 we see that the representations π_P and π of $\mathfrak{A}(\mathscr{K}_1)\odot\mathfrak{A}(\mathscr{K}_2')$ are not disjoint. We have thus recovered scenario P for this particular pair of regions.

It is clear from the preceding discussion that the existence of a nontrivial weak intertwiner between the representations π_P and π of $\mathfrak{P}$ is a necessary and sufficient condition for scenario P. As discussed in Sec. II, this condition can also be stated in terms of the seminorm τ_P on $\mathfrak{P}$ given by

$$\tau_P(\mathbf{Z}) = \inf\left\{\frac{1}{2}\sum_i (\omega_P(\mathbf{X}_i^\dagger\mathbf{X}_i) + \omega_0(\mathbf{Y}_i^\dagger\mathbf{Y}_i)) : \mathbf{Z} = \sum_i \mathbf{X}_i^\dagger\mathbf{Y}_i\right\}. \qquad (3.21)$$

Here the infimum refers to all decompositions of $\mathbf{Z}$ in $\mathfrak{P}$ and ω_P, ω_0 are the states induced by the vectors $\Omega_P = \Omega\otimes\Omega$ and Ω in the representations π_P and π, respectively. (Note that the expression for τ_P involves only the underlying Wightman functions.)

According to Proposition 2.4 $\tau_P \neq 0$ holds iff there exists a weak intertwiner between the representations π_P and π. In fact, it suffices to determine the value of $\tau_P(1)$. Namely, if scenario G holds we have, for any double cone $\mathscr{K}_3 \subset \mathscr{K}_1' \cap \mathscr{K}_2$, the inclusion $\mathfrak{A}(\mathscr{K}_3) \subset \mathfrak{A}(\mathscr{K}_1)' \cap \mathfrak{A}(\mathscr{K}_2')'$ because of locality: Hence if T is an intertwiner between the representations π_P and π of $\mathfrak{A}(\mathscr{K}_1)\odot\mathfrak{A}(\mathscr{K}_2')$ the same holds true for the operators in $\mathfrak{A}(\mathscr{K}_3)\cdot T$. In the presence of scenario G each intertwiner between π_P and π is a weak intertwiner between the representations π_P and π of $\mathfrak{P}$, as we have shown before. Applying Lemma 2.2 we conclude that $\tau_P(1) = 0$ is only possible if $(CT\Omega_P, \Omega) = 0$ for all $C \in \mathfrak{A}(\mathscr{K}_3)$, which in turn implies that $T = 0$ since Ω is cyclic for $\mathfrak{A}(\mathscr{K}_3)$ and Ω_P is separating for the set of weak intertwiners between π_P and π. Summarizing these results we have established the following proposition.

Proposition 3.4: Consider the following statements about a quantum field theory for a given pair of double cones $\mathscr{K}_1 \subset\subset \mathscr{K}_2$:

(i) scenario P holds,

(ii) the seminorm τ_P on $\mathfrak{P}(\mathscr{K}_1)\odot\mathfrak{P}(\mathscr{K}_2')$ satisfies $\tau_P(1) \neq 0$,

(iii) the seminorm τ_P on $\mathfrak{P}(\mathscr{K}_1)\odot\mathfrak{P}(\mathscr{K}_2')$ is nontrivial,

(iv) there exists a vector state ω in the vacuum representation of $\mathfrak{P}(\mathscr{K}_1)\odot\mathfrak{P}(\mathscr{K}_2')$ which is weakly correlated. The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Leftrightarrow$ (iv) hold. If condition (a) holds, then all the above statements are equivalent.

On the basis of the results in Refs. 19 and 23 one may expect that the existence of weakly correlated vector states in the vacuum representation is intimately related to phase space properties of the theory. As in the algebraic framework, one can characterize theories with physically acceptable phase space properties with the help of a nuclearity condition involving the underlying quantum field ϕ. It would be desirable to establish scenario P starting from such a condition.

We conclude this discussion of product states in quantum field theory with the remark that if scenario P holds for the net $\mathscr{K}\to\mathfrak{A}(\mathscr{K})$ it may happen that its maximal local extension $\mathscr{K}\to\mathfrak{A}(\mathscr{K}')'$ (which satisfies the condition of duality[6]) does not admit any product state. Therefore, scenario P does not necessarily imply that the theory has the so-called split property.[24]

IV. CONCLUSIONS

We have shown in the present investigation that the seminorms τ_0 and τ_P provide relevant information about the nature of the weak commutants $\mathfrak{P}(\mathscr{R})^w$ and hence about the validity of scenarios G and P, respectively. It is a nice feature of these seminorms that they are directly related to the underlying Wightman functions and as such, fit naturally into the Wightman framework.

Concerning their potential utility as a tool of investigation, these seminorms might on one hand be suited for the possible refutation of a quantum field theory. By "refutation" we mean here the demonstration that scenario G, respectively P, cannot be obtained. From the physical point of view one may, on the other hand, regard these scenarios as indispensible requirements. Our results then in turn provide specific information about the underlying Wightman functions which could be used as a starting point for further investigations.

Another potential field of application for the concepts and results in the present investigation is the reformulation of the general analysis of superselection sectors expounded in Ref. 25 in terms of quantum fields. The states of interest representing local excitations of the vacuum of arbitrary charge can be characterized with the help of seminorms of the form (2.7). In this context there arises the interesting problem of whether there holds a composition law for the resulting weak intertwiners which corresponds to the composition of superselection sectors (addition of charges) in quantum field theory.

ACKNOWLEDGMENTS

This article contains results of a collaboration with Professor E. H. Wichmann in the spring of 1987. I would like to thank Professor Wichmann for valuable contributions, noted in the text as private communications, and for the preparation of a detailed exposé of some of the material presented in this article.

I am also grateful for the hospitality shown to me by the Department of Physics of the University of California, Berkeley, and I acknowledge a travel grant from the DFG, Federal Republic of Germany, as well as financial support from the NSF, United States of America.

[1] R. S. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[2] R. Haag and B. Schroer, J. Math. Phys. **3**, 248 (1962).
[3] J. Yngvason, in *Proceedings of the IXth International Congress on Mathematical Physics,* (Swansea, 1988).
[4] W. Driessler, S. J. Summers, and E. H. Wichmann, Commun. Math. Phys. **105**, 49 (1986).
[5] H. J. Borchers and W. Zimmermann, Nuovo Cimento **31**, 1047 (1963).
[6] J. J. Bisognano and E. H. Wichmann, J. Math. Phys. **17**, 303 (1976).
[7] J. Hertel, *Lokale Quantentheorie und Felder am Punkt,* Internal report DESY T-81/01 (1981).
[8] W. Driessler and J. Fröhlich, Ann. Inst. Henri Poincaré **27**, 221 (1977).
[9] R. Wanzenberg, *Energie und Präparierbarkeit von Zuständen in der lokalen Quantenfeldtheorie* (Diplomarbeit, Hamburg, 1987).
[10] H. J. Borchers and J. Yngvason, Commun. Math. Phys. **127**, 607 (1990).
[11] A. L. Licht, J. Math. Phys. **4**, 1443 (1963).
[12] R. Haag and J. A. Swieca, Commun. Math. Phys. **1**, 308 (1965).
[13] D. Buchholz, Commun. Math. Phys. **36**, 287 (1974).
[14] J. Yngvason, Commun. Math. Phys. **34**, 315 (1973).
[15] E. H. Wichmann (private communications).
[16] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[17] U. Haagerup, Int. CNRS **274**, 175 (1979).
[18] I am indebted to H. Kosaki for discussions on this point.
[19] D. Buchholz and E. H. Wichmann, Commun. Math. Phys. **106**, 321 (1986).
[20] D. Buchholz, S. Doplicher, and R. Longo, Ann. Phys. **170**, 1 (1986).
[21] R. Werner, Lett. Math. Phys. **13**, 325 (1987).
[22] H. Roos, Commun. Math. Phys. **16**, 238 (1970).
[23] D. Buchholz, C. D'Antoni, and K. Fredenhagen, Commun. Math. Phys. **111**, 123 (1987).
[24] S. Doplicher and R. Longo, Invent. Math. **73**, 493 (1984).
[25] S. Doplicher, R. Haag, and J. E. Roberts, Commun. Math. Phys. **23**, 199 (1971); Commun. Math. Phys. **35**, 49 (1974).

Trivector Fourier transformation and electromagnetic field

Bernard Jancewicz
Institute of Theoretical Physics, University of Wroclaw, Cybulskiego 36, PL-50-205 Wroclaw, Poland

(Received 28 March 1989; accepted for publication 28 March 1990)

A new kind of Fourier transformation is proposed for distributions taking values in the Clifford algebra of three-dimensional space, with the unit imaginary replaced by the unit trivector. The transformation is used to introduce special distributions that describe the free electromagnetic field.

I. INTRODUCTION

The real *Clifford algebra* of n-dimensional Euclidean space E^n [1-3] is 2^n-dimensional associative algebra with unity, containing $\mathbb{R}$ and E^n as subspaces so that the condition

$$ab + ba = 2a\cdot b, \tag{1}$$

is satisfied for all $a, b \in E^n$, where $a\cdot b$ denotes the scalar product in E^n.

Let $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ be an orthonormal basis in E^3. The Clifford algebra of E^3 is called the *Pauli algebra.*[2] Denote $\mathbf{e}_{ij} = \mathbf{e}_i\mathbf{e}_j$ for $i \neq j$ and $\mathbf{e}_{123} = \mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$. The sets

$$\{1\},\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\},\{\mathbf{e}_{12},\mathbf{e}_{23},\mathbf{e}_{13}\},\{\mathbf{e}_{123}\}, \tag{2}$$

span the subspaces of scalars, vectors, bivectors, and trivectors, respectively, or, in other words, *multivectors* of zeroth, first, second, and third grade. The elements of the basis (2) are written symbolically as $\mathbf{e}_A$, where A is an ordered set of indices, and $\mathbf{e}_\phi = 1$. An arbitrary element X of the algebra, that is a combination $X = \Sigma_A X_A e_A$, $X_A \in \mathbb{R}$, of multivectors of various grades, is referred to as the *Clifford number.* The *magnitude* $|X|$ of X is defined by

$$|X|^2 = \sum_A X_A^2, \tag{3}$$

which for vectors coincides with magnitude determined by the scalar product.

Pauli algebra furnishes a good language for describing the classical electromagnetic field.[2,4] This is done by a function F of the space-time variables $\mathbf{r}$, $x_0 = ct$, which assumes its values as sums of vectors and bivectors only. Moreover, the so-called *regularity condition* should be satisfied by

$$\nabla F + \frac{\partial F}{\partial x_0} = 0,$$

which is equivalent to the *Maxwell equations* in regions devoid of electric charges and currents in a given uniform, isotropic and stationary medium. Here, $\nabla = \mathbf{e}_i\,\partial/\partial x_i$ acts as differentiation and Clifford algebra multiplication by $\mathbf{e}_i$.

This paper uses the results of Ref. 4, Chap. 4, where it has been shown that the expressions

$$e^{\pm I(\mathbf{k}x_0 - \mathbf{k}\cdot\mathbf{r})}N, \tag{4}$$

$$e^{\pm I[kx_0 - \mathbf{n}(\mathbf{k}\cdot\mathbf{r})]}N, \tag{5}$$

for any constant vector $\mathbf{k} \in E^3$, $k = |\mathbf{k}|$, $\mathbf{n} = \mathbf{k}/k$, where $I = \mathbf{e}_{123}$ and N is a Clifford number such that $N\mathbf{k} = -\mathbf{k}N$, correspond to plane electromagnetic waves that are generalizations of the circularly polarized wave. The solutions (4) are superpositions of two circularly polarized waves of opposite polarizations (left and right), traveling in opposite directions $\pm\mathbf{n}$—they are called waves with *round polarization.* The expressions (5), called waves with *spiral polarization,* are sums of two circularly polarized waves of the same (left or right) polarizations, traveling in opposite directions $\pm\mathbf{n}$.

Since the unit trivector I commutes with all Clifford numbers and $I^2 = -1$, the expression $e^{I\mathbf{k}\cdot\mathbf{r}}$ can play the role of $e^{i\mathbf{k}\cdot\mathbf{r}}$ for i being the imaginary unit of the complex numbers. Therefore, we propose to consider an integral transformation with the exponentials $e^{I\mathbf{k}\cdot\mathbf{r}}$ that resembles the well-known Fourier transformation—we call it *trivector Fourier transformation.* We apply this to *Pauli algebra valued distributions* introduced in Sec. II in close analogy to the tempered distributions of Schwartz.[5] Special kinds of such distributions, considered in Sec. III, correspond to the electromagnetic fields in empty space (we call them free electromagnetic fields)—they can be treated as superpositions of fields (4).

II. DISTRIBUTIONS

The Pauli algebra can be treated as an eight-dimensional normed space equipped with norm (3). Each finite-dimensional normed space is complete, hence we ascertain that the Pauli algebra is a complete normed space, i.e., a *Banach space.* But the question arises whether it is a *Banach algebra.* According to Ref. 6 the condition $|X_1X_2| \leqslant |X_1||X_2|$ should be satisfied for all elements X_1, X_2 of the algebra. This condition, however, generally is not satisfied for the Pauli algebra with norm (3). Yet one may introduce another norm. Each Clifford number Y can be treated as a linear operator acting according to $X \to YX$ and the following *operator norm*:

$$\|Y\| = \sup_{|X|=1} |YX|, \tag{6}$$

can be defined. This expression has all the desired properties of the norm and additionally satisfies

$$\|Y_1Y_2\| \leqslant \|Y_1\|\|Y_2\| \tag{7}$$

(see Theorem 10.2 in Ref. 6). Thus we may claim that the Pauli algebra with norm (6) is a *Banach algebra.*

As an easy exercise, one may verify that $|e^{I\alpha}X| = |X|$ for any $\alpha \in \mathbb{R}$ and any Clifford number X, hence it follows from (6) that $\|e^{I\alpha}\| = 1$.

Now we consider functions φ of three variables $\mathbf{r} = (x_1,x_2,x_3)$ taking values in the Pauli algebra. We call φ

 0022-2488/90/081847-06$03.00

a *rapidly decreasing function* if it is differentiable arbitrarily many times and for an arbitrary multi-index $q = (q_1, q_2, q_3)$ with natural q_i satisfies the condition

$$\sup_{|q|<n} \sup_{\mathbf{r}\in E^3} (1 + |\mathbf{r}|^2)^m \|\varphi^{(q)}\| < \infty,$$

for arbitrary natural n, m, where $|q| = q_1 + q_2 + q_3$,

$$\varphi^{(q)} = \frac{\partial^{|q|}\varphi}{\partial x_1^{q_1}\partial x_2^{q_2}\partial x_3^{q_3}}, \tag{8}$$

and $\|\varphi\|$ is the norm (6) of the Clifford number φ. The set $\mathcal{S}$ of all rapidly decreasing functions is a linear space. The topology in $\mathcal{S}$ is defined in close analogy to the classical one:[5] A sequence of functions $\varphi_n \in \mathcal{S}$ is convergent to zero if for arbitrary multi-indices p, q the sequence $x_1^{p_1} x_2^{p_2} x_3^{p_3} \varphi_n^{(q)}$ tends to zero uniformly on E^3.

A linear continuous mapping from $\mathcal{S}$ into the Pauli algebra is called the *Pauli algebra valued tempered distribution*, or for short the *distribution*. The value of F on a function φ is denoted (φ, F). The following integral form for distributions will be used:

$$(\varphi, F) = \int d^3\mathbf{r}\, \varphi(\mathbf{r}) F(\mathbf{r}),$$

in analogy to what is often done for ordinary distributions. A distribution F is *k-vector valued* if its value on each scalar function φ is a k-vector.

The qth *partial derivative of a distribution* F is defined by

$$(\varphi, F^{(q)}) = (-1)^{|q|}(\varphi^{(q)}, F),$$

where $\varphi^{(q)}$ is given by (8). One may check that this is a distribution. The *Clifford derivative of a distribution* F is defined by

$$(\varphi, \nabla F) = -(\varphi\nabla, F) = -\left(\frac{\partial\varphi}{\partial x_i}\mathbf{e}_i, F\right),$$

with the summation convention over repeated indices. The Clifford derivative coincides with the gradient when acting on a scalar-valued distribution, but is a combination of the divergence and curl for a vector-valued distribution.

Now we introduce the main notion. The mapping $\varphi \to \hat{\varphi}$ of $\varphi \in \mathcal{S}$, defined by

$$\hat{\varphi}(\mathbf{k}) = (2\pi)^{-3/2} \int d^3\mathbf{r}\, \varphi(\mathbf{r}) e^{-I\mathbf{k}\cdot\mathbf{r}},$$

is called the *trivector Fourier transformation*. The novelty in comparison with the usual Fourier transformation is the replacement of the complex imaginary i by the unit trivector $I = \mathbf{e}_{123}$. Function $\hat{\varphi}$ is called the *trivector Fourier transform* of φ. One may check, similarly as in Ref. 5 for ordinary complex functions, that $\hat{\varphi} \in \mathcal{S}$ with respect to the variable $\mathbf{k} \in E^3$. The commutativity of I with all Clifford numbers and the fact that $\|e^{I\mathbf{k}\cdot\mathbf{r}}\| = 1$ are essential for this. These facts also allow us to apply all theorems and results established for ordinary complex-valued distributions.

An inverse of the trivector Fourier transformation exists and is given by $\psi \to \tilde{\psi}$, where

$$\tilde{\psi}(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, \psi(\mathbf{k}) e^{I\mathbf{k}\cdot\mathbf{r}}.$$

As one can easily check, the identity

$$(\varphi\nabla)^{\sim}(\mathbf{k}) = -\tilde{\varphi}(\mathbf{k}) I\mathbf{k}, \tag{9}$$

is satisfied.

We define the *trivector Fourier transform* $\hat{F}$ *of a distribution* F by

$$(\psi, \hat{F}) = (\hat{\psi}, F), \tag{10}$$

which can be written in the integral form:

$$\int d^3\mathbf{k}\, \psi(\mathbf{k})\hat{F}(\mathbf{k}) = \int d^3\mathbf{r}\, \hat{\psi}(\mathbf{r}) F(\mathbf{r}).$$

This implies that $\hat{F}$ can be written symbolically as

$$\hat{F}(\mathbf{k}) = (2\pi)^{-3/2} \int d^3\mathbf{r}\, e^{-I\mathbf{k}\cdot\mathbf{r}} F(\mathbf{r}).$$

The inverse of this transformation is given by $G \to \tilde{G}$, where

$$\tilde{G}(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, e^{I\mathbf{k}\cdot\mathbf{r}} G(\mathbf{k}). \tag{11}$$

Equation (10) has its analog:

$$(\varphi, \tilde{G}) = (\tilde{\varphi}, G). \tag{12}$$

The distributions can be multiplied by functions of a special class. We say that an infinitely differentiable Pauli algebra valued function α belongs to $\mathcal{O}_M$ if each its derivative $\alpha^{(q)}$ is majorized by a polynomial, i.e.,

$$\|\alpha^{(q)}(\mathbf{k})\| \leqslant C(1 + |\mathbf{k}|^m),$$

where $C > 0$ and natural m depend on q. Then, due to (7), for any $\varphi \in \mathcal{S}$ the product $\varphi\alpha$ belongs to $\mathcal{S}$ and we define the product αF for a distribution F as

$$(\varphi, \alpha F) = (\varphi\alpha, F). \tag{13}$$

Let α_{x_0} be a function defined by

$$\alpha_{x_0}(\mathbf{k}) = e^{-I\mathbf{k}x_0} = \cos kx_0 - I\mathbf{n} \sin kx_0, \tag{14}$$

where $\mathbf{n} = \mathbf{k}/k$. Its value is a combination of scalar and bivector, it is infinitely differentiable:

$$\frac{\partial^{|q|}\alpha_{x_0}(\mathbf{k})}{\partial k_1^{q_1}\partial k_2^{q_2}\partial k_3^{q_3}} = (-I\mathbf{e}_1 x_0)^{q_1}(-I\mathbf{e}_2 x_0)^{q_2}(-I\mathbf{e}_3 x_0)^{q_3} e^{I\mathbf{k}x_0},$$

and $\|\alpha_{x_0}^{(q)}(\mathbf{k})\| = |x_0|^{|q|}$, hence $\alpha_{x_0} \in \mathcal{O}_M$. Therefore, we may consider the product $\alpha_{x_0} G$ for any $G \in \mathcal{S}$.

III. REGULAR TIME-DEPENDENT DISTRIBUTIONS

We assume that the distributions may depend on a real parameter x_0 to be interpreted as time. We assume, moreover, that such time-dependent distributions F_{x_0} are differentiable with respect to x_0 in the sense that the function $x_0 \to (\varphi, F_{x_0})$ is differentiable for each $\varphi \in \mathcal{S}$. We define symbol $\partial_0 F_{x_0} = \partial F_{x_0}/\partial x_0$ by

$$(\varphi, \partial_0 F_{x_0}) := \frac{d}{dx_0}(\varphi, F_{x_0}).$$

This equality determines obviously a linear functional on $\mathcal{S}$. If it is continuous on $\mathcal{S}$, it defines a tempered distribution—in this case we call F_{x_0} the *time-dependent differentiable distribution* and call $\partial_0 F_{x_0}$ the *derivative of* F_{x_0}.

Now we define the following time-dependent distribution:

$$F_{x_0} = (\alpha_{x_0} G)^{\sim}, \tag{15}$$

for α_{x_0} given in (14). Using the symbolic expression (11), one can write this as

$$F_{x_0}(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, e^{I\mathbf{k}\cdot\mathbf{r}} e^{-I\mathbf{k}x_0} G(\mathbf{k}). \tag{16}$$

Proposition 1: F_{x_0} defined in (16) is a time-dependent differentiable distribution.

Proof: We have by virtue of (15) and (12)

$$\frac{d}{dx_0}(\varphi, F_{x_0}) = \frac{d}{dx_0}(\varphi, (\alpha_{x_0} G)^{\sim}) = \frac{d}{dx_0}(\tilde{\varphi}\alpha_{x_0}, G).$$

This derivative exists for each $\varphi \in \mathcal{S}$, is equal to $(\tilde{\varphi}\,\partial_0 \alpha_{x_0}, G)$ and defines a linear and continuous functional on $\mathcal{S}$, hence the statement is true. □

If a time-dependent differentiable distribution F_{x_0} satisfies the equation

$$(\varphi, \partial_0 F_{x_0}) + (\varphi, \nabla F_{x_0}) = 0, \tag{17}$$

for any $\varphi \in \mathcal{S}$, we call F_{x_0} the *regular time-dependent distribution* or, for short, the *regular distribution*. When F_{x_0} is represented by a locally integrable function $F(x_0, \mathbf{r})$, Eq. (17) implies

$$\partial_0 F(x_0, \mathbf{r}) + \nabla F(x_0, \mathbf{r}) = 0.$$

Clifford algebra valued functions satisfying this condition are referred to as *regular*[7] or *monogenic*.[8]

Proposition 2: F_{x_0} given by (16) is a regular distribution.

Proof: We have

$$\begin{aligned}
&(\varphi, \partial_0 F_{x_0}) + (\varphi, \nabla F_{x_0}) \\
&\quad = \frac{d}{dx_0}(\varphi, F_{x_0}) - (\varphi\nabla, (\alpha_{x_0} G)^{\sim}) \\
&\quad = (\tilde{\varphi}\,\partial_0 \alpha_{x_0}, G) - ((\varphi\nabla)^{\sim} \alpha_{x_0}, G) \\
&\quad = (\tilde{\varphi}\,\partial_0 \alpha_{x_0} - (\varphi\nabla)^{\sim} \alpha_{x_0}, G).
\end{aligned} \tag{18}$$

Now, using (9) and (14),

$$\begin{aligned}
&\tilde{\varphi}(\mathbf{k})\partial_0 \alpha_{x_0}(\mathbf{k}) - (\varphi\nabla)^{\sim}(\mathbf{k})\alpha_{x_0}(k) \\
&\quad = -\tilde{\varphi}(\mathbf{k}) I\mathbf{k} e^{-I\mathbf{k}x_0} + \tilde{\varphi}(\mathbf{k}) I\mathbf{k} e^{-I\mathbf{k}x_0} = 0.
\end{aligned}$$

This, substituted into (18), yields

$$\partial_0 F_{x_0} + \nabla F_{x_0} = 0, \tag{19}$$

in the distribution sense. □

The value of F_{x_0} at $x_0 = 0$, denoted F, is a tempered distribution. In this case F_{x_0} is called a *regular extension of* F. Since each tempered distribution F has its Fourier transform $\hat{F}$, related through the formula

$$F(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, e^{I\mathbf{k}\cdot\mathbf{r}} \hat{F}(\mathbf{k}), \tag{20}$$

one can substitute $G = \hat{F}$ into (16) and obtain, in this manner, the regular extension of F:

$$F_{x_0}(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, e^{I\mathbf{k}\cdot\mathbf{r}} e^{-I\mathbf{k}x_0} \hat{F}(\mathbf{k}). \tag{21}$$

For the three-dimensional Dirac delta $\delta^3(\mathbf{r})$ its Fourier transform is $(2\pi)^{-3/2}$, hence we can write regular extension of the Dirac delta:

$$\delta^3_{x_0}(\mathbf{r}) = (2\pi)^{-3} \int d^3\mathbf{k}\, e^{I\mathbf{k}\cdot\mathbf{r} - I\mathbf{k}x_0}. \tag{22}$$

We shift to Appendix A the calculations leading to the result

$$\delta^3_{x_0}(\mathbf{r}) = \frac{1}{4\pi}\left(\nabla - \frac{\partial}{\partial x_0}\right)\frac{1}{r}\left[\delta(r + x_0) - \delta(r - x_0)\right]. \tag{23}$$

The regular extension of the two-dimensional Dirac delta is

$$\delta^2_{x_0}(\mathbf{r}) = -\left[(\operatorname{sgn} x_0)/2\pi\right](x_0 + \mathbf{r})\mathrm{PV}(x_0^2 - r^2)^{-3/2}, \tag{24}$$

as shown in Appendix B. Here, $\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ and PV means the principal value. The regular extension of the one-dimensional Dirac delta is

$$\delta_{x_0}(\mathbf{r}) = \tfrac{1}{2}(1 + \mathbf{e}_1)\delta(x_0 - x_1) + \tfrac{1}{2}(1 - \mathbf{e}_1)\delta(x_0 + x_1), \tag{25}$$

which has been demonstrated in Ref. 4, Chap. 4.

Comparison of formulas (20) and (22) helps us to notice what is the Fourier transform of the regular extension of three-dimensional delta: $(\delta^3_{x_0})^{\wedge}(\mathbf{k}) = (2\pi)^{-3/2} e^{-I\mathbf{k}x_0}$. Now Eq. (21) can be written as

$$F_{x_0}(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, e^{I\mathbf{k}\cdot\mathbf{r}} (2\pi)^{3/2} (\delta^3_{x_0})^{\wedge}(\mathbf{k}) \hat{F}(\mathbf{k}),$$

and shows that the Fourier transform of F_{x_0} is the product of two distributions: $(F_{x_0})^{\wedge} = (2\pi)^{3/2}(\delta^3_{x_0})^{\wedge}\hat{F}$. The distribution $(\delta^3_{x_0})^{\wedge}$ belongs to $\mathcal{O}_M$, hence $\delta^3_{x_0} \in \mathcal{O}'_C$, so we can use the famous "multiplication and convolution theorem" (Ref. 5, Chap. VII, Theorem XV) and write $F_{x_0} = (2\pi)^{3/2}\delta^3_{x_0} * F$, or more explicitly:

$$F_{x_0}(\mathbf{r}) = \int d^3\mathbf{r}'\, \delta^3_{x_0}(\mathbf{r} - \mathbf{r}') F(\mathbf{r}'). \tag{26}$$

We can interpret formula (26) as the solution of the *Cauchy problem* for the operator $\partial_0 + \nabla$: find a solution of Eq. (19) when the initial value of F_{x_0} is F. It is the reason why $\delta^3_{x_0}(\mathbf{r} - \mathbf{r}')$ is called the *Cauchy kernel* of the operator $\partial_0 + \nabla$.

It is interesting to notice that the cases of odd and even dimensions differ considerably. The regular extensions of one- and three-dimensional deltas have their support on the light cone, i.e., at points $r = |x_0|$, whereas the extension of the two-dimensional delta has its support in the full (solid) light cone, that is at points $r \leqslant |x_0|$. This is a manifestation of more general regularity for the *Cauchy kernels* as discussed in Ref. 9. A similar distinction between even and odd number of independent variables occurs also for traditional hyperbolic partial differential equations as one can notice for instance in a classical work of Hadamard.[10]

IV. THE ELECTROMAGNETIC FIELD

As shown in Refs. 2, 4, and 7, the regularity condition (17) or (19) is equivalent to homogeneous Maxwell equations if the distributions consist only of vector and bivector parts, hence we claim that such distributions describe free

electromagnetic fields. Expression (21) helps us to see that F_{x_0} contains only vectors and bivectors iff $e^{-I\mathbf{k}x_0}\hat{F}$ fulfills the same condition. After using (14) and decomposing $\hat{F}$ into its vector and bivector parts $\hat{F}=\mathbf{G}_v+\mathbf{G}_b$ we get

$$\begin{aligned} e^{-I\mathbf{k}x_0}\hat{F}(\mathbf{k}) = & -I(\mathbf{n}\wedge\mathbf{G}_b)\sin kx_0 \\ & +[\mathbf{G}_v\cos kx_0 - I(\mathbf{n}\wedge\mathbf{G}_v)\sin kx_0] \\ & +[\mathbf{G}_b\cos kx_0 - I(\mathbf{n}\cdot\mathbf{G}_b)\sin kx_0] \\ & -I(\mathbf{n}\cdot\mathbf{G}_v)\sin kx_0. \end{aligned}$$

The first term is a scalar, the last term is a trivector—they should be absent in this combination, which implies

$$\mathbf{n}\wedge\mathbf{G}_b=0, \quad \mathbf{n}\cdot\mathbf{G}_v=0. \tag{27}$$

This can be united into a single condition:

$$\mathbf{k}\hat{F}(\mathbf{k}) = -\hat{F}(\mathbf{k})\mathbf{k}. \tag{28}$$

The factor $e^{I\mathbf{k}\cdot\mathbf{r}}$ does not produce any scalar nor trivector terms, so it does not impose any new condition on $\hat{F}$. Thus we may summarize our present discussion: *For any vector-plus-bivector valued distribution $\hat{F}$ satisfying* (28) *the expression* (21) *yields a time-dependent distribution F_{x_0} describing a free electromagnetic field.*

As proven in Ref. 11, Sec. 16.5, the time-dependent electromagnetic field F_{x_0} is unique for any initial electromagnetic field F and is related to F through $F_{x_0}(\mathbf{r})|_{x_0=0}=F(\mathbf{r})$. We may express this in other words: *Each time-dependent free electromagnetic field is the regular extension of its initial value at $x_0=0$.* This is uniqueness of the Cauchy problem solution for the operator $\partial_0+\nabla$. In this manner the prescription of obtaining time-dependent electromagnetic fields through (21) or, equivalently, through (26), is unique.

In particular, when

$$\hat{F}(\mathbf{k})=(2\pi)^{3/2}\delta^3(\mathbf{k}-\mathbf{q})N, \tag{29}$$

one gets

$$F_{x_0}(\mathbf{r})=e^{-I\mathbf{q}x_0+I\mathbf{q}\cdot\mathbf{r}}N, \tag{30}$$

which is the example (4) of (not necessarily traveling) plane harmonic electromagnetic wave with round polarization in direction $-I\mathbf{q}/q$, discussed in Ref. 4. By demanding $\mathbf{q}N=-N\mathbf{q}$, condition (28) is satisfied due to the fact that (29) has its support in single point $\mathbf{k}=\mathbf{q}$. If, in addition, N satisfies the eigenequation

$$\mathbf{q}N=qN, \tag{31}$$

for $q=|\mathbf{q}|$, one gets $F_{x_0}(\mathbf{r})=e^{-I(qx_0-\mathbf{q}\cdot\mathbf{r})}N$, which is the traveling (in direction $\mathbf{q}/q$) plane wave with the left circular polarization. Alternatively, if

$$\mathbf{q}N=-qN, \tag{32}$$

one gets $F_{x_0}(\mathbf{r})=e^{I(qx_0+\mathbf{q}\cdot\mathbf{r})}N$, which is the traveling (in direction $-\mathbf{q}/q$) plane wave with the right circular polarization.

In this manner we infer from (21) that each free electromagnetic field is a (generalized, because integral) superposition of plane waves with the round polarization.

If neither of conditions (31) and (32) is met, wave (30) is a superposition of two waves traveling in opposite directions $\pm\mathbf{q}/q$. This occurs in spite of the fact that its Fourier transform (29) has the support in one point $+\mathbf{q}$ only. This is something new in comparison with the situation known from quantum mechanics: If a wave function has its support at a single point in the space of wave vectors, it describes a state with a definite momentum.

The regular extensions (23)–(25) of various Dirac deltas do not represent any electromagnetic fields because they are scalar-plus-vector valued. One can improve this by multiplying them from the right by constant vector-plus-bivector Clifford numbers N [constant ones, since only such elements do not destroy the regularity condition (17)]. The constant is also present at the Fourier transform $\hat{F}N$ of the initial distribution FN and condition (28) takes the form $\mathbf{k}\hat{F}(\mathbf{k})N=-\hat{F}(\mathbf{k})N\mathbf{k}$. For all the named Dirac deltas $\hat{F}$ are scalar valued, so we obtain

$$\mathbf{k}N=-N\mathbf{k}, \tag{33}$$

for all $\mathbf{k}\in\operatorname{supp}\hat{F}$.

For the three-dimensional delta its Fourier transform is $\widehat{(\delta^3)}(\mathbf{k})=(2\pi)^{-3/2}$, so $\operatorname{supp}\widehat{(\delta^3)}=E^3$ and, therefore, one cannot find constant vector-plus-bivector Clifford number N satisfying (33) for all $\mathbf{k}\in\operatorname{supp}\widehat{(\delta^3)}$. [Remember that, as seen from (27), $\mathbf{N}_v$ must be perpendicular to $\mathbf{k}$ and $\mathbf{N}_b$ parallel to $\mathbf{k}$.] Hence we conclude that the regular extension of the three-dimensional Dirac delta can not be used to represent any electromagnetic field. (This fact probably has a bearing on the impossibility of localization of photons in a point, see Refs. 12–14.)

The situation is better for lower-dimensional Dirac deltas. For instance, Fourier transform of $\delta^2(x_1,x_2)$:

$$\widehat{(\delta^2)}(\mathbf{k})=(2\pi)^{-1/2}\delta(k_3), \tag{34}$$

has its support on the (k_1,k_2) plane, so after multiplying (34) by a constant element $N=\mathbf{N}_v+\mathbf{N}_b$ with $\mathbf{N}_v\|\mathbf{e}_3$ and $\mathbf{N}_b\perp\mathbf{e}_3$ one gets $\mathbf{k}N=-N\mathbf{k}$ and then (28) for all $\mathbf{k}\in\operatorname{supp}\widehat{(\delta^2)}$, which is sufficient. In this way we obtain the regular extension of $\delta^2 N$ as (24) multiplied by the above N, and this describes a free electromagnetic field.

The one-dimensional Dirac delta $\delta(x_1)$ has the Fourier transform $\hat{\delta}(\mathbf{k})=(2\pi)^{1/2}\delta(k_2,k_3)$ with the support on the k_1 axis. In this case the element $N=\mathbf{N}_v+\mathbf{N}_b$ with $\mathbf{N}_v\perp\mathbf{e}_1$, $\mathbf{N}_b\|\mathbf{e}_1$ ensures (33) and (28) for all $\mathbf{k}\in\operatorname{supp}\hat{\delta}$. The regular extension of $\delta(x_1)N$ is (25) multiplied by N and is a particular case of the plane electromagnetic field discussed in Ref. 4, Chap. 4.

V. CONCLUSION

Due to the fact that free electromagnetic fields can be identified with vector-plus-bivector valued time-dependent regular distributions, we have shown that they can be uniquely represented as regular extensions of their initial values at time $x_0=0$. Two ways of expressing them are given: first one based on the trivector Fourier transformation formula (21), second one using the convolution formula (26) of the initial field with the regular extension of the three-dimensional Dirac delta.

The lower-dimensional Dirac deltas can be immediately used to describe electromagnetic fields after multiplying them by an appropriate constant Clifford number. These

time-dependent distributions can be interpreted as electromagnetic fields initially localized on planes or straight lines.

When contemplating expansion (16) or (21) of a time-dependent electromagnetic field F_{x_0} into plane waves with the round polarization, one naturally asks the question: Can similar expansion:

$$T_{x_0}(\mathbf{r}) = (2\pi)^{-3/2} \int d^3\mathbf{k}\, e^{-I\mathbf{n}(\mathbf{k}\cdot\mathbf{r})} e^{Ikx_0} G(\mathbf{k}), \qquad (35)$$

be considered into plane waves with the spiral polarization? The answer to this question must be negative, because the mapping $G \to T_{x_0}$ defined by (35) is not one-to-one. It yields the same image for G and G', where $G'(\mathbf{k}) = G(-\mathbf{k})$.

ACKNOWLEDGMENT

I am grateful to Franciscus Sommen for discussions on regular extensions of the Dirac deltas and to Beata Wilczyńska for conversations on the early version of this work.

This work is supported by the Ministry of National Education, project CPBP 01.03.

APPENDIX A

Due to (14), the regular extension of δ^3 can be written as

$$\delta^3{}_{x_0}(\mathbf{r}) = (2\pi)^{-3} \int d^3\mathbf{k} \left(\cos kx_0 - \frac{I\mathbf{k}}{k} \sin kx_0\right) e^{I\mathbf{k}\cdot\mathbf{r}}.$$

Extract from this the scalar and vector parts:

$$A(x_0,\mathbf{r}) = (2\pi)^{-3} \int d^3\mathbf{k} \cos kx_0 e^{I\mathbf{k}\cdot\mathbf{r}},$$

$$\mathbf{B}(x_0,\mathbf{r}) = -I(2\pi)^{-3} \int d^3\mathbf{k} \frac{\mathbf{k} \sin kx_0}{k} e^{I\mathbf{k}\cdot\mathbf{r}}.$$

Change Cartesian coordinates $k_1 k_2, k_3$ into spherical ones k, ϑ, ϕ, where ϑ is the angle between $\mathbf{k}$ and $\mathbf{r}$:

$$A = (2\pi)^{-3} \int dk \int d\vartheta \int d\phi\, k^2 \sin\vartheta \cos kx_0 e^{Ikr\cos\vartheta}$$

$$= (2\pi)^{-2} \int dk\, k^2 \cos kx_0 \int d\vartheta \sin\vartheta e^{Ikr\cos\vartheta}$$

$$= \frac{2}{(2\pi)^2 r} \int_0^\infty dk\, k \cos kx_0 \sin kr$$

$$= \frac{1}{2(2\pi)^2 r} \left[\int_{-\infty}^{\infty} dk\, k \sin k(r - x_0) + \int_{-\infty}^{\infty} dk\, k \sin k(r + x_0)\right].$$

Use the known identities from the distribution theory

$$A = -\frac{1}{4\pi r} [\delta'(r + x_0) + \delta'(r - x_0)]$$

$$= -\frac{1}{4\pi r} \frac{\partial}{\partial x_0} [\delta(r + x_0) - \delta(r - x_0)].$$

We similarly integrate the vector part

$$\mathbf{B} = -(2\pi)^{-3} \nabla \int d^3\mathbf{k} \frac{\sin kx_0}{k} e^{I\mathbf{k}\cdot\mathbf{r}}$$

$$= -(2\pi)^{-3} \nabla \int dk \int d\vartheta \int d\phi\, k^2 \times \sin\vartheta \frac{\sin kx_0}{k} e^{Ikr\cos\vartheta}$$

$$= -(2\pi)^{-2} \nabla \int dk\, k \sin kx_0 \int d\vartheta \sin\vartheta\, e^{Ikr\cos\vartheta}$$

$$= -\nabla \frac{2}{(2\pi)^2 r} \int_0^\infty dk \sin kx_0 \sin kr$$

$$= -\nabla \frac{1}{2(2\pi)^2 r} \left[\int_{-\infty}^{\infty} dk \cos k(r - x_0) - \int_{-\infty}^{\infty} dk \cos k(r + x_0)\right].$$

Again, by known facts from the distribution theory;

$$\mathbf{B} = \nabla \frac{1}{4\pi r} [\delta(r + x_0) - \delta(r - x_0)].$$

Thus at last

$$\delta^3{}_{x_0} = A + \mathbf{B}$$

$$= \frac{1}{4\pi} \left(\nabla - \frac{\partial}{\partial x_0}\right) \frac{1}{r} [\delta(r + x_0) - \delta(r - x_0)].$$

APPENDIX B

We apply (26) to find the regular extension of two-dimensional Dirac delta:

$$\delta^2{}_{x_0}(\mathbf{r}) = \int d^3\mathbf{r}'\, \delta^3{}_{x_0}(\mathbf{r} - \mathbf{r}')\delta^2(\mathbf{r}'). \qquad (36)$$

Explicitly,

$$\delta^2{}_{x_0} = \frac{-1}{4\pi} \int d^3\mathbf{r}' \left\{\left(\nabla' + \frac{\partial}{\partial x_0}\right) \times \frac{\delta(|\mathbf{r} - \mathbf{r}'| + x_0) - \delta(|\mathbf{r} - \mathbf{r}'| - x_0)}{|\mathbf{r} - \mathbf{r}'|}\right\} \delta^2(\mathbf{r}'). \qquad (37)$$

Relation (36) becomes trivial for $x_0 = 0$. If $x_0 > 0$, only second term in the numerator gives nonzero contribution:

$$\delta^2{}_{x_0}(\mathbf{r}) = \frac{-1}{4\pi} \int \frac{d^3\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \delta'(|\mathbf{r} - \mathbf{r}'| - x_0)\delta(x_1')\delta(x_2') + \frac{1}{4\pi} \int d^3\mathbf{r}' \left\{\left(\mathbf{e}_1 \frac{\partial}{\partial x_1'} + \mathbf{e}_2 \frac{\partial}{\partial x_2'}\right) \times \frac{\delta(|\mathbf{r} - \mathbf{r}'| - x_0)}{|\mathbf{r} - \mathbf{r}'|}\right\} \delta(x_1')\delta(x_2').$$

The last term $\mathbf{e}_3 \partial/\partial x_3'$ in the nabla was omitted since it gives zero after integration. We select the scalar and vector parts:

$$A = \frac{-1}{4\pi} \int_{-\infty}^{\infty} \frac{dz'}{[x^2 + y^2 + (z' - z)^2]^{1/2}} \times \delta'([x^2 + y^2 + (z' - z)^2]^{1/2} - x_0),$$

$$\mathbf{B} = \frac{-1}{4\pi} \int_{-\infty}^{\infty} dz' \left(\mathbf{e}_1 \frac{\partial}{\partial x} + \mathbf{e}_2 \frac{\partial}{\partial y}\right) \times \frac{\delta([x^2 + y^2 + (z' - z)^2]^{1/2} - x_0)}{[x^2 + y^2 + (z' - z)^2]^{1/2}}.$$

The integrands are symmetric with respect to point $z' = z$, hence

$$A = \frac{-1}{2\pi}\int_z^\infty \frac{dz'}{[x^2+y^2+(z'-z)^2]^{1/2}} \times \delta'([x^2+y^2+(z'-z)^2]^{1/2} - x_0),$$

$$\mathbf{B} = \frac{-1}{2\pi}\int_z^\infty dz'\left(\mathbf{e}_1\frac{\partial}{\partial x} + \mathbf{e}_2\frac{\partial}{\partial y}\right) \times \frac{\delta([x^2+y^2+(z'-z)^2]^{1/2} - x_0)}{[x^2+y^2+(z'-z)^2]^{1/2}}.$$

It is clear that for $x^2+y^2 > x_0^2$ both expressions are zero, so the result should be multiplied by $\theta(x_0^2 - x^2 - y^2)$ where θ is the Heaviside or step function:

$$\theta(x) = \begin{cases} 1, & \text{for } x \geqslant 0, \\ 0, & \text{for } x < 0. \end{cases}$$

Change the variables:

$$u = \sqrt{x^2+y^2+(z'-z)^2}, \quad du/\sqrt{u^2-x^2-y^2} = dz'/u.$$

Then

$$A = \frac{-1}{2\pi}\int_0^\infty \frac{du}{(u^2-x^2-y^2)^{1/2}}\delta'(u-x_0) = \frac{-1}{2\pi}\int_0^\infty \frac{u\,du}{(u^2-x^2-y^2)^{3/2}}\delta(u-x_0).$$

Thus we get the result for the scalar part

$$A = \frac{-x_0}{2\pi(x_0^2-x^2-y^2)^{3/2}}\theta(x_0^2-x^2-y^2).$$

We proceed to the vector part:

$$\mathbf{B} = \frac{-1}{2\pi}\int_z^\infty dz' \frac{\mathbf{e}_1 x + \mathbf{e}_2 y}{x^2+y^2+(z'-z)^2} \times \left\{\delta'([x^2+y^2+(z'-z)^2]^{1/2} - x_0) - \frac{\delta([x^2+y^2+(z'-z)^2]^{1/2} - x_0)}{[x^2+y^2+(z'-z)^2]^{1/2}}\right\}.$$

Change the variables as previously:

$$\mathbf{B} = \frac{\mathbf{e}_1 x + \mathbf{e}_2 y}{2\pi}\int_0^\infty \frac{du}{(u^2-x^2-y^2)^{1/2}} \times \left[\frac{1}{u}\delta(u-x_0) - \delta'(u-x_0)\right]$$

$$= \frac{\mathbf{e}_1 x + \mathbf{e}_2 y}{2\pi}\int_0^\infty du\left\{\frac{1}{u^2(u^2-x^2-y^2)^{1/2}} + \left[\frac{1}{u(u^2-x^2-y^2)^{1/2}}\right]'\right\}\delta(u-x_0)$$

$$= \frac{\mathbf{e}_1 x + \mathbf{e}_2 y}{2\pi}\left[\frac{1}{x_0^2(x_0^2-x^2-y^2)^{1/2}} - \frac{2x_0^2 - x^2 - y^2}{x_0^2(x_0^2-x^2-y^2)^{3/2}}\right]$$

$$= \frac{\mathbf{e}_1 x + \mathbf{e}_2 y}{2\pi}\frac{-1}{(x_0^2-x^2-y^2)^{3/2}}.$$

Hence the result for the vector part is

$$\mathbf{B} = -\frac{\mathbf{e}_1 x + \mathbf{e}_2 y}{2\pi(x_0^2-x^2-y^2)^{3/2}}\theta(x_0^2-x^2-y^2).$$

Having denoted $\mathbf{r} = \mathbf{e}_1 x + \mathbf{e}_2 y$ we write the net result:

$$\delta^2_{x_0}(\mathbf{r}) = \frac{-x_0 - \mathbf{r}}{2\pi(x_0^2-r^2)^{3/2}}\theta(x_0^2-r^2), \quad \text{for } x_0 > 0. \tag{38}$$

For $x_0 < 0$ the other term in the numerator of (37) remains, but this can be reduced to the previous case by writing $|\mathbf{r}-\mathbf{r}'| + x_0 = |\mathbf{r}-\mathbf{r}'| - |x_0|$ as the argument of the delta. This yields only opposite sign for the vector part, hence

$$\delta^2_{x_0}(\mathbf{r}) = \frac{-|x_0| + \mathbf{r}}{2\pi(x_0^2-r^2)^{3/2}}\theta(x_0^2-r^2), \quad \text{for } x_0 < 0.$$

After combining this with (38) we get

$$\delta^2_{x_0}(\mathbf{r}) = -\frac{x_0 + \mathbf{r}}{2\pi(x_0^2-r^2)^{3/2}}\operatorname{sgn} x_0\,\theta(x_0^2-r^2).$$

This result is not correct for distributions because of too large power $\frac{3}{2}$ in the denominator. One should rather use the principal value distribution

$$\mathrm{PV}(x^{-3/2}) = \lim_{\epsilon\to 0}\tfrac{1}{2}[(x+i\epsilon)^{-3/2} - (x-i\epsilon)^{-3/2}].$$

In this case the Heaviside function is not needed because the complex function z^α has singularities on the positive half-line. So $\mathrm{PV}(x^{-3/2})$ vanishes for negative values of x. In this manner we obtain at last

$$\delta^2_{x_0}(\mathbf{r}) = -(\operatorname{sgn} x_0/2\pi)(x_0 + \mathbf{r})\ \mathrm{PV}\ (x_0^2 - r^2)^{-3/2}.$$

[1] I. R. Porteous, *Topological Geometry* (Cambridge U.P., Cambridge, 1981), 2nd ed.
[2] D. Hestenes, *Space-Time Algebra* (Gordon and Breach, New York, 1966).
[3] D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus* (Reidel, Dordrecht, 1984).
[4] B. Jancewicz, *Multivectors and Clifford Algebra in Electrodynamics* (World Scientific, Singapore, 1988).
[5] L. Schwartz, *Theorie des Distributions* (Hermann, Paris, 1966).
[6] W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973), Chap. 10.
[7] K. Imaeda, *Quaternionic Formulation of Classical Electrodynamics* (Okayama Univ. of Science, Okayama, 1983).
[8] F. Brackx, R. Delanghe, and F. Sommen, *Clifford Analysis*, Research Notes in Mathematics Vol. 76 (Pitman, London, 1982).
[9] F. Sommen, Q. J. Pure Appl. Math. **62**, 293 (1988).
[10] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Yale U.P., New Haven, 1923).
[11] R. D. Richtmyer, *Principles of Advanced Mathematical Physics* (Springer, New York, 1978), Vol. 1.
[12] J. M. Jauch and C. Piron, Helv. Phys. Acta **40**, 559 (1967).
[13] A. Z. Jadczyk and B. Jancewicz, Bull. Acad. Pol. Sci. Ser. Math. Astr. Phys. **21**, 477 (1973).
[14] K. Kraus, in *The Uncertainty Principle and Foundations of Quantum Mechanics*, edited by V. C. Price and S. Chissick (Wiley, New York, 1977).

Representations of affine Kac–Moody algebras and the affine scalar product

Richard H. Capps
Department of Physics, Purdue University, West Lafayette, Indiana 47907

(Received 18 July 1989; accepted for publication 11 April 1990)

Simple procedures are given for finding all the dominant weights in a highest weight representation of an affine algebra, for finding the Weyl orbit of an arbitrary weight, and for determining whether or not any given weight is in any given representation. A simple definition of congruency is given that applies to all affine algebras. The standard indefinite scalar product is generalized; the generalization is used in the procedures.

I. INTRODUCTION

Affine Kac–Moody algebras are important in many branches of modern physics, including string theories, conformal field theories, and other field theories. This paper is concerned primarily with one type of irreducible representation (irrep) of simple affine algebras, highest weight irreps. In many papers such irreps are treated within the context of some field theoretic model, so that field theory concepts are used. The point of view taken here is that it is useful to construct and discuss irreps with purely mathematical concepts, in analogy with well-known procedures used for finite algebras. The approach used is motivated by Kac, who points out that the properties of an affine algebra may be determined from the generalized Cartan matrix.[1] When one uses this approach, the twisted affine algebras are not much more complicated than the untwisted algebras.

Every highest weight irrep except the identity irrep contains an infinite number of weights. Since the multiplicities of all weights in the same Weyl orbit are the same in any irrep, it is logical to organize the weights by their orbits. However, each irrep contains an infinite number of orbits, and each orbit contains an infinite number of weights. Therefore, even if one knows the basic construction rules,[2] one cannot construct an entire irrep, and it is difficult to picture an irrep, except in the simplest cases. One of the two main purposes of this paper is to help illuminate this situation by finding a simple procedure for answering the following key question: Given an arbitrary affine algebra, an arbitrary highest weight irrep Λ and an arbitrary weight M, is M in Λ? Of course the procedure given, when the order of operations is changed, provides a simple method for constructing the top weights of an irrep. We are not concerned here with the multiplicities of weights; multiplicity calculations are discussed in the literature.[3]

Every simple affine algebra $\mathscr{A}$ of rank $n-1$ is represented by an indecomposable, generalized Dynkin diagram with n vertices. If any one of these vertices R_i is deleted (with its connecting lines), the resulting diagram represents a finite subalgebra $\mathscr{S}_i$, called here a basic subalgebra. An irrep of $\mathscr{A}$ is a sum of irreps of any of the $\mathscr{S}_i$, and is most easily understood in terms of the $\mathscr{S}_i$. The second main purpose of this paper is to treat the n basic subalgebras on equal footing, as much as possible. This leads to a generalization of the standard affine scalar product.

Section II contains most of the basic concepts and formulas used. Two important concepts are introduced in Sec. III, extended subweights and the generalized affine scalar product. A simple criterion for the Weyl orbits (each characterized by its dominant weight) in a highest weight irrep is given in Sec. IV. Section V is concerned with the construction of Weyl orbits and the determination of the orbit of an arbitrary weight. The generalized scalar product is useful for these constructions. Section VI summarizes the procedure given for answering the key question mentioned earlier.

II. BASIC FORMALISM

A. Fundamental formulas for both finite and affine algebras

The basic rules given in this section can be found in many references, for example Cahn[4] and Slansky.[5] If there are n vertices in the Dynkin diagram for a simple finite or affine algebra, there are n simple roots, denoted by R_i. A weight vector M may be characterized by n integral coefficients m_j (called here Dynkin coefficients), defined by the scalar product equation,

$$m_j = \langle M,R_j\rangle (2/R_j^2). \tag{2.1}$$

A weight M is dominant iff all m_j are non-negative.

Some weights may be written as linear combinations of the simple roots. (In the case of finite algebras all weights may be written in this manner.) The root-basic components of such a weight Q are denoted by Q_i, i.e.,

$$Q = \sum_i Q_i R_i. \tag{2.2}$$

Substitution of Eq. (2.2) into Eq. (2.1) (with m_j and M replaced by q_j and Q) yields the relation

$$q_j = \sum_{i=1}^{n} Q_i A_{ij} \quad \text{or} \quad q = QA, \tag{2.3}$$

where q and Q in the last equation are row vectors, and A is the Cartan matrix, defined by $A_{ij} = \langle R_i,R_j\rangle (2/R_j^2)$. If the weight Q may be written as a linear combination of the simple roots, and M is any weight vector, it follows from Eqs. (2.1) and (2.2) that

$$\langle M,Q\rangle = \sum_i m_i Q_i \frac{R_i^2}{2}. \tag{2.4}$$

 0022-2488/90/081853-06$03.00

We take as a set of basis vectors the coroots $R_i^\vee$ associated with the simple roots, defined by $R_i^\vee = R_i(2/R_i^2)$. If the weight Q is a linear combination of the simple roots, the components $Q_i^\vee$ are defined by

$$Q = \sum_i Q_i^\vee R_i^\vee. \tag{2.5}$$

Thus $Q_i^\vee = Q_i(R_i^2/2)$. It is seen from Eq. (2.4) that

$$\langle M,Q\rangle = \sum_{i=1}^{n} m_i Q_i^\vee. \tag{2.6}$$

The analog of Eq. (2.3) is $q = Q^\vee S$, where S is the symmetric, metric matrix defined by $S_{ij} = \langle R_i^\vee, R_j^\vee\rangle$. The matrix S is related to the Cartan matrix by the matrix equation $S = 2(R^2)^{-1}A$, where R^2 is the diagonal positive-definite matrix with elements $(R^2)_{ij} = \delta_{ij}R_i^2$.

If S^{-1} exists, we define $G = S^{-1}$. The coroot-basis components and Dynkin components are then related by

$$M_i^\vee = \sum_j m_j G_{ij}. \tag{2.7}$$

The fundamental weight ω_i is the weight with Dynkin components $(\omega_i)_j = \delta_{ij}$. It is seen from Eq. (2.6) that the coroots $R_i^\vee$ and fundamental weights ω_i are dual sets of basis vectors.

The scalar product $\langle M,Q\rangle$ may be written either in the form of Eq. (2.6) or in one of the forms,

$$\langle M,Q\rangle = \sum_{ij} m_i q_j G_{ij} = \sum_{ij} M_i^\vee Q_j^\vee S_{ij}. \tag{2.8}$$

The Weyl reflection $W_\alpha(M)$ of a weight M associated with a nonzero root α is defined by

$$W_\alpha = M - \langle M,\alpha\rangle(2/\alpha^2)\alpha. \tag{2.9}$$

If α is the simple root R_j the reflection is denoted by W_j and is called simple. It is seen from Eqs. (2.1) and (2.9) that the reflected weight is

$$W_j(M) = M - m_j R_j. \tag{2.10}$$

The Weyl group consists of all sequences of zero or more Weyl reflections, and may be generated by the simple reflections alone. All weights related to a weight M by sequences of Weyl reflections comprise the Weyl orbit of M.

B. Further properties of affine algebras

For a simple affine algebra the Coxeter–Dynkin diagram is indecomposable and $|A|$, the determinant of the Cartan matrix, is zero.[1,6] Four of these diagrams are shown in Fig. 1; diagrams for all the affine algebras are given in Kac[7] and in Ref. 2. Since the vertices are treated here on equal footing, they will be numbered 1 to n.

Since $|A| = 0$, there is an eigenvector δ that satisfies the matrix equation $\delta A = 0$. The root-basis components of this null vector are denoted by c_i, i.e., $\delta = \Sigma_i\, c_i R_i$, so that $\Sigma_i c_i A_{ij} = 0$. It is seen from Eq. (2.3) that the Dynkin components of δ are all zero.

It is well known that the components c_i are either all positive or all negative, and that their ratios are rational. I make the conventional definition that these c_i (called marks) are positive, and are as small as possible consistent with all being integers.

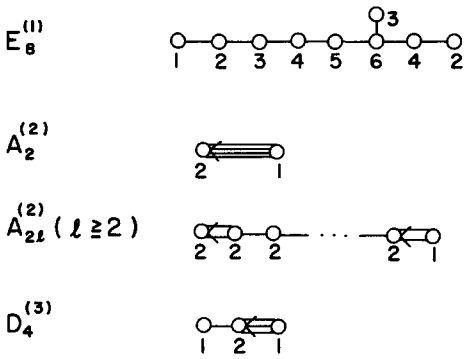


FIG. 1. Coxeter–Dynkin diagrams for some affine algebras. An arrow points to the smaller of two connecting roots, and the integer ℓ is one less than the number of vertices. The twist index is the number in superscript parentheses. The numbers by the vertices are the marks.

The level L of a weight M is defined by the equation

$$L(M) = \langle M,\delta\rangle = \sum_{i=1}^{n} m_i c_i^\vee, \tag{2.11}$$

where the coefficients $c_i^\vee$ are called comarks and are given by $c_i^\vee = c_i R_i^2/2$. It is desirable that the comarks be integers. The Cartan matrix specifies the ratios of the lengths of the simple roots, but not the overall normalization. I use the common convention that the roots are as short as possible, consistent with all the comarks being integers. This is equivalent to the convention of Ref. 1, although the wording is different. The twist index k may then be defined by the condition

$$k = \tfrac{1}{2}R_\ell^2, \tag{2.12}$$

where R_ℓ is the longest simple root.

Since the Dynkin components of δ are all zero it follows from Eq. (2.11) that the level of all simple roots is zero.

There is a well-known algorithm for constructing a highest weight irrep by subtracting simple roots from the highest weight, employing the same rules used in the algorithm for finite algebras.[8] Since the simple roots are of level 0, all weights in the irrep are of the same level.

The weights M and $M - \delta$ are different weights, yet they have the same Dynkin components. Consequently, one parameter in addition to the Dynkin components is needed to specify a weight completely. A convenient parameter is the displacement P_k, defined to increase by 1 when the simple root R_k is subtracted from a weight to form another. One can use any simple root for this purpose. For simplicity, I take all P_i to be zero for the highest weight of an irrep. If $c_k = 1$, P_k is the null depth discussed in Ref. 2.

If M is an affine weight, the quantity $M^{(h)}$ is called a subweight and refers to the corresponding weight of the basic subalgebra $\mathcal{S}_h$. The Dynkin components $m^{(h)}$ are obtained simply by ignoring the h component of M. Similarly, $A^{(h)}$ and $S^{(h)}$ are $(n-1)$-dimensional square matrices, obtained by deleting the h rows and columns of A and S. These matrices do have inverses. The root-basis and coroot-basis

components $M^{(h)}$ and $M^{\vee(h)}$ are defined by the matrix equations,

$$M^{(h)} = m^{(h)} A^{(h)-1}, \tag{2.13}$$

$$M^{\vee(h)} = m^{(h)} G^{(h)}, \tag{2.14}$$

where $G^{(h)} = S^{(h)-1}$.

III. EXTENDED SUBWEIGHTS AND THE AFFINE SCALAR PRODUCT

The vectors $M^{(h)}$ and $M^{\vee(h)}$ each have $(n-1)$ components. I define an n-component vector $M^{(h)*}$ by the equations,

$$M^{(h)*}{}_i = M^{(h)}{}_i \text{ for } i \neq h, \tag{3.1a}$$

$$M^{(h)*}{}_h = 0. \tag{3.1b}$$

An n-component vector $M^{\vee(h)*}$ is defined in the same way. These vectors are called extended subweights. In a similar fashion $G^{(h)-1*}$ and $A^{(h)-1*}$ are defined to be $n \times n$ matrices, with the h rows and columns consisting entirely of zeros, and the other elements equal to those of $G^{(h)}$ and $A^{(h)-1}$.

In the case of affine algebras, the Cartan matrix A and symmetrized matrix S do not have inverses. Therefore, one cannot define $G = S^{-1}$ and use Eq. (2.7) to find a basis dual to the coroots. Kac overcame this problem by introducing an additional basis vector and adding an $(n+1)$th dimension.[1] In Ref. 1, one of the simple roots for each affine algebra (always one with a comark of 1) is given a special role. We generalize this procedure by allowing any of the simple roots R_k to play the special role. If the new vector is denoted by $R_0^\vee$ (the numbering convention is different from that of Ref. 1), the $(n+1)$-dimensional, extended S matrix S_x is defined by

$$S_{x,00} = \langle R_0^\vee, R_0^\vee \rangle = 0, \quad S_{x,0k} = c_k / c_k^\vee, \tag{3.2}$$

$$S_{x,0i} = 0 \text{ for } i \neq k, \quad S_{x,ij} = S_{ij},$$

where $i > 0$ and $j > 0$. This matrix reduces to that obtainable from Ref. 1 if the comark $c_k^\vee = 1$. The matrix S_x has an inverse, and also has a negative eigenvalue, so the metric is indefinite. If $G_x = S_x^{-1}$, the elements of G_x are

$$G_{x,00} = 0, \quad G_{x,0i} = c_i^\vee / c_k, \quad G_{x,ij} = G^{(k)*}{}_{ij}. \tag{3.3}$$

The Dynkin components of the $(n+1)$ component, extended weight vector M_x are the integers,

$$(-P_k, m_1 \cdots m_n), \tag{3.4}$$

i.e., $m_0 = -P_k$. The components of a weight vector in the extended coroot basis may be determined by letting the subscripts run from 0 to n in Eq. (2.7). The result is

$$M_0^\vee = L / c_k, \tag{3.5a}$$

$$M_i^\vee = -P_k (c_i^\vee / c_k) + M^{\vee(k)*}{}_i \text{ for } i \neq 0. \tag{3.5b}$$

If one uses the extended dual form [the extended form of Eq. (2.6)] for the scalar product, it is seen from Eqs. (3.4) and (3.5a) and (3.5b) that

$$\langle M_x, Q_x \rangle = -[P_k(M) L(Q) + P_k(Q) L(M)] c_k^{-1} + \langle M, Q \rangle_k. \tag{3.6}$$

In order to write Eq. (3.6) I have used the level equation for M, i.e., $L(M) = \Sigma_{i=1}^n m_i c_i^\vee$. (The quantity $\langle M, Q \rangle_k$ is the scalar product associated with the subalgebra $\mathcal{S}_k$, i.e., $\langle M^{(k)}, Q^{(k)} \rangle$).

The value of P_k is 0 for the zero roots, and so is -1 for R_k and 0 for the other simple roots. If Q is a weight of level 0, it is seen from Eq. (3.5a) that $Q_0^\vee = 0$. Furthermore, it can be shown that in this case the $Q_i^\vee$ defined in Eq. (3.5b) is equal to those taken from the coroot expansion $Q = \Sigma_i Q_i^\vee R_i^\vee$. Therefore, the scalar product is that of Eq. (2.6).

IV. THE DOMINANT WEIGHTS IN A HIGHEST WEIGHT IRREP

A. A useful formula

In this section root-basis components are more useful than coroot-basis components; the $Q_i (i > 0)$ characterize a level-0 weight completely. We ask the question: Given a set of Dynkin components $q_1 \cdots q_n$ that correspond to level 0, what vectors Q in the root basis lead to these Dynkin components, i.e., are solutions of Eq. (2.3)? One solution is the extended subweight $Q^{(h)*}$ of Eqs. (3.1a) and (3.1b), for any $h \leqslant n$. This follows because $Q^{(h)*}$ was defined so that its ith Dynkin component is equal to that of Q, if $i \neq h$. Furthermore, any vector of the form of Eq. (2.3) must be a level-0 vector, so the h Dynkin component is also that of Q. It follows that a general solution is

$$Q = Q^{(h)*} + \mu_h \delta, \tag{4.1}$$

where μ_h is any real number. This set of solutions is independent of the choice of h. The allowed weights correspond to an infinite set of rational values of μ_h.

B. Congruency

In both finite algebras and affine algebras, weights in different congruency classes cannot be in the same irrep. It is known that the congruency label of an affine weight has two parts.[2] The first part is the level number. The second part is the congruency class of the underlying subalgebra. The congruency class of a weight depends only on the Dynkin components.

Since one of the goals of this paper is to treat the basic subalgebras in the same manner as much as possible, we must express the second part differently. Furthermore, in the cases of the twisted algebras, the identity of the underlying subalgebra is not obvious. I classify the basic subalgebra $\mathcal{S}_h$ as fundamental if and only if the mark c_h is equal to unity. One may use the congruency classes of any fundamental subalgebra to describe a weight because of the following congruency equivalence principle: two weights that are congruent (in the same class) with respect to one fundamental subalgebra are congruent with respect to any other fundamental subalgebra. The proof is given below. It should be noted from Fig. 1 that in the algebras $A_2^{(2)}$ and $A_{2\ell}^{(2)}$, the subalgebra given special treatment in Ref. 1, which corresponds to omitting the left-hand vertex, is not fundamental and cannot be used to determine congruency.

In the case of a finite algebra, two weights A and B are congruent if and only if the root-basis components of $Q = A - B$ are all integers.[9] I will show that for affine alge-

bras the fundamental subalgebra criterion given above is equivalent to the following condition; if two sets of Dynkin components are a_j and b_j, and if $q_j = a_j - b_j$, A and B are congruent if and only if a set of integral root-basis components Q_i exist that satisfy Eq. (2.3). Since this latter criterion does not refer to any basic subalgebra, its proof will establish the congruency equivalence principle.

The proof makes use of Eq. (4.1). Since the roots are of level zero, it is clear that the above condition requires that A and B be on the same level. Part of the demonstration is trivial. For any basic subalgebra $\mathscr{S}_h$ (fundamental or not), if A and B are congruent with respect to $\mathscr{S}_h$ then the components of $Q^{(h)*}$ are integers, so the μ_h in Eq. (4.1) may be chosen to be any integer. The requirement that the subalgebra be fundamental comes into play in the next part of the demonstration. If A and B are not congruent with respect to the subalgebra $\mathscr{S}_h$, then the components of $Q^{(h)*}$ are not all integers. Since the h component is zero, the requirement that Q_h be an integer limits μ_h to integral values (if $\mathscr{S}_h$ is fundamental), so the components of Q cannot all be made integers.

The subalgebra definition of congruency is useful in applications, since the congruency relations for finite algebras are well known.[10]

For some of the twisted affine algebras, certain choices of the fundamental subalgebra appear to lead to more congruency classes than other choices. In these cases the extra congruency conditions are superfluous. I will illustrate this for the affine algebra $D_4^{(3)}$ of Fig. 1. The fundamental subalgebra $\mathscr{S}_1$ is G_2, for which there is only one congruency class. On the other hand, the fundamental algebra $\mathscr{S}_3$ is A_2, for which there are three triality classes. However, the level equation for $D_4^{(3)}$ is $m_1 + 2m_2 + 3m_3 = L$. The A_2 triality may be written $m_1 + 2m_2$, mod 3. This is $L - 3m_3$, mod 3, so that all weights of the same level have the same A_2 triality automatically.

C. The containment criterion

The following criterion is valid for affine algebras. If Λ is the highest weight in an irrep, and if the dominant set of Dynkin components m^{++} is congruent to Λ, then weights with the components m^{++} will appear in the irrep with an infinite series of displacements P_h (for any h) given by

$$P_h = P_h^{\min} + \ell c_h, \tag{4.2}$$

where ℓ runs through all the non-negative integers, and the minimum displacement $P_h^{\min}$ may be computed from a simple procedure, given below. This rule has a reciprocity property not possessed by the corresponding rule for finite algebras. The criterion follows from a generalization of a theorem known for finite algebras.[11,12] The generalized theorem is: Let Λ be the highest weight in an irrep and M^{++} a dominant weight of the same level. Denote the difference by $Q = \Lambda - M^{++}$. Then M^{++} is in the irrep Λ if and only if the root-basis components Q_i are all non-negative integers.

The containment criterion follows almost immediately. Let $Q = \Lambda - M^{++}$, where Λ and M^{++} are congruent, and assume that the displacement P_k of M^{++} is not known *a priori*. For simplicity we take $\mathscr{S}_h$ to be a fundamental subalgebra, although it is easy to generalize the procedure to include nonfundamental $\mathscr{S}_h$. The root-basis components of $Q^{(h)*}$ may be computed from the Dynkin components by using Eqs. (2.13), (3.1a), and (3.1b). These root-basis components are all integers, although some may be negative. One then chooses the μ_h in Eq. (4.1) to be the smallest non-negative integer such that all the root-basis components Q_i are non-negative. The resulting value of Q_k is $P_k^{\min}$.

I take as an example some level-2 weights of affine E_8, the algebra $E_8^{(1)}$ of Fig. 1. It is convenient to number the nine roots of the Coxeter–Dynkin diagram according to the scheme,

$$\begin{array}{ccccccccc} & & & & & & 7 & & \\ 1 & 2 & 3 & 4 & 5 & 6 & & 8 & 9 \end{array} \tag{4.3}$$

It is seen from Fig. 1 that the null vector δ is given by $\delta = \{123\ 456\ 342\}$. (Consistently, I use curly brackets to denote root-basis components and ordinary parentheses to denote Dynkin components.) This affine algebra is simply laced, so the marks and comarks are the same. Therefore, the level of a weight M is

$$L = m_1 + 2m_2 + 3m_3 + 4m_4 + 5m_5 + 6m_6 + 3m_7 + 4m_8 + 2m_9. \tag{4.4}$$

The only mark that is unity is c_1, so $\mathscr{S}_1$ is the only fundamental subalgebra. Since $\mathscr{S}_1$ is E_8, for which all weights are congruent, all weights of the same level of affine E_8 are congruent.

One can look up the matrix $G = A^{-1}$ for finite E_8 in the literature.[13] The result is

$$G^{(1)} = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 3 & 4 & 2 \\ 3 & 6 & 8 & 10 & 12 & 6 & 8 & 4 \\ 4 & 8 & 12 & 15 & 18 & 9 & 12 & 6 \\ 5 & 10 & 15 & 20 & 24 & 12 & 16 & 8 \\ 6 & 12 & 18 & 24 & 30 & 15 & 20 & 10 \\ 3 & 6 & 9 & 12 & 15 & 8 & 10 & 5 \\ 4 & 8 & 12 & 16 & 20 & 10 & 14 & 7 \\ 2 & 4 & 6 & 8 & 10 & 5 & 7 & 4 \end{pmatrix}, \tag{4.5}$$

where the rows and columns correspond to the roots 2 through 9 in the scheme of Eq. (4.3).

It is seen from Eq. (4.4) that there are three level-2 dominant sets of Dynkin components. These are

$$\begin{aligned} A &= (200\ \ 000\ \ 000)\ \ \mathbf{0}, \\ B &= (010\ \ 000\ \ 000)\ \ \mathbf{2}, \\ C &= (000\ \ 000\ \ 001)\ \ \mathbf{4}. \end{aligned} \tag{4.6}$$

The boldface number to the right is the E_8 subnorm N_1 of each weight, determined by setting $Q = M$ in Eq. (2.8) and using the matrix $G^{(1)}$ of Eq. (4.5). Thus

$$N_1 = \sum_{i,j=2}^{9} m_i m_j G_{ij}^{(1)}. \tag{4.7}$$

Let us consider the irrep with highest weight A and find the minimum displacement $P_1^{\min}$ for the Dynkin components C. The extended subweights $A^{(1)*}$ and $C^{(1)*}$ may be determined from Eqs. (3.1a) and (3.1b). [Of course, $A^{(1)*}$

is the zero weight and the last eight components of $C^{(1)*}$ are the coefficients of the row labeled 9 (the eighth row) of the matrix of Eq. (4.5).] If $Q = A - C$, then

$$Q^{(1)*} = (0 - 2 - 4 - 6 - 8 - 10 - 5 - 7 - 4). \quad (4.8)$$

The minimum number of δ's that one needs to add to $Q^{(1)*}$ so that none of the resulting components is negative is 2. One finds

$$Q^{(1)*} + 2\delta = \{222 \quad 222 \quad 110\}. \quad (4.9)$$

Since the first component of this weight is 2, $P_1^{\min}$ is 2 for the Dynkin components C.

V. THE CONTENTS OF WEYL ORBITS

We consider the problems of constructing Weyl orbits and identifying the orbit (represented by its dominant weight) of an arbitrary weight. If there is exactly one negative Dynkin component m_h for an affine weight M, then the subweight $M^{(h)}$ (defined at the end of Sec. II B) is dominant for the subalgebra $\mathcal{S}_h$; such a weight is called a subdominant. If a subdominant weight of $\mathcal{S}_h$ occurs in an affine irrep with a particular value of the displacement P_h, then the entire irrep of $\mathcal{S}_h$ occurs with the same P_h so these weights are important. I will emphasize subdominant weights in this section, although most of the results apply to all nondominant weights.

A positive (negative) simple Weyl reflection is one that leads to a more positive (negative) weight. It is seen from Eq. (2.10) that a simple reflection is positive if the Dynkin component of the original weight is negative, and vice versa. The standard method for finding the orbit of an arbitrary weight is to make a series of positive simple reflections until the dominant weight is obtained.[14] Similarly, one may construct the first layers (up to some finite limit) of an orbit from the dominant weight by making negative simple reflections.

However, if the rank of the affine algebra is not small, this process is too cumbersome to be practical. The generalized affine scalar product of Sec. III is useful for constructing and picturing orbits. It follows from Eq. (2.9) that the scalar product is invariant to Weyl reflections, i.e.,

$$\langle W_\alpha(M), W_\alpha(Q)\rangle = \langle M, Q\rangle, \quad (5.1)$$

and hence is invariant to all Weyl transformations. We consider the norm of M, obtained by substituting M for Q in Eq. (3.6). The only quantities in this equation that may change in a Weyl transformation are the displacement P_k and the subnorm $N_k = \langle M, M\rangle_k$. It follows from the invariance of the norm that if M and M' are two weights in the same orbit,

$$P_k(M) - P_k(M') = (c_k/2L)[N_k(M) - N_k(M')]. \quad (5.2)$$

Equation (5.2) not only gives a simple formula for the displacement, but in many cases helps one identify the orbit of a particular weight. I illustrate with the algebra affine E_8. The generalized Dynkin diagram for this algebra is given in Fig. 1 and the root-numbering scheme and level formula are shown in Eqs. (4.3) and (4.4). As pointed out in Sec. IV C all weights of the same level are congruent.

Let us consider the weight M with Dynkin components (-310000100); this weight is subdominant with respect to the E_8 subalgebra $\mathcal{S}_1$. We want to find the dominant orbit weight m^{++}, as well as the P_1 difference between M and M^{++}. It is seen from Eq. (4.4) that this weight is of level 2. Consequently, M^{++} must be one of the three weights A, B, and C listed in Eq. (4.6). From Eqs. (4.7) and (4.5) the subnorm of M is $N_1(M) = 16$. Since $L = 2$ and $c_1 = 1$, it is seen that if M' is identified with M^{++}, the requirement that the right-hand side of Eq. (5.2) be an integer implies that $N_1(M) - N_1(M^{++})$ is divisible by 4. This rules out weight B. Since the Dynkin components of M are not all divisible by 2, A cannot be the dominant orbit weight, so M^{++} is the weight C. Furthermore, from Eq. (5.2), the difference in P_1 between M and C is 3.

One cannot always identify the orbit so easily, if the level is not small. A rapid algorithm for treating such cases is given in Ref. 12.

VI. SUMMARY OF PROCEDURE

This section summarizes the procedure developed for answering the question raised in Sec. I, namely: given an arbitrary irrep with highest weight Λ of a simple affine algebra and an arbitrary weight M, is M in the irrep Λ? Definitions and proofs will not be repeated.

A weight of an affine algebra of n vertices may be characterized by n Dynkin components and a displacement parameter P_k, where k labels any of the vertices. Congruency depends only on the Dynkin components. Two weights are congruent if and only if they are on the same level and are congruent with respect to any fundamental subalgebra. The weight M must be congruent to Λ to be in the irrep Λ. If they are congruent then an infinite set of weights with the Dynkin components of M are in Λ; these correspond to the values of displacement $P_k = P_k^{\min} + \ell c_k$, where ℓ is any non-negative integer. The minimum displacement may be written

$$P_k^{\min}(M) = P_k^{\min}(M^{++}) + \Delta P_k.$$

One may use the procedure of Sec. IV to find $P_k^{\min}(M^{++})$ and Eq. (5.2) to find ΔP_k.

ACKNOWLEDGMENT

This work was supported, in part, by the U.S. Department of Energy.

[1] V. G. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U.P., Cambridge, 1985), second edition.

[2] R. Slansky, Comm. Nucl. Part. Phys. A XVIII, 175 (1988). The mathematical approach to constructing irreps of affine algebras is taken in this paper.

[3] See, for example, V. G. Kac and D. H. Peterson, Adv. Math. 53, 125 (1984). See also Ref. 2.

[4] R. N. Cahn, *Semi-Simple Lie Algebras and Their Representations* (Benjamin/Cummings, Reading, MA, 1984).

[5] R. Slansky, Phys. Rep. 79, 1 (1981).

[6] An early discussion of these matrices is given by S. Berman, R. Moody, and M. Wonenburger, Indiana Univ. Math. J. 21, 1091 (1972).

[7] V. G. Kac, op. cit. The diagrams for affine algebras are listed on pp. 48 and 49.

[8] This construction procedure is discussed in Refs. 4 and 5. It is applied to affine algebras in Ref. 2.

[9] The congruency relations for all the finite algebras may be derived from this principle. See R. H. Capps, J. Math. Phys. **27**, 914 (1986), p. 920.

[10] The congruency relations for the finite algebras are listed by F. Lemire and J. Patera, J. Math. Phys. **21**, 2026 (1980). See also Ref. 5, p. 37.

[11] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972). The finite algebra theorem follows from Secs. 13.4 and 21.3.

[12] R. H. Capps, "Representations of Affine Kac-Moody Algebras and the Affine Scalar Product," Purdue University Preprint, July, 1989. This paper contains a generalization of the theorem of Ref. 11 to affine algebras.

Nontransitive imprimitivity systems for the Galilei group

Gianni Cassinelli and Pekka J. Lahti[a)]
Dipartimento di Fisica, Universitá di Genova, I.N.F.N. Sezione di Genova, Genova, Italy

(Received 4 January 1990; accepted for publication 4 April 1990)

A class of nontransitive imprimitivity systems and the corresponding projective unitary representations for the inhomogeneous Galilei group are worked out with the Mackey–Varadarajan method of group representations.

I. INTRODUCTION

In his "extended footnote" (p. 846, Ref. 1) to a paper of Newton and Wigner, Wightman[1] worked out a class of (projective unitary) representations of the Euclidean group and the corresponding representations for the inhomogeneous Galilei group, which, he says, describe the most general localizable Galilei invariant particles. From the mathematical point of view the paper of Wightman "merely writes out Mackey's theory (of imprimitivity representations) in detail for the Euclidean group" (p. 847, Ref. 1), and applies the methods of Bargmann and Loomis to obtain the resulting representations for the Galilei group. In the same spirit the present paper is nothing but an appendix to Sec. IX.8 of the monograph of Varadarajan[2] in which we study a wider class of (projective unitary) representations of the inhomogeneous Galilei group. We construct a nontransitive family of imprimitivity systems for the (extension of the covering group of the) homogeneous Galilei group that induces the representations for the (inhomogeneous) Galilei group. Since this Mackey–Varadarajan method is very powerful and elegant, we think it is worthwhile to work out this "Appendix."

II. THE PRELIMINARIES

A. The group

Consider the Newtonian space-time $\mathfrak{R}^3\times\mathfrak{R}$, and let $(\mathbf{x},t)$, $\mathbf{x}=(x_1,x_2,x_3)$, denote the coordinates of its points. The inhomogeneous Galilei group G consists of the space-time transformations

$$(\mathbf{x},t)\mapsto(\mathbf{x}',t')\doteq(R\mathbf{x}+\mathbf{v}t+\mathbf{a},t+b), \tag{1}$$

where $R\in\mathrm{SO}(3)$ is a (proper) rotation of the space $\mathfrak{R}^3$, $\mathbf{v}\in\mathfrak{R}^3$ is a velocity boost, $\mathbf{a}\in\mathfrak{R}^3$ is a space translation, and $b\in\mathfrak{R}$ is a time translation. Denoting the generic element of G as $g=(R,b,\mathbf{v},\mathbf{a})$ the group multiplication law has the form

$$gg'=(R,b,\mathbf{v},\mathbf{a})(R',b',\mathbf{v}',\mathbf{a}')$$
$$=(RR',b+b',\mathbf{v}+R\mathbf{v}',\mathbf{a}+R\mathbf{a}'+b'\mathbf{v}). \tag{2}$$

The identity element e of G is $(I,0,\mathbf{0},\mathbf{0})$ and the inverse g^{-1} of a g is $(R^{-1},-b,-R^{-1}\mathbf{v},-R^{-1}(\mathbf{a}-b\mathbf{v}))$.

The group of space-time translations $A=\{(I,b,\mathbf{0},\mathbf{a}):b\in\mathfrak{R},\mathbf{a}\in\mathfrak{R}^3\}$ is a normal, closed, Abelian subgroup of G. The group of space rotations and the velocity boosts $G_0=\{(R,0,\mathbf{v},\mathbf{0}):R\in\mathrm{SO}(3),\mathbf{v}\in\mathfrak{R}^3\}$ is another closed subgroup of G. Since

$$(I,b,\mathbf{0},\mathbf{a})(R,0,\mathbf{v},\mathbf{0})=(R,b,\mathbf{v},\mathbf{a}) \tag{3}$$

and

$$(R,0,\mathbf{v},\mathbf{0})(I,b,\mathbf{0},\mathbf{a})(R,0,\mathbf{v},\mathbf{a})^{-1}=(I,b,\mathbf{0},R\mathbf{a}+b\mathbf{v}), \tag{4}$$

G is a semidirect product of A and G_0, i.e., $G=A\times' G_0$ (Ref. 2). The action of G_0 on A is explicitly given by (4).

B. Its covering group

The inhomogeneous Galilei group G is connected but not simply connected. Let G^* denote its universal covering group, and let $g=(h,b,\mathbf{v},\mathbf{a})$ with $h\in\mathrm{SU}(2)$ denote its generic element. If δ is the covering homomorphism $\mathrm{SU}(2)\to\mathrm{SO}(3)$, then the group structure of G^* is given as

$$(h,b,\mathbf{v},\mathbf{a})(h',b',\mathbf{v}',\mathbf{a}')$$
$$=(hh',b+b',\mathbf{v}+\delta(h)\mathbf{v}',\mathbf{a}+\delta(h)\mathbf{a}'+b'\mathbf{v}). \tag{5}$$

Again, G^* is a semidirect product of A and G_0^*, $G^*=A\times' G_0^*$. The kernel of the covering homomorphism $G^*\to G$, $(h,b,\mathbf{v},\mathbf{a})\mapsto(\delta(h),b,\mathbf{v},\mathbf{a})$ consists of the two elements $(\pm I,0,\mathbf{0},\mathbf{0})$. Thus the projective representations of G are in one-to-one correspondence to the projective representations of G^*. Since G^* is a connected and simply connected Lie group its multipliers can be determined exactly by the Lie algebraic methods.[2–4]

C. The m_M extensions

Let $\mathbf{T}=\{c\in\mathbf{C}:|c|=1\}$ denote the torus group. For each real number M the mapping $m_M:G^*\times G^*\to\mathbf{T}$, with

$$m_M(g,g')=m_M((h,b,\mathbf{v},\mathbf{a}),(h',b',\mathbf{v}',\mathbf{a}'))$$
$$\doteq\exp[iM(\tfrac{1}{2}b'\mathbf{v}\cdot\mathbf{v}+\mathbf{v}\cdot\delta(h)\mathbf{a}')], \tag{6}$$

is a multiplier for G^*, and any of its multiplier is similar to one of this form.[2,5]

Let G_M^* denote the m_M extension of G^*, so that for any $(g;z),(g';z')\in G_M^*=G^*\times_M\mathbf{T}$ the multiplication is

$$(g;z)(g';z')=(gg';m_M(g,g')zz'). \tag{7}$$

Let $V:G_M^*\to\mathcal{U}(\mathcal{H})$ be a unitary representation of G_M^* acting on a complex separable Hilbert space $\mathcal{H}$ such that $V_{(e;z)}=z^{-1}I$ for each $z\in\mathbf{T}$. Then $U:G^*\to\mathcal{U}(\mathcal{H})$, with $U_g\doteq V_{(g;1)}$ is a projective representation of G^* with the multiplier m_M, i.e.,

$$U_{gg'}=m_M(g,g')U_gU_{g'},\quad g,g'\in G^*. \tag{8}$$

Conversely, if $U:G^*\to\mathcal{U}(\mathcal{H})$ is a projective representation of G^* with a multiplier m_M, $V_{(g,z)}=z^{-1}U_g$ defines a repre-

[a)] Permanent address: Department of Physical Sciences, University of Turku, Turku, Finland.

 0022-2488/90/081859-03$03.00

sentation of G_M^* that has the property $V_{(e,z)} = z^{-1}I$ (Refs. 2 and 4).

The group G_M^* is, again, a semidirect product of the subgroups A_M and G_0^* consisting of the elements of the form $(I,b,0,\mathbf{a};z)$ and $(h,0,\mathbf{v},0;1)$, respectively. To simplify the notations we denote the generic elements of A_M and G_0^* as $(b,\mathbf{a};z)$ and $(h,\mathbf{v})$, respectively. Then the action of G_0^* on A_M has the form

$$\begin{aligned}(h,\mathbf{v})[(b,\mathbf{a};z)] &\doteq (h,\mathbf{v})(b,\mathbf{a};z)(h,\mathbf{v})^{-1} \\ &= (b,\delta(h)\mathbf{a} + b\mathbf{v};z \\ &\quad \times \exp[iM(\tfrac{1}{2}b\mathbf{v}\cdot\mathbf{v} + \mathbf{v}\cdot\delta(h)\mathbf{a})]). \quad (9)\end{aligned}$$

D. Unitary representations of G_M^*

Since G_M^* is a semidirect product of A_M and G_0^*, the unitary representations of G_M^* are in one-to-one correspondence with the imprimitivity systems for G_0^* based on the dual $\hat{A}_M$ of A_M. This will be described next briefly.[2]

The elements of the dual group $\hat{A}_M$ are the characters x of A_M. They are given by

$$x((b,\mathbf{a};z)) = z^n \exp[i(p_0 b - \mathbf{p}\cdot\mathbf{a})], \quad (10)$$

where $p_0 \in \mathfrak{R}, \mathbf{p} \in \mathfrak{R}^3$, and $n \in \mathbf{Z}$. The action $(b,\mathbf{a};z) \mapsto (h,\mathbf{v})[(b,\mathbf{a};z)]$ of G_0^* on A_M induces the dual action $x \mapsto (h,\mathbf{v})^*[x]$ of G_0^* on $\hat{A}_M$ defined via

$$(h,\mathbf{v})^*[x]((b,\mathbf{a};z)) \doteq x((h,\mathbf{v})^{-1}[(b,\mathbf{a};z)]). \quad (11)$$

This makes $\hat{A}_M$ a G_0^* space.

If T is a unitary representation of G_M^* acting in a (complex separable) Hilbert space $\mathscr{H}$, its restrictions $U: A_M \to \mathscr{U}(\mathscr{H})$ and $V: G_0^* \to \mathscr{U}(\mathscr{H})$ are unitary representations. Let $P: \mathscr{B}(\hat{A}_M) \to \mathscr{P}(\mathscr{H})$ be the (unique) projection-valued measure associated with U via the SNAG theorem. Then (V,P) is a system of imprimitivity for G_0^* based on $\hat{A}_M$, acting on $\mathscr{H}$. Conversely, let (V,P) be an imprimitivity system for G_0^* based of $\hat{A}_M$, acting on a (complex separable) Hilbert space $\mathscr{H}$. By the SNAG theorem, the projection operator-valued measure $P: \mathscr{B}(\hat{A}_M) \to \mathscr{P}(\mathscr{H})$ determines a unique unitary representation $U: A_M \to \mathscr{U}(\mathscr{H})$, and the composition $T = UV$ is a unitary representation of $G_M^* = A_M \times' G_0^*$ acting in $\mathscr{H}$. Moreover, the commuting rings of T and (V,P) are the same.

E. Transitive imprimitivity systems

Among the imprimitivity systems for G_0^* based on $\hat{A}_M$ there are the transitive ones. They are those (V,P) for which the measure class of P lives on an orbit for G_0^* on $\hat{A}_M$. According to the imprimitivity theorem all transitive imprimitivity systems are induced, and they all can be described explicitly.[2]

Let $x_0 \in \hat{A}_M$, and let $O = \{(h,\mathbf{v})^*[x_0] : (h,\mathbf{v}) \in G_0^*\}$ be the orbit of x_0. Let μ be a G_0^*-invariant measure on O. Let $H = \{(h,\mathbf{v}) \in G_0^* : (h,\mathbf{v})^*[x_0] = [x_0]\}$ be the stabilizer of x_0. Fix a (complex separable) Hilbert space $\mathscr{K}$, and choose a unitary representation L of H acting in $\mathscr{K}$. Let φ^L be a (G_0^*, O) cocycle taking values in $\mathscr{U}(\mathscr{K})$. Denote $\mathscr{H} = L^2(\hat{A}_M, \mathscr{K}, \mu)$. The pair (V,P), defined via

$$P(E)f = \chi_{E\cap O} f, \quad (12)$$

and

$$(V_{(h,\mathbf{v})}f)(x) = \varphi^L((h,\mathbf{v}),(h,\mathbf{v})^{-1^*}[x])f((h,\mathbf{v})^{-1^*}[x]), \quad (13)$$

for $f \in \mathscr{H}$, $E \in \mathscr{B}(\hat{A}_M)$, is an imprimitivity system for G_0^* based on $\hat{A}_M$ acting on $\mathscr{H}$. Here, χ_E denotes the characteristic function of the set E. By the recipe of Sec. II D the corresponding unitary representation T of G_M^* is then given by

$$\begin{aligned}(T_{(h,b,\mathbf{v},\mathbf{a};z)}f)(x) &= x((b,\mathbf{a});z))\varphi^L((h,\mathbf{v}),(h,\mathbf{v})^{-1^*}[x]) \\ &\quad \times f((h,\mathbf{v})^{-1^*}[x]). \quad (14)\end{aligned}$$

Due to (10), the condition $T_{(e,z)} = z^{-1}I$ can be fullfilled only if $n = -1$, which then is the only case that gives rise to projective representations of G^*.

In particular, all irreducible unitary representations of G_M^* come from transitive imprimitivity systems. This fact leads to the well-known classification of irreducible projective unitary representations of the inhomogeneous Galilei group.[2,5-7]

In the next section we shall work out a more general class of projective unitary representations of G^*. According to Wightman,[1] this class exhausts the localizable Galilean objects.

III. A CLASS OF NONTRANSITIVE IMPRIMITIVITY SYSTEMS

A. Some invariant subspaces and measures

Consider, again, the dual group $\hat{A}_M$. By (10), its elements x can be labeled as triples $(p_0,\mathbf{p};n)$, with $p_0 \in \mathfrak{R}$, $\mathbf{p} \in \mathfrak{R}^3$, and $n \in \mathbf{Z}$. The dual action of G_0^* on $\hat{A}_M$ takes then the explicit form

$$\begin{aligned}(h,\mathbf{v})^*[(p_0,\mathbf{p};n)] &= (p_0 + \tfrac{1}{2}Mn\mathbf{v}\cdot\mathbf{v} + \mathbf{v}\cdot\delta(h)\mathbf{p}, nM\mathbf{v} \\ &\quad + \delta(h)\mathbf{p};n). \quad (15)\end{aligned}$$

For fixed $n \in \mathbf{Z}$, consider the set $O^n = \{(p_0,\mathbf{p};n) \in \hat{A}_M : p_0 \in \mathfrak{R}, \mathbf{p} \in \mathfrak{R}^3\}$, so that $\hat{A}_M = \cup_{n\in\mathbf{Z}} O^n$. The action of G_0^* on O^n is given by (15). From that one may verify that for each $(p_0,\mathbf{p};n) \in O^n$ the quantity $K = p_0 - (2nM)^{-1}\mathbf{p}\cdot\mathbf{p}$ (with $n \neq 0 \neq M$) remains unchanged under the action of G_0^*. Hence the points of O^n can be parametrized as $(K,\mathbf{p})$, with $K \in \mathfrak{R}$ and $\mathbf{p} \in \mathfrak{R}^3$. With this parametrization the action of G_0^* on O^n obtains the following form:

$$(h,\mathbf{v})^*[(K,\mathbf{p})] = (K, nM\mathbf{v} + \delta(h)\mathbf{p}). \quad (16)$$

Thus, G_0^* has no affect on K, and its action on $\mathbf{p}$ is Euclidean. Thus, if $d\rho$ is any measure on $\mathfrak{R}$ and if $d^3\mathbf{p}$ is the Lebesgue measure in $\mathfrak{R}^3$,

$$d\rho(K)d^3\mathbf{p}, \quad (17)$$

is an invariant measure on O^n.

As in Sec. II E, the only systems of imprimitivity for G_0^* living on O^n that lead to projective representations of G^* (with multipliers m_M) are those living on O^{-1}. Hence, we consider here only the case $n = -1$. [Note that the case $M = 0$ leads to nonprojective representations of the Galilei

group, and they are considered to be nonphysical.[1,2] The pairs $(M, -M)$ lead to pairs of physically equivalent representations.[2]]

B. Cocycles

The orbits $O_K^{-1} = \{(p_0,\mathbf{p};-1): p_0 + p^2/2M = K\}$, $K\in\Re$, of G_0^* on $\hat{A}_M$ are all contained in O^{-1}, $O^{-1} = \cup_{K\in\Re} O_K^{-1}$. Moreover, G_0^* leaves the coordinate K of a point of O^{-1} unchanged (since G_0^* acts transitively on each O_K^{-1}). Hence any (G_0^*, O_K^{-1}) cocycle φ that takes values in $\mathcal{U}(\mathcal{H})$,

$$\varphi: G_0^* \times O_K^{-1} \to \mathcal{U}(\mathcal{H}), \tag{18}$$

defines a cocycle

$$\varphi': G_0^* \times O^{-1} \to \mathcal{U}(\mathcal{H}), \tag{19}$$

via the relation

$$\varphi'((h,\mathbf{v}),(K,\mathbf{p})) = \varphi((h,\mathbf{v}),\mathbf{p})),$$
$$(h,\mathbf{v})\in G_0^*, \quad (K,\mathbf{p})\in O^{-1}. \tag{20}$$

Let $L: \mathrm{SU}(2) \to \mathcal{U}(\mathcal{H})$ be a unitary representation. Since G_0^* is also a semidirect product of the (normal, closed, Abelian) subgroup of the velocity boosts and SU(2), the mapping

$$\varphi^L: G_0^* \times O_K^{-1} \to \mathcal{U}(\mathcal{H}), \quad ((h,\mathbf{v}),\mathbf{p}) \mapsto L(h), \tag{21}$$

is a cocycle for G_0^* on the orbit O_K^{-1} acting in $\mathcal{U}(\mathcal{H})$. Hence, the mapping

$$((h,\mathbf{v}),\mathbf{p}) \mapsto L(h), \tag{22}$$

defines a cocycle $G_0^* \times O^{-1} \to \mathcal{U}(\mathcal{H})$.

Any such a cocycle and any invariant measure $d\rho\, d^3\mathbf{p}$ allows one to construct a system of imprimitivity for G_0^*, based on $\hat{A}_M$, and living on O^{-1}. Such imprimitivity systems are clearly nontransitive, since O^{-1} is not an orbit of G_0^*. The corresponding unitary representations of G_M^* are thus also more general than those considered in II E. We shall now work out a class of them.

C. Representations

We define the pair (U,P) via

$$P(E)f = \chi_{E\cap O^{-1}} f \tag{23}$$

$$(U_{(h,\mathbf{v})}f)(K,\mathbf{p}) = L(h) f((h,\mathbf{v})^{-1*}[(K,\mathbf{p})])$$
$$= L(h) f(K, \delta(h)^{-1}(\mathbf{p} - M\mathbf{v})), \tag{24}$$

for $f\in\mathcal{H} \doteq L^2(\Re\times\Re^3, d\rho(K)d^3\mathbf{p}, \mathcal{K})$, $E\in\mathcal{B}(\hat{A}_M)$. This is a system of imprimitivity for G_0^* based on $\hat{A}_M$, and that acts on $\mathcal{H}$. In defining $\mathcal{H}$ we have used the fact that the set O^{-1} can, in fact, be identified with $\Re^4$. The unitary representation of G_M^*, which corresponds to this nontransitive system of imprimitivity, is given by

$$(U_{(h,b,\mathbf{v},\mathbf{a};z)}f)(K,\mathbf{p})$$
$$= z^{-1}\exp[ib(K + (b/2M)\mathbf{p}\cdot\mathbf{p} - \mathbf{p}\cdot\mathbf{a})]$$
$$\times L(h) f(K, \delta(h)^{-1}(\mathbf{p} - M\mathbf{v})), \tag{25}$$

for any $(h,b,\mathbf{v},\mathbf{a};z)\in G_M^*$ and for all $f\in\mathcal{H}$.

The projective representation of G^* (with the multiplier m_M) corresponding to the representation (25) of G_M^* is now obtained putting $z = 1$ in this equation. The representation $U: G^* \to \mathcal{U}(\mathcal{H})$, with $\mathcal{H} = L^2(\Re\times\Re^3, d\rho(K)d^3\mathbf{p}, \mathcal{K})$, takes then the explicit form:

$$(U_{(h,b,\mathbf{v},\mathbf{a})}f)(K,\mathbf{p})$$
$$= \exp[i(bK + (b/2M)\mathbf{p}\cdot\mathbf{p} - \mathbf{p}\cdot\mathbf{a})]$$
$$\times L(h) f(K, \delta(h)^{-1}(\mathbf{p} - M\mathbf{v})). \tag{26}$$

The above family of projective representations for G^* can also be obtained as follows. Consider, for each $K\in\Re$, the representation

$$(U^K_{(h,b,\mathbf{v},\mathbf{a})}f)(\mathbf{p})$$
$$= \exp[i(bK + (b/2M)\mathbf{p}\cdot\mathbf{p} - \mathbf{a}\cdot\mathbf{p})]$$
$$\times L(h) f(\delta(h)^{-1}(\mathbf{p} - M\mathbf{v})), \tag{27}$$

which acts (for each fixed K) in the Hilbert space $L^2(\Re^3, d^3\mathbf{p}, \mathcal{K})$. Choose any measure $d\rho$ on $\Re$, and define

$$U = \int_{\Re}^{\oplus} U^{(K)}\, d\rho(K). \tag{28}$$

Since $\int_{\Re}^{\oplus} L^2(\Re^3, d^3\mathbf{p}, \mathcal{K}) d\rho(K)$ can be identified with $L^2(\Re\times\Re^3, d\rho(K)d^3\mathbf{p}, \mathcal{K})$, (27) and (28) constitute, in fact, the same (i.e., equivalent) representation than (26).

Usually in the physical applications of the group representations one considers only those coming from *transitive* imprimitivity systems. Here we have exhibited a class of representations that corresponds to *nontransitive* imprimitivity systems. Nevertheless, they seem to be of physical interest. Indeed, in his classic paper Wightman[1] studied representations of the form (28) as the most general localizable Galilei objects.

[1] A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).
[2] V. S. Varadarajan, *Geometry of Quantum Theory*, 2nd ed. (Springer, New York, 1985).
[3] V. Bargmann, Ann. Math. **59**, 1 (1954).
[4] K. R. Parthasarathy, *Lecture Notes in Mathematics*, No. 93 (Springer, Berlin, 1969).
[5] J.-M. Lévy-Leblond, *Group Theory and its Applications, II*, Edited by E. M. Loeble (Academic, New York, 1971), pp. 221–299.
[6] J.-M. Levy-Leblond, J. Math. Phys. **4**, 776 (1963).
[7] J. Voisin, J. Math. Phys. **6**, 1519 (1965).

Some remarks on the Gupta–Bleuler triplet

Dario Pierotti
International School for Advanced Studies (ISAS), Strada Costiera 11, I-34014, Trieste, Italy

(Received 12 October 1989; accepted for publication 14 February 1990)

Indecomposable representations of the Poincaré group associated to infrared singular field theory models are discussed in the framework of the general theory of the Gupta–Bleuler triplet formulated by Araki. It is shown that the definition of maximal Hilbert space structures, related to the infrared properties of the states of the models, can be exploited to construct representation spaces for the Gupta–Bleuler triplet. The examples of the two-dimensional massless scalar field and of the electromagnetic field in the Landau gauge are discussed. In particular, in the first example, the relation between the Gupta–Bleuler triplet and the algebraic treatment of the massless scalar field is investigated. In the case of electromagnetism, the structure of the representation of the Poincaré group in the Landau gauge is clarified. The explicit form of the corresponding Gupta–Bleuler triplet for the one-particle space of the electromagnetic field is exhibited.

I. INTRODUCTION

Indefinite metric structures of degenerate systems[1] have been widely investigated in recent years, both in the framework of local and covariant Wightman formulation,[2,3] and in a C^*-algebraic context.[4–6] At the same time, the study of mass zero representations of the Poincaré group[7–9] stimulated the investigation of indecomposable representations π of a Lie group, on a space $\mathscr{H}$ with an invariant indefinite inner product, taking the following form:

$$\pi_n \to \pi_{n-1} \to \cdots \to \pi_1,$$
$$\mathscr{H} \equiv \mathscr{H}_n \supset \mathscr{H}_{n-1} \supset \cdots \supset \mathscr{H}_1, \tag{1.1}$$

where $\mathscr{H}_j$ is a π invariant subspace of $\mathscr{H}_{j+1}$ without any π invariant complement and π_j is constructed on the quotient space $\mathscr{H}_j/\mathscr{H}_{j-1}$.

As a result, a general theory of the Gupta–Bleuler triplet was formulated by Araki,[10] who also gave a necessary and sufficient condition for (1.1) to have invariant nondegenerate inner product in terms of cohomologies related to π_j.

However, the realization of a Gupta–Bleuler triplet in the framework of specific field theoretical models is faced by a mathematical obstruction connected to the following structural problem. The representation space $\mathscr{H}$ of Ref. 10 is an indefinite inner product space[11] provided with a locally convex topology generated by the family of seminorms

$$p_\phi(\psi) \equiv |\langle \phi, \psi \rangle|, \tag{1.2}$$

where $\phi, \psi \in \mathscr{H}$, and $\langle\ ,\ \rangle$ denotes the inner product. The notions of closed subspace and irreducible or indecomposable representation are defined by Araki in terms of the above topology (see Ref. 10, remark 2.2 and definition 2.3). However, in the case of indefinite metric field theory models, the Wightman functions define a vector space $\mathscr{D}_0$, an inner product $\langle\ ,\ \rangle$ on $\mathscr{D}_0$, and a locally convex topology by the family of seminorms (1.2) (with $\phi, \psi \in \mathscr{D}_0$) with respect to which $\mathscr{D}_0$ is, in general, *not complete*. Moreover, one cannot uniquely extend the inner product to the closure of $\mathscr{D}_0$ in the above topology, since the form $\langle\ ,\ \rangle$ is not jointly continuous.

A possible strategy for the construction of a (complete) space of states is to look for a topology induced by a *maximal Hilbert structure* or *Krein structure*[12] defined by a positive scalar product that bounds the indefinite inner product, and to consider the Hilbert space closure of $\mathscr{D}_0$, where the form $\langle\ ,\ \rangle$ can be extended by joint continuity. It can be shown[12] that in a maximal Hilbert space one can always define the positive scalar product $(\ ,\ )$ in such a way that the *metric operator* η defined by the relation $\langle\ ,\ \rangle = (\ , \eta)$, satisfies $\eta^2 = 1$. The resulting Hilbert space is a particular type of indefinite inner product space called Krein space.[11]

It is worthwhile to stress that the maximal Hilbert structure associated to a given set of Wightman functions, is not, in general, uniquely determined. This arbitrariness in some sense parametrizes different representations of the field algebra corresponding to different infrared properties of the states.[12,13]

The nice property of Krein spaces is that the topology defined by the seminorms (1.2) is equivalent to the weak Hilbert topology, so that the notions of closed subspace and irreducible representation of Ref. 10 are reduced to the usual Hilbert space definitions.

In the present work, the above strategy is applied to define the Gupta–Bleuler triplet for the representations of the Poincaré group in the one-particle space of the two-dimensional massless scalar field and Landau gauge electromagnetism.

In Sec. II, after a short review of the basic definitions and results of Ref. 10, we describe the Krein space quantization of the two-dimensional massless scalar field and the related representation of the Poincaré group. The discussion of the corresponding Gupta–Bleuler triplet is simplified by choosing a suitable orthogonal decomposition (with respect to the positive scalar product) of the Krein space. From the resulting structure, one easily recovers the $\mathbb{R}^2$ cohomology of the Poincaré group (in two dimensions) generated by the soliton solutions of the wave equation.[14–16]

In the subsequent sections, the Gupta–Bleuler triplet in electromagnetism is investigated. The free electromagnetic potential in the Gupta–Bleuler gauge[1] provides a well-known example of Krein space quantization. The one-particle space is defined by

 0022-2488/90/081862-10$03.00

$$\mathscr{H}=\left\{\psi_\mu(p),\quad \psi_\mu\in L^2\left(\frac{d^3p}{p_0},C_+\right),\quad \mu=0,1,2,3\right\} \tag{1.3}$$

(C_+ is the forward light cone) with the positive scalar product

$$(\phi,\psi)=\sum_{\mu=0}^{3}\int\frac{d^3p}{p_0}\bar{\phi}_\mu(p)\psi_\mu(p). \tag{1.4}$$

The metric operator is defined by

$$(\eta\psi)_\mu=-\psi^\mu, \tag{1.5}$$

so that one obtains the usual Lorentz invariant inner product

$$\langle\phi,\psi\rangle=-\int\frac{d^3p}{p_0}\bar{\phi}_\mu(p)\psi^\mu(p). \tag{1.6}$$

We stress that the construction of Krein space structures associated to infrared singular field theory models (like e.g., the two-dimensional massless scalar field[12,17] and the Schwinger model[18] is, in general, more elaborated than the simple structure (1.3)–(1.6). However, one can show that the Landau gauge electromagnetic potential (characterized by a propagator which is infrared singular as $p_\mu p_\nu/p^4$) can be defined as an operator-valued distribution acting in the same Fock space as the Gupta–Bleuler gauge potential and with the same Hilbert structure.[19]

As a result, one can define a Landau gauge representation of the Poincaré group in the one-particle space $\mathscr{H}$ defined by (1.3)–(1.6) (see Sec. III). This representation can be investigated from the point of view of the general theory given by Araki and related to the corresponding representation in the Gupta–Bleuler gauge. In particular, by suitably choosing an orthogonal decomposition of the space $\mathscr{H}$, it is shown (Sec. IV) that the Landau gauge representation define a new triplet which, by the group property, differs by a cocycle from the triplet in the Gupta–Bleuler gauge. In the last section (Sec. V) the properties of this cocycle and the nontriviality of its cohomology class are discussed.

II. THE GUPTA–BLEULER TRIPLET FOR THE TWO-DIMENSIONAL MASSLESS SCALAR FIELD

We start by recalling some definitions and results of Ref. 10. A representation π^* of a Lie group G on a space $\mathscr{H}^*$ is called conjugate to a representation π of G on $\mathscr{H}$ if there is a sesquilinear form $\langle\xi,\zeta\rangle$, $\xi\in\mathscr{H}^*$, $\zeta\in\mathscr{H}$, such that (1) $\mathscr{H}$ and $\mathscr{H}^*$ separate each other, i.e., $\langle\xi,\zeta\rangle=0$ for any $\zeta\in\mathscr{H}$ implies $\xi=0$ and $\langle\xi,\zeta\rangle=0$ for any $\xi\in\mathscr{H}^*$ implies $\zeta=0$. (2) For any $g\in G$,

$$\langle\pi^*(g^{-1})\xi,\zeta\rangle=\langle\xi,\pi(g)\zeta\rangle.$$

The following is the main result of Ref. 10. Let π_1 be an irreducible subrepresentation of an indecomposable representation π of G on a space $\mathscr{H}$ with a G-invariant (indefinite) inner product. Then (1) π is of the form (1.1) with $n=2$ or 3 such that $\pi_n=(\pi|_{\mathscr{H}_n})$ mod $\mathscr{H}_{n-1}$ on $\mathscr{H}_n/\mathscr{H}_{n-1}$ is conjugate to π_1 on $\mathscr{H}_1$; (2) $\mathscr{H}_1$ is a null space; (3) the induced inner product on $\mathscr{H}_2/\mathscr{H}_1$ is nondegenerate and invariant for the representation π_2.

Let us suppose that $\mathscr{H}_j$ has a closed complement $\mathscr{R}_{j+1}$ in $\mathscr{H}_{j+1}$. Then $\pi(g)$ can be written on $\mathscr{H}=\mathscr{H}_1+\mathscr{R}_2+\cdots+\mathscr{R}_n$ in the matrix form

$$\pi(g)=\begin{pmatrix}\pi_1(g) & c_{12}(g) & \cdots & c_{1n}(g)\\ 0 & \pi_2(g) & \cdots & c_{2n}(g)\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \pi_n(g)\end{pmatrix}.$$

A *Gupta–Bleuler triplet in the standard form* is defined by

$$\pi(g)=\pi_1^*(g)\to\pi_2(g)\to\pi_1(g),$$

$$\pi(g)=\begin{pmatrix}\pi_1(g) & c_{12}(g) & c_{13}(g)\\ 0 & \pi_2(g) & c_{23}(g)\\ 0 & 0 & \pi_1^*(g)\end{pmatrix}, \tag{2.1}$$

on the space

$$\mathscr{H}=\mathscr{R}_1+\mathscr{R}_2+\mathscr{R}_3,$$

with $\mathscr{R}_1=\mathscr{H}_1$, $\mathscr{R}_3=\mathscr{H}_1^*$, $\mathscr{R}_2=\mathscr{R}_2^*$, and the inner product

$$\langle\xi_1+\xi_2+\xi_3,\zeta_1+\zeta_2+\zeta_3\rangle=\langle\xi_1,\zeta_3\rangle+\langle\xi_3,\zeta_1\rangle+\langle\xi_2,\zeta_2\rangle,\quad(\xi_j,\zeta_j\in\mathscr{R}_j). \tag{2.2}$$

The off-diagonal elements of the matrix (2.1)

$$c_{ij}(g):\ \mathscr{R}_j\to\mathscr{R}_i,$$

are cochains for the representations π_i^*,π_j^*. The group property of the representation (2.1) is expressed by the condition

$$\delta c_{kl}=-\sum_{j=k+1}^{l-1}c_{kj}c_{jl},\quad l\neq k+1,\quad \delta c_{kk+1}=0, \tag{2.3}$$

where δ is the coboundary operation defined by

$$\delta c_{kl}(g_1,g_2)=\pi_k(g_1)c_{kl}(g_2)-c_{kl}(g_1g_2)+c_{kl}(g_1)\pi_l(g_2),\quad g_1,g_2\in G. \tag{2.4}$$

The cochain c_{ij} is said to be *bounded* if there exists a cochain

$$c_{ij}^*:\ \mathscr{R}_i^*\to\mathscr{R}_j^*,$$

such that

$$\langle\xi,c_{ij}(g)\zeta\rangle=\langle c_{ij}^*(g^{-1})\xi,\zeta\rangle, \tag{2.5}$$

for all $\zeta\in\mathscr{R}_j$, $\xi\in\mathscr{R}_i^*$. If $c_{ij}=c_{ij}^*$, we have a *self-adjoint* cochain.

By defining in the usual way cocycles, coboundaries, and cohomologies $H(\pi_j,\pi_i)$, the existence of a Gupta–Bleuler triplet in the standard form is equivalent to the condition that there are nonzero bounded elements $c_{12}\in H(\pi_2,\pi_1)$ and $c_{23}=c_{12}^*\in H(\pi_1^*,\pi_2)$ (see Ref. 10, Theorem 2).

As discussed in the previous section, it may be interesting to discuss the realization of the above general structures in the framework of infrared singular local field theory models. The simplest example is provided by the massless scalar field in two space-time dimensions.[12,17] The two-point function of the model defines an indefinite inner product on the test function space $\mathscr{S}(\mathbf{R}^2)$ by

$$\langle f,g\rangle=\int d^2p\,\hat{W}(p)\bar{f}(p)g(p), \tag{2.6}$$

where

$$\hat{W}(p) = \left[\left(\frac{1}{p^0+p^1}\right)_+ \delta(p^0-p^1) + \left(\frac{1}{p^0-p^1}\right)_+ \delta(p^0+p^1)\right] \tag{2.7}$$

and

$$\left(\frac{1}{p}\right)_+ \equiv \frac{d}{dp}(\theta(p)\log(p)).$$

We recall that the distribution (2.7) is a regularization of $1/|p|$ and it is positive on the test function space $\mathcal{S}_0 = \{f \in \mathcal{S}, f(0) = 0\}$. According to the discussion in the previous section, we want to define a positive scalar product on $\mathcal{S}$ that bounds the indefinite inner product (2.6) and so that the norm closure $\overline{\mathcal{S}}$ provides a maximal set of states.

In this case, the problem has a unique solution, in the sense that the maximal Hilbert closure $\overline{\mathcal{S}}$ is uniquely determined. The Hilbert scalar product can be given in the form[12]

$$(f,g)_K = \langle f^0, g^0\rangle + \langle f,\chi\rangle\langle\chi,g\rangle + \overline{f}(0)g(0), \tag{2.8}$$

where we have decomposed the generic $f \in \mathcal{S}$ as

$$f(p) = f(0)\chi(p) + f^0(p), \quad f^0 \in \mathcal{S}_0$$

and the function $\chi(p)$ satisfies $\chi(0) = 1$, $\langle\chi,\chi\rangle = 0$.

The closure of $\mathcal{S}$ in the norm defined by the above scalar product is a Krein space $\overline{\mathcal{S}} = K$, with a metric operator η satisfying $\eta^2 = 1$. We denote by $\oplus$ the orthogonal sum (with respect to the positive scalar product) in K. It can be shown[12] that the space K has the following orthogonal decomposition:

$$K = \mathcal{R}_1 \oplus \mathcal{R}_2 \oplus \mathcal{R}_3, \tag{2.9}$$

where $\mathcal{R}_1$ and $\mathcal{R}_3$ are one-dimensional subspaces given by

$$\mathcal{R}_1 = \{\alpha v_0, \alpha \in \mathbb{C}, v_0 = \eta\chi\},$$
$$\mathcal{R}_3 = \{\beta\chi, \beta \in \mathbb{C}\}, \tag{2.10}$$

while

$$\mathcal{R}_2 \cong L^2\left(\frac{dp_1}{|p_1|}, \mathbb{R}\right), \tag{2.11}$$

with $\eta|_{\mathcal{R}_2} = 1$.

From (2.9)–(2.11) it follows that the eigenspace corresponding to the negative eigenvalue of η is the one-dimensional subspace generated by the vector $v_- = (1/\sqrt{2}) \times (\chi - v_0)$. When the metric operator (in a Krein space) has a finite-dimensional eigenspace corresponding to the negative eigenvalue, the indefinite inner space is called a *Pontrjagin space*.[11]

The representation of the Poincaré group (a,Λ) is defined on the dense set $\mathcal{S}$ in the usual way:

$$U(a,\Lambda)f(p) = f_{a,\Lambda}(p) = e^{ipa}f(\Lambda^{-1}p). \tag{2.12}$$

From the Poincaré invariance of the two-point function (2.7), it follows that $U(a,\Lambda)$ is an isometric operator, i.e.,

$$\langle U(a,\Lambda)f,\ U(a,\Lambda)g\rangle = \langle f,g\rangle, \quad f,g \in S. \tag{2.13}$$

We can now extend the representation (2.12) to all K by the following result.[11]

Let an isometric operator U be defined on a dense domain of a Pontrjagin space and suppose that the indefinite inner product is nondegenerate on the closures of the domain and range of U. Then U is continuous with its inverse.

In our case, the closures of the domain and range of $U(a,\Lambda)$ are equal to the space K and therefore the above theorem can be applied to extend the representation to the bounded operators on K. We still denote by $U(a,\Lambda)$ this extension. It is worthwhile to stress that the so-obtained representation is η unitary, that is

$$U(a,\Lambda)^\dagger = U(a,\Lambda)^{-1},$$

where $\dagger$ denotes the adjoint operation with respect to the indefinite inner product $\langle\ ,\ \rangle$.

An interesting feature of the Krein space quantization is the existence of translationally invariant states. It can be shown that the vector $v_0 = \eta\chi \in K$, satisfies the relation

$$U(a,\Lambda)v_0 = v_0 \tag{2.14}$$

and it is the unique vector in K with this property. Moreover, the vector v_0 has zero norm:

$$\langle v_0, v_0\rangle = \langle\eta\chi, \eta\chi\rangle = \langle\chi,\chi\rangle = 0. \tag{2.15}$$

Hence, the one-dimensional subspace $\mathcal{R}_1$ is a Poincaré invariant null subspace of K. Furthermore, it can be easily seen that the subspace

$$\mathcal{H}_2 = \mathcal{R}_1 \oplus \mathcal{R}_2 = \overline{\mathcal{S}}_0$$

is Poincaré invariant.

Then, we have the following proposition.

Proposition 2.1: The representation of the Poincaré group in the one-particle Krein space of the two-dimensional massless scalar field defines a Gupta–Bleuler triplet in the standard form

$$\pi(g) = \pi_1^*(g) \to \pi_2(g) \to \pi_1(g)$$

on the space

$$\mathcal{H} = \mathbb{C} \oplus L^2\left(\frac{dp_1}{|p_1|}, \mathbb{R}\right) \oplus \mathbb{C}, \tag{2.16}$$

with the nonzero elements of the matrix (2.1) given by

$$\pi_1(a,\Lambda) = \pi_1^*(a,\Lambda) = 1, \tag{2.17}$$

$$\pi_2(a,\Lambda)\psi = \psi_{a,\Lambda}, \quad \psi \in L^2\left(\frac{dp_1}{|p_1|}, \mathbb{R}\right), \tag{2.18}$$

$$c_{12}(a,\Lambda)\psi = \langle\chi, \psi_{a,\Lambda} - \psi\rangle, \tag{2.19}$$

$$c_{23}(a,\Lambda)\alpha = \alpha(\chi_{a,\Lambda} - \chi), \quad \alpha \in \mathbb{C}, \tag{2.20}$$

$$c_{13}(a,\Lambda)\alpha = \alpha\langle\chi, \chi_{a,\Lambda} - \chi\rangle = \alpha\langle\chi, \chi_{a,\Lambda}\rangle, \tag{2.21}$$

where $\psi_{a,\Lambda}$, $\chi_{a,\Lambda}$ are defined as in (2.12).

The matrix elements c_{12} and c_{23} are cocycles whose cohomology classes do not depend on the shape of the function χ, satisfying $\chi(0) = 1$ and $\langle\chi,\chi\rangle = 0$, which is chosen in the definition (2.8) of the Krein norm.

Proof: By (2.9) and (2.10), we can decompose any $f \in \mathcal{S} \subset K$ in the following way:

$$f = (v_0, f)_K v_0 \oplus f_2 \oplus (\chi, f)_K \chi, \tag{2.22}$$

where f_2 is the projection of f on the subspace $\mathcal{R}_2$.

By definition (2.8) and the relations $v_0 = \eta\chi, \eta^2 = 1$, we can also write

$$f = \langle \chi, f\rangle v_0 \oplus f_2 \oplus f(0)\chi. \tag{2.23}$$

Then, by using $\langle \chi,\chi\rangle = 0$, we get

$$f_2 = f^0 - \langle \chi, f^0\rangle v_0, \tag{2.24}$$

with

$$f^0 = f - f(0)\chi, \quad f^0 \in \mathscr{S}_0.$$

Thus, the Krein norm (2.8) of f_2 is given by

$$(f_2, f_2)_K = \int \frac{dp_1}{|p_1|} |f^0(p)|^2, \tag{2.25}$$

and we can identify f_2 with the function $f^0 \in L^2(dp_1/|p_1|,\mathbf{R})$.

By (2.12), (2.14), and the identity (2.24), the action of the Poincaré group on the components of the decomposition (2.22) can be written

$$U(a,\Lambda)v_0 = v_0, \tag{2.26}$$

$$\begin{aligned} U(a,\Lambda)f_2 &= f^0_{a,\Lambda} - \langle \chi, f^0\rangle v_0 \\ &= \langle \chi, f^0_{a,\Lambda} - f^0\rangle v_0 \oplus \{f^0_{a,\Lambda} - \langle \chi, f^0_{a,\Lambda}\rangle v_0\} \\ &= \langle \chi, f^0_{a,\Lambda} - f^0\rangle v_0 \oplus (f^0_{a,\Lambda})_2, \end{aligned} \tag{2.27}$$

$$\begin{aligned} U(a,\Lambda)\chi &= \chi_{a,\Lambda} \\ &= \langle \chi, \chi_{a,\Lambda} - \chi\rangle v_0 \oplus \{(\chi_{a,\Lambda} - \chi) \\ &\quad - \langle \chi, \chi_{a,\Lambda} - \chi\rangle v_0\} \oplus \chi \\ &= \langle \chi, \chi_{a,\Lambda}\rangle v_0 \oplus (\chi_{a,\Lambda} - \chi)_2 \oplus \chi. \end{aligned} \tag{2.28}$$

Now, the relations (2.17)–(2.21) follow from (2.26)–(2.28) and by the remark that the vectors $\{f_2\}$ span a dense subspace in $\mathscr{R}_2 \cong L^2(dp_1/|p_1|,\mathbf{R})$.

The cocycle condition $\delta c_{12} = \delta c_{23} = 0$ and the relation $\delta c_{13} = c_{12}c_{23}$ can be easily verified.

Finally, let us consider a different function $\chi' \in \mathscr{S}$, with $\chi'(0) = \chi(0)$, in the definition of the Krein structure (2.8). Then, one gets the cocycles

$$c'_{12}(a,\Lambda)\psi = \langle \chi', \psi_{a,\Lambda} - \psi\rangle, \tag{2.29}$$

$$c'_{23}(a,\Lambda) = \chi'_{a,\Lambda} - \chi', \tag{2.30}$$

which differ from c_{12}, c_{23} by a coboundary, since

$$(\chi'_{a,\Lambda} - \chi') - (\chi_{a,\Lambda} - \chi) = (\chi' - \chi)_{a,\Lambda} - (\chi' - \chi),$$

and $(\chi' - \chi) \in \mathscr{S}_0$. □

Remark 1: It is possible to prove that a change in the shape of the test function χ as in the above proposition corresponds to the transformation

$$\pi(a,\Lambda) \to W\pi(a,\Lambda)W^{-1},$$

where W is an operator (in a triangular matrix form) preserving the indefinite inner product.[10]

The effect of the above transformation is to change c_{12} and c_{23} by a coboundary and to transform the cochain c_{13} in such a way to preserve the group property $\delta c_{13} = c_{12} \times c_{23}$. □

The above discussion can be extended to the case of the scalar and pseudoscalar fields, $\varphi, \tilde{\varphi}$, related by $\partial_\mu \tilde{\varphi} = \epsilon_{\mu\nu}\partial^\nu \varphi$ with $\epsilon_{\mu\nu}$ antisymmetric and $\epsilon_{01} = 1$. It can be shown[12,17,20] that, in this case, the one-particle Krein space must be enlarged with one more dimension, allowing the definition of the vector $\tilde{\varphi}(\chi)\Psi_0 \equiv \tilde{\chi}$ out of the vacuum Ψ_0.

In the enlarged Krein space $\tilde{K}$, the vector $\tilde{\chi}$ is normalized, orthogonal to χ and is represented by the discontinuous function $\tilde{\chi}(p) = \epsilon(p_1)\chi(p)$, where $\epsilon(p_1) = 1$ for $p_1 > 0$, $\epsilon(p_1) = -1$ for $p_1 < 0$.

The new space $\tilde{K}$ contains another translationally invariant vector $\tilde{v}_0 = \eta\tilde{\chi}$. Moreover, it is possible to define on $\tilde{K}$ an indecomposable representation of the Poincaré group, which preserves the indefinite inner product.[12,17]

The resulting triplet is a generalization of the structure (2.16)–(2.21) on the space

$$\mathbf{C}^2 \oplus L^2(dp_1/|p_1|,\mathbf{R}) \oplus \mathbf{C}^2. \tag{2.31}$$

In particular, one obtains from the above construction the cocycle

$$\tilde{c}_{23}(a,\Lambda)\{\alpha,\beta\} = \alpha(\chi_{a,\Lambda} - \chi) + \beta(\tilde{\chi}_{a,\Lambda} - \tilde{\chi}), \quad \alpha,\beta \in \mathbf{C}. \tag{2.32}$$

It is worthwhile to discuss the relation between the above construction and the known algebraic treatment of the two-dimensional massless scalar field.[14,15]

As it is known, the space $L^2(dp_1/|p_1|,\mathbf{R})$ is isomorphic to the completion $\bar{\mathscr{M}}$ of the space $\mathscr{M}$ defined in the following way: The elements of $\mathscr{M}$ are the real solutions $\xi(x_0,x_1)$ of the wave equation with localized Cauchy data $\xi(0,x_1)$, $\partial_0\xi(0,x_1)$, where $\int dx_1\, \partial_0\xi_1(0,x_1) = 0$. The space $\mathscr{M}$ is furnished with the Poincaré invariant (positive) scalar product

$$(\xi_1, \xi_2) = \int dp\ \overline{\psi_1(p)}\psi_2(p), \tag{2.33}$$

where

$$\psi(p) = |p|^{1/2}\hat{\xi}(p) - i|p|^{-1/2}\hat{\dot{\xi}}$$

and

$$\hat{\xi}(p) = \frac{1}{2\pi}\int dx_1\, e^{-ipx_1}\xi(0,x_1),$$

$$\hat{\dot{\xi}}(p) = \frac{1}{2\pi}\int dx_1\, e^{-ipx_1}\partial_0\xi(0,x_1).$$

Finally, the completion $\bar{\mathscr{M}}$ is endowed with a complex structure J [making the scalar product (2.33) sesquilinear] by

$$J\{\xi\} = J\{\hat{\xi}, \hat{\dot{\xi}}\} = \{(1/|p|)\hat{\dot{\xi}}, -|p|\hat{\xi}\}.$$

From the above definition, one easily checks that J maps $\mathscr{M}$ to $\mathscr{M}$ and $J^2 = -1$. The space $\mathscr{M}$ is the basic test function space in the algebraic description of the mass zero field.[14,15]

Let us consider the two vectors $\chi, \tilde{\chi}$ which define the cocycle (2.32). We can associate to these vectors the following solutions of the wave equation:

$$\phi(x) = \int d^2p\, \hat{W}(p)e^{-ipx}\chi(p),$$

$$\tilde{\phi}(x) = \int d^2p\, \epsilon(p_1)\hat{W}(p)e^{-ipx}\chi(p), \tag{2.34}$$

with $\hat{W}(p)$ defined by (2.7).

The solutions (2.34) can be seen as elements of the algebraic dual of $\mathscr{M}$, defining the cocycle (2.32) on the closure $\overline{\mathscr{M}}$.

Let us now consider the real functions $\phi_1 = \mathrm{Im}\,\phi$ and $\phi_2 = \mathrm{Im}\,\tilde{\phi}$. By choosing a real and symmetric function $\chi(p)$ in (2.34), we get two real solutions of the wave equation ϕ_1, ϕ_2, with Cauchy data satisfying the relations

$$\phi_1(0,x_1) = 0,$$

$$\partial_0\phi_1(0,x_1) = \check{\chi}(x_1) = \frac{1}{\sqrt{2\pi}}\int dp\, e^{ipx_1}\,\chi(|p|,p);$$

$$\phi_2(0,x_1) = \frac{1}{2\sqrt{2\pi}}\int dy\, \epsilon(x_1 - y)\check{\chi}(y),$$

$$\partial_0\phi_2(0,x_1) = 0.$$

Hence, the real cohomology of the Poincaré group can be described by the soliton solutions of the wave equation

$$\phi_{q,q_5}(x) = q\phi_1(x) + q_5\phi_2(x),$$

where the real parameters

$$q = \frac{1}{\sqrt{2\pi}}\int dx_1\, \partial_0\phi_{q,q_5}(0,x_1),$$

$$q_5 = \frac{1}{\sqrt{2\pi}}\int dx_1\, \partial_1\phi_{q,q_5}(0,x_1),$$

are the "cohomological" charge and axial charge defined in the algebraic treatment.[15]

By taking the real parts of the solutions (2.34), one gets another pair of real solutions $\phi_3 = \mathrm{Re}\,\phi$ and $\phi_4 = \mathrm{Re}\,\tilde{\phi}$ corresponding to the initial conditions

$$\phi_3(0,x_1) = \int dy\,(\log|x_1 - y| - \Gamma'(1))\check{\chi}(y),$$

$$\partial_0\phi_3(0,x_1) = 0;$$

$$\phi_4(0,x_1) = 0, \quad \partial_0\phi_4(0,x_1) = \int dy\, \mathscr{P}\left(\frac{1}{x_1 - y}\right)\check{\chi}(y),$$

where Γ' is the derivative of the Euler function and the distribution $\mathscr{P}(1/x)$ is the Cauchy principal part of the integral. It can be shown that, as elements of the algebraic dual of $\mathscr{M}$, these two vectors are given by

$$(\phi_3, \xi) = (\phi_1, J\xi); \quad (\phi_4, \xi) = (\phi_2, J\xi),$$

for any $\xi \in \mathscr{M}$. Thus, the two pairs of solutions $\{\phi_1, \phi_2\}$ and $\{\phi_3, \phi_4\}$ are related by the induced complex structure on the algebraic dual of $\mathscr{M}$.

III. THE REPRESENTATION OF THE POINCARÈ GROUP IN THE LANDAU GAUGE

The present section describes the main features of the Krein space realization of the Landau gauge given in Ref. 19. (For a C^*-algebraic treatment, see Refs. 5 and 21.) The explicit expression of the Landau gauge representation of the Poincaré group on the space (1.3) is then deduced from the definition of the electromagnetic potential and the covariance requirement.

The Landau gauge electromagnetic potential is defined, in the Fock space over the indefinite inner product space $\mathscr{H}$, by the following expression:

$$A^L_\mu(f^\mu) = (1/\sqrt{2})\{a_\mu((\pi^L f)^\mu) + a^\dagger_\mu((\pi^L f)^\mu)\},$$
$$f^\mu \in \mathscr{S}(\mathbb{R}^4), \tag{3.1}$$

where $a, a^\dagger$ are the annihilation and creation operators and the projection π^L is defined by

$$\pi^L(f)_\mu = \sqrt{2\pi}\left\{f_\mu(p) - \frac{p_\mu}{2p_0}\left[\frac{\partial}{\partial p_0}(p^\nu f_\nu(p)) - \left|\frac{p^\nu f_\nu(p)}{2p_0}\right|\right]\right\}\Big|_{C_+}. \tag{3.2}$$

From the above definition, we get

$$p_\mu f(p) \in \mathrm{Ker}\,\pi^L$$

for any $f \in \mathscr{S}(\mathbb{R}^4)$, so that the transversality condition $\partial^\mu A^L_\mu = 0$ holds as an operator equation.

Moreover, one can show that the set

$$\{\pi^L(f)_\mu,\ f_\mu \in \mathscr{S}(\mathbb{R}^4)\}, \tag{3.3}$$

is dense in $\mathscr{H}$, so that the Fock vacuum Ψ_0 is a cyclic vector in the state space.

By the covariance requirement and the Poincaré invariance of the vacuum Ψ_0, we find that the Landau gauge representation of the Poincaré group $U^L(a,\Lambda)$ is defined on the dense domain (3.3) by

$$U^L(a,\Lambda)\pi^L(f) = \pi^L(f_{a,\Lambda}), \tag{3.4}$$

where $(f_{a,\Lambda})_\mu(x) = \Lambda_\mu{}^\nu f_\nu(\Lambda^{-1}(x-a))$.

Let us first consider the subgroup of pure space-time translations ($\Lambda = 1$). By the expression (3.2) and the identity

$$p^\mu\pi^L(f)_\mu|_{C_+} = p^\mu f_\mu|_{C_+},$$

the operator $U^L(a,1)$ can be extended to the following closed operator in $\mathscr{H}$:

$$U^L(a,1)\psi_\mu(p) = e^{ipa}(\psi_\mu(p) - ia_0(p_\mu/2p_0)\,p^\nu\psi_\nu(p)), \tag{3.5}$$

where

$$\psi_\mu \in \mathscr{D}_a \equiv \{\psi_\mu \in \mathscr{H},\ p^\mu\psi_\mu \in L^2(d^3p/p_0, C_+)\}.$$

As can be seen from the expression (3.5), the operator $U^L(a,1)$ is not bounded for $a_0 \neq 0$ (i.e., for temporal translations). However, for any $a \in \mathbb{R}^4$, it is η-unitary, i.e., $U^L(a,1)^\dagger = U^L(a,1)^{-1}$, where $U^\dagger$ is the adjoint of U with respect to the indefinite inner product (1.6).

To find the explicit expression of the Lorentz transformations $U(0,\Lambda)$, some elaboration is necessary. From definitions (3.2), (3.4) we get

$$(U(0,\Lambda)\pi^L f)_\mu(p)$$
$$= (\pi^L f_\Lambda)_\mu(p)$$
$$= \Lambda_\mu{}^\nu f_\nu(\Lambda^{-1}p) - (p_\mu/2p_0)[\partial_0(pf)(\Lambda^{-1}p) - (1/2p_0)(pf)(\Lambda^{-1}p)]|_{C_+}, \tag{3.6}$$

where pf denotes the Lorentz scalar product $p^\mu f_\mu$. Let us consider the following identity:

$$(\pi^L f_\Lambda)_\mu(p) = \Lambda_\mu{}^\nu(\pi^L f)_\nu(\Lambda^{-1}p) + (p_\mu/2np) \times [n^\nu \partial_\nu (pf)(\Lambda^{-1}p) - (1/2np)(pf)(\Lambda^{-1}p)]|_{C_+} - (p_\mu/2p_0)[\partial_0(pf)(\Lambda^{-1}p) - (1/2p_0)(pf)(\Lambda^{-1}p)]|_{C_+}, \quad (3.7)$$

where we have defined

$$n^\mu = (\Lambda^{-1})_0{}^\mu. \quad (3.8)$$

The operator $n^\mu \partial_\mu$ restricted to the light cone satisfies the following relation [for any $f \in \mathcal{S}(\mathbb{R}^4)$]:

$$(n^\mu \partial_\mu f)|_{C_+} = (n^0 \partial_0 f)|_{C_+} + \left(\frac{n^i p_i}{p_0} \partial_0 f\right)\Big|_{C_+} + n^i \partial_i (f|_{C_+}) = \left(\frac{n^\mu p_\mu}{p_0} \partial_0 f\right)\Big|_{C_+} + n^i \partial_i (f|_{C_+}). \quad (3.9)$$

By defining $\psi_\mu = (\pi^L f)_\mu$ and by using (3.9) in (3.7) we find

$$(U(0,\Lambda)\psi)_\mu(p) = \Lambda_\mu{}^\nu \psi_\nu(\Lambda^{-1}p) + \frac{p_\mu}{2(np)} n^i \partial_i (p\psi)(\Lambda^{-1}p) - \frac{p_\mu}{4}\left(\frac{1}{(np)^2} - \frac{1}{p_0^2}\right)(p\psi)(\Lambda^{-1}p). \quad (3.10)$$

Finally, by (3.8) we obtain the expression

$$(U(0,\Lambda)\psi)_\mu(p) = \Lambda_\mu{}^\nu \psi_\nu(\Lambda^{-1}p) + \frac{p_\mu}{2(\Lambda^{-1}p)_0} (\Lambda^{-1})_0{}^i \partial_i (p\psi)(\Lambda^{-1}p) + \frac{p_\mu}{4}\left(\frac{1}{p_0^2} - \frac{1}{(\Lambda^{-1}p)_0^2}\right)(p\psi)(\Lambda^{-1}p). \quad (3.11)$$

It is not difficult to see that the above operator is well defined on the dense domain:

$$\mathcal{D}_\Lambda \equiv \left\{\psi_\mu \in \mathcal{H},\ p^\mu \partial_i \psi_\mu \in L^2\left(\frac{d^3p}{p_0}, C_+\right),\ i = 1,2,3\right\}.$$

If the transformation contains Lorentz boosts ($(\Lambda^{-1})_0{}^i \neq 0$), the operator (3.11) is unbounded. It can be checked by explicit calculation that the operator $U(0,\Lambda)$ is η-unitary in $\mathcal{H}$ for any Lorentz transformation Λ.

IV. THE GUPTA–BLEULER TRIPLET IN THE LANDAU GAUGE

We begin this section with the discussion of the Gupta–Bleuler triplet for electromagnetism in the Gupta–Bleuler gauge. By the results of the previous section, the structure of the triplet in this gauge will provide a suitable framework for the analogous discussion in the Landau gauge.

The space $\mathcal{H}_1$ of definition (1.1) is the space of longitudinal photons;

$$\mathcal{H}_1 \equiv \mathcal{R}_1 = \left\{\psi_\mu^{(1)} \in \mathcal{H},\ \psi_\mu^{(1)}(p) = \frac{p_\mu}{p_0}\psi(p),\ \psi \in L^2\left(\frac{d^3p}{p_0}, C_+\right)\right\}. \quad (4.1)$$

The space $\mathcal{H}_2$ is the space defined by the subsidiary condition

$$\mathcal{H}_2 = \{\psi_\mu \in \mathcal{H},\ p^\mu \psi_\mu(p) = 0 \text{ for almost all } p \in C_+\}. \quad (4.2)$$

We choose as closed complement $\mathcal{R}_2$ of $\mathcal{H}_1$ in $\mathcal{H}_2$ the orthogonal complement (with respect to the positive scalar product) $\mathcal{H}_2 \ominus \mathcal{H}_1$. Then

$$\mathcal{R}_2 = \left\{\psi_\mu^{(2)} \in \mathcal{H}_2,\ \sum_{\mu=0}^{3} p_\mu \psi_\mu^{(2)}(p) = 0 \text{ for almost all } p \in C_+\right\}. \quad (4.3)$$

The space $\mathcal{R}_2$ is the space of the physical photons. In fact, by (4.2), we have the equivalent definition

$$\mathcal{R}_2 = \left\{\psi_\mu^{(2)} \in \mathcal{H},\ \psi_0^{(2)} = 0,\ \sum_{i=1}^{3} p_i \psi_i^{(2)}(p) = 0 \text{ for almost all } p \in C_+\right\}.$$

Similarly, the subspace $\mathcal{R}_3$ is defined by

$$\mathcal{R}_3 = \mathcal{H} \ominus \mathcal{H}_2,$$

i.e.,

$$\mathcal{R}_3 = \left\{\psi_\mu^{(3)} \in \mathcal{H},\ \sum_{\mu=0}^{3} \int \frac{d^3p}{p_0} \bar{\phi}_\mu(p) \psi_\mu^{(3)}(p) = 0,\ \forall \phi_\mu \in \mathcal{H}_2\right\}. \quad (4.4)$$

It can be shown[19] that the above definition is equivalent to the following one:

$$\mathcal{R}_3 = \left\{\psi_\mu^{(3)} \in \mathcal{H},\ \psi_\mu^{(3)}(p) = \frac{(Ip)_\mu}{p_0}\psi(p), \psi \in L^2\left(\frac{d^3p}{p_0}, C_+\right)\right\}, \quad (4.5)$$

where $(Ip)_0 = p_0$, $(Ip)_i = -p_i$, $i = 1,2,3$.

In the following, we will still denote by $\psi^{(1)}, \psi^{(2)}, \psi^{(3)}$, the vectors belonging to $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, respectively.

With the above choice for $\mathcal{R}_2$ and $\mathcal{R}_3$, we investigate the structure of the matrix (2.1) in the case of the representation of the Poincaré group in the Gupta–Bleuler gauge:

$$(\pi^{GB}(a,\Lambda)\psi)_\mu(p) = e^{ipa}\Lambda_\mu{}^\nu \psi_\nu(\Lambda^{-1}p), \quad \forall \psi \in \mathcal{H}. \quad (4.6)$$

Proposition 4.1: The matrix elements of the Gupta–Bleuler triplet defined by the representation (4.6) and the decomposition (4.1)–(4.5), can be written in the following form:

$$(\pi_1^{GB}(a,\Lambda)\psi^{(1)})_\mu(p) = \frac{p_\mu}{p_0} e^{ipa} \frac{p_0}{(\Lambda^{-1}p)_0}\psi(\Lambda^{-1}p), \quad (4.7)$$

$$(\pi_2^{GB}(a,\Lambda)\psi^{(2)})_\mu(p) = \begin{cases} 0 & \text{if } \mu = 0, \\ e^{ipa}(\Lambda_i{}^j - (p_i/p_0)\Lambda_0{}^j)\psi_j^{(2)}(\Lambda^{-1}p), & \text{if } \mu = i = 1,2,3. \end{cases} \quad (4.8)$$

$$(\pi_3^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) = \frac{(Ip)_\mu}{p_0} e^{ipa} \frac{(\Lambda^{-1}p)_0}{p_0} \psi(\Lambda^{-1}p), \tag{4.9}$$

$$(c_{12}^{GB}(a,\Lambda)\psi^{(2)})_\mu(p) = \frac{p_\mu}{p_0} e^{ipa} \Lambda_0{}^i \psi_i^{(2)}(\Lambda^{-1}p), \tag{4.10}$$

$$(c_{23}^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) = \begin{cases} 0, & \text{if } \mu = 0, \\ 2 & e^{ipa}((\Lambda^{-1})^0{}_i + (p_i/p_0^2)(\Lambda^{-1})_0{}^j p_j)\psi(\Lambda^{-1}p), \\ & \text{if } \mu = i = 1,2,3. \end{cases} \tag{4.11}$$

$$(c_{13}^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) = \frac{p_\mu}{p_0} e^{ipa}\left(2\Lambda_0{}^0 - \frac{p_0}{(\Lambda^{-1}p)_0} - \frac{(\Lambda^{-1}p)_0}{p_0}\right)\psi(\Lambda^{-1}p). \tag{4.12}$$

Proof: Equation (4.7) follows immediately from (4.1) and (4.6). Furthermore, from (4.2), (4.3), and (4.6) it follows that

$$(\pi_2^{GB}(a,\Lambda)\psi^{(2)})_\mu(p) = e^{ipa}\Lambda_\mu{}^\nu \psi_\nu^{(2)}(\Lambda^{-1}p) - (c_{12}^{GB}(a,\Lambda)\psi^{(2)})_\mu(p), \tag{4.13}$$

for any $\psi^{(2)} \in \mathscr{R}_2$, where we have defined

$$(c_{12}^{GB}(a,\Lambda)\psi^{(2)})_\mu(p) = \frac{p_\mu}{2p_0^2} e^{ipa} \sum_{\nu=0}^{3} p_\nu \Lambda_\nu{}^\rho \psi_\rho^{(2)}(\Lambda^{-1}p).$$

Thus, recalling the relations $\psi_0^{(2)} = 0$ and $p^\nu \Lambda_\nu{}^\rho \psi_\rho^{(2)} \times (\Lambda^{-1}p) = 0$, the above expression can be written in the form (4.10). Now, by inserting (4.10) in (4.13), we easily find the expression (4.8).

Similarly, the action of the Poincaré group on the elements of the subspace $\mathscr{R}_3$ can be decomposed in the following way:

$$\begin{aligned}(\pi^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) &= \frac{(\Lambda I \Lambda^{-1}p)_\mu}{(\Lambda^{-1}p)_0} e^{ipa}\psi(\Lambda^{-1}p) \\ &= \frac{(Ip)_\mu}{2p_0^2} e^{ipa} \frac{p^\nu(\Lambda I\Lambda^{-1}p)_\nu}{(\Lambda^{-1}p)_0}\psi(\Lambda^{-1}p) \\ &\quad + \left[\frac{(\Lambda I\Lambda^{-1}p)_\mu}{(\Lambda^{-1}p)_0} e^{ipa}\psi(\Lambda^{-1}p) \right. \\ &\quad - \frac{(Ip)_\mu}{2p_0^2} e^{ipa} \\ &\quad \left. \times \frac{p^\nu(\Lambda I\Lambda^{-1}p)_\nu}{(\Lambda^{-1}p)_0}\psi(\Lambda^{-1}p)\right], \end{aligned} \tag{4.14}$$

where the first term in the above decomposition belongs to $\mathscr{R}_3$ and the last term is in $\mathscr{H}_2 = \mathscr{R}_1 \oplus \mathscr{R}_2$.

By means of the identity

$$p^\nu(\Lambda I\Lambda^{-1}p)_\nu = \sum_{\nu=0}^{3} (\Lambda^{-1}p)_\nu (\Lambda^{-1}p)_\nu = 2(\Lambda^{-1}p)_0^2,$$

we can rewrite Eq. (4.14) as

$$\begin{aligned}(\pi^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) &= \frac{(Ip)_\mu}{p_0} e^{ipa} \frac{(\Lambda^{-1}p)_0}{p_0}\psi(\Lambda^{-1}p) \\ &\quad + \left[\frac{(\Lambda I\Lambda^{-1}p)_\mu}{(\Lambda^{-1}p)_0} e^{ipa}\psi(\Lambda^{-1}p) \right. \\ &\quad \left. - \frac{(Ip)_\mu}{p_0} e^{ipa} \frac{(\Lambda^{-1}p)_0}{p_0}\psi(\Lambda^{-1}p)\right]. \end{aligned} \tag{4.15}$$

The first term on the right-hand side of the above relation gives representation (4.9). The last term can be further decomposed in the orthogonal sum of vectors belonging to $\mathscr{R}_1$ and $\mathscr{R}_2$.

Thus, we obtain

$$\begin{aligned}(c_{23}^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) &= \left[\frac{(\Lambda I\Lambda^{-1}p)_\mu}{(\Lambda^{-1}p)_0} e^{ipa}\psi(\Lambda^{-1}p) - \frac{(Ip)_\mu}{p_0} e^{ipa} \right. \\ &\quad \left. \times \frac{(\Lambda^{-1}p)_0}{p_0}\psi(\Lambda^{-1}p)\right] - (c_{13}^{GB}(a,\Lambda)\psi^{(3)})_\mu(p), \end{aligned} \tag{4.16}$$

where

$$(c_{13}^{GB}(a,\Lambda)\psi^{(3)})_\mu(p) = \frac{p_\mu}{2p_0^2} e^{ipa} \frac{(Ip)^\nu(\Lambda I\Lambda^{-1}p)_\nu}{(\Lambda^{-1}p)_0}\psi(\Lambda^{-1}p). \tag{4.17}$$

By using the identities

$$\begin{aligned}(Ip)_\mu &= -p_\mu + 2\delta_\mu{}^0 p_0; \\ (\Lambda I\Lambda^{-1}p)_\mu &= -p_\mu + 2\Lambda_\mu{}^0(\Lambda^{-1}p)_0, \end{aligned} \tag{4.18}$$

we readily obtain

$$(Ip)^\nu(\Lambda I\Lambda^{-1}p)_\nu = 2(2\Lambda_0{}^0 p_0(\Lambda^{-1}p)_0 - p_0^2 - (\Lambda^{-1}p)_0^2).$$

By using the above identity in the definition (4.17), we get the relation (4.12).

Finally, we obtain the cocycle (4.11) by inserting (4.12) and (4.18) in the relation (4.16). □

Remark I: It is not difficult to check that the expressions (4.7)–(4.12) satisfy the relations

$$\pi_3^{GB}(a,\Lambda) = \pi_1^{*GB}(a,\Lambda), \quad \pi_2^{GB}(a,\Lambda) = \pi_2^{*GB}(a,\Lambda),$$

$$c_{23}^{GB}(a,\Lambda) = c_{12}^{*GB}(a,\Lambda), \quad c_{13}{}^{GB}(a,\Lambda) = c_{13}^{*GB}(a,\Lambda).$$

□

Given the explicit expression of the elements of the matrix (2.1) for the Gupta–Bleuler gauge representation in the one-particle space $\mathscr{H} = \mathscr{R}_1 \oplus \mathscr{R}_2 \oplus \mathscr{R}_3$, we can now investigate the structure of this matrix for the Landau gauge representation in the same one-particle space (and with the same orthogonal decomposition).

By (3.5) and (3.11), the Landau gauge representation of the Poincaré group can be written in the form

$$(U^L(a,\Lambda)\psi)_\mu(p) = (U^{GB}(a,\Lambda)\psi)(p) - ia_0(p_\mu/2p_0)e^{ipa}(p\psi)(\Lambda^{-1}p)$$

$$+p_\mu e^{ipa}\left\{\frac{(\Lambda^{-1})_0{}^i\partial_i(p\psi)(\Lambda^{-1}p)}{2(\Lambda^{-1}p)_0}\right.$$

$$\left.+\frac{1}{4}\left(\frac{1}{p_0^2}-\frac{1}{(\Lambda^{-1}p)_0^2}\right)(p\psi)(\Lambda^{-1}p)\right\}, \tag{4.19}$$

where $p\psi\equiv p^\mu\psi_\mu$.

From Eq. (4.19) we see that, for any (a,Λ), the range of the operator $U^L(a,\Lambda)-U^{GB}(a,\Lambda)$ is contained in $\mathscr{H}_1$. Furthermore, by recalling that $p^\mu\psi_\mu^{(2)}=0$ for any $\psi^{(2)}\in\mathscr{H}_2$, we have

$$U^L(a,\Lambda)|_{\mathscr{H}^{(2)}}=U^{GB}(a,\Lambda)|_{\mathscr{H}^{(2)}}.$$

Therefore, in the chosen basis, the following relations hold:

$$\pi_j^L(a,\Lambda)=\pi_j^{GB}(a,\Lambda),\quad j=1,2,3,$$

$$c_{12}^L(a,\Lambda)=c_{12}^{GB}(a,\Lambda),\quad c_{23}^L(a,\Lambda)=c_{23}^{GB}(a,\Lambda), \tag{4.20}$$

i.e., the only matrix elements that are different in the two representations are the cochains $c_{13}^L(a,\Lambda)$ and $c_{13}^{GB}(a,\Lambda)$.

Let us now investigate the explicit expression of the operators $c_{13}^L(a,\Lambda)$. From (4.5), (4.12), and (4.19) we find after some calculations

$$\begin{aligned}&(c_{13}^L(a,\Lambda)\psi^{(3)})_\mu(p)\\&\quad=(p_\mu/p_0)e^{ipa}[p_0(\Lambda^{-1})_0{}^i\partial_i\psi(\Lambda^{-1}p)\\&\qquad-ia_0(\Lambda^{-1}p)_0\psi(\Lambda^{-1}p)]+\tfrac{1}{2}(c_{13}^{GB}(a,\Lambda)\psi^{(3)})_\mu(p).\end{aligned} \tag{4.21}$$

We can write the above expression in the following compact form:

$$c_{13}^L(a,\Lambda)=L(a,\Lambda)+\tfrac{1}{2}c_{13}^{GB}(a,\Lambda), \tag{4.22}$$

where the operators $L(a,\Lambda)$ are defined by

$$\begin{aligned}&L(a,\Lambda):\ \mathscr{R}_3\to\mathscr{R}_1,\\&(L(a,\Lambda)\psi^{(3)})_\mu(p)\\&\quad=(p_\mu/p_0)e^{ipa}[p_0(\Lambda^{-1})_0{}^i\partial_i\psi(\Lambda^{-1}p)\\&\quad-ia_0(\Lambda^{-1}p)_0\psi(\Lambda^{-1}p)].\end{aligned} \tag{4.23}$$

Remark II: From the above expressions, we see that the operators $L(a,\Lambda)$ are unbounded operators, densely defined on the domain $\mathscr{D}_3^L\subset\mathscr{R}_3$ given by

$$\begin{aligned}&\mathscr{D}_3^L\equiv\{\psi_\mu^{(3)}=[(Ip)_\mu/p_0]\psi(p),\\&\psi,\partial_i\psi\in L^2(p_0\,d^3p,C_+),\quad i=1,2,3\}.\end{aligned} \tag{4.24}$$

It is not difficult to see that $L(a,\Lambda)$ are bounded, everywhere defined operators on the Sobolev space

$$H\equiv\{\psi_\mu^{(3)}=[(Ip)_\mu/p_0]\psi(p),\quad\|\psi\|_H<\infty\},$$

with the norm

$$\|\psi\|_H=\left[\int d^3p\,p_0\left(|\psi(p)|^2+\sum_{i=1}^3|\partial_i\psi(p)|^2\right)\right]^{1/2}.$$

Then, the operators $c_{13}^L(a,\Lambda)$ defined by (4.21) are bounded and everywhere defined on the space

$$\mathscr{R}_3^L=\mathscr{R}_3\cap H,$$

endowed with the norm defined by the sum of the norms of $\mathscr{R}_3$ and H. □

Remark III: From definition (4.23) one can easily check the relation $L(a,\Lambda)=L^*(a,\Lambda)$, which implies $c_{13}^L(a,\Lambda)=c_{13}^{*L}(a,\Lambda)$, as it is required by the invariance of the inner product (1.6) under the representation. □

From the explicit expressions (4.20)–(4.22) of the matrix elements π_i^L, c_{ij}^L, it is possible to check the validity of the group property (2.3) for the above defined Landau gauge representation.

Proposition 4.2: The Landau gauge representation of the Poincaré group defined by (4.20)–(4.22) satisfies the group property, i.e.,

$$\delta c_{13}^L=c_{12}^L\times c_{23}^L. \tag{4.25}$$

Proof: By (4.20) and the group property of the representation in the Gupta–Bleuler gauge we find

$$c_{12}^L\times c_{23}^L=c_{12}^{GB}\times c_{23}^{GB}=\delta c_{13}^{GB}.$$

Therefore, the relation (4.25) is equivalent to

$$\delta c_{13}^L=\delta c_{13}^{GB}.$$

By (4.22), the above equation can also be written

$$\delta L=\tfrac{1}{2}\delta c_{13}^{GB}. \tag{4.26}$$

Now, the identity (4.26) can be checked by direct calculation from (4.12), (4.23), and the definition (2.4). □

Corollary 4.3: The two cochains c_{13}^{GB} and c_{13}^L differ by a cocycle.

Remark IV: By the identity (4.26), one can obtain a one-parameter family of representations by defining

$$c_{13}^\beta=(1-\beta/2)c_{13}^{GB}+\beta L,\quad\beta\in\mathbb{R}$$

with the other matrix elements unchanged as in (4.20).

The values $\beta=0$ and $\beta=1$ in the above definition correspond to the Gupta–Bleuler gauge and Landau gauge representations, respectively. □

V. CONCLUDING REMARKS

In the previous sections, we discussed the explicit construction of the triplet (2.1) in some simple models of indefinite metric quantum field theory. In particular, we showed that the Gupta–Bleuler triplet for the one-particle space of these models can be defined provided a Hilbert space structure is associated to the corresponding two-point functions.

Therefore, the structure of the matrix (2.1) depends on the explicit definition of the Hilbert scalar product in the one particle space. As we remarked in the Introduction, this definition is not, in general, unique, allowing different representations of the field algebra that transform covariantly under different representations of the Poincaré group.

An elementary example is provided by the two-dimensional massless scalar field discussed in Sec. II. In that case, the representation space for the Gupta–Bleuler triplet (2.1) is uniquely determined, as there is a unique maximal Hilbert topology on the test function space.Nevertheless, the cohomology class of the cocycles c_{12}, c_{23}, is determined by the infrared behavior of the test function χ in the explicit definition (2.8) of the Krein norm on the space $\mathscr{S}$.

The representations of the Poincaré group in the

Gupta–Bleuler gauge and Landau gauge discussed in Sec. IV are examples of representations related by a *gauge transformation* according to the definition given in Ref. 1. The remarkable fact is that both these representations can be defined in the same (maximal) Hilbert space, the one-particle space of the free electromagnetic field in the Gupta–Bleuler formalism.

Moreover, by a suitable choice of the subspaces $\mathscr{R}_1$, $\mathscr{R}_2$, $\mathscr{R}_3$ in the decomposition of the single-particle space, we exhibited the "gauge transformation" of the resulting triplets in the form (4.20), (4.21). Therefore, recalling (4.26), we see that the triplets in the two gauges differ by the cocycle

$$d(a,\Lambda)\equiv c^L_{13}(a,\Lambda)-c^{GB}_{13}(a,\Lambda)=L(a,\Lambda)-\tfrac{1}{2}c^{GB}_{13}(a,\Lambda). \tag{5.1}$$

We now discuss the cohomology of $d(a,\Lambda)$. We start by considering the subgroup of space-time translations $(a,1)$.

Proposition 5.1: The cocycle $d(a,1)$ defined by (5.1) has nontrivial cohomology.

Proof: Suppose that $d(a,1)$ is a coboundary. Then, there exists an operator R_{13}: $\mathscr{R}_3\to\mathscr{R}_1$, such that

$$d(a,1)=\delta R_{13}(a,1). \tag{5.2}$$

Let us consider the subgroup $(\mathbf{a},1)$ of pure spatial translation. By (5.1) we have $d(\mathbf{a},1)=0$ and therefore the operator R_{13} must satisfy $\delta R_{13}(\mathbf{a},1)=0$ for any $\mathbf{a}\in\mathbb{R}^3$. It follows that the operator R_{13} commutes with any multiplication operator by a function of $\mathbf{p}$.

In particular, we get $\delta R_{13}(a_0,1)=0$ for any element $(a_0,1)$ of the subgroup of time translations. Hence, the relation (5.2) cannot hold, since $d(a_0,1)\neq 0$. □

It is worthwhile to remark that the relation (5.2) holds when we consider the representations $\pi^{GB}(a_0)$, $\pi^L(a_0)$ of the one-parameter subgroup of time translations.

Proposition 5.2: The cocycle $d(a_0,1)$ is a coboundary for the representation of the pure time translations in the Gupta–Bleuler gauge, and there exists an η-unitary operator W_0 in $\mathscr{H}$ satisfying

$$\pi^L(a_0,1)=W_0\pi^{GB}(a_0,1)W_0^{-1}. \tag{5.3}$$

Proof: From (5.1) and (4.23) we have

$$(d(a_0,1)\psi^{(3)})_\mu(p)=-ia_0p_\mu e^{ip_0a_0}\psi(p). \tag{5.4}$$

Let us define the operator

$$R_{13}\colon \mathscr{R}^L_3\to\mathscr{R}_1,$$
$$(R_{13}\psi^{(3)})_\mu(p)=(p_\mu/p_0)(1+p^i\partial_i)\psi(p). \tag{5.5}$$

From the above definition we obtain

$$\begin{aligned}(\delta R_{13}(a_0,1)\psi^{(3)})_\mu(p)&=(p_\mu/p_0)[e^{ip_0a_0}(1+p^i\partial_i)-(1+p^i\partial_i)e^{ip_0a_0}]\psi(p)\\&=-ia_0p_\mu e^{ip_0a_0}\psi(p),\end{aligned} \tag{5.6}$$

so that the cocycle (5.4) is a coboundary.
Furthermore, the operator (5.5) satisfies the relation

$$R^*_{13}=-R_{13}. \tag{5.7}$$

By (5.6) and (5.7), the matrix

$$W_0=\begin{pmatrix}1&0&-R_{13}\\0&1&0\\0&0&1\end{pmatrix}$$

satisfies (5.3) and $W^\dagger=W^{-1}$, i.e., the representations $\pi^L(a_0,1)$ and $\pi^{GB}(a_0,1)$ are η-unitary equivalent. □

We finally discuss the subgroup of pure Lorentz transformations. In this case, the cochain transformation

$$c^{GB}_{13}(0,\Lambda)\to c^L_{13}(0,\Lambda)=c^{GB}_{13}(0,\Lambda)+d(0,\Lambda), \tag{5.8}$$

can be characterized in the following way

Proposition 5.3: The cocycle $d(0,\Lambda)$ in the transformation (5.8) can be written in the form

$$\begin{aligned}d(0,\Lambda)=&-\{c^{GB}_{12}(0,\Lambda)R^*_{12}+R_{12}c^{GB}_{23}(0,\Lambda)\}\\&+\tfrac{1}{2}\delta R_{13}(0,\Lambda),\end{aligned} \tag{5.9}$$

where R_{13} is given by (5.5) and the operator R_{12} is defined (on a dense domain in $\mathscr{R}_2$) by

$$(R_{12}\psi^{(2)})_\mu(p)=-\tfrac{1}{4}p_\mu\,\partial^i\Psi^{(2)}_i(p). \tag{5.10}$$

Proof: By definition (5.10) and the relation

$$\langle\psi^{(3)},R_{12}\psi^{(2)}\rangle=\langle R^*_{12}\psi^{(3)},\psi^{(2)}\rangle,$$

we obtain the expression

$$(R^*_{12}\psi^{(3)})_i(p)=\tfrac{1}{2}[p_0\,\partial_i\psi(p)+(p_i/p_0)p^j\,\partial_j\psi(p)], \tag{5.11}$$

with $i=1,2,3$.

Then, the result follows by direct calculation from (4.10), (4.11) and the above expressions for R_{12}, R^*_{12}, and R_{13}. □

Remark I: By considering the subgroup of pure Lorentz transformations, it follows from the above proposition that the cohomology of $d(0,\Lambda)$ is not trivial.

In fact, if the term $d(0,\Lambda)$ is a coboundary, we get from (5.9)

$$c_{12}(0,\Lambda)R^*_{12}+R_{12}c_{23}(0,\Lambda)=\delta S_{13}(0,\Lambda), \tag{5.12}$$

where

$$\delta S_{13}(0,\Lambda)\equiv\pi_1(0,\Lambda)S_{13}-S_{13}\pi_3(0,\Lambda)$$

for some operator S_{13}: $\mathscr{R}_3\to\mathscr{R}_1$.

Hence, the left-hand side of (5.12) must be invariant under the coordinate transformation:[10]

$$\pi_j\to\pi_j,\quad j=1,2,3,$$
$$c_{12}\to c_{12}+\delta W_{12},\quad c_{23}\to c_{23}+\delta W_{23},$$

with W_{12}: $\mathscr{R}_2\to\mathscr{R}_1$ and W_{23}: $\mathscr{R}_3\to\mathscr{R}_2$ arbitrarily chosen operators.

By choosing $W_{12}=R_{12}$, $W_{23}=R^*_{12}$, we get the condition

$$\delta(R_{12}R^*_{12})(0,\Lambda)=0, \tag{5.13}$$

which is not satisfied by (5.10) and (5.11). □

Remark II: It is possible to relate the transformation (5.9) to a coordinate change of the form

$$\pi(a,\Lambda)\to W\pi(a,\Lambda)W^{-1},$$

where W is the matrix defined by

$$W=\begin{pmatrix}1 & -R_{12} & -R_{12}R^{*}_{12}/2-R_{13}/2\\ 0 & 1 & R^{*}_{12}\\ 0 & 0 & 1\end{pmatrix}, \tag{5.14}$$

The above operator preserves the indefinite inner product and implements the transformation:[10]

$$c_{12}\to c_{12}+\delta R_{12},\quad c_{23}\to c_{23}+\delta R^{*}_{12},$$
$$c_{13}\to c_{13}-\{(\delta R_{12})R^{*}_{12}-R_{12}(\delta R^{*}_{12})\}/2$$
$$-c_{12}R^{*}_{12}-R_{12}c_{23}+\delta R_{13}/2.$$

Then, by using (5.9), it is not difficult to check that the gauge transformation

$$\pi^{\mathrm{GB}}(0,\Lambda)\to\pi^{L}(0,\Lambda),$$

can be represented in terms of the matrix (5.14) in the form

$$\pi^{L}(0,\Lambda)=W\pi^{\mathrm{GB}}(0,\Lambda)W^{-1}+\delta W(0,\Lambda)W^{-1},$$

where

$$\delta W\equiv\begin{pmatrix}0 & -\delta R_{12} & -\delta(R_{12}R^{*}_{12})/2\\ 0 & 0 & \delta R^{*}_{12}\\ 0 & 0 & 0\end{pmatrix}. \qquad \square$$

[1] F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[2] F. Strocchi, Commun. Math. Phys. **56**, 57 (1977).
[3] F. Strocchi, Phys. Rev. D **17**, 2010 (1978).
[4] H. Grundling and C. A. Hurst, Commun. Math. Phys. **98**, 369 (1985).
[5] H. Grundling and C. A. Hurst, J. Math. Phys. **28**, 559 (1987).
[6] H. Grundling, Commun. Math. Phys. **114**, 69 (1988).
[7] G. Rideau, J. Math. Phys. **19**, 1627 (1978).
[8] G. Rideau, Rep. Math. Phys. **16**, 251 (1979).
[9] G. Rideau, *Lectures Notes in Physics* (Springer-Verlag, Berlin, 1982), Vol. 153, p. 302.
[10] H. Araki, Commun. Math. Phys. **97**, 149 (1985).
[11] J. Bognar, *Indefinite Inner Product Spaces* (Springer-Verlag, Berlin, 1974).
[12] G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré A **33**, 251 (1980); SNS 9/79, Pisa, Italy (1979); G. Morchio, D. Pierotti, and F. Strocchi, ISAS 84/87/EP, Trieste, Italy (1987).
[13] G. Morchio and F. Strocchi, Nucl. Phys. B **211**, 471 (1983); *Erice Lectures in Fundamental Problems of Gauge Theory*, edited by G. Velo and A. S. Wightman (Plenum, New York, 1986).
[14] R. F. Streater and I. F. Wilde, Nucl. Phys. B **24**, 561 (1970).
[15] R. F. Streater, *Many Degrees of Freedom in Field Theory*, edited by L. Streit (Plenum, New York, 1978).
[16] P. Basarab-Horwath, R. F. Streater, and J. Wright, Commun. Math. Phys. **68**, 195 (1979).
[17] L. K. Hadjiivanov, JINR Commun. E2-80-445, Dubna, USSR (1980).
[18] G. Morchio, D. Pierotti, and F. Strocchi, Ann. Phys. **188**, 217 (1988).
[19] D. Pierotti, J. Math. Phys. **26**, 143 (1985).
[20] L. K. Hadjiivanov and D. T. Stoyanov, Rep. Math. Phys. **15**, 361 (1979).
[21] A. L. Carey and C. A. Hurst, Lett. Math. Phys. **2**, 227 (1978).

All $\overline{SL}(3,R)$ ladder representations

Dj. Šijački
Institute of Physics, P.O. Box 57, Belgrade, Yugoslavia

(Received 11 October 1989; accepted for publication 14 March 1990)

All unitary and nonunitary ladder (multiplicity free with respect to the $\overline{SO}(3)$ maximal compact subgroup) representations of the double covering group $\overline{SL}(3,R)$ of the $SL(3,R)$ group are cataloged and presented explicitly. The list of nonunitary representations is corrected and completed, and a new insight concerning the $\Delta j = 2$ ladder representation starting with $j = \frac{3}{2}$ is obtained.

I. INTRODUCTION

The interest in the $SL(3,R)$ group, i.e., its unitary irreducible representations (unirreps), was initiated in the physics literature by Dothan, Gell-Mann, and Ne'eman.[1] They suggested an assignment of the particles laying on a single Regge trajectory to the $SL(3,R)$ $\Delta j = 2$ ladder [$SO(3)$ subgroup multiplicity-free] unirreps. At that time only the ladder unirreps starting with $j = 0$ and $j = 1$ were known, and allowed for an assignment of the π and ρ Regge trajectory recurrences. It turned our that the relativistic extension of this model is given by the $SL(4,R)$ symmetry,[2] and that its affine extension and the quantum field theory framework provide for a classification of all known hadron resonances.[3]

Biedenharn and Weaver[4] promoted the application of the $SL(3,R)$ symmetry and its unirreps in the field of nuclear physics leading subsequently to various generalizations.

Hehl and collaborators formulated an affine generalization[5–9] of the Poincaré gauge theory of gravity based on the $GA(4,R)$ gauge symmetry. The fermionic matter fields of this theory are determined by the $SL(3,R)$ stability subgroup spinorial unirreps.

The nonunitary $SL(3,R)$ representations are essential for the three-dimensional relativistic quantum field theory (RQFT) as well as for the extended object theory development: (a) these representations determine the form of the 3-D quantum fields, and are of inevitable importance when the fields are defined in a curved space-time, (b) they define both the tensorial (bosonic) and spinorial (fermionic) fields of a 3-D affine gauge theory of gravity, and finally (c) the nonunitary $SL(3,R)$ representations play an important role in the membrane and supermembrane[10] theories, in the program of formulating a spinning membrane theory,[11] as well as in the theory of 3-D world spinors and realizations of $Diff(3,R)$ transformations.[12,13]

The $SL(3,R)$ ladder unirreps, both tensorial and spinorial, were a subject of numerous papers and various approaches. There were two questions of special interest: the construction of all ladder unirreps, and the existence, i.e., the reason of nonexistence, of a ladder unirrep starting with the rotational subgroup label $j = \frac{3}{2}$. The first question has been settled quite some time ago (though reappeared and resettled again), while the second question kept the attention of the workers in the field.

The study of the $SL(3,R)$ unirreps included the following approaches: (a) in simultaneous eigenvalues problem solution for the Casimir operators and a certain paritylike operator,[14] (b) an explicit solution of the algebraic equations for the noncompact operators matrix elements that are obtained from the commutation relations,[15–17] (c) an analytic continuation of the $SU(3)$ representation labels to the labels of the corresponding noncompact $SL(3,R)$ group unirreps,[18] (d) a (rather remarkable) noncompact operator construction in terms of the forth order "spin $\frac{1}{2}$" boson operators,[19] (e) a general method based on the Harish–Chandra – Kihlberg work,[20,21] (f) an $O(3)$ shift operator technique,[22] (g) an enveloping algebra method,[23,24] (h) a decontraction of the $T_5 \times SO(3)$ noncompact operators into the $SL(3,R)$ ones,[25] (i) a geometric quantization method,[26,27] (j) a determination of the unitary dual,[28,29] and (k) a so-called "constructive" method (with numerous incorrect and confusing results).[30,31]

The nonunitary $SL(3,R)$ ladder representations were treated by the enveloping algebra method,[23] and by the constructive method.[31] In Ref. 23 the $\Delta j = 2$ nonunitary ladder representations were studied, and thus only partial results were obtained, while the results of Ref. 31 are mostly incorrect.

II. PRELIMINARIES

The $SL(3,R)$ group is a group of R^3 space volume-preserving transformations. Its maximal compact subgroup is the $SO(3)$ group. Let us denote the $SL(3,R)$ group generators, in the so-called spherical basis, by J_0, $J_\pm$; T_μ ($\mu = 0, \pm 1, \pm 2$), where the three J operators form an $SO(3)$ vector-operator ($j = 1$), while the remaining five noncompact T operators form an $SO(3)$ quadrupole operator ($j = 2$). The minimal set of the $SL(3,R)$ commutation relations is given as follows:

$$[J_0, J_\pm] = \pm J_\pm, \quad [J_+, J_-] = 2J_0,$$

$$[J_0, T_\mu] = \mu T_\mu, \quad [J_\pm, T_\mu] = \sqrt{6 - \mu(\mu \pm 1)}\, T_{\mu \pm 1},$$

$$[T_{+2}, T_{-2}] = -4J_0.$$

Let us denote by $\overline{SL}(3,R)$ the double-covering (universal-covering) group of the $SL(3,R)$ group. The $SU(2)$ is the maximal compact subgroup of the $\overline{SL}(3,R)$ group. One has the following minimal sequence of relevant groups:

 0022-2488/90/081872-05$03.00

$$
\begin{array}{ccccccc}
1 & - & Z_2 & \rightarrow & \overline{\mathrm{SL}}(3,R) & \rightarrow & \mathrm{SL}(3,R) \rightarrow 1 \\
 & & & & \cup & & \cup \\
1 & - & Z_2 & \rightarrow & \mathrm{SU}(2) & \rightarrow & \mathrm{SO}(3) \rightarrow 1,
\end{array}
$$

where Z_2 is the center of the $\overline{\mathrm{SL}}(3,R)$ group. It is well known that there are no finite-dimensional spinorial representations (representations reducing to the half-integer labeled SO(3) representations) of the SL(3,R) group, as well as that the $\overline{\mathrm{SL}}(3,R)$ group is a group of infinite-dimensional complex matrices. Both single-valued (tensorial) and double-valued (spinorial) SL(3,R) representations are obtained as single-valued representations of the $\overline{\mathrm{SL}}(3,R)$ group.

In general we make use of the Hilbert space defined as a symmetric homogeneous space over the maximal compact subgroup, i.e., over the $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \supset \mathrm{U}(1)_L \times \mathrm{U}(1)_R$ group (L,R refer to the left, right group action, respectively).[21] The basis elements are $|{}_k{}^j{}_m\rangle$, where $j = j_L = j_R$ is the SU(2) label, while k, m are the $\mathrm{U}(1)_L$, $\mathrm{U}(1)_R$ labels, respectively. When the multiplicity is known *a priori* to be trivial, one starts with a homogeneous space over the $\mathrm{SU}(2) \supset \mathrm{U}(1)$ group with the usual basis vectors $|{}^j_m\rangle$.

III. REPRESENTATION INVARIANT SPACES

There are, as it was shown in the general analysis,[21] two principal types of the $\overline{\mathrm{SL}}(3,R)$ representation invariant spaces. They are distinguished according to the multiplicity of their SU(2) subgroup invariant subspaces: (a) the *a priori* ladder-type $\overline{\mathrm{SL}}(3,R)$ invariant spaces [each SU(2) invariant subspace appears at most once], and (b) the *a priori* generic-type $\overline{\mathrm{SL}}(3,R)$ invariant spaces [the SU(2) invariant subspaces could appear in the reduction more than once]. We consider these two cases separately.

A. Ladder-type invariant spaces

The noncompact operator matrix elements in the $|{}^j_m\rangle$ basis are given by the following expression:[21]

$$
\left\langle {}^{j'}_{m'} \middle| T_\mu \middle| {}^{j}_{m} \right\rangle = (-)^{j-m'} \begin{bmatrix} j' & 2 & j \\ -m' & \mu & m \end{bmatrix} \langle j' \| T \| j \rangle, \quad (1)
$$

where

$$
\begin{aligned}
\langle j' \| T \| j \rangle = & -i(-)^{j'} \sqrt{\tfrac{2}{3}} \sqrt{(2j'+1)(2j+1)} \\
& \times \{ \sigma_1 + i\sigma_2 - \tfrac{1}{2}[j'(j'+1) - j(j+1)] \} \\
& \times \begin{bmatrix} j' & 2 & j \\ 0 & 0 & 0 \end{bmatrix}.
\end{aligned}
$$

In this expression j is an integer and σ_1, σ_2 are two real $\overline{\mathrm{SL}}(3,R)$ representation parameters. One can replace the 3-j symbol in Eq. (1) by its explicit analytic expression and then allow j to be half-integer as well. In this manner one can treat simultaneously both tensorial and spinorial representations.

The noncompact operators matrix elements read now as follows:

$$
\left\langle {}^{j'}_{m'} \middle| T_\mu \middle| {}^{j}_{m} \right\rangle = (-)^{j-m'} \begin{bmatrix} j' & 2 & j \\ -m' & \mu & m \end{bmatrix} \langle j' \| T \| j \rangle, \quad (2)
$$

where

$$
\begin{aligned}
\langle j-2 \| T \| j \rangle = & -i(-)^{2j} (\sigma_1 + i\sigma_2 \\
& + 2j - 1) \left[\frac{j(j-1)}{(2j-1)} \right]^{1/2}, \\
\langle j \| T \| j \rangle = & +i(-)^{2j} (\sigma_1 + i\sigma_2) \left[\frac{2j(j+1)(2j+1)}{3(2j+3)(2j-1)} \right]^{1/2}, \\
\langle j+2 \| T \| j \rangle = & -i(\sigma_1 + i\sigma_2 - 2j - 3) \\
& \times \left[\frac{(j+1)(j+2)}{(2j+3)} \right]^{1/2}.
\end{aligned}
$$

The $\overline{\mathrm{SL}}(3,R)$ invariant spaces of the ladder type are now defined as the subspaces of the space $H = \Sigma^{\oplus} H(j)$ which are invariant with respect to the action of the $\overline{\mathrm{SL}}(3,R)$ operators, where $H(j) = \{|{}^j_m\rangle;\ |m| < j\}$ are SU(2) invariant spaces determined by the label j, i.e., $J^2 \rightarrow j(j+1)$. In order to construct explicitly an $\overline{\mathrm{SL}}(3,R)$ invariant space one collects together all $H(j)$ subspaces that are connected mutually by the T_μ operator action, verifies the validity of the $\overline{\mathrm{SL}}(3,R)$ commutation relations, and determines the spaces $H(j_{\min})$ and $H(j_{\max})$, as well as the $\overline{\mathrm{SL}}(3,R)$ group labels σ_1 and σ_2, in such a way that T_μ cannot connect $H(j_{\min})$, $H(j_{\max})$ to $H(j < j_{\min})$, $H(j > j_{\max})$, respectively.

It is obvious from (1), (2) that the T_μ operators connect the $H(j)$ spaces that differ in j by 0, ± 2, and it is rather straightforward to check that the $\overline{\mathrm{SL}}(3,R)$ commutation relations are satisfied. The $j_{\min}$ values are obtained from (2) when the requirement that $\langle j_{\min} - 2 \| T \| j_{\min} \rangle = 0$ is satisfied, i.e., when

$$
(\sigma_1 + i\sigma_2 + 2j_{\min} - 1) \left[\frac{j_{\min}(j_{\min} - 1)}{2j_{\min} - 1} \right]^{1/2} = 0.
$$

The allowed $j_{\min}$ values are now as follows:

$$
\begin{aligned}
& j_{\min} = 0, 1, && \text{for any } \sigma_1 \text{ and } \sigma_2, \\
& j_{\min} = \tfrac{1}{2}, && \text{for } \sigma_1 = \sigma_2 = 0, \\
& j_{\min} = \tfrac{3}{2}, 2, \tfrac{5}{2}, 3, \ldots, && \text{for } \sigma_1 = 1 - 2j_{\min}, \sigma_2 = 0.
\end{aligned} \quad (3)
$$

The $j_{\max}$ values are obtained either from (2) provided $\langle j_{\max} + 2 \| T \| j_{\max} \rangle = 0$, i.e., when

$$
(\sigma_1 + i\sigma_2 - 2j_{\max} - 3) \left[\frac{(j_{\max} + 2)(j_{\max} + 1)}{2j_{\max} + 3} \right]^{1/2} = 0,
$$

or from the requirement that j increases indefinitely. The allowed $j_{\max}$ values are now as follows:

$$
\begin{aligned}
& j_{\max} = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots, && \text{for } \sigma_1 = 2j_{\max} + 3, \quad \sigma_2 = 0, \\
& j_{\max} \text{ is unlimited} && \text{for } \sigma_2 \neq 0.
\end{aligned} \quad (4)
$$

Finally, one determines the $\overline{\mathrm{SL}}(3,R)$ invariant spaced by

selecting compatible values for σ_1, σ_2, $j_{\min}$, and $j_{\max}$ from (3) and (4).

B. Generic type invariant spaces

The noncompact operator matrix elements in the $|_k{}^j{}_m\rangle$ basis are given by the following expression:[21]

$$\left\langle \begin{matrix} j' \\ k' \end{matrix} \middle| T \middle| \begin{matrix} j \\ k \end{matrix} \right\rangle = -i(-)^{j'-k'}\sqrt{(2j'+1)(2j+1)}$$
$$\times\left\{\left(\left(\frac{2}{3}\right)^{1/2}(\sigma_1+i\sigma_2)-\frac{1}{\sqrt{6}}[j'(j'+1)-j(j+1)]\right)\begin{bmatrix} j' & 2 & j \\ -k' & 0 & k \end{bmatrix}\right.$$
$$\left.-(\delta_1+i\delta_2+k+1)\begin{bmatrix} j' & 2 & j \\ -k' & 2 & k \end{bmatrix}-(\delta_1+i\delta_2-k+1)\begin{bmatrix} j' & 2 & j \\ -k' & -2 & k \end{bmatrix}\right\}. \tag{5}$$

In this expression, obtained from a general analysis, j is either integer or half-integer, and σ_1, σ_2, δ_1, and δ_2 are real $\overline{\mathrm{SL}}(3,R)$ representation labels (two of them suffice).

The $\overline{\mathrm{SL}}(3,R)$ invariant spaces of the generic type are defined as the subspaces of the space $H=\Sigma^{\oplus}H(j,k)$ that are invariant with respect to the action of the $\overline{\mathrm{SL}}(3,R)$ operators, where

$$H(j,k)=\left\{\left|\begin{matrix} j \\ k \end{matrix}\ \begin{matrix} \\ m \end{matrix}\right\rangle;\quad |m|<j,j>|k|\right\}$$

are SU(2) invariant spaces. In general, there is a nontrivial multiplicity of the $H(j,k)$ subspaces in H that is determined by, and increases with the label k. In order to obtain explicitly an $\overline{\mathrm{SL}}(3,R)$ invariant space, one collects together all $H(j,k)$ subspaces that are connected mutually by the T_μ operator action, verifies the validity of the $\overline{\mathrm{SL}}(3,R)$ commutation relations, and determines the spaces $H(j_{\min},k)$ and $H(j_{\max},k)$.

It is obvious from (5) that the T_μ operators connect the $H(j,k)$ spaces that differ in k by 0, ± 2, and in j by 0, ± 1, ± 2, while the $\overline{\mathrm{SL}}(3,R)$ commutation relations are satisfied by construction [Eq. (5) is defined for both integer and half-integer j—there is no need for a continuation in j]. Moreover, it is clear from Eq. (5) that the requirement

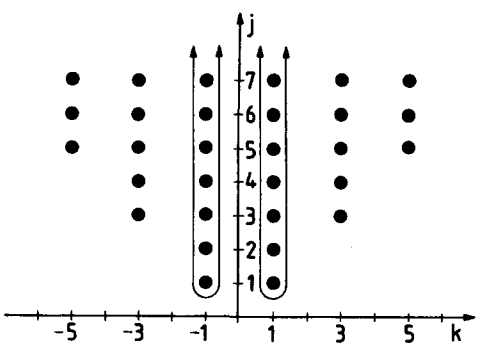


FIG. 1. The $k=\pm 1$ subspace of the $\overline{\mathrm{SL}}(3,R)$ generic type $j_{\min}=1$ invariant space.

$$\left\langle \begin{matrix} j' \\ k' \end{matrix}\ \begin{matrix} \\ m' \end{matrix}\middle| T_\mu \middle| \begin{matrix} j \\ k \end{matrix}\ \begin{matrix} \\ m \end{matrix}\right\rangle = (-)^{j'-m'}\begin{bmatrix} j' & 2 & j \\ -m' & \mu & m \end{bmatrix}\left\langle \begin{matrix} j' \\ k' \end{matrix}\middle|\middle| T \middle|\middle| \begin{matrix} j \\ k \end{matrix}\right\rangle, \tag{5}$$

where

$\langle {}_{k+2}{}^{j'}\|T\|{}_k{}^{j}\rangle=0$, $k>0$ implies that $\langle {}_{-k-2}{}^{j'}\|T\|{}_{-k}{}^{j}\rangle=0$, $k>0$ as well. The only possibility now to obtain a ladderlike invariant subspace within a class of the generic-type invariant spaces [of a new kind, not contained in the ladder-type case $(k=0)$] is achieved when $k=1$. One finds, by demanding that $\langle {}^{j'}_3\|T\|{}^{j}_1\rangle=0$, i.e. [cf. Eq. (5)], that

$$(\delta_1+i\delta_2+2)=0.$$

This implies that $\delta_1=-2$, $\delta_2=0$, and that $\langle {}^{j'}_{-3}\|T\|{}^{j}_{-1}\rangle=0$ as well. The T_μ operators acts within a subspace defined by $k=\pm 1$, while $j=1,2,3,\ldots$, see Fig. 1. At first sight each $H(j)$ appears twice, i.e., as $H(j,+1)$ and as $H(j,-1)$. However, the space $H=\Sigma^{\oplus}(H(j,+1)\oplus H(j,-1))$ is not irreducible. It can be slit into two irreducible spaces[21] $H(j,+)$ and $H(j,-)$ with the bases vectors given as follows:

$$H(j,+):\ \left|\begin{matrix} j \\ m \end{matrix}\right]=(2)^{-1/2}\left[\left|\begin{matrix} j \\ 1 \end{matrix}\ \begin{matrix} \\ m \end{matrix}\right\rangle+(-)^j\left|\begin{matrix} j \\ -1 \end{matrix}\ \begin{matrix} \\ m \end{matrix}\right\rangle\right],$$
$$H(j,-):\ \left|\begin{matrix} j \\ m \end{matrix}\right]=(2)^{-1/2}\left[\left|\begin{matrix} j \\ 1 \end{matrix}\ \begin{matrix} \\ m \end{matrix}\right\rangle-(-)^j\left|\begin{matrix} j \\ -1 \end{matrix}\ \begin{matrix} \\ m \end{matrix}\right\rangle\right].$$

Obviously, each $H(j)$ subspace, $j=1,2,3,\ldots$, appear in both $H(j,+)$ and $H(j,-)$ spaces only once. A natural $H(j_{\min})$ subspace is found for $j_{\min}=1$ when the $\overline{\mathrm{SL}}(3,R)$ commutation relations are satisfied without any restrictions on σ_1 and σ_2. Nontrivial $j_{\min}$ values are obtained by imposing the following conditions:

$$\left\langle \begin{matrix} j_{\min} \\ 1 \end{matrix}\ \begin{matrix} -n \\ \end{matrix}\middle|\middle| T \middle|\middle| \begin{matrix} j_{\min} \\ -1 \end{matrix}\right\rangle=\left\langle \begin{matrix} j_{\min} \\ 1 \end{matrix}\ \begin{matrix} -n \\ \end{matrix}\middle|\middle| T \middle|\middle| \begin{matrix} j_{\min} \\ 1 \end{matrix}\right\rangle, n=1,2$$

which are satisfied provided $\sigma_1=-2j_{\min}$, $\sigma_2=0$. Thus the possible $j_{\min}$ values read

$$j_{\min}=1,\quad \sigma_1\in R,\quad \sigma_2\in R,\quad \delta_1=-2,\quad \delta_2=0.$$
$$j_{\min}=2,3,4,\ldots,\quad \sigma_1=-2j_{\min},\quad \sigma_2=0,\quad \delta_1=-2,\quad \delta_2=0. \tag{6}$$

On the other hand, the $j_{\max}$ value is unlimited unless we impose the following conditions:

$$\left\langle \begin{matrix} j_{\max} \\ 1 \end{matrix}\ \begin{matrix} +n \\ \end{matrix}\middle|\middle| T \middle|\middle| \begin{matrix} j_{\max} \\ -1 \end{matrix}\right\rangle=\left\langle \begin{matrix} j_{\max} \\ 1 \end{matrix}\ \begin{matrix} +n \\ \end{matrix}\middle|\middle| T \middle|\middle| \begin{matrix} j_{\max} \\ 1 \end{matrix}\right\rangle,$$

$n=1,2$, which are satisfied if $\sigma_1=2j_{\max}+2$, $\sigma_2=0$. Thus one has

$$j_{\max} = 2,3,4,\ldots, \quad \text{for } \sigma_1 = 2j_{\max} + 2, \quad \sigma_2 = 0,$$
$$j_{\max} \quad \text{is unlimited for} \begin{cases} \sigma_1 \in R, \ \sigma_2 \in R \setminus \{0\}, \\ \sigma_1 \in R \setminus \{\pm (2j+3)\}, \quad \sigma_2 \in R. \end{cases} \tag{7}$$

Finally, one determines the $\overline{\mathrm{SL}}(3,R)$ invariant spaces by selecting compatible values for $\sigma_1, \sigma_2, j_{\min}$, and $j_{\max}$ from (6) and (7).

IV. CONTRAGRADIENCE AND UNITARITY

Let $D(g)$ be an $\overline{\mathrm{SL}}(3,R)$ representation. The corresponding contragradient representation $\overline{D}(g)$ is given by $\overline{D}(g) = D^t(g^{-1})$, implying that the $\overline{D}(g)$ noncompact generators $\overline{T}_\mu$ are given in terms of the corresponding $D(g)$ generators T_μ by $\overline{T}_\mu = -(T_{-\mu})^t$.

It is rather straightforward to find from Eq. (5) that

$$\left\langle \begin{matrix} j' \\ k' \quad m' \end{matrix} \middle| \overline{T}_\mu [\sigma_1,\sigma_2,\delta_1,\delta_2] \middle| \begin{matrix} j \\ k \quad m \end{matrix} \right\rangle = \left\langle \begin{matrix} j' \\ k' \quad m' \end{matrix} \middle| T_\mu [-\sigma_1,-\sigma_2,-\delta_1,-\delta_2] \middle| \begin{matrix} j \\ k \quad m \end{matrix} \right\rangle,$$

i.e., that the contragradient representation of some $\overline{\mathrm{SL}}(3,R)$ representation is given by the same analytic expression with the representation parameters substituted by their negative values,

$$\overline{D}[\sigma_1,\sigma_2,\delta_1,\delta_2] = D[-\sigma_1,-\sigma_2,-\delta_1,-\delta_2]. \tag{8}$$

The unitarity of the $\overline{\mathrm{SL}}(3,R)$ representations, i.e., the Hermiticity of the corresponding generator representations is a condition that in general requires a rather complex type of the representation Hilbert space scalar product.[21] The representation space scalar product takes, in the case of ladder representations, the usual form—there is no need for a nontrivial scalar product kernel.

The noncompact operators Hermiticity condition,

$$T_\mu^+ = (-)^\mu T_{-\mu},$$

for the matrix elements of Eqs. (2) and (5) implies the following possibilities:[21]

A. Ladder unirreps Eq. (2)

$$\text{(a)} \quad \sigma_1 = 0, \ \sigma_2 \begin{cases} \in R, & j_{\min} = 0,1 \\ = 0, & j_{\min} = \frac{1}{2}. \end{cases}$$

B. Nontrivial-multiplicity unirreps Eq. (5)

$$\text{(b)} \quad \sigma_1 = 0, \ \sigma_2 \in R, \ \delta_1 = 0, \ \delta_2 \in R, \tag{9}$$

$$\text{(c)} \quad \sigma_1 = 0, \ \sigma_2 \in R, \ |\delta_1| < \begin{cases} \frac{1}{2}, & j_{\min} = \frac{1}{2}, \\ 1, & j_{\min} = 0,1, \end{cases} \quad \delta_2 = 0.$$

$$\text{(d)} \quad \sigma_1 = 0, \ \sigma_2 \in R, \ \delta_1 = \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots, \ \delta_2 = 0.$$

V. UNITARY AND NONUNITARY LADDER REPRESENTATIONS SUMMARY

The list of all ladder unitary and nonunitary $\overline{\mathrm{SL}}(3,R)$ irreducible representations can now be obtained easily by combining the results of (3) and (4), and of (6) and (7) with the contragradience (8) and unitarity (9) relations. The list of all unitary (infinite-dimensional) and nonunitary (finite- and infinite-dimensional) $\overline{\mathrm{SL}}(3,R)$ representations is given in the following. Representations are characterized by the minimal set of labels, when necessary by the minimal and maximal $\overline{\mathrm{SO}}(3)$ label values $j_{\min}$ and $j_{\max}$, respectively, and the j content (reduction to the $\overline{\mathrm{SO}}(3)$ subrepresentations) is given as well.

A. Unitary irreducible representations

U1. Scalar representation;
$\{j\} = \{0\}$.

U2. $D[\sigma_1 = 0, \sigma_2 = 0]$, $j_{\min} = 1/2$;
$\{j\} = \{\frac{1}{2}, \frac{5}{2}, \frac{9}{2}, \ldots\}$.

U3. $D[\sigma_1 = 0, \sigma_2 \in R]$, $j_{\min} = 0$;
$\{j\} = \{0,2,4,\ldots\}$.

U4. $D[\sigma_1 = 0, \sigma_2 \in R]$, $j_{\min} = 1$;
$\{j\} = \{1,3,5,\ldots\}$.

B. Nonunitary finite-dimensional irreducible representations

F1. $D[\sigma_1 = \pm (2j_{\max} + 3), \sigma_2 = 0]$, $j_{\min} = 0$, $j_{\max} = 2,4,6,\ldots$;
$\{j\} = \{0,2,4,\ldots j_{\max}\}$.

F2. $D[\sigma_1 = \pm (2j_{\max} + 3), \sigma_2 = 0]$, $j_{\min} = 1$, $j_{\max} = 1,3,5,\ldots$;
$\{j\} = \{1,3,5,\ldots j_{\max}\}$.

F3. $D[\sigma_1 = \pm 2(j_{\max} + 1), \sigma_2 = 0, \delta_1 = -2, \delta_2 = 0]$, $j_{\min} = 1, j_{\max} = 2,3,4,\ldots$;
$\{j\} = \{1,2,3,\ldots j_{\max}\}$.

C. Nonunitary infinite-dimensional irreducible representations

N1. $D[\sigma_1 = \pm |1 - 2j_{\min}|, \sigma_2 = 0]$, $j_{\min} = \frac{1}{2}, 2, \frac{5}{2}, \ldots$;
$\{j\} = \{j_{\min}, j_{\min} + 2, j_{\min} + 4, \ldots\}$.

N2. $D[\sigma_1 \in R, \sigma_2 \in R \setminus \{0\}, \delta_1 = -2, \delta_2 = 0]$;
$\{j\} = \{1,2,3,\ldots\}$.

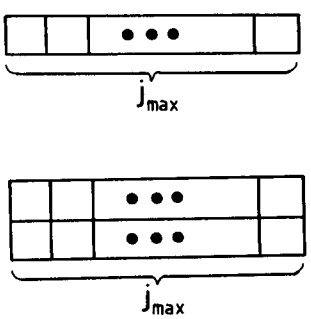


FIG. 2. Young tableaux of the $D[\sigma_1 = \pm (2j_{\max} + 3), \sigma_2 = 0]$ $\overline{\mathrm{SL}}(3,R)$ $\Delta j = 2$ ladder representations.

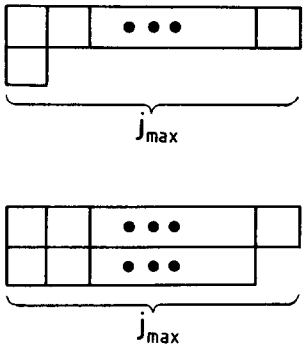


FIG. 3. Young tableaux of the $D[\sigma_1 = \pm 2(j_{max}+1), \sigma_2 = 0, \delta_1 = -2, \delta_2 = 0]$ $\overline{SL}(3,R)$ $\Delta j = 1$ ladder representations.

N3. $D[\sigma_1 \in R \setminus \{\pm 2(n+1); n = 2,3,...\}, \sigma_2 \in R, \delta_1 = -2, \delta_2 = 0]$:
$\{j\} = \{1,2,3,...\}$.

N4. $D[\sigma_1 = \pm 2j_{min}, \sigma_2 = 0, \delta_1 = -2, \delta_2 = 0]$, $j_{min} = 2,3,4,...$;
$\{j\} = \{j_{min}, j_{min}+1, j_{min}+2,...\}$.

Representation U1 is a trivial one, (scalar), the representations U2, U3, U4, F1, F2, and N1 are constructed by starting with the ladder-type invariant spaces, while the remaining representations, F3, N2, N3, and N4, are constructed by starting with the generic-type invariant spaces. The unirrep U2 starting with $j_{min} = \frac{1}{2}$ is unique, while the remaining nontrivial ladder representations characterized by $\pm|\sigma_1|$ or by $\pm|\sigma_2|$, $\sigma_2 \neq 0$ are pairs of mutually contragradient representations. The Young tableaux corresponding to the finite-dimensional nonunitary ladder representations F1, F2, and to F3 are given on Fig. 2 and on Fig. 3, respectively.

The two $\Delta j = 2, j_{min} = \frac{3}{2}$ representations,[2,15,16,19,21,22,27] as well as any other pair of N1 representations as well as any of her pair of N1 representations, starting with $j_{min} = 2, \frac{5}{2}, 3,...$, are bona fide representations, though nonunitary. It is the unitarity condition (a), $\sigma_1 = 0, \sigma_2 = \in R$, of Eq. (9) that in a clash is, in these cases, with the invariant space condition $\sigma_1 = 1 - 2j_{min}, \sigma_2 = 0$ of Eq. (3); unless, of course, $j_{min} = \frac{1}{2}$ when one recovers the unique $\Delta j = 2$ spinorial ladder representation of the $\overline{SL}(3,R)$ group.

ACKNOWLEDGMENT

This work is supported in part by RZNS (Belgrade).

[1]Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Lett. **17**, 148 (1965).
[2]Dj. Šijački, Ph. D. thesis, Duke University, 1974.
[3]Y. Ne'eman and Dj. Šijački, Phys. Lett. B **157**, 267 (1985); Phys. Rev. D **37**, 3261 (1988).
[4]L. Weaver and L. C. Biedenharn, Nucl. Phys. A **185**, 1 (1972).
[5]F. W. Hehl, G. D. Kerlick, and P. von der Heyde, Phys. Lett. B **63**, 446 (1976).
[6]F. W. Hehl, E. A. Lord, and Y. Ne'eman, Phys. Lett. B **71**, 432 (1977); Phys. Rev. D **17**, 428 (1978).
[7]Y. Ne'eman and Dj. Šijački, Ann. Phys. (N.Y.) **120**, 292 (1979); Proc. Natl. Acad. Sci. (USA) **76**, 561 (1979).
[8]Dj. Šijački, Phys. Lett. B **109**, 435 (1982).
[9]Y. Ne'eman and Dj. Šijački, Phys. Lett. B **200**, 489 (1988).
[10]E. Bergshoeff, E. Sezgin, and P. K. Townsend, Phys. Lett. B **189**, 75 (1987).
[11]Y. Ne'eman and Dj. Šijački, J. Math. Phys. **21**, 1312 (1980).
[12]Y. Ne'eman and Dj. Šijački, Phys. Lett. B **157**, 275 (1985); Int. J. Mod. Phys. A **2**, 1655 (1987); Phys. Lett. B **206**, 458 (1988).
[13]Dj. Šijački, Ann. Isr. Phys. **3**, 35 (1980); in *Frontiers in Particle Physics '83* (World Scientific, Singapore, 1984).
[14]G. Rosen, J. Math. Phys. **7**, 1284 (1966).
[15]D. W. Joseph, University of Nebraska preprint, 1970.
[16]V. I. Ogievetsky and E. Sokachev, Teor. Mat. Fiz. **23**, 214 (1975).
[17]L. Weaver and L. C. Biedenharn, J. Math. Phys. **26**, 18 (1985).
[18]N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Lett. **15**, 1041 (1965).
[19]L. C. Biedenharn, R. Y. Cusson, M. Y. Han, and O. L. Weaver, Phys. Lett. B **42**, 257 (1972), and Erratum, Phys. Lett. B **73**, 503 (1978).
[20]I. Hulthen, Ark. Fys. **38**, 175 (1968).
[21]Dj. Šijački, J. Math. Phys. **16**, 298 (1975).
[22]J. W. B. Hughes, J. Phys. A **7**, 779 (1974).
[23]E. Angelopoulos, Rep. Math. Phys. **11**, 239 (1977).
[24]E. Angelopoulos, J. Math. Phys. **19**, 2108 (1978).
[25]Y. Ne'eman and Dj. Šijački, J. Math. Phys. **21**, 1312 (1980).
[26]P. Torasso, C. R. Acad. Sci. Paris **291**, Series A-185 (1980); Acta Math. **150**, 153 (1983).
[27]J. Rawnsley and S. Sternberg, Am. J. Math. **104**, 1153 (1982).
[28]B. Speh, Math. Ann. **258**, 113 (1981).
[29]D. Vogan, Invent. Math. **83**, 449 (1986).
[30]Y. Guler, J. Math. Phys. **19**, 508 (1978).
[31]Y. Guler, J. Math. Phys. **26**, 21 (1985).

Symmetries of hyper-Kähler (or Poisson gauge field) hierarchy

K. Takasaki
Research Institute for Mathematical Sciences, Kyoto University, Sakyo-ku, Kyoto-shi 606, Japan

(Received 18 October 1989; accepted for publication 14 March 1990)

Symmetry properties of the space of complex (or formal) hyper-Kähler metrics are studied in the language of hyper-Kähler hierarchies. The construction of finite symmetries is analogous to the theory of Riemann–Hilbert transformations, loop group elements now taking values in a (pseudo-) group of canonical transformations of a simplectic manifold. In spite of their highly nonlinear and involved nature, infinitesimal expressions of these symmetries are shown to have a rather simple form. These infinitesimal transformations are extended to the Plebanski key functions to give rise to a nonlinear realization of a Poisson loop algebra. The Poisson algebra structure turns out to originate in a contact structure behind a set of symplectic structures inherent in the hyper-Kähler hierarchy. Possible relations to membrane theory are briefly discussed.

I. INTRODUCTION

This is a sequel of a previous paper[1] on hidden variables of hyper-Kähler geometry. Three types of hidden variables are introduced therein and shown to form altogether an infinite system of differential equations, which we call a "hyper-Kähler hierarchy." This is fully parallel to the case of various nonlinear integrable systems such as the KdV and KP hierarchies.[2,3]

In the theory of nonlinear integrable systems, such hidden variables also play a basic role in studying symmetries (hidden or manifest) of a system. It has been observed for a number of cases that these symmetries give a realization of various interesting Lie algebras such as Kac–Moody algebras. This is a place where the theory of nonlinear integrable systems exhibits a deep connection with representation theory. A systematic classification of nonlinear integrable systems has indeed been worked out from this standpoint for so called "soliton equations" (see, for example, Ref. 3).

For the hyper-Kähler hierarchy, a loop group (algebra) of canonical transformations in a finite-dimensional symplectic manifold plays the role of such a Lie group (algebra). In the case of self-dual metrics, for example, as pointed out by Boyer and Plebanski,[4] the target group is a (pseudo-) group Γ of canonical transformations in two dimensions. In the $4r$-dimensional hyper-Kähler case, Γ is a (pseudo-) group of $2r$-dimensional canonical transformations. Loop groups relevant to these gravitational cases are thus not of matrix type as in the case of a number of nonlinear integrable systems known until now, but of *diffeomorphism type*. This is one of the most remarkable characteristics of the hyper-Kähler hierarchy, reflecting its *gravitational nature*.

More precisely, Boyer and Plebanski in Ref. 4 derived such a group structure from the "nonlinear graviton construction" of Penrose[5] and called it a "nonlinear superposition principle." Although their results are formulated in a more geometric language (a kind of infinite-dimensional symplectic geometry), the group thus obtained is a loop (pseudo-) group of canonical transformations in the sense mentioned above. The present author[6] attempted to find a more explicit description of these symmetries, but that was not very successful. A central difficulty is that these transformations are in general extremely complicated; it seems almost hopeless to have a concise formula of transformation laws.

In this paper, we reexamine this issue in more detail and from several new aspects. First, we mostly focus our attention on infinitesimal transformations rather than finite ones. As to finite transformations, as mentioned above, we still have to overcome a number of difficulties, whose resolution is beyond our present scope. Infinitesimal transformations, to the contrary, turn out to have a much simpler, and even beautiful structure. In particular, we shall see that the structure of differential equations in the hierarchy itself has its origin in these symmetries.

Second, we now consider the hierarchy. This allows us, as discussed in Ref. 1, to consider the Plebanski key functions[7] as analogs of the τ functions.[2,3] It turns out that symmetries constructed above can be further extended to the Plebanski key functions. For the τ functions, a similar construction causes a nonzero central charge (or commutator anomaly) in the extended symmetries;[2,3] one then naturally has to consider a central extension of original symmetry algebras. For the Plebanski key functions, we shall see that the hidden symmetries give a nonlinear realization of a Poisson algebra rather than the corresponding Lie algebra of Hamiltonian vector fields, but the Poisson algebra requires no central charge.

Such a Poisson algebra structure suggests a very interesting possibility that was previously not noticed, i.e., some relation with membrane theory. In the recent progress of membrane theory (see, e.g., Ref. 8) gauge groups and algebras of a Poisson algebra are expected to play an important role like Kac–Moody and Virasoro algebras in string theory. Further, several groups,[9] along that line, discussed self-duality equations in such "Poisson gauge theories." This is evidently a system of the same kind as we have pursued in the previous paper. Our method and results might find new applications in membrane theory.

This paper is organized as follows. In Sec. II, we show a hierarchy version of the construction of finite transformations along the lines of Ref. 6. In Sec. III, we derive their

 0022-2488/90/081877-12$03.00

infinitesimal form. Infinitesimal transformations are first obtained without putting symplectic constraints (a pseudo-group arising here is, therefore, of general type), then in the symplectic situation. A final answer is to be stated in terms of Hamiltonian vector fields. Section IV deals with symmetries extended to the second Plebanski key function. This is the place where a Poisson algebra becomes relevant. In Sec. V, we argue the origin of the Poisson algebra structure in the context of contact geometry. Appendix A presents an attempt at a purely algebraic reformulation of the contents of Sec. II. Appendix B is a summary of parallel results on the first Plebanski key function.

Notations in the text obey the convention of Ref. 1 as far as possible. Here, ϵ_{AB} and ϵ^{AB} denote the symplectic ϵ symbols with indices $A,B = 1,\ldots,2r$ (r is related to the dimensionality of hyper-Kähler manifolds) normalized as

$$\epsilon_{1,2} = \epsilon^{1,2} = \cdots = \epsilon_{2r-1,2r} = \epsilon^{2r-1,2r} = 1,$$

$$\epsilon_{2,1} = \epsilon^{2,1} = \cdots = \epsilon_{2r,2r-1} = \epsilon^{2r,2r-1} = -1,$$

other components $= 0$. Symplectic indices $A,B,\ldots$ are to be raised and lowered as

$$\xi_A = \epsilon_{AB}\xi^B, \quad \eta^B = \eta_A\epsilon^{AB}.$$

In general, d stands for the total differentiation with respect to both space-time variables and other independent variables of the hierarchy, but excluding the "spectral parameter" λ. The Einstein summation convention is applied only for symplectic indices; for other indices, we shall write the summation sign explicitly in every event.

II. LOOP GROUPS AND NONLINEAR GRAVITON CONSTRUCTION

The nonlinear graviton construction, as Boyer and Plebanski stressed in Ref. 4, is very similar to the Riemann–Hilbert problem. Their formulation, however, as opposed to the Riemann–Hilbert problem, still does not take the form of a factorization problem in the loop (pseudo-)group. This issue is settled in Ref. 6. An advantage of such a formulation as a factorization problem is that it is also readily applicable to hierarchies like our hyper-Kähler hierarchy.[1] Penrose's original method[5] is based on the deformation theory of Kodaira and Spencer; such a geometric framework breaks down in the presence of an infinite number of flows that the hierarchy describes. (See Appendix A for an algebraic framework for justifying the arguments below along the line of Ref. 6.)

The Riemann–Hilbert problem has long been used as a powerful solution technique in the theory of nonlinear integrable systems. Hauser and Ernst[10] exploited it to the issue of hidden symmetries (of the Ernst equation). Ueno and Nakamura,[11] deeply influenced by that work, developed the theory of "Riemann–Hilbert transformations." Our construction of symmetries for the hyper-Kähler hierarchy lies in basically the same direction.

The essence of the Riemann–Hilbert problem method is to factorize a loop group element $g(\lambda)$ [i.e., an analytic map: $S^1 \to G$, where G is a matrix Lie group like GL $(r,\mathbb{C})$, SL $(r,\mathbb{C})$, etc.] into two pieces as:

$$g(\lambda) = h(\lambda)^{-1}\hat{h}(\lambda), \tag{2.1}$$

where h and $\hat{h}$ are required to be holomorphic maps

$$h:\{\lambda\in\mathbb{P}^1;|\lambda| > 1-\varepsilon\}\to G, \tag{2.2a}$$

$$\hat{h}:\{\lambda\in\mathbb{P}^1;|\lambda| < 1+\varepsilon\}\to G, \tag{2.2b}$$

for some small $\varepsilon > 0$, where $S^1 = \{\lambda\in\mathbb{P}^1;|\lambda| = 1\}$.

The nonlinear graviton construction, in the formulation of Ref. 4 (see, also, Ref. 6), requires one to solve functional equations of the form

$$u^A(\lambda) = f^A(\hat{u}^1(\lambda),\ldots,\hat{u}^{2r}(\lambda),\lambda) \quad (1\leqslant A\leqslant 2r), \tag{2.3}$$

where $f^A(x,\lambda)$, $x = (x^1,\ldots,x^{2r})$, are the coordinate components of a loop $f(\lambda)$ in the (pseudo-)group Γ_{can} of canonical transformations, i.e., a holomorphic (local) diffeomorphism with a loop parameter $\lambda\in S^1$ sending

$$x\to f(\lambda) = (f^1(x,\lambda),\ldots,f^{2r}(x,\lambda)),$$

and satisfying the symplectic condition

$$\epsilon^{CD}\frac{\partial f^A}{\partial x^C}\frac{\partial f^B}{\partial x^D} = \epsilon^{AB}. \tag{2.4}$$

The unknown functions $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ of the functional equations are required to have Laurent expansion of the form

$$u^A(\lambda) = \sum_{n=-\infty}^{\infty} u_n^A\lambda^n, \tag{2.5a}$$

$$\hat{u}^A(\lambda) = \sum_{n=-\infty}^{\infty} \hat{u}_n^A\lambda^n, \tag{2.5b}$$

in an annular neighborhood of S^1. Such solutions $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ obviously satisfy the exterior differential equation

$$\epsilon_{AB}\,du^A(\lambda)\wedge du^B(\lambda) - \epsilon_{AB}\,d\hat{u}^A(\lambda)\wedge d\hat{u}^B(\lambda) = 0, \tag{2.6}$$

or, equivalently, the system of equations

$$\sum_m \epsilon_{AB}\,du_{n+m}^A\wedge du_m^B - \sum_m \epsilon_{AB}\,d\hat{u}_{n+m}^A\wedge d\hat{u}_m^B = 0 \quad (-\infty < n < \infty). \tag{2.6'}$$

Recall[1] that Eqs. (2.6) and (2.6') give basic expressions of the hyper-Kähler hierarchy. Actually, we consider *only such solutions for which* $u_n^A(n\geqslant 0)$ *and* $\hat{u}_n^A(n\leqslant -1)$ *are functionally independent.* This condition is, in fact, included in the definition of the hyper-Kähler hierarchy,[1] and also a basic postulate of the following description of symmetries.

Remark: We call the set of functions f^A "patching functions" after the terminology in the literature. Our usage is, however, slightly different from the conventional one. In the self-dual case, for example, patching functions in Ref. 4 are required to satisfy a relation like Eq. (2.4) but with the right-hand side replaced by $\lambda^{-2}\epsilon^{AB}$. This simply means that the patching functions in Ref. 4 are given by $\lambda^{-1}f^A(\lambda)$. The seemingly somewhat strange multiplication rule of patching functions in Ref. 4 then changes into the ordinary composition of maps:

$$f_1\circ f_2(\lambda) = (f_1^1(f_2(x,\lambda),\lambda),\ldots,f_1^r(f_2(x,\lambda),\lambda)). \tag{2.7}$$

The present convention is thus more suited from a group-theoretic view.

A key idea of Ref. 6 is to regard $u^A(\lambda)$ and $\hat{u}^A(\lambda)$, too, as loops in Γ_{can} *by identifying* u_0^A *with* x^A. This is indeed reasonable, because $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ also obey the symplectic condition

$$\epsilon^{CD}\frac{\partial u^A(\lambda)}{\partial u_0^C}\frac{\partial u^B(\lambda)}{\partial u_0^D}=\epsilon^{AB}, \tag{2.8a}$$

$$\epsilon^{CD}\frac{\partial \hat{u}^A(\lambda)}{\partial u_0^C}\frac{\partial \hat{u}^B(\lambda)}{\partial u_0^D}=\epsilon^{AB}. \tag{2.8b}$$

[Recall that Eqs. (2.8) are derived from Eq. (2.6).[1,6]] Therefore, $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ define canonical transformations

$$u(\lambda):x=(u_0^A)\mapsto(u^A(\lambda)), \tag{2.9a}$$

$$\hat{u}(\lambda):x=(u_0^A)\mapsto(\hat{u}^A(\lambda)). \tag{2.9b}$$

They are further factorized as

$$u(\lambda)=u_+(\lambda)\circ\varphi(\lambda), \tag{2.10a}$$

$$\hat{u}(\lambda)=\hat{u}_-(\lambda)\circ\hat{\varphi}(\lambda), \tag{2.10b}$$

where

$$u_+(\lambda):x=(u_0^A)\mapsto x+\left(\sum_{n\geq 1}u_n^A\lambda^n\right), \tag{2.11a}$$

$$\hat{u}_-(\lambda):x=(u_0^A)\mapsto x+\left(\sum_{n\leq -1}\hat{u}_n^A\lambda^n\right), \tag{2.11b}$$

$$\varphi(\lambda):x=(u_0^A)\mapsto\left(\sum_{n\leq 0}u_n^A\lambda^n\right), \tag{2.11c}$$

$$\hat{\varphi}(\lambda):x=(u_0^A)\mapsto\left(\sum_{n\geq 0}\hat{u}_n^A\lambda^n\right). \tag{2.11d}$$

The maps in (2.11) all become canonical transformations. The previous functional equations can be rewritten into a factorization problem for loops in Γ_{can} as follows:

$$u_+(\lambda)^{-1}\circ f(\lambda)\circ\hat{u}_-(\lambda)=\varphi(\lambda)\circ\hat{\varphi}(\lambda)^{-1}. \tag{2.12}$$

It is now obvious that $u_n^A(n\geq 1)$ and $\hat{u}_n^A(n\leq -1)$, in the above picture, simply play the role of *deformation parameters* in the factorization problem. Half of the $4r$ coordinates in the original geometrical setting are included in these deformation parameters. Their group-theoretic meaning is thus essentially different from the other half, i.e., x^A's.

A bisided translation

$$f(\lambda)\mapsto g(\lambda)\circ f(\lambda)\circ\hat{g}(\lambda)^{-1} \tag{2.13}$$

on the group manifold now induces a transformation $T_{g(\lambda),\hat{g}(\lambda)}$ of solutions of the hyper-Käler hierarchy. Let $T_{g(\lambda)}$ and $\hat{T}_{\hat{g}(\lambda)}$, respectively, denote the transformations caused by the left and right translations in (2.13). Evidently these two transformations commute with each other and factorize the full transformation:

$$T_{g(\lambda),\hat{g}(\lambda)}=T_{g(\lambda)}\hat{T}_{\hat{g}(\lambda)}=\hat{T}_{\hat{g}(\lambda)}T_{g(\lambda)}. \tag{2.14}$$

The solution space of the hyper-Kähler hierarchy thus allows a pseudogroup action by the direct product $\Gamma_{\text{can}}\times\Gamma_{\text{can}}$ of two copies of Γ_{can}. The transformations obtained above thus have a "chiral structure" just as in the case of gauge fields.[12]

III. DERIVATION OF INFINITESIMAL TRANSFORMATIONS

We first derive an infinitesimal form of $T_{g(\lambda),\hat{g}(\lambda)}$ without putting the symplectic condition on $f(\lambda)$, $\varphi(\lambda)$, and $\hat{\varphi}(\lambda)$, which are accordingly loops in $\Gamma_{\text{gen}}=\{\text{general local diffeomorphisms}:x\to f(x)\}$. The construction in the previous section can carry over to that case. An infinitesimal transformation will be associated with a pair of vector fields

$$\xi(\lambda,\partial)=\xi^A(x,\lambda)\frac{\partial}{\partial x^A}, \tag{3.1a}$$

$$\hat{\xi}(\lambda,\partial)=\hat{\xi}^A(x,\lambda)\frac{\partial}{\partial x^A}. \tag{3.1b}$$

on the x space which also depend on λ.

To make contact with the previous construction, we consider the exponentials

$$g(\varepsilon,\lambda)\underset{\text{def}}{=}\exp\varepsilon\xi(\lambda,\partial), \tag{3.2a}$$

$$\hat{g}(\varepsilon,\lambda)\underset{\text{def}}{=}\exp\varepsilon\hat{\xi}(\lambda,\partial), \tag{3.2b}$$

where ε is an infinitesimal parameter. Then $g(\varepsilon,\lambda)$ and $\hat{g}(\varepsilon,\lambda)$ become loop group elements of Γ_{gen}, and accordingly give rise to a one-parameter family of solutions of the hierarchy:

$$u(\varepsilon,\lambda)\underset{\text{def}}{=}T_{g(\varepsilon,\lambda),\hat{g}(\varepsilon,\lambda)}u(\lambda), \tag{3.3a}$$

$$\hat{u}(\varepsilon,\lambda)\underset{\text{def}}{=}T_{g(\varepsilon,\lambda),\hat{g}(\varepsilon,\lambda)}\hat{u}(\lambda). \tag{3.3b}$$

From the previous construction [see, in particular, (2.12) and (2.13)] one finds that the transformed solutions are connected with the original solutions by the relation

$$u(\lambda)^{-1}\circ g(\varepsilon,\lambda)^{-1}\circ u(\varepsilon,\lambda)=\hat{u}(\lambda)^{-1}\circ\hat{g}(\varepsilon,\lambda)^{-1}\circ\hat{u}(\varepsilon,\lambda). \tag{3.4}$$

We now calculate both sides of Eq. (3.4) to the first order of ε. This should yield some equations on the infinitesimal transformations

$$\delta u^A(\lambda)=\delta_{\xi,\hat{\xi}}u^A(\lambda)\underset{\text{def}}{=}\left.\frac{\partial u^A(\varepsilon,\lambda)}{\partial\varepsilon}\right|_{\epsilon=0}, \tag{3.5a}$$

$$\delta\hat{u}^A(\lambda)=\delta_{\xi,\hat{\xi}}\hat{u}^A(\lambda)\underset{\text{def}}{=}\left.\frac{\partial\hat{u}^A(\varepsilon,\lambda)}{\partial\varepsilon}\right|_{\epsilon=0}. \tag{3.5b}$$

One can indeed obtain, with careful manipulations of derivatives of composite functions, the equation

$$\left(\frac{\partial u(\lambda)}{\partial x}\right)^{-1}\cdot[\delta u(\lambda)-\xi(u(\lambda),\lambda)]=\left(\frac{\partial\hat{u}(\lambda)}{\partial x}\right)^{-1}\cdot[\delta\hat{u}(\lambda)-\hat{\xi}(\hat{u}(\lambda),\lambda)], \tag{3.6}$$

where the first factors on both sides denote the $2r\times 2r$ matrices

$$\frac{\partial u(\lambda)}{\partial x}=\left(\frac{\partial u^A(\lambda)}{\partial x^B}\right), \tag{3.7a}$$

$$\frac{\partial\hat{u}(\lambda)}{\partial x}=\left(\frac{\partial\hat{u}^A(\lambda)}{\partial x^B}\right), \tag{3.7b}$$

and the second factors are the $2r\times 1$ column vectors

$$\delta u(\lambda) - \xi(u(\lambda),\lambda) = (\delta u^A(\lambda) - \xi^A(u(\lambda),\lambda)), \quad (3.7c)$$

$$\delta \hat{u}(\lambda) - \hat{\xi}(\hat{u}(\lambda),\lambda) = (\delta \hat{u}^A(\lambda) - \hat{\xi}^A(\hat{u}(\lambda),\lambda)). \quad (3.7d)$$

Remember that here we still identify u_0^A with x^A. Let us recall the following projection operators that played a very important role in Ref. 1:

$$\left(\sum_{n\in\mathbb{Z}} a_n\lambda^n\right)_+ \underset{\text{def}}{=} \sum_{n\geqslant 0} a_n\lambda^n, \quad (3.8a)$$

$$\left(\sum_{n\in\mathbb{Z}} a_n\lambda^n\right)_- \underset{\text{def}}{=} \sum_{n\leqslant -1} a_n\lambda^n. \quad (3.8b)$$

With these projectors, one can split Eq. (3.6) into two pieces. Note, on the other hand, that $(\delta u^A(\lambda))_+ = (\delta \hat{u}^A(\lambda))_- = 0$; $u_n (n\geqslant 0)$ and $\hat{u}_n (n\leqslant -1)$ are parameters that the transformations leave invariant. Bearing this fact in mind, one can easily deduce from (3.6) an explicit expression of the infinitesimal transformations for the case of Γ_{gen}.

Proposition 1: The infinitesimal transformation $\delta = \delta_{\xi,\hat{\xi}}$ for general vector fields $\xi = \xi(\lambda,\partial)$ and $\hat{\xi} = \hat{\xi}(\lambda,\partial)$ acts on $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ as:

$$\frac{\partial u^A(\lambda)}{\partial x^B}\delta u^B(\lambda) = \left(\frac{\partial u^A(\lambda)}{\partial x^B}\xi^B(u(\lambda),\lambda) - \frac{\partial \hat{u}^A(\lambda)}{\partial x^B}\hat{\xi}^B(\hat{u}(\lambda),\lambda)\right)_-, \quad (3.9a)$$

$$\frac{\partial \hat{u}^A(\lambda)}{\partial x^B}\delta \hat{u}^B(\lambda) = \left(\frac{\partial \hat{u}^A(\lambda)}{\partial x^B}\hat{\xi}^B(u(\lambda),\lambda) - \frac{\partial u^A(\lambda)}{\partial x^B}\xi^B(\hat{u}(\lambda),\lambda)\right)_+. \quad (3.9b)$$

Remark: Equation (3.4) is, for a number of reasons, more fundamental than the factorization problem in the previous section. An advantage is that, as we have observed above, one can deduce an explicit form of infinitesimal transformations with only algebraic calculations. Equation (3.4) is also suited for theoretical considerations of finite transformations. In the consideration of formal power series solutions, for example, the naive picture of the previous section breaks down but Eq. (3.4) is still meaningful, and one can make sense of various calculations in this and the next sections; see Appendix A. Basically, the same situation occurs in the case of integrable systems of gauge fields.[12] The method of calculations above is mostly borrowed from experience therein.

Our next task is to to put the symplectic condition to go to the hyper-Kähler case. The derivative matrices $(\partial u^A(\lambda)/\partial x^B)$ etc., are then required to take values in $\mathrm{Sp}(r,\mathbb{C})$. Accordingly the vector fields ξ and $\hat{\xi}$ must be Hamiltonian:

$$\xi(\lambda,\partial) = H(F) \underset{\text{def}}{=} \epsilon^{AB}\frac{\partial F}{\partial x^A}\frac{\partial}{\partial x^B}, \quad (3.10a)$$

$$\hat{\xi}(\lambda,\partial) = H(\hat{F}) \underset{\text{def}}{=} \epsilon^{AB}\frac{\partial \hat{F}}{\partial x^A}\frac{\partial}{\partial x^B}, \quad (3.10b)$$

with λ-dependent generating functions $F = F(x,\lambda)$ and $\hat{F} = \hat{F}(x,\lambda)$. Poisson brackets are also defined as

$$\{F,G\} \underset{\text{def}}{=} \epsilon^{AB}\frac{\partial F}{\partial x^A}\frac{\partial G}{\partial x^B} = H(F)G. \quad (3.11)$$

They satisfy the familiar commutation relations

$$[H(F),H(G)] = H(\{F,G\}). \quad (3.12)$$

Substituting the vector fields in Eqs. (3.9) with these expressions, one finds the following proposition.

Proposition 2: The infinitesimal transformation $\delta = \delta_{H(F),H(\hat{F})}$ acts on $u^A(\lambda)$ and $\hat{u}(\lambda)$ as:

$$\delta u^A(\lambda) = \{(F(u(\lambda),\lambda) - \hat{F}(\hat{u}(\lambda),\lambda))_-, u^A(\lambda)\}, \quad (3.13a)$$

$$\delta \hat{u}^A(\lambda) = \{(\hat{F}(\hat{u}(\lambda),\lambda) - F(u(\lambda),\lambda))_+, \hat{u}^A(\lambda)\}. \quad (3.13b)$$

Remark: The right-hand side of Eqs. (3.13) takes the form of Hamiltonian vector fields acting on u^A and $\hat{u}^A$ [cf. (3.11)]. Further, it is not hard to see that this expression is independent of the choice of the generating functions F and $\hat{F}$ and determined by Hamiltonian vector fields $H(F)$ and $H(\hat{F})$.

Having obtained these formulas, one can find their commutation relations as follows.

Proposition 3: The infinitesimal transformations $\delta_{H(F),H(\hat{F})}$ satisfy the commutation relations

$$[\delta_{H(F),H(\hat{F})},\delta_{H(G),H(\hat{G})}] = \delta_{H(\{F,G\}),H(\{\hat{F},\hat{G}\})}, \quad (3.14a)$$

$$[\delta_{H(F),H(\hat{F})},\delta_{H(G),H(\hat{G})}] = \delta_{H(\{F,G\}),H(\{\hat{F},\hat{G}\})}. \quad (3.14b)$$

The infinitesimal transformations thus respect the Lie algebra structure of Hamiltonian vector fields.

To prove Proposition 3, it is convenient to use a contour integral representation of the right-hand side of (3.13); techniques available are to be provided in the next section. Since necessary calculations are actually very similar to that case, we omit details of the proof of Proposition 3.

Finally, for illustration, let us consider the following two cases:

case (i)

$$F = x_A\lambda^n, \quad \hat{F} = 0 (n\geqslant 1), \quad (3.15)$$

case (ii)

$$F = 0, \quad \hat{F} = x_A\lambda^n (n\leqslant -1). \quad (3.16)$$

The corresponding infinitesimal transformations can be written, for case (i),

$$\delta_{H(x^A\lambda^n),0}u^B(\lambda) = -\{(u_A(\lambda)\lambda^n)_-, u^B(\lambda)\} - \delta_A^B\lambda^n, \quad (3.17a)$$

$$\delta_{H(x^A\lambda^n),0}\hat{u}^B(\lambda) = -\{(u_A(\lambda)\lambda^n)_+, \hat{u}^B(\lambda)\}, \quad (3.17b)$$

and, for case (ii),

$$\delta_{0,H(x^A\lambda^n)}u^B(\lambda) = -\{(\hat{u}_A(\lambda)\lambda^n)_-, u^B(\lambda)\}, \quad (3.18a)$$

$$\delta_{0,H(x^A\lambda^n)}\hat{u}^B(\lambda) = -\{(\hat{u}_A(\lambda)\lambda^n)_+, \hat{u}^B(\lambda)\} - \delta_A^B\lambda^n. \quad (3.18b)$$

In the course of the above calculation the following relations, which are simply a restatement of Eqs. (2.8), are also used:

$$\{u^A(\lambda),u^B(\lambda)\} = \epsilon^{AB}, \quad (3.19a)$$

$$\{\hat{u}^A(\lambda),\hat{u}^B(\lambda)\} = \epsilon^{AB}. \tag{3.19b}$$

Recall[1] that the hyper-Kähler hierarchy consists of Eqs. (3.19a) and (3.19b) and the evolution equations

$$\frac{\partial u^B(\lambda)}{\partial u_n^A} + \{(\lambda^n u_A(\lambda))_+, u^B(\lambda)\} = 0 \quad (n \geqslant 1), \tag{3.19c}$$

$$\frac{\partial \hat{u}^B(\lambda)}{\partial u_n^A} + \{(\lambda^n u_A(\lambda))_+, \hat{u}^B(\lambda)\} = 0 \quad (n \geqslant 1), \tag{3.19d}$$

$$\frac{\partial u^B(\lambda)}{\partial \hat{u}_n^A} + \{(\lambda^n \hat{u}_A(\lambda))_-, u^B(\lambda)\} = 0 \quad (n \leqslant -1), \tag{3.19e}$$

$$\frac{\partial \hat{u}^B(\lambda)}{\partial \hat{u}_n^A} + \{(\lambda^n \hat{u}_A(\lambda))_-, \hat{u}^B(\lambda)\} = 0 \quad (n \leqslant -1). \tag{3.19f}$$

The right-hand side of Eqs. (3.17) almost reproduces these evolution equations, except for the presence of extra-terms proportional to δ_B^A. This discrepancy is due to the fact that $\partial/\partial u_n^A (n \geqslant 0)$ and $\partial/\partial \hat{u}_n^A (n \leqslant -1)$ act nontrivially on $(u^A(\lambda))_+$ and $(\hat{u}^A(\lambda))_-$, which are, by definition, annihilated by $\delta_{H(F),H(\hat{F})}$. Thus the infinitesimal transformations with generating functions in (3.15) and (3.16) can be, in essence, identified with translational symmetries in the space of independent variables. In other words, generating functions in (3.15) and (3.16) exactly give infinitesimal generators of the time evolutions of the hierarchy.

Remark: More precisely only half, (3.19c) and (3.19d), of these evolution equations are included in the hierarchy of Ref. 1. This part closes within itself and forms a *subtheory* of the present setting. A similar subtheory is contained in the other half, (3.19e) and (3.19f), of the full hierarchy. As discussed in detail in Ref. 1 for the case of these subtheories, full system (3.19) likewise gives an equivalent expression of Eq. (2.6) [of course, under the previously remarked requirement that $u_n^A (n \geqslant 0)$ and $\hat{u}_n^A (n \leqslant -1)$ be independent variables].

IV. INFINITESIMAL TRANSFORMATION OF SECOND KEY FUNCTION

We now consider the second Plebanski key function Θ defined by the Pfaffian equation

$$d\Theta = \sum_{n>0} \epsilon_{AB} u^A_{-n-1}\, du_n^B - \sum_{n<-1} \epsilon_{AB} \hat{u}^A_{-n-1}\, d\hat{u}_n^B, \tag{4.1}$$

or, equivalently, by the system of partial differential equations

$$\frac{\partial \Theta}{\partial u_n^A} = -u_{A,-n-1} \quad (n \geqslant 0), \tag{4.2a}$$

$$\frac{\partial \Theta}{\partial \hat{u}_n^A} = \hat{u}_{A,-n-1} \quad (n \leqslant -1). \tag{4.2b}$$

This is a natural extension of the definition in Ref. 1. The closedness of the right-hand side of Eq. (4.1) [i.e., the integrability condition of Eqs. (4.2)] is an immediate consequence of Eq. (2.6) [see Eq. (2.6') for $n = -1$]. Thus Θ is uniquely determined except for an integration constant, $\Theta \mapsto \Theta + \text{const}$.

Our task here is to find a transformation law of Θ that consistently reproduces the results of the previous section. We first show the answer.

Proposition 4: Such an infinitesimal transformation of Θ is given by

$$\delta\Theta = \delta_{F,\hat{F}}\Theta \underset{\text{def}}{=} \operatorname*{res}_{\lambda=\infty} F(u(\lambda),\lambda) + \operatorname*{res}_{\lambda=0} \hat{F}(\hat{u}(\lambda),\lambda). \tag{4.3}$$

Remark: The residues in the above formula are defined as

$$\operatorname*{res}_{\lambda=\infty} \sum a_n \lambda^n = -a_{-1}, \tag{4.4a}$$

$$\operatorname*{res}_{\lambda=0} \sum b_n \lambda^n = b_{-1}. \tag{4.4b}$$

As opposed to the case of u^A and $\hat{u}^A$, the right-hand side of (4.3) is determined not by Hamiltonian vector fields but their generating functions. Having this fact in mind, we write $\delta_{F,\hat{F}}$ rather than $\delta_{H(F),H(\hat{F})}$.

Let us give a proof of the above proposition. For the moment let $\delta u^A(\lambda)$ and $\delta \hat{u}^A(\lambda)$ denote the infinitesimal variations of $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ induced by (4.3) via (4.1) or (4.2). We have to show that they agree with the previous result, i.e., (3.13).

These infinitesimal variations, by definition, leave invariant the independent variables $u_n^A (n \geqslant 0)$ and $\hat{u}_n^A (n \leqslant -1)$. [This is included in the construction. Accordingly, they commute with the action of $\partial/\partial u_n^A (n \geqslant 0)$ and $\partial/\partial \hat{u}_n^A (n \leqslant -1)$.] Therefore,

$$\delta u_{A,-n-1} = -\frac{\partial}{\partial u_n^A}\delta\Theta \quad (n \geqslant 0), \tag{4.5a}$$

$$\delta \hat{u}_{A,-n-1} = \frac{\partial}{\partial \hat{u}_n^A}\delta\Theta \quad (n \leqslant -1). \tag{4.5b}$$

It is more convenient to rewrite the residues in (4.3) into contour integrals:

$$\delta\Theta = -\oint F(u(\mu),\mu)\frac{d\mu}{2\pi i} + \oint \hat{F}(\hat{u}(\mu),\mu)\frac{d\mu}{2\pi i}, \tag{4.6}$$

where the contours both encircle the origin $\mu = 0$ in the anticlockwise direction (along, say, a circle $|\mu| \sim 1$). All integration contours in the following calculations are supposed to be oriented in such a way. [If one attempts to extend the present consideration to formal power series solutions, these integrals must be replaced with the original algebraic residue operation as in (4.4). Nevertheless one can easily check that all formulas in the following are still correct under that interpretation (see, also, Appendix A).] Now, for $n \geqslant 0$,

$$\delta u_{A,-n-1} = \oint \frac{\partial}{\partial u_n^A} F(u(\mu),\mu)\frac{d\mu}{2\pi i} - \oint \frac{\partial}{\partial u_n^A}\hat{F}(\hat{u}(\mu),\mu)\frac{d\mu}{2\pi i}, \tag{4.7}$$

and from evolution equations (3.19), further,

$$= \oint \{(\lambda^n u_A(\mu))_+, \hat{F}(\hat{u}(\mu),\mu) - F(u(\mu),\mu)\} \frac{d\mu}{2\pi i}. \tag{4.8}$$

The next step is to use the general formulas

$$(f(\mu))_+ = \oint \frac{f(\nu)}{\nu(1-\mu/\nu)} \frac{d\nu}{2\pi i}, \tag{4.9a}$$

$$(f(\mu))_- = \oint \frac{f(\nu)}{\mu(1-\nu/\mu)} \frac{d\nu}{2\pi i}. \tag{4.9b}$$

Then, Eqs. (4.7) and (4.8) can be rewritten

$$\delta u_{A,-n-1} = \oint\oint_{|\mu|<|\nu|} \{\nu^n u_A(\nu), \hat{F}(\hat{u}(\mu),\mu) - F(u(\mu),\mu)\} \frac{d\mu\, d\nu}{(2\pi i)^2 \nu(1-\mu/\nu)}. \tag{4.10}$$

In terms of the generating functions $u_A(\lambda)$,

$$\begin{aligned}\delta u_A(\lambda) &= \sum_{n=0}^{\infty} \delta u_{A,-n-1} \lambda^{-n-1} \\ &= \oint\oint_{|\mu|<|\nu|} \{u_A(\nu), \hat{F}(\hat{u}(\mu),\mu) - F(u(\mu),\mu)\} \\ &\quad \times \frac{d\mu\, d\nu}{(2\pi i)^2 \nu(1-\mu/\nu)\lambda(1-\nu/\lambda)}.\end{aligned} \tag{4.11}$$

Now, one can perform the μ integration first, using formulas (4.9) twice, to find that

$$\begin{aligned}\delta u_A(\lambda) &= \oint \{u_A(\nu), (\hat{F}(\hat{u}(\mu),\mu) - F(u(\mu),\mu))_-\} \\ &\quad \times \frac{d\nu}{2\pi i \lambda(1-\nu/\mu)} \\ &= (\{u_A(\lambda), (\hat{F}(\hat{u}(\lambda),\lambda) - F(u(\lambda),\lambda))_-\})_-.\end{aligned} \tag{4.12}$$

Because of the evident identity $\{u_A(\lambda),...\} = \{\varphi_A(\lambda),...\}$, the ingredient of the outer $(\cdots)_-$ in the last line contains no non-negative powers of λ. Therefore,

$$\delta u_A(\lambda) = \{u_A(\lambda), (\hat{F}(\hat{u}(\lambda),\lambda) - F(u(\lambda),\lambda))_-\}. \tag{4.13}$$

This exactly agrees with Eq. (3.13a). Basically, the same calculations show that $\delta\hat{u}^A(\lambda)$ also coincides with the result given by Eq. (3.13b). Proposition 4 is thus proven.

One can also deduce commutation relations of the infinitesimal transformations. First, from the construction and from what we have just checked above,

$$\begin{aligned}\delta_{F,\hat{F}}\delta_{G,\hat{G}}\Theta &= \oint \delta_{F,\hat{F}}(\hat{G}(\hat{u}(\lambda),\lambda) - G(u(\lambda),\lambda)) \frac{d\lambda}{2\pi i} \\ &= \oint (\{\hat{\mathcal{F}}_+ - \mathcal{F}_+, \hat{\mathcal{G}}\} \\ &\quad - \{\mathcal{F}_- - \hat{\mathcal{F}}_-, \mathcal{G}\}) \frac{d\lambda}{2\pi i},\end{aligned} \tag{4.14}$$

where, for simplifying notations, we use the following abbreviation:

$$\mathcal{F} \underset{\text{def}}{=} F(u(\lambda),\lambda), \quad \hat{\mathcal{F}} \underset{\text{def}}{=} \hat{F}(\hat{u}(\lambda),\lambda),$$

$$\mathcal{F}_\pm \underset{\text{def}}{=} (\mathcal{F})_\pm, \quad \mathcal{G}_\pm \underset{\text{def}}{=} (\mathcal{G})_\pm. \tag{4.15}$$

$\delta_{G,\hat{G}}\delta_{F,\hat{F}}\Theta$, too, has a similar expression. Therefore,

$$\begin{aligned}[\delta_{F,\hat{F}}, \delta_{G,\hat{G}}]\Theta = \oint (&\{\hat{\mathcal{F}}_+, \hat{\mathcal{G}}\} + \{\hat{\mathcal{F}}, \hat{\mathcal{G}}_+\} - \{\mathcal{F}_-, \mathcal{G}\} \\ &- \{F, \mathcal{G}_-\} + \{\mathcal{F}, \hat{\mathcal{G}}_-\} \\ &- \{\mathcal{F}_+, \hat{\mathcal{G}}\} + \{\hat{\mathcal{F}}_-, \mathcal{G}\} \\ &- \{\hat{\mathcal{F}}, \mathcal{G}_+\}) \frac{d\lambda}{2\pi i}.\end{aligned} \tag{4.16}$$

The first two terms in the integrand can be gathered up into a single Poisson bracket as

$$\oint (\{\hat{\mathcal{F}}_+, \hat{\mathcal{G}}\} + \{\hat{\mathcal{F}}, \hat{\mathcal{G}}_+\}) \frac{d\lambda}{2\pi i} = \oint \{\hat{\mathcal{F}}, \hat{\mathcal{G}}\} \frac{d\lambda}{2\pi i}, \tag{4.17}$$

because $\{\hat{\mathcal{F}}_+, \hat{\mathcal{G}}_+\}$ and $\{\hat{\mathcal{F}}_-, \hat{\mathcal{G}}_-\}$ disappear after integration. A similar reasoning shows that the sum of the third and fourth terms becomes $\oint\{\mathcal{F},\mathcal{G}\}d\lambda/(2\pi i)$. The other terms cancel out. Thus the right-hand side of (4.16) reduces to

$$\oint (\{\hat{\mathcal{F}}, \hat{\mathcal{G}}\} - \{\mathcal{F}, \mathcal{G}\}) \frac{d\lambda}{2\pi i} = \delta_{\{F,G\},\{\hat{F},\hat{G}\}}\Theta. \tag{4.18}$$

To summarize, we have the following proposition.

Proposition 5: The infinitesimal transformations $\delta_{F,\hat{F}}$ of Θ obey the commutation relations

$$[\delta_{F,\hat{F}}, \delta_{G,\hat{G}}] = \delta_{\{F,G\},\{\hat{F},\hat{G}\}}. \tag{4.19}$$

Thus we find that the infinitesimal transformations extended to Θ give a nonlinear realization of a Poisson loop algebra. Further, as (4.19) clearly tells, the Poisson algebra is actually a direct sum of two Poisson algebras carrying, respectively, $F = F(x,\lambda)$ and $\hat{F} = \hat{F}(x,\lambda)$. The "chiral structure" in the symmetries of $u^A(\lambda)$ and $\hat{u}^A(\lambda)$ are thus also retained at this level.

V. ORIGIN OF POISSON ALGEBRA AND CONTACT GEOMETRY

For the moment, we forget of the presence of the loop parameter λ and review several basic facts in symplectic geometry. Suppose that a symplectic manifold with canonical coordinates $(p,x) = (p_j, x^j)$ is given. Naturally, one can introduce the notion of Hamiltonian vector fields and Poisson brackets:

$$H(F) \underset{\text{def}}{=} \frac{\partial F}{\partial p_j}\frac{\partial}{\partial x^j} - \frac{\partial F}{\partial x^j}\frac{\partial}{\partial p_j}, \tag{5.1a}$$

$$\{F,G\} \underset{\text{def}}{=} \frac{\partial F}{\partial p_j}\frac{\partial G}{\partial x^j} - \frac{\partial F}{\partial x^j}\frac{\partial G}{\partial p_j} = H(F)G. \tag{5.1b}$$

They represent infinitesimal canonical transformations, under which the canonical coordinates transform as

$$\delta_F(x^j) = \{F, x^j\} = \frac{\partial F}{\partial p_j}, \tag{5.2a}$$

$$\delta_F(p_j) = \{F, p_j\} = -\frac{\partial F}{\partial x^j}. \tag{5.2b}$$

In particular, a generating function F depending only on p, $F=F(p)$, gives rise to a transformation that leaves x^j's invariant:

$$\delta_F(x^j)=0, \tag{5.3a}$$

$$\delta_F(p_j)=-\frac{\partial F}{\partial x^j}. \tag{5.3b}$$

It is well known that infinitesimal canonical transformations can be consistently extended to the contact manifold with canonical coordinates (z,p,x) and contact form $dz-p_j\,dx^j$. In the above situation, the z coordinate transforms as

$$\delta_F(z)=-F. \tag{5.3c}$$

Their commutation relations become very simple; transformations with p-independent generating functions commute with each other:

$$[\delta_F,\delta_G]=0 \quad (\forall F=F(p),G=G(p)).$$

In the hyper-Kähler hierarchy, the x coordinates correspond to the independent variables $u_n^A(n\geqslant 0)$ and $\hat{u}_n^A(n\leqslant -1)$ of the hierarchy. The p coordinates are to be identified with the other variables. Here, z can be compared to the key functions; note that Eq. (4.1) can be viewed as a Pfaffian equation defining "contact elements" (or their integral manifolds). Thus, schematically, we have the following correspondence:

$$z\leftrightarrow\Theta,$$

$$x^j\leftrightarrow u_n^A(n\geqslant 0),\quad \hat{u}_n^A(n\leqslant -1),$$

$$p_j\leftrightarrow u_n^A(n\leqslant -1),\quad \hat{u}_n^A(n\geqslant 0).$$

An unnegligible difference is that the hyper-Kähler hierarchy is related to *an infinite number of* symplectic forms rather than a single one [see Eq. (2.6′)]. Boyer and Plebanski[4] considered this as defining a generalized symplectic structure and presented a reinterpretation of the nonlinear graviton construction. In their framework, a solution of the nonlinear graviton construction may be identified with a Lagrangian submanifold under some regularity condition. Canonical transformations in the generalized symplectic structure will naturally cause transformations on the set of all such Lagrangian submanifolds. This is also exactly what we have described in Sec. II in a form more suited for the group-theoretic view. [In our formulation, the regularity condition corresponds to the requirement that $u_n^A(n\geqslant 0)$ and $\hat{u}_n^A(n\geqslant -1)$ be functionally independent.] Replacing one of the infinitely many symplectic forms in Eq. (2.6′) by an associated contact form, one can introduce a key function as the z coordinate. Thus, as also noted in Ref. 4, various key functions such as Θ, $\hat{\Theta}$, and Ω (Refs. 1 and 4) are treated on an equal footing.

VI. CONCLUDING REMARKS

(1) Recently Boyer and Winternitz[13] constructed an infinite dimensional symmetry group for the self-dual Einstein case. They take the Plebanski first heavenly equation as the basic field equation. Their symmetries consist of two parts, one of which forms a Poisson algebra, and the other forms a special class of coordinate transformations of independent variables. It seems likely that the first part is included in our symmetries as those whose generating functions F and $\hat{F}$ are independent of λ. The second part is a kind of residual gauge freedom of the first heavenly equation, which is fixed in our formulation. [Recall that our symmetries $\delta_{F,\hat{F}}$ (or their counterparts in the formulation of Appendix B) are, from the very beginning, understood to leave the independent variables invariant.]

(2) We have seen in Sec. VI that symmetries of the second Plebanski key function, in contrast to the original variables of the hierarchy, are not determined by Hamiltonian vector fields but their generating functions, and that their commutation relations respect the Poisson algebra structure of generating functions. In the context of commutator anomalies, this result seems to allow two different interpretations.

On the one hand, one can understand this as implying the absence of commutator anomalies. The geometric consideration in Sec. V will also support this standpoint. As we have seen therein, the Poisson algebra structure has its origin in an interrelation of symplectic geometry and contact geometry. From such a point of view this interpretation looks very natural.

On the other hand, it seems also possible to assert that commutator anomalies do exist but are simply hidden. This is due to the observation that the Poisson algebra is a central extension of the Lie algebra of Hamiltonian vector fields. The central part consists of constant loops, i.e., constant functions on the symplectic manifold that are allowed to depend only on the loop variable. This is thus actually an *infinite dimensional* central extension. The passage from Hamiltonian vector fields to their generating functions is certainly a nontrivial central extension. This observation is also advocated in part by the fact that the first stage of the central extension in the case of τ functions is to assign to an Abelian loop algebra a Heisenberg (CCR) algebra.[3]

It seems plausible that this is simply two manifestations of a more fundamental structure that lies behind. This issue seems to deserve further study.

(3) We do not know whether there is any possibility of commutator anomalies in these Poisson loop algebras. Presumably, as mentioned in Sec. I, membrane theory will provide some hints to this question. For the case without the loop variable (i.e., a pure Poisson algebra), indeed, central extensions are already discovered.[14] Such commutator anomalies are obviously of quantum nature. Our analysis in this paper seems to show that the notion of the Plebanski key functions still belongs, at most, to classical field theory. At the present stage, it is absolutely unclear whether any structure like the hyper-Kähler hierarchy survives quantization, though persuing such a possibility would be an interesting problem. This is in sharp contrast to the case of τ functions, which are already of quantum nature, reflecting the presence of quantum field theory behind.[3]

(4) Nevertheless, we still expect our method to contribute to membrane theory, mostly in its classical aspects.

One possible direction is to study structural aspects of equations of self-duality type as discussed in Ref. 9. A remarkable observation in Ref. 9 is that there is an unexpected relation of membrane theory with self-dual gauge fields and Toda fields. This fact is also very suggestive from our stand-

point, because we already know a number of common structural characteristics among the hyper-Kähler hierarchy, a gauge field hierarchy[12] and a Toda field hierarchy[15] (see, also, Appendix A).

Another related issue of importance is to develop a superspace version of our method. A naive idea is to make a theory based upon supersymplectic geometry [i.e., integrable $OSp(s|r)$ structures] and super-Poisson algebras; such an extension is a rather easy task. This subject will also have implications to the work of Chau and Milewski.[16]

(5) After all, it would be better to call our hierarchy a "Poisson gauge field hierarchy" rather than the "hyper-Kähler hierarchy." Hyper-Kähler geometry is certainly an origin of the present subject, but now we know that mathematical structures inherent in the hierarchy are of more universal nature. In this respect, it would be an interesting problem to search for some relation with recent work of Mason and Newman.[17]

ACKNOWLEDGMENTS

The author is grateful to Izumi Ojima, Hiroaki Kanno, and Hiroshi Yamada for discussions and encouragement. This work was supported in part by Grants-in-Aid for Scientific Researches, the Ministry of Education.

APPENDIX A: FACTORIZATION IN FORMAL GROUPS

The contents of Sec. II are stated within the language of complex analytic functions and loops in a pseudogroup. Basically, the same results can be derived for formal power series solutions. The algebraic meaning of various formulas in Secs. III and IV become rather transparent in such a formal framework.

This Appendix is intended to give an outline of that approach. This will also serve as a brief review and erratum of Sec. 5 of Ref. 6. As mentioned in Ref. 18, this part of Ref. 6 contains several conceptual errors (which are all related to an infinite dimensional Grassmannian manifold and dynamical flows on it). The first three subsections therein are, however, absolutely correct, and give a rigorous justification of our derivation of finite and infinitesimal symmetries.

Actually the present setting requires a slight extension of the formulation in Ref. 6. Namely, we now have to deal with both u^A and $\hat{u}^A$, and the associated two types of transformations caused by the left and right translations in (2.13). (In Ref. 6, only u^A and left transformations are considered.) Such complexity is already observed in the case of gauge fields[12] and the Toda fields.[15] Mathematical tools for justifying formal calculations in Ref. 12 are provided in Refs. 19 (in particular, in Part II). We now reconstruct the results of Ref. 6 along the line of Refs. 19 so as to fit into the the present setting. This formal theory should, in principle, include the analytical case; we shall come back to that case in the end of this Appendix.

We have to start with several mathematical definitions. Let $\mathscr{A}$ denote the set of all formal power series of the independent variables $u_n^A (n \geqslant 1)$ and $\hat{u}_n^A (n \leqslant -1)$ with complex coefficients:

$$\mathscr{A} \underset{\text{def}}{=} \mathbb{C}[[u_n^A(n \geqslant 1), \hat{u}_n^A(n \leqslant -1)]]. \tag{A1}$$

This is the most basic ingredient in the present formulation. To put such formal power series under good mathematical control, we assign to each monomial therein an integer called *weight* (or *weighted degree*) in such a way that: (i) the weights of u_n^A and $\hat{u}^A{}_{-n}$ for $n \geqslant 1$ are both equal to n; (ii) the weight of any monomial of these variables is the sum of the weights of all factors. The minimum value of the weights of all monomials in a formal power series $a \in \mathscr{A}$ is called the *order* of a, and written $\mathrm{ord}(a)$. Conventionally, we put $\mathrm{ord}(0) = \infty$. For each $n \geqslant 0$, elements of $\mathscr{A}$ of order $\leqslant n$ form a vector subspace

$$\mathscr{A}_n \underset{\text{def}}{=} \{a \in \mathscr{A}; \mathrm{ord}(a) \geqslant n\}. \tag{A2}$$

Evidently,

$$\mathscr{A}_n \supset \mathscr{A}_{n+1}, \tag{A3a}$$

$$\bigcap_{n>0} \mathscr{A}_n = \{0\}, \tag{A3b}$$

$$\bigcup_{n>0} \mathscr{A}_n = \mathscr{A}, \tag{A3c}$$

$$\mathscr{A}_m \mathscr{A}_n \subset \mathscr{A}_{m+n}. \tag{A3d}$$

Mathematically, these properties ensure that one can introduce the notion of convergence in $\mathscr{A}$, and $\mathscr{A}$ becomes a complete topological algebra. By definition, a sequence $a_n \in \mathscr{A}$ is said to converge to $a \in \mathscr{A}$ if $\mathrm{ord}(a_n - a) \to \infty$ as $n \to \infty$. Consequently, an infinite series of the form $\Sigma_{n>0} a_n$ in $\mathscr{A}$ converges if $\mathrm{ord}(a_n) \to \infty$ as $n \to \infty$. The subspaces $\mathscr{A}_n$ play the role of *scales* measuring convergence. (Compare these notions with the case of the construction of p-adic number fields;[20] both have basically the same mathematical structure.)

Formal power series solutions of the hyper-Kähler hierarchy are, by definition, solutions whose coordinate components are such that

$$\varphi^A(\lambda) \in \mathscr{A}[[x, \lambda^{-1}]], \quad \hat{\varphi}^A(\lambda) \in \mathscr{A}[[x, \lambda]] \tag{A4}$$

and also satisfy the condition

$$\varphi(\lambda = \infty) = \text{identity}, \quad \hat{\varphi}(\lambda = 0) = \text{invertible}. \tag{A5}$$

[For the definition of $\varphi(\lambda)$ and $\hat{\varphi}(\lambda)$, see Sec. II.]

The problem is to find a framework to make sense of Eq. (3.4). Reducing it to the symplectic case is a rather easy task. Equation (2.12) is inadequate as a theoretical foundation within the present setting, because the right-hand side becomes meaningless if $\varphi(\lambda)$ and $\hat{\varphi}(\lambda)$ are formal series as in (A4). Equation (3.4), on the other hand, turns out to have a definite meaning even in that case.

Actually, since Eq. (3.4) contains the group parameter ϵ as well, we are forced to replace $\mathscr{A}$ by

$$\mathscr{A}_\epsilon \underset{\text{def}}{=} \mathscr{A}[[\epsilon]] \tag{A6}$$

and repeat the same topological argument as has been done above, where ϵ is to be given an appropriate positive weight. [This is a new circumstance that is absent in the subject of Ref. 6.] The weight of ϵ should be chosen in such a way that $g(\epsilon, \partial)$ and $\hat{g}(\epsilon, \partial)$, respectively, belong to the following sets of *formal maps*:

$$\Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)} \underset{\text{def}}{=} \{f(\lambda); f^A_{A_1,\ldots,A_p n} \in (\mathscr{A}_\varepsilon)_{-p+n+1} \cap (\mathscr{A}_\varepsilon)_n,$$

and $(f^A_{B,0}) \in \mathrm{GL}(r,\mathbb{C})\}$, (A7a)

$$\hat{\Gamma}_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)} \underset{\text{def}}{=} \Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(1/\lambda)} = \{\hat{f}(\lambda); \hat{f}(\lambda^{-1}) \in \Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}\}, \quad \text{(A7b)}$$

where $f^A_{A_1,\ldots,A_p n}$ denote the numerical coefficients of the Laurent expansion of the coordinate components of $f(\lambda)$ as

$$f^A(\lambda) = \sum_{n=-\infty}^{\infty} \sum_{p=0}^{\infty} \sum_{1 \le A_1 \le \cdots \le A_p \le r} f^A_{A_1,\ldots,A_p n} \times x^{A_1} \cdots x^{A_p} \lambda^n. \quad \text{(A8)}$$

Coefficients with $p=0$ in (A7) and (A8) simply mean those with no lower indices $A_1,\ldots,A_p$; all conditions given therein are also imposed on these $p=0$ coefficients. The following fact is fundamental.

Proposition A1: The above two sets $\Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$ and $\hat{\Gamma}_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$ are, respectively, closed under composition and inversion, and therefore become groups.

Proof (sketch): To prove this proposition, we use an infinite-matrix representation of these formal maps. We shall see that this technique is also a key to prove the existence of a solution of Eq. (3.4). (This is exactly what was intended to be done in Ref. 6.) More precisely, we define for each $f(\lambda) \in \Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$ a matrix $\rho(f)$ of the form

$$\rho(f) = (\rho(f)^{A_1,\ldots,A_p m}_{B_1,\ldots,B_q n}), \quad \text{(A9)}$$

where $(A_1,\ldots,A_p,m)$ (row index) and $(B_1,\ldots,B_q,n)$ (column index) range over the same index set as occurs in the expansion of $f(\lambda)$, as follows:

$$f^{A_1}(\lambda) \cdots f^{A_p}(\lambda) \lambda^m = \sum_n \sum_q \sum_{B_1 \le \cdots \le B_q} \rho(f)^{A_1,\ldots,A_p m}_{B_1,\ldots,B_q n} x^{B_1} \cdots x^{B_q} \lambda^n. \quad \text{(A10)}$$

Equation (A10) should be treated carefully, because coefficients resulting from the Taylor and Laurent expansion of the right-hand side will become a sum of an infinite number of elements of $\mathscr{A}_\varepsilon$. In fact, the somewhat involved conditions in (A7) guarantee that they fall into "convergent" series in the sense mentioned above. One can likewise check with lengthy tedious calculations that the composition $f \circ g$ and the corresponding matrix product $\rho(f)\rho(g)$ are rigorously defined for any pair of two formal maps f and g taken from one of the two sets in (A7). Once this fact is established, it will be immediate from the construction that the matrix representation obeys the composition rule

$$\rho(f \circ g) = \rho(f)\rho(g). \quad \text{(A11)}$$

Invertibility of $\rho(f)$ is a far more technical issue. The essence is, however, basically the same as a method developed in Part I of Ref. 19 for gauge fields: As in that case, divide the infinite matrix $\rho(f)$ into the sum of two parts consisting of matrix elements with, say, $m \geqslant n$ and $m < n$, respectively (actually, one may also take some other way of splitting); then show that the former itself has an inverse (this can be rather easily seen because the resulting matrix is blockwise triangular); finally, construct an inverse of the full matrix using the familiar equality

$$(A+B)^{-1} = A^{-1} - A^{-1}BA^{-1} + A^{-1}BA^{-1}BA^{-1} - \cdots,$$

and carefully check the convergence of infinite series which arise from these constructions. Proposition A1 is thus proven.

Having introduced the above groups, we now turn to the issue of giving a precise setting in which to consider Eq. (3.4). The vector fields $\xi(\lambda,\partial)$ and $\hat{\xi}(\lambda,\partial)$ [see Eqs. (3.1)] are assumed to be of the form

$$\xi^A(x,\lambda) = \sum_{n=-\infty}^{m} \xi^A_n(x)\lambda^n, \quad \xi^A_n(x) \in \mathscr{A}, \quad \text{(A12a)}$$

$$\hat{\xi}^A(x,\lambda) = \sum_{n=-\hat{m}}^{\infty} \hat{\xi}^A_n(x)\lambda^n, \quad \hat{\xi}^A_n(x) \in \mathscr{A}, \quad \text{(A12b)}$$

where m and $\hat{m}$ are some integers. If one assign to ε a weight as

$$\mathrm{ord}(\varepsilon) = \min\{m,\hat{m},0\}, \quad \text{(A13)}$$

then $g(\varepsilon,\lambda)$ and $\hat{g}(\varepsilon,\lambda)$, respectively, belong to $\Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$ and $\hat{\Gamma}_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$. The other ingredients of Eq. (3.4) such as $u(\varepsilon,\lambda)$, etc., are further factorized as in Eqs. (2.10), and $u_+(\lambda)$ and $\hat{u}_-(\lambda)$ are evidently included in the above two groups.

To summarize, all ingredients on the left- (right-) hand side of Eq. (3.4) are to be taken from the first (resp. second) group in (A7). The equation then requires that both sides be in the intersection of these two groups and coincide. (Note that this intersection becomes a subgroup of these two groups.) This is a precise reformulation of the "factorization problem" of Sec. II in the present formal setting.

Remark: To formulate Eq. (3.4) within the language of the groups defined in (A7), one first has to fix $\xi(\lambda,\partial)$ and $\hat{\xi}(\lambda,\partial)$ to define $\mathrm{ord}(\varepsilon)$. The definition of the above formal groups thus essentially depends on how $\xi(\lambda,\partial)$ and $\hat{\xi}(\lambda,\partial)$ are given. This is an unpleasant, but inevitable feature of the present approach. A more elegant language for describing this situation will be provided by the abstract notion of "categories and functors" in mathematics.

To state a result on the above problem, we further introduce the following subgroups of the above groups:

$$\Gamma_{\text{gen},\pm}^{\mathscr{A}_\varepsilon(\lambda)} = \{f(\lambda) \in \Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}; f^A_{A_1,\ldots,A_p n} = 0, \quad \text{if} \quad \pm n < 0\}, \quad \text{(A14a)}$$

$$\hat{\Gamma}_{\text{gen},\pm}^{\mathscr{A}_\varepsilon(\lambda)} = \{\hat{f}(\lambda) \in \hat{\Gamma}_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}; \hat{f}^A_{A_1,\ldots,A_p} = 0, \quad \text{if} \quad \pm n < 0\}. \quad \text{(A14b)}$$

Then, the existence of a unique solution of Eq. (3.4) in the present formulation is a corollary of the following more general result.

Proposition A2: For any $f(\lambda) \in \Gamma_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$ and $\hat{f}(\lambda) \in \hat{\Gamma}_{\text{gen}}^{\mathscr{A}_\varepsilon(\lambda)}$ there are a unique pair of elements $\varphi(\lambda) \in \Gamma_{\text{gen},-}^{\mathscr{A}_\varepsilon(\lambda)}$ and $\hat{\varphi}(\lambda) \in \hat{\Gamma}_{\text{gen},+}^{\mathscr{A}_\varepsilon(\lambda)}$ that satisfy the relation

$$f^{-1} \circ \varphi(\lambda) = \hat{f}^{-1} \circ \hat{\varphi}(\lambda) \quad \text{(A15a)}$$

and the normalization condition

$$\varphi(\lambda = \infty) = \text{identity}. \quad \text{(A15b)}$$

We now sketch a proof of this result. A prototype of this proof is given in Part II of Ref. 19 for the case of gauge field.

Following that idea, we construct a solution of Eq. (3.4) step by step by, first, solving a factorization problem in $\Gamma_{\rm gen}^{\mathscr{A}_\epsilon(\lambda)}$ and, then, another one in $\hat{\Gamma}_{\rm gen}^{\mathscr{A}_\epsilon(\lambda)}$. At each step, we can use the following result, which is mentioned (but with an incomplete proof) in Ref. 6.

Proposition A3: (Ref. 6) For any $f(\lambda)\in\Gamma_{\rm gen}^{\mathscr{A}_\epsilon(\lambda)}$ there is a unique pair of elements $\varphi(\lambda)\in\Gamma_{{\rm gen},-}^{\mathscr{A}_\epsilon(\lambda)}$ and $\psi(\lambda)\in\Gamma_{{\rm gen},+}^{\mathscr{A}_\epsilon(\lambda)}$ that satisfy the conditions

$$f(\lambda)=\varphi(\lambda)\circ\psi(\lambda)^{-1}, \tag{A16a}$$

$$\varphi(\lambda=\infty)=\text{identity}. \tag{A16b}$$

Remark: One may find various variations of this result. For example, a similar result is also true for $\hat{\Gamma}_{\rm gen}^{\mathscr{A}_\epsilon(\lambda)}$; the order of factors on the right-hand side of (A16a) may be exchanged; the role of the two groups may be interchanged, etc. The second condition in (A16) is imposed just for ensuring the uniqueness of solution. If this condition is removed, solutions are not unique, but the arbitrariness is limited to the obvious one: $\varphi(\lambda)\to\varphi(\lambda)\circ h$, $\psi(\lambda)\to\psi(\lambda)\circ h$, where $h=(h^A(x))$ is independent of λ. Equation (A15a) also has similar arbitrariness, which is suppressed by normalization condition (A15b).

Proof of Proposition A3 (sketch): The uniqueness part of Proposition A3 is rather immediate: If there is another solution pair $\varphi_1(\lambda)$ and $\psi_1(\lambda)$, one has the obvious relation

$$\varphi(\lambda)^{-1}\circ\varphi_1(\lambda)=\psi(\lambda)^{-1}\circ\psi_1(\lambda). \tag{A17}$$

This implies that both sides are independent of λ. Hence, the two solutions exactly differ by the arbitrariness mentioned in the above remark. Actually, we also put normalization condition (A16b), by which the difference disappears. Thus follows the uniqueness. To see the existence part, let us consider the corresponding factorization problem of infinite matrices:

$$F=\Phi\Psi^{-1}, \tag{A18}$$

where

$$F=\rho(f),\quad \Phi=\rho(\varphi),\quad \Psi=\rho(\psi). \tag{A19}$$

Let us disregard the presence of the first set of indices $A_1,\ldots,A_p$ (row) and $B_1,\ldots,B_q$ (column) in (A9); more precisely, consider F, Φ, and Ψ as $\infty\times\infty$ matrices whose matrix elements are also matrices with indices $(A_1,\ldots,A_p,B_1,\ldots,B_q)$. Then, (A18) is a factorization of a matrix (F) into a lower triangular matrix (Φ) and an upper triangular matrix (Ψ). This type of factorization is universal in the theory of nonlinear integrable systems; the method of Ref. 19 is applicable to the present case. To see this, we divide F into four blocks as

$$F=\begin{pmatrix}F_{--} & F_{-+}\\ F_{+-} & F_{++}\end{pmatrix}, \tag{A20}$$

where F_{--} denotes the submatrix in which the second indices (m,n) [cf. (A9)] are limited to such regions as $m<0$ (row) and $n<0$ (column); the other blocks are also defined in a similar way. The proof of the invertibility of $\rho(f)$ (see the comments after Proposition A1) carries over to F_{--} without any essential change, and one can show that the matrix product $F_{+-}(F_{--})^{-1}$ is meaningful within $\mathscr{A}_\epsilon$. On the other hand, in view of the lower- or upper-triangular form of the factors on the right-hand side, one deduces from (A18) the following equation:

$$(\Phi^{-1})_{0-}F_{--}+F_{0-}=0, \tag{A21}$$

where $(\cdots)_{0\pm}$ means the submatrices whose second indices (m,n) range over $m=0$ and $n\geqslant 0$ or $n\leqslant -1$ according to the signature $\pm$. One can now solve this linear equation to determine all Taylor and Laurent coefficients of φ^{-1} [written as in (A8)] in the following form:

$$(\varphi^{-1})^{A,0}_{B_1,\ldots,B_q,n}=-\left[F_{0-}(F_{--})^{-1}\right]^{A,0}_{B_1,\ldots,B_q,n}\quad(n\leqslant -1). \tag{A22}$$

At this stage, we have only derived a necessary condition for φ to be a solution of (A16a); it is still not clear whether Φ agrees with $\rho(\varphi)$. This is, however, sufficient: One can check from the above construction that $\psi(\lambda)\underset{\rm def}{=}f(\lambda)^{-1}\circ\varphi(\lambda)$ contains no negative powers of λ. Proposition A3 is thus proven.

We can now complete the proof of Proposition A2. Checking the uniqueness part is just the same as in the proof of Proposition A3. For the proof of the existence part, we first factorize f as in Proposition A2:

$$f(\lambda)=\varphi_1(\lambda)\circ\psi_1(\lambda)^{-1}, \tag{A23}$$

where $\varphi_1(\lambda)\in\Gamma_{{\rm gen},-}^{\mathscr{A}_\epsilon(\lambda)}$ and $\psi_1(\lambda)\in\Gamma_{{\rm gen},+}^{\mathscr{A}_\epsilon(\lambda)}$. We next factorize $\hat{f}(\lambda)\circ\psi_1(\lambda)\in\hat{\Gamma}_{\rm gen}^{\mathscr{A}_\epsilon(\lambda)}$ in a reversed order:

$$\hat{f}(\lambda)\circ\psi_1(\lambda)=\psi_2(\lambda)\circ\varphi_2(\lambda)^{-1}, \tag{A24}$$

where $\varphi_2(\lambda)\in\hat{\Gamma}_{{\rm gen},-}^{\mathscr{A}_\epsilon(\lambda)}$ and $\psi_2(\lambda)\in\hat{\Gamma}_{{\rm gen},+}^{\mathscr{A}_\epsilon(\lambda)}$. Then,

$$\varphi(\lambda)\underset{\rm def}{=}\varphi_1(\lambda)\circ\varphi_2(\lambda),$$

$$\hat{\varphi}\psi(\lambda)\underset{\rm def}{=}\psi_2(\lambda) \tag{A25}$$

give a solution of (A15a). Suitably adjusting the arbitrariness as mentioned in the remark after Proposition A3, one can obtain a solution that also satisfies (A15b). This completes the proof of Proposition A3.

Remarks: (i) The construction of the groups of formal maps and the results on factorization problems in this Appendix can be readily extended to the symplectic case. Actually, all that we need are just corollaries of them. For example, if $f(\lambda)$ in Proposition A3 is assumed to satisfy the symplectic condition as in (2.4), then $\varphi(\lambda)$ and $\psi(\lambda)$ therein also turn out to obey the same condition. This is rather immediate from the construction (and equally applicable to more general cases) as follows: The chain rule of differentiation applied to (A16a) gives the relation

$$\left(\frac{\partial f(\lambda)}{\partial x}\right)\Big|_{x\to\psi(\lambda)}\left(\frac{\partial\psi(\lambda)}{\partial x}\right)=\left(\frac{\partial\varphi(\lambda)}{\partial x}\right), \tag{A26}$$

where the notations are the same as in (3.7). Let ϵ denote the $2r\times 2r$ skew-symmetric matrix (ϵ^{AB}); (2.4) then can be written

$$\left(\frac{\partial f(\lambda)}{\partial x}\right)\epsilon\left(\frac{\partial f(\lambda)}{\partial x}\right)^{\rm tr}\epsilon=-1, \tag{A27}$$

where "tr" stands for transposed matrices. From (A26) and (A27), one can see that

$$\epsilon\left(\frac{\partial\psi(\lambda)}{\partial x}\right)^{\mathrm{tr}}\epsilon\left(\frac{\partial\psi(\lambda)}{\partial x}\right)=\epsilon\left(\frac{\partial\varphi(\lambda)}{\partial x}\right)^{\mathrm{tr}}\epsilon\left(\frac{\partial\varphi(\lambda)}{\partial x}\right). \tag{A28}$$

In fact, the left- (right-) hand side of (A28) contains only nonnegative (nonpositive) powers of λ. Both sides should accordingly be independent of λ; because of (A16b) they must be equal to -1. This means that $\varphi(\lambda)$ and $\psi(\lambda)$ are actually (formal) canonical transformations. A similar argument applies to Proposition A2.

(ii) To go into an analytic world from the above formal setting, one just has to consider such elements of the formal groups in (A7) that are analytic functions of all parameters contained in $\mathscr{A}_\varepsilon$ [precisely, $u_n^A(n\geqslant 0)$ and $\hat{u}_n^A(n\leqslant -1)$] with a nonempty domain of convergence as Laurent series of λ. Then, one can show (with many steps of lengthy estimates) that the resulting formal solutions of Eq. (3.4), etc., have the same property, thus giving an analytic solution.

(iii) The method presented in this Appendix is developed just for proving the existence of solutions, and not very suited for practical use.

APPENDIX B: TRANSFORMATION THEORY OF FIRST KEY FUNCTION

The results in Secs. II, III, and IV all have analogs for the first Plebanski key function. The first key function is the "heavenly" counterpart of the Kähler potential.[7] In our formulation based on a hierarchy structure it is to be defined by the following equation:[1]

$$d\Omega=-\sum_{n>1}\epsilon_{AB}u^A_{-n+1}\,du^B_n+\sum_{n\leq 0}\epsilon_{AB}\hat{u}^A_{-n+1}\,d\hat{u}^B_n. \tag{B1}$$

The closedness of the right-hand side is again ensured by the hierarchy itself. As independent variables, we now choose $u_n^A(n\geqslant 1)$ and $\hat{u}_n^A(n\leqslant 0)$ rather than $u_n^A(n\geqslant 0)$ and $\hat{u}_n^A(n\leqslant -1)$. Such a change of independent variables is permitted because of the relation

$$\bigwedge_{A=1}^{2r}du_0^A\equiv\pm\bigwedge_{A=1}^{2r}d\hat{u}_0^A \quad \mathrm{mod}\{\mathrm{Eqs.}(2.6'),du_n^A(n\neq 0),d\hat{u}_n^A(n\neq 0)\} \tag{B2}$$

that follows from Eq. (2.6). In terms of these new independent variables, the above definition of Ω can also be rewritten

$$u_{A,-n+1}=\frac{\partial\Omega}{\partial u_n^A}\quad(n\geqslant 1), \tag{B3a}$$

$$\hat{u}_{A,-n+1}=-\frac{\partial\Omega}{\partial\hat{u}_n^A}\quad(n\leqslant 0). \tag{B3b}$$

The next step is to single out an appropriate symplectic structure. Actually, there are two different options. The first one is to use the $p^A(\underset{\mathrm{def}}{=}u_1^A)$ space with the symplectic form $\epsilon_{AB}dp^A\wedge dp^B$; the other is due to the $\hat{p}^A(\underset{\mathrm{def}}{=}\hat{u}_0^A)$ space with $\epsilon_{AB}\,d\hat{p}^A\wedge d\hat{p}^B$. Both lead to essentially the same theory, simply the representation being different. We now adopt the second framework (see Ref. 1 for the case of the first choice). Hamiltonian vector fields and Poisson brackets in this setting are then given by

$$\hat{H}(F)\underset{\mathrm{def}}{=}\epsilon_{AB}\frac{\partial F}{\partial\hat{p}^A}\frac{\partial}{\partial\hat{p}^B}, \tag{B4a}$$

$$\{F,G\}\hat{}\underset{\mathrm{def}}{=}\epsilon_{AB}\frac{\partial F}{\partial\hat{p}^A}\frac{\partial G}{\partial\hat{p}^B}=\hat{H}(F)G. \tag{B4b}$$

Carets are used to distinguish the new Poisson brackets and Hamiltonian vector fields from previous ones.

The hierarchy then takes the following Hamiltonian form, and, like Eqs. (3.19), give an equivalent expression of Eq. (2.6):

$$\{u^A(\lambda),u^B(\lambda)\}\hat{}=\epsilon^{AB}, \tag{B5a}$$

$$\{\hat{u}^A(\lambda),\hat{u}^B(\lambda)\}\hat{}=\epsilon^{AB}, \tag{B5b}$$

$$\frac{\partial u^B(\lambda)}{\partial u_n^A}+\{(\lambda^n u_A(\lambda))_{>1},u^B(\lambda)\}\hat{}=0\quad(n\geqslant 1), \tag{B5c}$$

$$\frac{\partial\hat{u}^B(\lambda)}{\partial u_n^A}+\{(\lambda^n u_A(\lambda))_{>1},\hat{u}^B(\lambda)\}\hat{}=0\quad(n\geqslant 1), \tag{B5d}$$

$$\frac{\partial u^B(\lambda)}{\partial\hat{u}_n^A}+\{(\lambda^n u_A(\lambda))_{<0},u^B(\lambda)\}\hat{}=0\quad(n\leqslant -1), \tag{B5e}$$

$$\frac{\partial\hat{u}^B(\lambda)}{\partial\hat{u}_n^A}+\{(\lambda^n u_A(\lambda))_{<0},\hat{u}^B(\lambda)\}\hat{}=0\quad(n\leqslant -1), \tag{B5f}$$

where $(\cdots)_{>1}$ and $(\cdots)_{<0}$ stand for the projectors

$$\left(\sum a_n\lambda^n\right)_{>1}\underset{\mathrm{def}}{=}\sum_{n>1}a_n\lambda^n, \tag{B6a}$$

$$\left(\sum a_n\lambda^n\right)_{<0}\underset{\mathrm{def}}{=}\sum_{n<0}a_n\lambda^n. \tag{B6b}$$

The previous projectors $(\cdots)_+$ and $(\cdots)_-$, in this notation, may be likewise written $(\cdots)_{\geqslant 0}$ and $(\cdots)_{\leqslant -1}$.

We now show below a set of results that correspond to the case discussed in Secs. II, III, and IV. A point is that we now take $\hat{p}=(\hat{u}_0^A)$, rather than $x=(u_0^A)$, as a coordinate system for which a (pseudo-)group structure of the nonlinear graviton construction is to be identified.

Proposition B1: Given a pair of functions (Hamiltonians) $F=F(\hat{p},\lambda)$ and $\hat{F}=\hat{F}(\hat{p},\lambda)$, an infinitesimal symmetry $\hat{\delta}=\hat{\delta}_{F,\hat{F}}$ of the hierarchy is defined by

$$\hat{\delta}u^B(\lambda)=\{(F(u(\lambda),\lambda)-\hat{F}(\hat{u}(\lambda),\lambda))_{<0},u^B(\lambda)\}\hat{}, \tag{B7a}$$

$$\hat{\delta}\hat{u}^B(\lambda)=\{(\hat{F}(u(\lambda),\lambda)-F(\hat{u}(\lambda),\lambda))_{>1},\hat{u}^B(\lambda)\}\hat{}. \tag{B7b}$$

Generating functions corresponding to the time evolutions of the hierarchy are given by the following:
case (i) for $\partial/\partial u_n^A$ $(n\geqslant 1)$,

$$F=\lambda^n\hat{p}_A,\quad\hat{F}=0; \tag{B8a}$$

case (ii) for $\partial/\partial\hat{u}_n^A$ $(n\leqslant -1)$,

$$F=0,\quad\hat{F}=\lambda^n\hat{p}_A. \tag{B8b}$$

Proposition B2: An infinitesimal transformation of Ω consistent with (B7) is given by

$$\hat{\delta}\Omega = -\operatorname*{res}_{\lambda=\infty} \lambda^{-2} F(u(\lambda),\lambda) - \operatorname*{res}_{\lambda=0} \lambda^{-2} \hat{F}(\hat{u}(\lambda),\lambda). \tag{B9}$$

Proposition B3: The infinitesimal transformations $\hat{\delta} = \hat{\delta}_{F,\hat{F}}$ of $u^A(\lambda)$, $\hat{u}^A(\lambda)$ and Ω obey the commutation relations

$$[\hat{\delta}_{F,\hat{F}}, \hat{\delta}_{G,\hat{G}}] = \hat{\delta}_{\{F,G\}\hat{},\{\hat{F},\hat{G}\}\hat{}}. \tag{B10}$$

[1] K. Takasaki, J. Math. Phys. **30**, 1515 (1989).
[2] M. Sato and Y. Sato, in *Nonlinear Partial Differential Equations in Applied Sciences*, Proc. U.S.–Japan seminar, Tokyo, 1982, edited by P. D. Lax, H. Fujita, and G. Strang (North-Holland, Amsterdam/Kinokuniya, Tokyo, 1982).
[3] E. Date, M. Kashiwara, M. Jimbo, and T. Miwa, in *Nonlinear Integrable Systems—Classical Theory and Quantum Theory*, Kyoto 1981, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).
[4] C. P. Boyer and J. F. Plebanski, J. Math. Phys. **26**, 229 (1985).
[5] R. Penrose, Gen. Rel. Grav. **7**, 31 (1976).
[6] K. Takasaki, Publ. RIMS, Kyoto Univ. **22**, 949 (1986).
[7] J. F. Plebanski, J. Math. Phys. **16**, 2395 (1975).
[8] E. Bergshoeff, E. Sezgin, Y. Tanii, and P. K. Townsend, "Super p-brane as gauge theories of volume preserving diffeomorphisms," ICTP preprint IC/88/186 (October 1988); B. de Wit, J. Hoppe, and H. Nicolai, Nucl. Phys. B **305** [FS23], 545 (1988).
[9] B. Biran, E. G. F. Floratos, and G. K. Savvidy, Phys. Lett. B **198**, 329 (1987); R. P. Zaikov, *ibid.* **211**, 281 (1988); F. G. Floratos, and G. K. Leontaris, *ibid.* **223**, 153 (1989).
[10] I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 362 (1979).
[11] K. Ueno and Y. Nakakmura, Phys. Lett. B **109**, 273 (1982).
[12] K. Takasaki, "Hierarchy structure in integrable systems of gauge fields and underlying Lie algebras," Commun. Math. Phys. **127**, 225 (1990).
[13] C. P. Boyer and P. Winternitz, J. Math. Phys. **30**, 1081 (1989).
[14] T. A. Arakelyan and G. K. Savvidy, Phys. Lett. B **214**, 350 (1988); I. Bars, C. N., Pope, and E. Sezgin, *ibid.* **210**, 85 (1988); F. G. Floratos and J. Iliopoulos, *ibid.* **201**, 237 (1988); J. Hoppe, *ibid.* **215**, 706 (1988).
[15] K. Ueno and K. Takasaki, in *Group Representations and Systems of Differential Equations*, edited by K. Okamoto (North-Holland, Amsterdam/Kinokuniya, Tokyo, 1984) Advanced Studies in Pure Math., Vol. 4, pp. 1–94; K. Takasaki, in the same volume, pp. 139–163.
[16] L.-L. Chau and B. Milewski, Phys. Lett. **216**, 330 (1989).
[17] L. J. Mason and E. T. Newman, Commun. Math. Phys. **121**, 659 (1989).
[18] K. Takasaki, "Nonlinear integrable systems as deformations of $\mathscr{D}$-modules," in the Proceedings of the A. M. S. Summer Institute *Theta Functions*, Bowdoin College, 1987, A. M. S. Proc. Symp. Pure Math. **49**, Part 1, 143 (1989).
[19] K. Takasaki, Commun. Math. Phys. **94**, 35 (1984) Part II, Saitama Math. J. **3**, 11 (1985).
[20] N. Koblitz, *p-adic Numbers, p-adic Analysis, and Zeta Functions* (Springer-Verlag, Berlin, 1984).

Classification of star and grade-star representations of $C(n+1)$

R. B. Zhang and M. D. Gould
Department of Mathematics, University of Queensland, St. Lucia, Queensland 4067, Australia

(Received 15 September 1989; accepted for publication 21 March 1990)

The two types of * and grade-* representations of the Lie superalgebra $C(n+1)$ are classified. A type (1) irreducible *-representation is characterized by the single condition $(\Lambda,\alpha_s)\geqslant 0$, Λ and α_s being the highest weight and the odd simple root, respectively, while the type (2) *-representations are duals of type (1) *-representations. For $n>1$, only the identity and vector representations of $C(n+1)$ are shown to be grade -*. The superalgebra $C(2)$ proves to be an exception and admits several classes of nontrivial irreducible grade -* representations.

I. INTRODUCTION

The theory of Lie superalgebras first arose in physical applications in the context of elementary particle physics and Fermi–Bose supersymmetry,[1-4] and has since been applied in a variety of other areas including nuclear physics[5] and condensed matter physics:[6-8] for a comprehensive review of the subject see Kostelecky and Campbell.[9] Those representations of Lie superalgebras most likely to be of interest in physical applications are the * and grade-* representations that were first introduced by Scheunert *et al.*[10] as a natural generalization of Hermitian representations of Lie algebras. However, despite several case studies,[11-15] the important problem of classifying the * and grade-* finite-dimensional *irreducible representations* (irreps) of a simple (basic classical) Lie superalgebra has so far received comparatively little attention.

This is the second paper in a series devoted to the classification of * and grade-* modules for a type I basic classical Lie superalgebra. In the first paper[16] of the series (herein after referred to as I) some general properties of * and grade-* irreps for the Lie superalgebra $\mathrm{gl}(m|n)$ were investigated and applied to give a classification, in terms of highest weights, of the * and grade-* irreps of $\mathrm{gl}(n|1)$. In particular, it was shown that while a large class of *-irreps exist, grade-* irreps for $\mathrm{gl}(n|1)$ are comparatively rare, and as such are unlikely to be of importance in applications: indeed for $n\neq 2$, all grade-* irreps are also *-irreps. It is the aim of this paper to extend the approach of I to investigate the * and grade-* irreps of the type I Lie superalgebra $C(n+1)=\mathrm{osp}(2|2n)$.

The Lie superalgebra $C(n+1)$, as for the case of $\mathrm{gl}(m|n)$, admits two types [herein referred to as type (1) and type (2)] of * and grade-* irreps. It is shown that, as for the $\mathrm{gl}(n|1)$ case, an irrep of $C(n+1)$ with highest weight Λ is type (1) * if and only if Λ is real and $(\Lambda,\alpha_s)\geqslant 0$, where α_s is the odd simple root, and that type (2) * irreps are duals of type (1) * irreps. The situation with grade-* irreps is quite different and it is demonstrated that for $n>1$, the only grade-* irreps of $C(n+1)$ are the identity and vector representations. The Lie superalgebra $C(2)$ proves to be a special case and admits two classes of atypical[17] grade-* irreps (which are also *-irreps) as well as a one-parameter family of four-dimensional typical grade-* irreps (which are not *-irreps): this latter class affords the only examples of grade-* irreps for $C(n+1)$ that are typical.

It follows from these results, as for the case of $\mathrm{gl}(n|1)$, that while *-irreps for $C(n+1)$ comprise a large class, the grade-* irreps are comparatively rare. From the point of view of extensions, it would clearly be of interest to determine whether a similar situation prevails for the Lie superalgebras $\mathrm{gl}(m|n)$ and $\mathrm{osp}(m|n)$ in general. It would also be of interest to investigate * and grade-* irreps arising from noncompact real forms of the basic classical Lie superalgebras, particularly those likely to be of interest in applications. Such representations will clearly be infinite dimensional, in which case the infinitesimal character[18] of a representation may be used in place of a highest weight label. Finally, we note that the tensor product of two type (1) [or (2)] * irreps again gives a (completely reducible) *-representation, which opens up the interesting possibility of determining the Wigner coefficients and Clebsch–Gordan series for *-representations.

II. PRELIMINARIES

The generators of the Lie superalgebra $C(n+1)$ are given by the even $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$ generators σ_j^i $(1\leqslant i,j\leqslant 2n)$, Ω, respectively, together with the odd generators σ^i, σ_i $(1\leqslant i\leqslant 2n)$ satisfying the relations

$$
\begin{aligned}
&[\sigma_j^i,\sigma_l^k]_- = \delta_j^k\sigma_l^i-\delta_l^i\sigma_j^k-\eta_i\eta_j(\delta_{\bar i}^k\sigma_l^{\bar j}-\delta_{\bar i}^{\bar j}\sigma_{\bar l}^k),\\
&\qquad\qquad \sigma_j^i = -\eta_i\eta_j\sigma_{\bar i}^{\bar j},\\
&[\sigma_j^i,\sigma^k]_- = \delta_j^k\sigma^i-\eta_i\eta_j\delta_{\bar i}^k\sigma^{\bar j},\quad [\Omega,\sigma^k]_- = \sigma^k,\\
&[\sigma_j^i,\sigma_k]_- = -\delta_k^i\sigma_j+\eta_i\eta_j\delta_k^{\bar j}\sigma_{\bar i},\\
&[\Omega,\sigma_k]_- = -\sigma_k,\\
&[\sigma^i,\sigma_j]_+ = \sigma_j^i-\delta_j^i\Omega,\\
&[\sigma^i,\sigma^j]_+ = [\sigma_i,\sigma_j]_+ = [\alpha_j^i,\Omega]_- = 0,
\end{aligned}
\tag{1}
$$

where we have employed the useful notation

$$\bar i = 2n+1-i,\quad -\eta_{\bar i}=\eta_i=1\quad (1\leqslant i\leqslant n).$$

In the above $[\ ,\ ]_-$ (resp. $[\ ,\ ]_+$) denotes the usual commutator (resp. anticommutator): these two cases are taken into account below by the graded bracket, denoted $[\ ,\ ]$.

As a basis for the Cartan subalgebra of $C(n+1)$ we choose the commuting operators

$$h_i=\sigma_i^i=-\sigma_{\bar i}^{\bar i},\quad 1\leqslant i\leqslant n,\quad \Omega,$$

 0022-2488/90/081889-09$03.00

whose eigenvalues serve to label the weights of the representations. We denote the weights of $C(n+1)$ by[17]

$$\Lambda = (\lambda\,|\omega) = \sum_{i=1}^{n} \lambda_i \epsilon_i + \omega\delta_1$$

so that, with this convention, the root system of $C(n+1)$ is given by the set of even roots

$$\pm\epsilon_i \pm \epsilon_j, \quad 1 \leqslant i < j \leqslant n, \quad \pm 2\epsilon_i, \quad 1 \leqslant i \leqslant n$$

together with the set of odd roots

$$\pm\epsilon_i \pm \delta_1, \quad 1 \leqslant i \leqslant n.$$

As a system of simple roots, we choose the distinguished set

$$\alpha_i = \epsilon_i - \epsilon_{i+1} \quad (1 \leqslant i < n),$$

$$\alpha_n = 2\epsilon_n, \quad \alpha_s = -\epsilon_1 + \delta_1,$$

so that the sets of even and odd positive roots are given, respectively, by

$$\Phi_0^+ = \{\epsilon_i \pm \epsilon_j \,|\, 1 \leqslant i < j \leqslant n\} \cup \{2\epsilon_i \,|\, 1 \leqslant i \leqslant n\},$$

$$\Phi_1^+ = \{\pm\epsilon_i + \delta_1 \,|\, 1 \leqslant i \leqslant n\}.$$

We denote the half-sum of the even and odd positive roots, respectively, by

$$\rho_0 = \frac{1}{2} \sum_{\alpha \in \Phi_0^+} \alpha = \sum_{i=1}^{n} (n - i + 1)\epsilon_i,$$

$$\rho_1 = \frac{1}{2} \sum_{\alpha \in \Phi_1^+} \alpha = n\delta_1,$$

and set

$$\rho = \rho_0 - \rho_1.$$

Throughout, we let $(\ ,\ )$ denote the nondegenerate bilinear form defined on the weights by[17]

$$(\Lambda, \Lambda') = \sum_{i=1}^{n} \lambda_i \lambda_i' - \omega\omega',$$

where $\Lambda = (\lambda\,|\omega)$, $\Lambda' = (\lambda'\,|\omega')$. We recall[17,18] that this form is invariant under the Weyl group of $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$, herein denoted by W. We note that if α_s is the simple odd root then

$$(\rho, \alpha_s) = 0,$$

and moreover

$$(\rho, \alpha) = (\rho_0, \alpha) > 0, \quad \forall \alpha \in \Phi_0^+.$$

Every finite-dimensional $C(n+1)$ module V admits a Z_2-grading

$$V = V_{\bar{0}} \oplus V_{\bar{1}},$$

where $V_{\bar{0}}$ (resp. $V_{\bar{1}}$) is referred to as the even (resp. odd) component of V. We then define, for homogeneous $v \in V$, the parity factor (v) by $(v) = 0$ (resp. 1) according to whether $v \in V_{\bar{0}}$ (resp. $V_{\bar{1}}$). Following Kac,[17] the finite-dimensional irreducible $C(n+1)$ modules are uniquely characterized by their highest weights Λ, where Λ is a dominant weight for the Lie algebra $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$: recall that $\Lambda = (\lambda\,|\omega)$, λ an $\mathrm{sp}(2n)$ weight, is dominant if $\omega \in \mathbb{C}$ and

$$\lambda_i \in Z^+ \quad (1 \leqslant i \leqslant n), \quad \lambda_i - \lambda_j \in Z^+ \quad (1 \leqslant i < j \leqslant n).$$

Throughout, we denote the set of dominant weights for $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ [and hence $C(n+1)$] by D_+, and for $\Lambda \in D_+$, we denote the finite-dimensional irreducible $C(n+1)$ module with highest weight Λ by $V(\Lambda)$. We denote the finite-dimensional irreducible $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ module with highest weight $\Lambda \in D_+$ by $V_0(\Lambda)$ and we let $\Pi(\Lambda)$ [resp. $\Pi_0(\Lambda)$] denote the set of *distinct* weights in $V(\Lambda)$ [resp. $V_0(\Lambda)$].

Following Kac[17] we say that $V(\Lambda)$, and its highest weight $\Lambda \in D_+$, are *typical* if

$$(\Lambda + \rho, \alpha) \neq 0, \quad \forall \alpha \in \Phi_1^+,$$

otherwise we say that Λ and $V(\Lambda)$ are *atypical*. For future reference we observe, for $\Lambda = (\lambda\,|\omega)$, that

$$(\Lambda + \rho, \epsilon_i + \delta_1) = \lambda_i - \omega + 2n - i + 1,$$

$$(\Lambda + \rho, -\epsilon_i + \delta_1) = i - 1 - \lambda_i - \omega,$$

so that Λ is typical if and only if

$$\lambda_i - \omega + 2n - i + 1 \neq 0, \quad \lambda_i + \omega - i + 1 \neq 0, \quad 1 \leqslant i \leqslant n. \tag{2}$$

Every finite-dimensional irreducible $C(n+1)$ module admits a natural Z-gradation[17,19]

$$V(\Lambda) = \bigoplus_{k=0}^{d} V_k(\Lambda) \tag{3}$$

in consistency with the Z-gradation of the algebra itself.[17,19] We assume $V_d(\Lambda) \neq (0)$ and say that $V(\Lambda)$ *admits* $d+1$ *levels*. The Z-gradation (3) induces the following partitioning of the weight spectrum $\Pi(\Lambda)$:

$$\Pi(\Lambda) = \bigcup_{k=0}^{d} \Pi_k(\Lambda), \tag{4}$$

where $\Pi_k(\Lambda)$ is the set of distinct weights in $V_k(\Lambda)$. Each component $V_k(\Lambda)$ occurring in the decomposition (3) is to constitute a module over the even subalgebra $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ from which it follows[18] that $\Pi_k(\Lambda)$ is stable under the Weyl group W.

Following I, in order to investigate * and grade-* modules, it is useful to consider a natural method for inducing nondegenerate invariant sesquilinear forms on $V(\Lambda)$ from a given invariant inner product on its maximal Z-graded component $V_0(\Lambda)$. To this end, let us assume that $V_0(\Lambda)$ is a Hermitian irreducible $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ module so that $V_0(\Lambda)$ is equipped with a positive-definite inner product $\langle|\rangle$ satisfying

$$\langle \sigma_j^i v | w \rangle = \langle v | \sigma_i^j w \rangle, \quad \langle \Omega v | w \rangle = \langle v | \Omega w \rangle,$$

$$\langle v | w \rangle^* = \langle w | v \rangle, \quad \forall v, w \in V_0(\Lambda).$$

We note that such a form can always be set up on $V_0(\Lambda)$, $\Lambda = (\lambda\,|\omega) \in D_+$, provided ω is real. As for the $\mathrm{gl}(m|n)$ case,[16] the above inner product can be uniquely extended to a form on $V(\Lambda)$ in four different ways.

To treat these possibilities in a unified way, it is convenient to introduce two grading parameters θ, ϵ that can take values 0 or 1. For a given pair of values (θ, ϵ) we then define

$$\langle V_0(\Lambda) | V_k(\Lambda) \rangle = 0, \quad 0 < k \leqslant d,$$

$$\langle \sigma_i v | w \rangle = (-1)^{\theta \cdot (v) + \epsilon} \langle v | \sigma^i w \rangle,$$

$$\langle \alpha v_1 + \beta v_2 | w \rangle = \alpha^* \langle v_1 | w \rangle + \beta^* \langle v_2 | w \rangle,$$

which, as in the $\mathrm{gl}(m|n)$ case, enables a recursive definition

for $\langle|\rangle$ on all of $V(\Lambda)$. By exactly the same method as that used in I, it is easily verified that the above determines a well defined sesquilinear form on $V(\Lambda)$ with the following properties:

$$\langle\sigma_j^i v|w\rangle = \langle v|\sigma_i^j w\rangle, \quad \langle\Omega v|w\rangle = \langle v|\Omega w\rangle,$$
$$\langle\sigma^i v|w\rangle = (-1)^{\theta\cdot(w)+\epsilon}\langle v|\sigma_i w\rangle,$$
$$\langle v|w\rangle^* = \langle w|v\rangle, \quad \forall v,w\in V(\Lambda). \tag{5}$$

Let us agree to call a sesquilinear form on $V(\Lambda)$, satisfying the properties of Eq. (5), *invariant* of type (θ,ϵ). Following the method of I, the above form induced on $V(\Lambda)$ is necessarily nondegenerate and is the unique (up to scalar multiples) invariant sesquilinear form of type (θ,ϵ) on $V(\Lambda)$: it is thus uniquely determined by its restriction to $V_0(\Lambda)$. We note that such a form has all the properties of an inner product except that it is not generally positive definite. As shall be discussed in Secs. III and IV, if the induced form $\langle|\rangle$ is positive definite, we say that $V(\Lambda)$ is a *-module of type (1) [resp. (2)] if $\theta=0$ and $\epsilon=0$ (resp. 1) while we say that $V(\Lambda)$ is a grade-* module of type (1) [resp. (2)] if $\theta=1$ and $\epsilon=0$ (resp. 1).

As discussed in I, the above procedure in fact extends to any finite-dimensional indecomposable $C(n+1)$ module $\overline{V}(\Lambda)$ generated by a highest weight vector of weight $\Lambda=(\lambda|\omega)\in D_+$, ω real. In such a case the induced form $\langle|\rangle$ is sesquilinear and invariant but is not nondegenerate unless $\overline{V}(\Lambda)=V(\Lambda)$ is irreducible: indeed the Kernel

$$K=\{v\in\overline{V}(\Lambda)|\langle v|w\rangle=0, \quad \forall w\in\overline{V}(\Lambda)\}$$

of $\langle|\rangle$ is the unique maximal (Z-graded) submodule of $\overline{V}(\Lambda)$ so that

$$V(\Lambda)\cong\overline{V}(\Lambda)/K.$$

In such a case, the form $\langle|\rangle$ induces a nondegenerate form on the above factor module and suggests a convenient way of extracting an irreducible module from an indecomposable one.

We conclude this section by noting that even when the above form $\langle|\rangle$ induced on $V(\Lambda)$ is not positive definite, it may still be of use for constructing basis states, etc. Indeed, the $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$ Gel'fand invariants are Hermitian under the above form, so that irreducible $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$ submodules of $V(\Lambda)$ with different highest weights are necessarily orthogonal. We further note that although $\langle|\rangle$ is not generally positive definite on $V(\Lambda)$, it is possible to decompose $V(\Lambda)$ into $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$ submodules on each of which $\langle|\rangle$ is either positive or negative definite. Presumably, these useful properties of the form can be utilized for the evaluation of generator matrix elements, etc.

III. CLASSIFICATION OF $C(n+1)$*-MODULES

In this section we investigate both types of irreducible *-modules for $C(n+1)$. In particular we will prove Theorems 1 and 2, which give the necessary and sufficient conditions on the highest (resp. lowest) weight of the irrep. in order that it be * of type (1) [resp. (2)].

As for the $\mathrm{gl}(m|n)$ case, the Lie superalgebra $C(n+1)$ admits two types of irreducible *-modules. Following Scheunert *et al.*[10] we say that $V(\Lambda)$ is an irreducible *-module of type (1) [resp. (2)] if $V(\Lambda)$ can be equipped with a (positive definite) inner product $\langle|\rangle$ satisfying the $\theta=0$, $\epsilon=0$ (resp. 1) case of Eq. (5). In view of the uniqueness of the induced form, we note that such an inner product $\langle|\rangle$ on $V(\Lambda)$ necessarily coincides with the form induced on $V(\Lambda)$ by the restriction of $\langle|\rangle$ to $V_0(\Lambda)$. We see, therefore, that $V(\Lambda)$ is a type (1) [resp. (2)] *-module if and only if $V_0(\Lambda)$ gives rise to a Hermitian irrep. of $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$ and the corresponding $\theta=0$, $\epsilon=0$ (resp. 1) induced form is positive definite.

We note that if $V(\Lambda)$ is a type (1) or (2) *-module with inner product $\langle|\rangle$, then the even generators satisfy

$$\langle\sigma_j^i v|w\rangle = \langle v|\sigma_i^j w\rangle, \quad \langle\Omega v|w\rangle = \langle v|\Omega w\rangle, \quad \forall v,w\in V(\Lambda), \tag{6}$$

which is just the condition that $V(\Lambda)$ constitute a Hermitian module for $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$: in particular we note that the components of Λ must be real. For the odd generators, we have

$$\langle\sigma_i v|w\rangle = \langle v|\sigma^i w\rangle,$$
$$\langle\sigma_i v|w\rangle = -\langle v|\sigma^i w\rangle, \quad \forall v,w\in V(\Lambda),$$

in the type (1) and (2) cases, respectively. Equivalently we have, for the type (1) and (2) cases, respectively,

$$\pi_\Lambda^\dagger(\sigma_j^i)=\pi_\Lambda(\sigma_i^j), \quad \pi_\Lambda^\dagger(\Omega)=\pi_\Lambda(\Omega), \tag{7a}$$

with

$$\pi_\Lambda^\dagger(\sigma_i)=\pi_\Lambda(\sigma^i) \quad [\text{resp.} -\pi_\Lambda(\sigma^i)], \tag{7b}$$

where π_Λ is the representation afforded by $V(\Lambda)$.

As for the $\mathrm{gl}(m|n)$ case,[16] it turns out that the type (1) and (2) *-cases are interchanged by duality. We recall[20] that the representation $\tilde{\pi}_\Lambda$ dual to π_Λ is defined by

$$\tilde{\pi}_\Lambda(x)=-\pi_\Lambda^T(x), \quad x\in C(n+1),$$

where T denotes the supertranspose. Using the method of Ref. 16, it is easily verified that if π_Λ satisfies Eq. (7a) and

$$\pi_\Lambda^\dagger(\sigma_i)=\pi_\Lambda(\sigma^i),$$

then $\tilde{\pi}_\Lambda$ also satisfies Eq. (7a) but now

$$\tilde{\pi}_\Lambda^\dagger(\sigma_i)=-\pi_\Lambda(\sigma^i).$$

In other words, the dual of a type (1) *-irrep. is a type (2) *-irrep (and conversely). Denoting the module dual to $V(\Lambda)$ by $V^*(\Lambda)$, we thus obtain the following.

Proposition 1: For $\Lambda\in D_+$, $V(\Lambda)$ is an irreducible type (1) *-module if and only if $V^*(\Lambda)$ is an irreducible *-module of type (2). □

The above result demonstrates that the classification of the irreducible $C(n+1)$ *-modules essentially reduces to a classification of the type (1) *-modules. To this end, let $V(\Lambda)$ be a type (1) *-module with inner product $\langle|\rangle$. We first find it convenient to define weights ϵ_i for $n<i\leqslant 2n$, by

$$\epsilon_{\bar{i}}=-\epsilon_i, \quad \bar{i}=2n+1-i, \tag{8}$$

in terms of which the odd positive roots are given by the set

$$\Phi_1^+=\{\epsilon_i+\delta_1|1\leqslant i\leqslant 2n\}.$$

Then, for v^Λ the highest weight vector of $V(\Lambda)$, we have

$$0 \leqslant \langle \sigma_i v^\Lambda | \sigma_i v^\Lambda \rangle = \langle v^\Lambda | \sigma^i \sigma_i v^\Lambda \rangle$$
$$= \langle v^\Lambda | (\sigma_i^i - \Omega) v^\Lambda \rangle$$
$$= (\Lambda, \epsilon_i + \delta_1) \langle v^\Lambda | v^\Lambda \rangle, \quad 1 \leqslant i \leqslant 2n,$$

where in the above we used the fact that $\sigma^i v^\Lambda = 0$. Thus in order for $V(\Lambda)$ to be a type (1) *-module, Λ must be real and satisfy

$$(\Lambda, \alpha) \geqslant 0, \quad \forall \alpha \in \Phi_1^+,$$

and in particular $(\Lambda, \alpha_s) \geqslant 0$. Following I, we call $\Lambda \in D_+$ *-*permissible* if Λ is real and $(\Lambda, \alpha_s) \geqslant 0$: throughout, we denote the set of *-permissible dominant weights by D_+^*. The above demonstrates that in order for $V(\Lambda)$ to be a type (1) *-module we must have $\Lambda \in D_+^*$. We now investigate the converse.

We first require the following result, which summarizes some properties of *-permissible dominant weights.

Lemma 1: For $\Lambda \in D_+^*$, we have, for all $\nu \in \Pi(\Lambda)$,

(i) $(\nu, \alpha) \geqslant 0, \quad \forall \alpha \in \Phi_1^+$,

(ii) $(\Lambda - \nu, \Lambda + \nu) \geqslant 0$,

(iii) $(\Lambda - \nu, \Lambda + \nu + 2\rho) \geqslant 0$.

Proof: To prove (i) we note that $(\alpha_s, \alpha_s) = 0$ and for the even simple roots

$$\alpha_i = \epsilon_i - \epsilon_{i+1} \quad (1 \leqslant i < n), \quad \alpha_n = 2\epsilon_n,$$

we have

$$(\alpha_i, \alpha_s) \leqslant 0, \quad 1 \leqslant i \leqslant n.$$

Now for $\nu \in \Pi(\Lambda)$, we may write

$$\nu = \Lambda - \sum_i n_i \alpha_i - n_s \alpha_s, \quad n_i, n_s \in Z^+,$$

so that

$$(\nu, \alpha_s) = (\Lambda, \alpha_s) - \sum_i n_i (\alpha_i, \alpha_s)$$
$$\geqslant (\Lambda, \alpha_s) \geqslant 0, \quad \forall \nu \in \Pi(\Lambda).$$

Part (i) then follows from the W-invariance of $\Pi(\Lambda)$, together with the fact that every root $\alpha \in \Phi_1^+$ is W-conjugate to α_s (cf. Proposition 3 of I).

As in Proposition 4 of I, part (ii) follows by induction on the Z-grading index k, the result being true for all $\nu \in \Pi_0(\Lambda)$ by a known[18] Lie algebra result. Employing the partitioning of Eq. (4), we proceed inductively and assume (ii) holds for all $\nu \in \Pi_{k-1}(\Lambda)$ and note that every $\mu \in \Pi_k(\Lambda)$ may be written

$$\mu = \nu - \alpha, \quad \text{for some } \alpha \in \Phi_1^+, \ \nu \in \Pi_{k-1}(\Lambda).$$

Then using (i) and the inductive hypothesis, we have

$$(\Lambda - \mu, \Lambda + \mu) = (\Lambda - \nu, \Lambda + \nu) + 2(\nu, \alpha) \geqslant 0,$$

and hence $(\Lambda - \mu, \Lambda + \mu) \geqslant 0$, $\forall \mu \in \Pi_k(\Lambda)$. This establishes (ii) by induction.

Finally as to (iii), we may write, in terms of simple roots,

$$\Lambda - \nu = \sum_i n_i \alpha_i + n_s \alpha_s, \quad n_i, n_s \in Z^+, \quad \nu \in \Pi(\Lambda).$$

Using

$$(\rho, \alpha_i) > 0 \quad (1 \leqslant i \leqslant n), \quad (\rho, \alpha_s) = 0,$$

it follows that

$$(\Lambda - \nu, \rho) \geqslant 0, \quad \forall \nu \in \Pi(\Lambda).$$

Thus, using (ii),

$$(\Lambda - \nu, \Lambda + \nu + 2\rho)$$
$$= (\Lambda - \nu, \Lambda + \nu) + 2(\Lambda - \nu, \rho) \geqslant 0, \quad \forall \nu \in \Pi(\Lambda),$$

which proves the result. □

Our aim now is to demonstrate that for $\Lambda \in D_+^*$, $V(\Lambda)$ is a type (1) *-module of $C(n+1)$. To see this, we assume that $V_0(\Lambda)$ gives rise to a Hermitian representation of $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ and we let $\langle | \rangle$ be the corresponding unique nondegenerate invariant sesquilinear form of type $(\theta, \epsilon) = (0,0)$ induced on $V(\Lambda)$. It then suffices to show that $\langle | \rangle$ is positive definite (i.e., gives rise to an inner product).

We first note, as mentioned previously, that the $C(n+1)$ and $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ Gel'fand invariants are Hermitian under the induced form so that $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ submodules of $V(\Lambda)$ with different highest weights are orthogonal. In particular, we note that the $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ invariant

$$\xi = \sum_{i=1}^{2n} \sigma_i \sigma^i$$

is Hermitian and is expressible in terms of the universal Casimir invariants of $C(n+1)$ and $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$, so its eigenvalues depend only on the highest weight labels of $C(n+1)$ and the subalgebra $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$. We have the following result concerning the eigenvalues of the $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ invariant ξ.

Lemma 2: For $\Lambda \in D_+^*$, the eigenvalues of ξ on $V(\Lambda)$ are non-negative.

Proof: The universal Casimir invariant of $C(n+1)$ is given by

$$I_2 = C_2 + \sum_i [\sigma_i \sigma^i - \sigma^i \sigma_i] - \Omega^2,$$

where C_2 is the quadratic Casimir element of $\mathrm{sp}(2n)$. Now let $V_0(\nu)$ be an irreducible $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ submodule of $V(\Lambda)$ and let v^ν be the maximal state of $V_0(\nu)$. Then we have, for the eigenvalue γ of ξ on $V_0(\nu)$,

$$\sum_i \sigma_i \sigma^i v^\nu = \gamma v^\nu, \quad \sum_i \sigma^i \sigma_i v^\nu = \bar{\gamma} v^\nu,$$

so that

$$(\gamma + \bar{\gamma}) v^\nu = \sum_i (\sigma_i \sigma^i + \sigma^i \sigma_i) v^\nu = (\nu, 2\rho_1) v^\nu,$$
$$(\gamma - \bar{\gamma}) v^\nu = (I_2 - C_2 + \Omega^2) v^\nu \qquad (9)$$
$$= [(\Lambda, \Lambda + 2\rho) - (\nu, \nu + 2\rho_0)] v^\nu,$$

where we have used the fact that C_2 takes the eigenvalue $(\nu, \nu + 2\rho_0) + \omega^2$ when acting on $V_o(\nu)$. Adding this last equation to Eq. (9) then gives, for the eigenvalue γ of the $\mathrm{sp}(2n) \oplus \mathrm{o}(2)$ invariant $\xi = \Sigma_i \sigma_i \sigma^i$,

$$2\gamma = (\Lambda, \Lambda + 2\rho) - (\nu, \nu + 2\rho)$$
$$= (\Lambda - \nu, \Lambda + \nu + 2\rho) \geqslant 0,$$

where we have employed part (iii) of Lemma 1. □

We are now in a position to prove that the above induced form $\langle|\rangle$ is positive definite on $V(\Lambda)$, $\Lambda\in D^*_+$, by induction on the Z-grading index k, the result holding for $k=0$ by construction. We now assume that the form is positive definite on $V_{k-1}(\Lambda)$ and note that for $v\in V_k(\Lambda)$, $\sigma^i v\in V_{k-1}(\Lambda)$, from which we obtain, using the inductive hypothesis,

$$\langle v|\sigma_i\sigma^i v\rangle = \langle \sigma^i v|\sigma^i v\rangle \geqslant 0, \quad \forall v\in V_k(\Lambda), \quad 1\leqslant i\leqslant 2n. \tag{10}$$

Now $V_k(\Lambda)$ can be decomposed into $\mathrm{sp}(2n)\oplus \mathrm{o}(2)$ primary components,

$$V_k(\Lambda) = \mathop{\oplus}_{\nu} \overline{V}_0(\nu),$$

where $\overline{V}_0(\nu)$ is the direct sum of all irreducible $\mathrm{sp}(2n)\oplus \mathrm{o}(2)$ submodules of $V_k(\Lambda)$ with highest weight ν. Let us assume initially that $v\neq 0$ belongs to such a primary component $\overline{V}_0(\nu)$. Then from Lemma 2 we have for the invariant $\xi = \Sigma_i \sigma_i\sigma^i$,

$$\xi v = \gamma v, \quad \gamma\geqslant 0$$

(noting that these eigenvalues depend only on the representations labels Λ,ν), so that, in view of Eq. (10),

$$\gamma\langle v|v\rangle = \langle v|\xi v\rangle = \sum_i \langle v|\sigma^i\sigma_i v\rangle \geqslant 0.$$

Hence if $\gamma\langle v|v\rangle\neq 0$, we must have $\gamma>0$ and thus $\langle v|v\rangle>0$. The only other possibility is $\gamma\langle v|v\rangle = 0$, which, in view of Eq. (10), implies

$$\langle \sigma^i v|\sigma^i v\rangle = 0, \quad 1\leqslant i\leqslant 2n,$$

so that, by the induction hypothesis, $\sigma^i v = 0$, $1\leqslant i\leqslant 2n$. Because of the irreducibility of $V(\Lambda)$, this can only occur if $0\neq v\in V_0(\Lambda)$, in which case $\langle v|v\rangle>0$ by construction. Thus, in either case, we have $\langle v|v\rangle>0$.

The above shows, that for $v\neq 0$ in an $\mathrm{sp}(2n)\oplus \mathrm{o}(2)$ primary component $\overline{V}_0(\nu)$ of $V_k(\Lambda)$, $\langle v|v\rangle>0$. On the other hand, every $0\neq v\in V_k(\Lambda)$ is expressible as a sum

$$v = \sum_\alpha v_\alpha,$$

where $v_\alpha\neq 0$ belongs to an $\mathrm{sp}(2n)\oplus \mathrm{o}(2)$ primary component of $V_k(\Lambda)$ and since these primary components are orthogonal under the form $\langle|\rangle$, we must have

$$\langle v|v\rangle = \sum_\alpha \langle v_\alpha|v_\alpha\rangle > 0.$$

This proves that the induced form $\langle|\rangle$ is positive definite on $V_k(\Lambda)$ and thus, by induction, is positive definite on all of $V(\Lambda)$.

Thus we have shown that if $\Lambda\in D^*_+$, then $V(\Lambda)$ is an irreducible type (1) *-module. We have already established the converse, so we arrive at the following.

Theorem 1: For $\Lambda\in D_+$, $V(\Lambda)$ is an irreducible type (1) *-module of $C(n+1)$ if and only if Λ is real and $(\Lambda,\alpha_s)\geqslant 0$. □

As to the type (2) case, we have, recalling proposition (1), that an irreducible $C(n+1)$ module $V(\Lambda)$ is type (2)-* if and only if $V^*(\Lambda)$ is type (1)-*. On the other hand, we note that the weights in $V^*(\Lambda)$ are the negative of the weights in $V(\Lambda)$, and if Λ^- is the lowest weight of $V(\Lambda)$, then $-\Lambda^-$ is the highest weight of $V^*(\Lambda)$. In view of Theorem 1, we thus arrive at the following classification in terms of lowest weights.

Theorem 2: For $\Lambda\in D_+$, $V(\Lambda)$ is an irreducible type (2) *-module of $C(n+1)$ if and only if Λ is real and $(\Lambda^-,\alpha_s)\leqslant 0$.

Corollary: If $\Lambda\in D_+$ is typical, then $V(\Lambda)$ is an irreducible type (2) *-module of $C(n+1)$ if and only if Λ is real and $(\Lambda+\rho,\epsilon_1+\delta_1)<0$.

Proof: For $\Lambda = (\lambda\,|\,\omega)$ typical, the lowest weight of $V(\Lambda)$ is given by

$$\Lambda^- = (-\lambda\,|\,\omega) - 2\rho_1 = (-\lambda\,|\,\omega - 2n),$$

and the condition of the theorem reduces to $\lambda_1 + 2n - \omega\leqslant 0$, which, in view of Eq. 2, may be written $(\Lambda+\rho,\epsilon_1+\delta_1)\leqslant 0$. Obviously, for Λ typical we cannot have $(\Lambda+\rho,\epsilon_1+\delta_1) = 0$. □

IV. CLASSIFICATION OF GRADE-* MODULES

Here we study the two types of grade-* irreps of $C(n+1)$. Our main results are summarized in Theorems 3 and 4, which give a classification, in terms of highest weights, of the type (1) and (2) irreducible grade-* modules, respectively. It turns out that for $n>1$, only the identity and vector representations are grade-*. The superalgebra $C(2)$ proves to be a special case and admits two classes of two level atypical grade-* modules as well as a one-parameter family of four-dimensional typical grade-* modules. In view of the similar corresponding results previously obtained[16] for $\mathrm{gl}(2|1)$, this latter situation is obviously related to the superalgebra isomorphism $C(2)\cong \mathrm{sl}(2|1)$.

Following Scheunert *et al.*,[10] we note, as for the $\mathrm{gl}(m|n)$ case,[16] that $C(n+1)$ admits two types of grade-* representations. We say that $V(\Lambda)$ is a type (1) [resp. (2)] grade-* module if the corresponding representation π_Λ satisfies the graded Hermiticity conditions

$$\pi^\ddagger_\Lambda(\sigma^i) = (-1)^\epsilon \pi_\Lambda(\sigma_i), \quad \epsilon = 0 \text{ (resp. 1)}, \tag{11a}$$

together with

$$\pi^\ddagger_\Lambda(\sigma^i_j) = \pi^\dagger_\Lambda(\sigma^i_j) = \pi_\Lambda(\sigma^i_j), \tag{11b}$$
$$\pi^\ddagger_\Lambda(\Omega) = \pi^\dagger_\Lambda(\Omega) = \pi_\Lambda(\Omega),$$

where ‡ denotes superadjoint: recall[10,20] that the grade adjoint of a homogeneous operator A on a Z_2-graded Hilbert space is defined by

$$A^\ddagger = (A^T)^*,$$

where T denotes the supertranspose, and that for even operators superadjoint corresponds to normal Hermitian conjugate (cf. I).

Equivalently, we say that $V(\Lambda)$ is a grade-* module if it can be equipped with a positive definite inner product satisfying the requirements of Eq. (6) for the even generators, together with

$$\langle \sigma_i v|\omega\rangle = (-1)^{(v)}\langle v|\sigma^i\omega\rangle, \tag{12a}$$

$$\langle \sigma_i v|\omega\rangle = -(-1)^{(v)}\langle v|\sigma^i\omega\rangle, \tag{12b}$$

in the type (1) and (2) grade-* cases, respectively. Such an inner product determines a positive definite invariant sesquilinear form of type $\theta = 1$, $\epsilon = 0$ (resp. 1) on $V(\Lambda)$. As for

the case of *-modules, it follows that $V(\Lambda)$ is an irreducible grade-* module of type (1) [resp. (2)] if and only if its maximal Z-graded component $V_0(\Lambda)$ is a Hermitian $\mathrm{sp}(2n)\oplus\mathrm{o}(2)$ module and the corresponding induced form of type $\theta=1$, $\epsilon=0$ (resp. 1) is positive definite: note that the components of the highest weight Λ must, under such conditions, be real.

As for the $\mathrm{gl}(m|n)$ case,[16] it is important to note that, unlike *-modules, the type of grade-* module depends on the choice of Z_2-grading. However, from Eq. (12), we see that the type (1) and (2) grade-* cases are simply interchanged by a reversal of Z_2-grading. In other words, a type (1) grade-* module $V(\Lambda)$, whose maximal Z-graded component $V_0(\Lambda)$ is chosen to have *odd* Z_2-grading, may be viewed as a type (2) grade-* module in which $V_0(\Lambda)$ is chosen to have *even* Z_2-grading. It thus suffices to consider only the case where $V_0(\Lambda)$ has even Z_2-grading, herein referred to as the *standard* choice of grading (which is adopted throughout the paper).

Following I, the study of grade-* modules is facilitated by noting that if π_Λ is a grade-* irrep, i.e., satisfies the conditions of Eq. (11), then the dual representation $\tilde{\pi}_\Lambda$ satisfies these same equations. In other words, as distinct from the situation for *-irreps, the dual of a type (1) [resp. (2)] grade-* irrep is again grade-* of type (1) [resp. (2)]. It is important, however, that this duality be interpreted with the Z_2-grading for the *minimal* Z-graded component of $V(\Lambda)$ being given by the grading of the *maximal* Z-graded component of the dual module $V^*(\Lambda)$. Thus, with the standard choice of Z_2-grading for $V(\Lambda), V^*(\Lambda)$, we obtain the following.

Proposition 2: For $\Lambda\in D_+$, if $V(\Lambda)$ has an even (resp. odd) number of levels, then $V(\Lambda)$ is a type (1) grade-* module if and only if $V^*(\Lambda)$ is grade-* of type (2) [resp. (1)]. □

Suppose now that $V(\Lambda)$ is a type (1) grade-* $C(n+1)$ module so that Λ is real and $V(\Lambda)$ is equipped with a (positive definite) inner product $\langle|\rangle$ satisfying the conditions of Eqs. (6) and (12a). Then, for v^Λ the maximal weight vector of $V(\Lambda)$, we have, with the convention of Eq. (8),

$$0\leqslant\langle\sigma_i v^\Lambda|\sigma_i v^\Lambda\rangle=\langle v^\Lambda|\sigma^i\sigma_i v^\Lambda\rangle$$
$$=(\Lambda,\epsilon_i+\delta_1)\langle v^\Lambda|v^\Lambda\rangle.$$

It follows that $(\Lambda,\alpha)\geqslant 0$, $\forall\alpha\in\Phi_1^+$, and in particular $(\Lambda,\alpha_s)\geqslant 0$. In view of Theorem 1, we see that $V(\Lambda)$ must also constitute a type (1) *-module so that we get no new type (1) grade-* modules that are not already *-modules.

Proceeding further, it follows, by exactly the same argument as that used in I for $\mathrm{gl}(n|1)$, that $V(\Lambda)=V_0(\Lambda)\oplus V_1(\Lambda)$ can have at most two Z-graded levels. This can in fact be seen by noting that the inner products

$$\langle\sigma_i\sigma_j v^\Lambda|\sigma_i\sigma_j v^\Lambda\rangle$$

have opposite signs in the type (1) * and grade-* cases, and hence must vanish. This latter requirement, together with the restriction that Λ is real and $(\Lambda,\alpha_s)\geqslant 0$, imposes stringent conditions on Λ. In fact (see Appendix A) for $n>1$, $\Lambda=(\dot{0}|0)$, and for $n=1$, Λ must have the special form

$$\Lambda=(\tau|-\tau),\quad \tau\in Z^+.$$

Moreover, such two level type (1) $C(2)$ *-modules are indeed type (1) grade-* since, with the above inner product,

$$(\sigma^i)^\ddagger=(\sigma^i)^\dagger=\sigma_i,\quad \sigma_i^\ddagger=-\sigma_i^\dagger=-\sigma^i.$$

We thus arrive at the following.

Theorem 3: For $n>1$, the only irreducible type (1) grade-* $C(n+1)$ module $V(\Lambda)$ is the trivial one with $\Lambda=(\dot{0}|0)$. For $C(2)$, $V(\Lambda)$ is an irreducible type (1) grade-* module if and only if Λ has the form

$$\Lambda=(\tau|-\tau),\quad \tau\in Z^+.$$

Corollary: All irreducible type (1) grade-* $C(n+1)$ modules are also type (1) *-modules and admit at most two Z-graded levels. □

We now consider the type (2) grade-* case. We note, from Appendix A, that $V(\Lambda)$ has exactly one level only in the trivial case $\Lambda=(\dot{0}|0)$. In such a case, $V(\Lambda)$ is trivially type (1) and (2) * and grade-*, so it remains to consider the case $\Lambda\neq(\dot{0}|0)$, in which case we can assume that $V(\Lambda)$ has at least two Z-graded levels.

Let $V(\Lambda)$ be such a type (2) grade-* module so that $V(\Lambda)$ is equipped with a positive definite inner product $\langle|\rangle$ satisfying the conditions of Eq. (6) and (12b). Then we have, for v^Λ the maximal weight vector of $V(\Lambda)$,

$$0\leqslant\langle\sigma_i v^\Lambda|\sigma_i v^\Lambda\rangle=-\langle v^\Lambda|\sigma^i\sigma_i v^\Lambda\rangle$$
$$=-(\Lambda,\epsilon_i+\delta_1)\langle v^\Lambda|v^\Lambda\rangle,\quad 1\leqslant i\leqslant 2n, \tag{13}$$

so that $(\Lambda,\alpha)\leqslant 0$, $\forall\alpha\in\Phi_1^+$. We further observe that

$$\sigma_i v^\Lambda\neq 0,\quad 1\leqslant i\leqslant n, \tag{14}$$

otherwise we would have, for some $i\in\{1,2,\ldots,n\}$,

$$0=\sigma^i\sigma_i v^\Lambda=(\Lambda,\epsilon_i+\delta_1)v^\Lambda,$$

and hence

$$(\Lambda,\epsilon_i+\delta_1)=0. \tag{15}$$

Also, applying the even raising generator $\sigma_j^i(1\leqslant j\leqslant n)$ to $\sigma_i v^\Lambda=0$ would yield

$$\sigma_{\bar{j}} v^\Lambda=0,\quad 1\leqslant j\leqslant n,$$

from which we similarly deduce

$$(\Lambda,-\epsilon_j+\delta_1)=0,\quad 1\leqslant j\leqslant n.$$

These equations together with Eq. (15) would then imply $\Lambda=(\dot{0}|0)$, contrary to our choice of Λ. This proves Eq. (14) from which we deduce, in view of Eq. (13),

$$(\Lambda,\epsilon_i+\delta_1)<0,\quad 1\leqslant i\leqslant n,$$

and hence $(\Lambda,-\epsilon_i+\delta_1)\leqslant(\Lambda,\epsilon_i+\delta_1)<0$. This last inequality and Eq. (13) then imply that

$$\sigma_{\bar{i}} v^\Lambda\neq 0,\quad 1\leqslant i\leqslant n,$$

so we may write

$$\sigma_i v^\Lambda\neq 0,\quad (\Lambda,\epsilon_i+\delta_1)<0,\quad 1\leqslant i\leqslant 2n. \tag{16}$$

It is now convenient to consider the cases $n>1$ and $n=1$ separately. We consider first the case $n>1$.

Condition (16) shows that $V(\Lambda)$ has at least two levels, from which we deduce that $V(\Lambda)$ must have at least three

levels, since, from Appendix A, two level irreps exist only for $n=1$. Now for $i<n$ we have

$$0\leqslant\langle\sigma_i\sigma_n v^\Lambda|\sigma_i\sigma_n v^\Lambda\rangle$$
$$=[1+(\Lambda,\epsilon_i+\delta_1)]\langle\sigma_n v^\Lambda|\sigma_n v^\Lambda\rangle,$$
$$0\leqslant\langle\sigma_{\bar n}\sigma_{\bar i}v^\Lambda|\sigma_{\bar n}\sigma_{\bar i}v^\Lambda\rangle$$
$$=[1+(\Lambda,-\epsilon_n+\delta_1)]\langle\sigma_{\bar i}v^\Lambda|\sigma_{\bar i}v^\Lambda\rangle,$$

from which we deduce, in view of Eq. (16),

$$(\Lambda,-\epsilon_n+\delta_1)\geqslant -1,$$

and, since $(\Lambda,\epsilon_n+\delta_1)\geqslant(\Lambda,-\epsilon_n+\delta_1)$,

$$(\Lambda,\epsilon_i+\delta_1)\geqslant -1,\quad 1\leqslant i\leqslant n.$$

These conditions, together with those of Eq. (16), then yield

$$-1\leqslant(\Lambda,\epsilon_i+\delta_1)<0,\quad 1\leqslant i\leqslant n, \tag{17a}$$
$$-1\leqslant(\Lambda,-\epsilon_n+\delta_1)<0. \tag{17b}$$

We are now in a position to show that $V(\Lambda)$ can have no more than three levels, and hence exactly three levels. To prove this, it suffices to demonstrate

$$\sigma_1\sigma_i\sigma_{\bar i}v^\Lambda=0,\quad 1<i\leqslant n, \tag{18a}$$

and, for $n>2$,

$$\sigma_1\sigma_2\sigma_3 v^\Lambda=0, \tag{18b}$$

because by successively applying the even raising operators to (18a), we obtain $\sigma_i\sigma_j\sigma_{\bar j}v^\Lambda=\sigma_{\bar j}\sigma_j\sigma_{\bar i}v^\Lambda=0\ \forall i,j$, and to (18b) we have $\sigma_i\sigma_j\sigma_k v^\Lambda=\sigma_{\bar i}\sigma_{\bar j}\sigma_{\bar k}v^\Lambda=0\ \forall i,j,k$.

Now note that $\sigma_1\sigma_i\sigma_{\bar i}v^\Lambda\neq 0$, for some $i>1$, implies

$$0<\langle\sigma_1\sigma_i\sigma_{\bar i}v^\Lambda|\sigma_1\sigma_i\sigma_{\bar i}v^\Lambda\rangle$$
$$=-[(\Lambda,\epsilon_1+\delta_1)+2]\langle\sigma_i\sigma_{\bar i}v^\Lambda|\sigma_i\sigma_{\bar i}v^\Lambda\rangle,$$

which can only occur if $(\Lambda,\epsilon_1+\delta_1)<-2$ in contradiction to Eq. (17). Similarly for $n>2$, $\sigma_1\sigma_2\sigma_3 v^\Lambda\neq 0$ implies

$$0<\langle\sigma_1\sigma_2\sigma_3 v^\Lambda|\sigma_1\sigma_2\sigma_3 v^\Lambda\rangle$$
$$=-[(\Lambda,\epsilon_1+\delta_1)+2]\langle\sigma_2\sigma_3 v^\Lambda|\sigma_2\sigma_3 v^\Lambda\rangle,$$

which can only occur if $(\Lambda,\epsilon_1+\delta_1)>-2$, again in contradiction to Eq. (17). We have thus proved Eqs. (18a), (18b), from which we deduce that $V(\Lambda)$ must have exactly three Z-graded levels.

To complete the argument, we note that since the components λ_i $(1\leqslant i\leqslant n)$ of the highest weight $\Lambda=(\lambda\,|\omega)$ must differ by integers, Eqs. (17a) imply that Λ has the special form

$$\Lambda=(\dot\tau|\omega),\quad \tau\in Z^+,\quad \tau+1\geqslant\omega>\tau.$$

On the other hand, Eq. (17b) implies $1\geqslant\tau+\omega>0$, which can only occur if $\tau=0$ and $1\geqslant\omega>0$. Thus we conclude that Λ must be of the form

$$\Lambda=(\dot 0|\omega),\quad 0<\omega\leqslant 1.$$

For $0<\omega<1$, Λ must be typical, and since three level typical irreps can only exist for $n=1$, we arrive at the final possibility

$$\Lambda=(\dot 0|1).$$

This is the highest weight of the vector representation of $C(n+1)$, which is indeed a three level irrep that is easily seen to be grade-* of type (2): note, however, that the vector representation is not a star representation. We have thus proved the following.

Proposition 3: For $n>1$, the irreducible $C(n+1)$ module $V(\Lambda)$ is type (2) grade-* only for the cases $\Lambda=(\dot 0|0)$ and $(\dot 0|1)$ corresponding to the identity and vector representations. □

It remains to consider the case $n=1$. We note that an irreducible $C(2)$ module $V(\Lambda)$, for $\Lambda\neq 0$, either has two levels and is atypical or else has three levels and is typical. Beginning with the former case, we note that the only two level irreducible $C(2)$ modules $V(\Lambda)$ with highest weights Λ satisfying the requirements of Eq. (16) have highest weights given by (see Appendix A)

$$\Lambda=(\tau|\tau+2),\quad \tau\in Z^+.$$

In such a case, $V(\Lambda)$ has lowest weight $(-1-\tau|\tau+1)$ and hence the dual module $V^*(\Lambda)$ has highest weight $(\tau+1|-\tau-1)$, which, in view of Theorem 3, is both a type (1)* and grade-* module. Thus, from Propositions 1 and 2, we arrive at the following.

Proposition 4: For $\Lambda\in D_+$, the irreducible $C(2)$ module $V(\Lambda)$ is atypical and grade-* of type (2) if and only if $\Lambda=(\dot 0|0)$ or

$$\Lambda=(\tau|\tau+2),\quad \tau\in Z^+.$$

In such a case, $V(\Lambda)$ is also a type (2) *-module.

It finally remains to consider the typical three-level type (2) grade-* $C(2)$ modules $V(\Lambda)$. For this case, our only odd lowering operators are $\sigma_1,\sigma_{\bar 1}$ $(\bar 1=2)$ and we must have, for v^Λ the highest weight vector of $V(\Lambda)$,

$$0<\langle\sigma_1\sigma_{\bar 1}v^\Lambda|\sigma_1\sigma_{\bar 1}v^\Lambda\rangle$$
$$=\langle\sigma_{\bar 1}v^\Lambda|\sigma^1\sigma_1\sigma_{\bar 1}v^\Lambda\rangle$$
$$=[(\Lambda,\epsilon_1+\delta_1)+2]\langle\sigma_{\bar 1}v^\Lambda|\sigma_{\bar 1}v^\Lambda\rangle,$$

from which we deduce $(\Lambda,\epsilon_1+\delta_1)>-2$. This condition, together with Eq. (16), implies that Λ must be of the form

$$\Lambda=(\tau|\omega),\quad \tau\in Z^+,\quad \tau+2>\omega>\tau. \tag{19}$$

On the other hand, $V^*(\Lambda)$ must also give rise to a typical three level irreducible type (2) grade-* module, and since the lowest weight of $V(\Lambda)$ is $\Lambda^-=(-\tau|\omega-2)$, it follows that $V^*(\Lambda)$ has highest weight $\Lambda^*=(\tau|2-\omega)$ whose components must also satisfy Eq. (19). This imposes the additional constraints

$$\tau+2>2-\omega>\tau,$$

from which we deduce that $\tau=0$ and Λ must have the special form

$$\Lambda=(0|\omega),\quad 0<\omega<2.$$

Conversely, it can be shown (see Appendix B) that with Λ as above, $V(\Lambda)$ indeed gives rise to a (four-dimensional) typical type (2) grade-* $C(2)$ module: we note that such a module $V(\Lambda)$ is not a *-module. We have thus proved the following.

Proposition 5: Irreducible typical type (2) grade-* $C(n+1)$ modules $V(\Lambda)$ exist only for $n=1$ and have highest weights of the form

$$\Lambda=(0|\omega),\quad 0<\omega<2.$$ □

Combining the results of Propositions 3–5, we thus arrive at the following classification scheme.

Theorem 4: For $\Lambda \in D_+$, $V(\Lambda)$ is a type (2) grade-* $C(n+1)$ module if and only if Λ has one of the following special forms:

(i) $n > 1$

$$\Lambda = (\dot{0}|\theta), \quad \theta = 0,1,$$

(ii) $n = 1$

$$\Lambda = (\tau|\tau+2), \quad \tau \in Z^+ \quad \text{or} \quad \Lambda = (0|\omega), \quad 0 \leqslant \omega < 2.$$

□

V. CONCLUSIONS

A classification of all * and grade-* irreps has been presented for the type I Lie superalgebra $C(n+1)$. The main results are summarized in Theorems 1–4, which give the necessary and sufficient conditions for an irreducible representation to be * or grade-* of type (1) or (2), respectively. These conditions, except in the case of atypical type (2) *-irreps (which are classified according to their lowest weight), are given in terms of the highest weight labels. It would be of interest to exploit the modified induced module construction of Ref. 21 to determine the $C(n+1)\downarrow \mathrm{Sp}(n) \oplus \mathrm{o}(2)$ branching rules, along the lines recently applied to $\mathrm{gl}(n|1)$ (Ref. 22), for atypical irreducible modules, and to obtain the minimal weights of such irreps. In particular, this would enable a classification of the atypical type (2) *-irreps in terms of highest weights.

Finally, we note that * and grade-* representations have the useful property of being completely reducible and in this sense are most like the representations of simple Lie algebras. It would clearly be of interest to determine the characters and dimensions of the * and grade-* irreps for the type I basic classical Lie superalgebras, and in particular to determine whether the recent character formula conjectured by Hughes and King,[23] or the earlier formula proposed by Bernstein and Leites[24] and Van der Jengt,[25] apply to these representations.

ACKNOWLEDGMENTS

The authors take great pleasure in thanking Dr. A. J. Bracken for many useful discussions.

They also gratefully acknowledge the financial support of an ARC research grant.

APPENDIX A: CLASSIFICATION OF TWO-LEVEL IRREPS

Here we classify all irreducible $C(n+1)$ modules $V(\Lambda)$ with at most two Z-graded levels: throughout, v^Λ denotes the maximal weight vector of $V(\Lambda)$.

We consider first the one level irreducible modules $V(\Lambda)$. In such a case, we have

$$0 = \sigma_i v^\Lambda = \sigma^i \sigma_i v^\Lambda = (\Lambda, \epsilon_i + \delta_1) v^\Lambda, \quad 1 \leqslant i \leqslant 2n,$$

which implies

$$(\Lambda, \pm \epsilon_i + \delta_1) = 0, \quad 1 \leqslant i \leqslant n.$$

This can only occur for $\Lambda = (\dot{0}|0)$ corresponding to the trivial identity representation.

Now let us suppose $V(\Lambda)$ has exactly two levels so that

$$\sigma_i \sigma_j v^\Lambda = 0, \quad 1 \leqslant i, j \leqslant 2n.$$

We now show that, for $n > 1$, such two level irreducible modules do not exist. Following the proof of Eq. (14), we have, for $n > 1$,

$$\sigma_i v^\Lambda \neq 0, \quad 1 \leqslant i \leqslant n, \tag{A1}$$

from which we obtain for $i < n$,

$$0 = \sigma^i \sigma_i \sigma_n v^\Lambda = [(\Lambda, \epsilon_i + \delta_1) + 1] \sigma_n v^\Lambda,$$

$$0 = \sigma^i \sigma_i \sigma_{\bar{i}} v^\Lambda = [(\Lambda, \epsilon_i + \delta_1) + 2] \sigma_{\bar{i}} v^\Lambda.$$

The first equation above yields, in view of Eq. (A1),

$$(\Lambda, \epsilon_i + \delta_1) = -1, \quad 1 \leqslant i < n, \tag{A2}$$

and substituting into the second equation, we arrive at

$$\sigma_{\bar{i}} v^\Lambda = 0, \quad 1 \leqslant i < n.$$

This last equation in turn implies

$$0 = \sigma^{\bar{i}} \sigma_{\bar{i}} v^\Lambda = (\Lambda, -\epsilon_i + \delta_1) v^\Lambda,$$

so that

$$(\Lambda, -\epsilon_i + \delta_1) = 0, \quad 1 \leqslant i < n. \tag{A3}$$

Equations (A2) and (A3) together yield

$$2(\Lambda, \epsilon_i) = -1, \quad 1 \leqslant i < n,$$

which is impossible, since the components $\lambda_i = (\Lambda, \epsilon_i)$ of $\Lambda = (\lambda|\omega)$ are non-negative integers. Hence we conclude that $C(n+1)$, for $n > 1$, admits no two level irreps.

For the case of $C(2)$, we observe that all typical modules have three levels, so the atypical $C(2)$ modules must all have at most two levels. In view of the typicality conditions of Eq. (2), we see that $V(\Lambda)$ is atypical only in the following cases:

$$\begin{aligned} &\text{(i)} \quad \Lambda = (\tau|-\tau), \quad \tau \in Z^+, \\ &\text{(ii)} \quad \Lambda = (\tau|\tau+2), \quad \tau \in Z^+. \end{aligned} \tag{A4}$$

Note that the first class above corresponds to the two level type (1) * and grade-* irreps of Theorem 3, while the second class above corresponds to the two level type (2) * and grade-* irreps of Proposition 4.

Thus, in conclusion, the (at most) two-level irreps of $C(n+1)$ have highest weights Λ satisfying Eq. (A4) in the case $n = 1$, and for $n > 1$, we have only the case $\Lambda = (\dot{0}|0)$.

APPENDIX B: TYPICAL GRADE-* IRREPS OF $C(2)$

Here we explicitly investigate the irreducible $C(2)$ modules $V(0|\omega)$, $0 < \omega < 2$. We have the Z-gradation

$$V(0|\omega) = V_0(0|\omega) \oplus V_0(1|\omega - 1) \oplus V_0(0|\omega - 2),$$

where the top level is a one-dimensional $\mathrm{sp}(2) \oplus \mathrm{o}(2)$ module with basis vector e_0 satisfying

$$\alpha_j^i e_0 = 0, \quad \Omega e_0 = \omega e_0, \quad i,j = 1,\bar{1}.$$

The next level gives a two-dimensional module with (normalized) basis vectors

$$e_i = \omega^{-1/2} \sigma_i e_0, \quad i = 1,\bar{1},$$

and the bottom level is again one-dimensional with basis vector

$$e_{\bar{0}} = [\omega(2-\omega)]^{-1/2}\sigma_1\sigma_{\bar{1}}e_0.$$

Thus we have a four-dimensional module with basis vectors $e_0, e_1, e_{\bar{1}}, e_{\bar{0}}$ in which $e_0, e_{\bar{0}}$ are regarded as even and $e_1, e_{\bar{1}}$ as odd vectors. The action of the odd generators in this basis is easily seen to be given by

$$\sigma^i e_0 = \sigma_i e_{\bar{0}} = 0, \quad i = 1,\bar{1};$$
$$\sigma^{\bar{1}} e_1 = \sigma^1 e_{\bar{1}} = \sigma_1 e_1 = \sigma_{\bar{1}} e_{\bar{1}} = 0,$$
$$\sigma_i e_0 = \omega^{1/2} e_i, \quad i = 1,\bar{1}; \quad \sigma^1 e_1 = -\omega^{1/2} e_0 = \sigma^{\bar{1}} e_{\bar{1}},$$
$$\sigma_{\bar{1}} e_1 = -\sigma_1 e_{\bar{1}} = -(2-\omega)^{1/2} e_0,$$
$$\sigma^1 e_{\bar{0}} = (2-\omega)^{1/2} e_{\bar{1}}, \quad \sigma^{\bar{1}} e_{\bar{0}} = -(2-\omega)^{1/2} e_1,$$

and for the even generators we obtain

$$h_1 e_1 = e_1, \quad h_1 e_{\bar{1}} = -e_{\bar{1}}, \quad h_1 = \sigma_1^1;$$
$$\Omega e_i = (\omega - 1) e_i, \quad i = 1,\bar{1},$$
$$\sigma_{\bar{1}}^1 e_1 = -2e_{\bar{1}}, \quad \sigma_1^{\bar{1}} e_{\bar{1}} = -2e_1, \quad \sigma_1^{\bar{1}} e_1 = \sigma_{\bar{1}}^1 e_{\bar{1}} = 0,$$
$$\alpha_j^i e_{\bar{0}} = 0, \quad i,j = 1,\bar{1}; \quad \Omega e_{\bar{0}} = (\omega - 2) e_{\bar{0}}.$$

Thus in the basis $e_0, e_1, e_{\bar{1}}, e_{\bar{0}}$ we have the following 4×4 matrix representation:

$$\pi(\sigma^1) = \begin{pmatrix} 0 & -\omega & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (2-\omega) \\ 0 & 0 & 0 & 0 \end{pmatrix},$$
$$\pi(\sigma^{\bar{1}}) = \begin{pmatrix} 0 & -\omega & 0 & 0 \\ 0 & 0 & 0 & -(2-\omega) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$
$$\pi(\sigma_1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \omega & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & (2-\omega) & 0 \end{pmatrix} = -\pi(\sigma^1)^T,$$
$$\pi(\sigma_{\bar{1}}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \omega & 0 & 0 & 0 \\ 0 & -(2-\omega) & 0 & 0 \end{pmatrix} = -\pi(\sigma^{\bar{1}})^T, \tag{B1}$$

where T denotes supertranspose. One may similarly obtain the matrices of the even generators and check directly that the above in fact determines a (real) faithful matrix representation. Such a representation, in view of the $\epsilon = 1$ case of Eq. (11) and Eq. (B1), is indeed irreducible and grade-* of type (2) as noted previously.

Finally, by comparing this matrix representation with the faithful matrix representation of gl(2|1) constructed in I (see Appendix B), we see explicitly that $C(2) \cong$ sl(2|1) and that the typical type (2) grade-* irreps constructed above are isomorphic to the corresponding irreps of sl(2|1) arising from the construction of I, Appendix B.

[1] J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[2] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[3] J. Scherk, Rev. Mod. Phys. **47**, 123 (1975).
[4] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[5] F. Iachello, Phys. Rev. Lett. **44**, 772 (1980).
[6] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[7] Y. Nambu, Physica D **15**, 147 (1985).
[8] A. Montorsi, M. Rasetti, and A. I. Solomon, Phys. Rev. Lett. **59**, 2243 (1987).
[9] V. A. Kostelecky and D. K. Campbell (Editors), Physica D **15**, 3–294 (1985).
[10] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[11] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[12] M. De Crombrugghe and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[13] W. Heidenreich, Phys. Lett. B **110**, 461 (1982).
[14] J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).
[15] J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984); **26**, 913 (1985); **28**, 758 (1987).
[16] M. D. Gould and R. B. Zhang (to be published).
[17] V. G. Kac, *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), p. 597.
[18] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, Berlin, 1972).
[19] V. G. Kac, Adv. in Math. **26**, 8 (1977).
[20] W. Nahm and M. Scheunert, J. Math. Phys. **6**, 868 (1976).
[21] M. D. Gould, J. Phys. A: Math. Gen. **22**, 1209 (1989).
[22] M. D. Gould, A. J. Bracken, and J. W. B. Hughes, J. Phys. A: Math. Gen. **22**, 2879 (1989).
[23] J. W. B. Hughes and R. C. King, J. Phys. A: Math. Gen. **20**, L1047 (1987).
[24] I. N. Bernstein and D. A. Leites, C. R. Acad. Bulgare Sci. **33**, 1049 (1980).
[25] J. Van der Jeugt, J. Phys. A: Math. Gen. **20**, 209 (1987).

Tight frames of compactly supported affine wavelets

Wayne M. Lawton
AWARE, Inc., 124 Mount Auburn Street, Suite 310, Cambridge, Massachusetts 02138

(Received 12 October 1989; accepted for publication 28 March 1990)

This paper extends the class of orthonormal bases of compactly supported wavelets recently constructed by Daubechies [Commun. Pure Appl. Math. **41**, 909 (1988)]. For each integer $N \geqslant 1$, a family of wavelet functions ψ having support $[0,2N-1]$ is constructed such that $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k) | j,k \in Z\}$ is a tight frame of $L^2(R)$, i.e., for every $f \in L^2(R)$, $f = c\Sigma_{jk} \langle \psi_{jk} | f \rangle \psi_{jk}$ for some $c > 0$. This family is parametrized by an algebraic subset V_N of R^{4N}. Furthermore, for $N \geqslant 2$, a proper algebraic subset W_N of V_N is specified such that all points in V_N outside of W_N yield orthonormal bases. The relationship between these tight frames and the theory of group representations and coherent states is discussed.

I. INTRODUCTION

This paper discusses families of functions $h_{a,b}(x) = a^{-1/2}h((x-b)/a), a \neq 0$, called *wavelets*,[1] which are generated from a single function $h \in L^2(R)$ by dilation, translation, and possibly reflection. Wavelets provide a means of representing functions as a continuous linear superposition that is analogous to the Fourier transform. This representation is related to group representations. Let G denote the *affine group* $\{x \to ax + b \,|\, a \neq 0\}$ and let U denote the unitary representation of G on $L^2(R)$ defined by $U(a,b)h = h_{a,b}$. In this framework wavelets are subsets of $L^2(R)$ having the form $D(h) = \{h_{a,b} | (a,b) \in D\}$ for some subset D of G.

Let G_s denote either G or the connected affine subgroup $G_c = \{(a,b), a > 0\}$ and let $L^2(G_s)$ denote the Hilbert space of measurable functions on G_s that are square integrable with respect to the left-invariant Haar measure $d\mu(a,b) = a^{-2}\,da\,db$ on G_s. A function $h \in L^2(R)$ is *admissible* if $h \neq 0$ and if

$$\int_{G_s} d\mu(a,b) |\langle h_{a,b}, h \rangle|^2 = c_h < \infty.$$

This condition is independent of the choice of G_s and is equivalent to the condition $\int dy |y|^{-1} |\hat{h}(y)|^2 < \infty$, where $\hat{h}(y) = \int dx\, h(x) \exp(-2\pi i x y)$ is the Fourier transform of h. The following result is a consequence of standard results in the theory of unitary group representations:[2] If $G_s = G$ or $G_s = G_c$ and $h \in L^2(R)$ is admissible, then

$$c_h^{-1} \int_{G_s} d\mu(a,b) (U_h f)(a,b) h_{a,b} = P(f), \quad \text{for all } f \in L^2(R), \tag{1}$$

where $P(f)$ denotes the orthogonal projection of f onto the closed subspace H spanned by the wavelets $G_s(h)$.

The wavelets $G_s(h)$ constitute a set of *coherent states* for the action of G_s on H if h is admissible and if H is irreducible under the action of G_s. Note that $L^2(R)$ is irreducible under the action of G but not irreducible under the action of G_c since the subspaces H^+ and H^- of $L^2(R)$ consisting of functions whose Fourier transforms are supported on the set of positive, negative real numbers that are closed and invariant under G_c. However, these subspaces are irreducible under the action of G_c. Therefore, wavelets generated by admissible functions are generalizations of coherent states. Coherent states have extensive applications in physics.[3]

Computational considerations suggest the utility of representing functions by expansions involving discrete wavelets. A *tight frame* of a Hilbert space H is a countable subset $\{h_\alpha\}$ of vectors such that for every $f \in H$, $f = c\Sigma_\alpha \langle h_\alpha | f \rangle h_\alpha$ for some $c > 0$. Therefore, the discrete wavelets $\{h_{a,b} | (a,b) \in D\}$ form a *tight frame* if:

$$f = c \sum_{(a,b) \in D} \langle h_{a,b} | f \rangle h_{a,b}, \quad \text{for every } f \in L^2(R). \tag{2}$$

This relation is a discrete analog of relation (1) in which $P(f) = f$. Daubechies *et al.*[4] constructed tight frames for $L^2(R)$ that consist of discrete subsets $D(h)$ of coherent states for the action of G. In their construction, the set D is a lattice in G having the form $L_{\alpha,\beta} = \{(\pm \exp(\alpha n), \beta m \exp(\alpha n)) | m,n \in Z\}$ for $\beta \neq 0$. The function h is an admissible bandlimited function contained in either H^+ or in H^- satisfying additional restrictions. Since D contains reflections, it is not a subset of G_c and therefore $D(h)$ is not a set of wavelets for G_c. Furthermore, h does not have compact support.

The first bases consisting of wavelets having compact support were orthonormal bases for which $D = \{(2^{-j}, 2^{-j}k) | j,k \in Z\} = L_{\alpha,\beta} \cap G_c$ where $\alpha = \ln(2)$ and $\beta = 1$. In 1910, Haar[5] constructed the following basis for $L^2(R)$: $\{2^{j/2}\psi(2^j x - k) | j,k \in Z\}$ where $\psi = 1$ for $[0,\frac{1}{2}]$, $\psi = -1$ on $[\frac{1}{2},1)$, and $\psi = 0$ on $(-\infty,0) \cup [1,\infty)$. In 1988, Daubechies[6] constructed, for each integer $N \geqslant 2$, a function ψ_N satisfying: (i) the support of ψ_N has length $2N-1$; (ii) the $0,\ldots,N-1$ moments of ψ_N equal 0; (iii) each ψ_N is continuous and its regularity increases indefinitely as N increases; (iv) $\{2^{j/2}\psi_N(2^j x - k) | j,k \in Z\}$ is an orthonormal wavelet basis for $L^2(R)$. Daubechies constructs each wavelet ψ_N from $2N$ parameters, called *scaling parameters*, which are constrained by a set of N linear and N quadratic equations. These constraints are necessary and sufficient to ensure that all four properties above are satisfied. However, they imply that there exist at most 2^N different wavelets and, furthermore, all these wavelets have identical Fourier moduli. In this paper we prove that if $N-1$ of the

 0022-2488/90/081898-04$03.00

linear constraints are removed, then Daubechies' construction results in tight frames of compactly supported wavelets. Furthermore it is proved that for "almost all" choices of scaling parameters, the resulting frame is in fact an orthonormal basis.

II. STATEMENT OF RESULTS

The Haar wavelet function can be constructed as $\psi(x) = \phi(2x) - \phi(2x-1)$, where ϕ is the characteristic function of the set $[0,1)$. The function ϕ satisfies and is uniquely defined in $L^2(R)$, up to normalization, by the scaling relation $\phi(x) = \phi(2x) + \phi(2x-1)$. This approach is generalized below.

Step 1. Choose an integer N and construct a sequence of complex numbers $\{C_0,...,C_{2N-1}\}$, called scaling parameters, that satisfy the following two equations:

$$\sum C_m C^*_{m+2n} = 2\delta_{0n}, \quad \text{for all } n, \tag{3}$$

$$\sum C_m = 2. \tag{4}$$

Here, $C_m = 0$ for $m < 0$ and for $m \geqslant 2N$, * denotes complex conjugation, and $\delta_{00} = 1$ and $\delta_{0n} = 0$ for $n \neq 0$. Let V_N denote the set of all such sequences. Then V_N is an algebraic subset of $C^N = R^{4N}$, $V_1 = \{(1\ 1)\}$, and V_N is uncountably infinite for $N \geqslant 2$.

Step 2: Construct a scaling function $\phi \in L^2(R)$ that satisfies the scaling relation

$$\phi(x) = \sum C_m \phi(2x-m). \tag{5}$$

Step 3: Construct the wavelet function $\psi \in L^2(R)$ by

$$\Psi(x) = \sum (-1)^m C^*_{1-m} \phi(2x-m). \tag{6}$$

The main objective of this paper is to prove the following three results.

Theorem 1: If $\{C_0,...,C_{2N-1}\} \in V_N$ and ϕ^0 denotes the characteristic function of $[0,1)$, then the sequence ϕ^j defined by

$$\phi^{j+1}(x) = \sum C_m \phi^j(2x-m), \tag{7}$$

converges weakly in $L^2(R)$ to a function ϕ. Furthermore, ϕ is the unique function that satisfies (5) and for which $\int \phi = 1$. Also, support $(\phi) = [0,2N-1]$.

Theorem 2: If $\{C_0,...,C_{2N-1}\} \in V_N$, $\phi \in L^2(R)$ satisfies (5), and ψ is constructed by (6), then:

(i) $\{\phi(x-m) \mid m \in Z\}$ forms a partition of unity, i.e., for almost all x, $\Sigma m \phi(x-m) = 1$,

(ii) $\{2^{-j/2}\psi(2^j x - k) \mid j,k \in Z\}$ is a tight frame of $L^2(R)$; in particular,

$$f(y) = \sum_{j,k} \langle 2^{-j/2}\psi(2^j x - k) \mid f \rangle 2^{-j/2}\psi 2^j y - k),$$

$$\text{for all } f \in L^2(R).$$

The next result expresses the fact that "most" of the tight frames constructed above are orthonormal bases. Fix an integer $N > 0$, let L_N denote the $4N-3$ dimensional subspace of $L^2(Z)$ consisting of all sequences $\{B_m\}$ such that $B_m = 0$ if $m < -2N+2$ or if $m > 2N-2$. For any $v = \{C_0,...,C_{2N-1}\} \in V_N$, construct the following:

(i) a linear mapping $S_v : L_N \to L_N$ defined by

$$S_v(A)_m = \sum_n \left[\frac{1}{2}\tilde{C} * C\right]_n A_{2m-n}, \tag{8}$$

where $\tilde{C}(k) = C^*(-k)$ for all k and $\tilde{C} * C$ denotes the convolution of $\tilde{C}$ and C;

(ii) a sequence $A_v \in L_N$ by

$$A_{v,m} = \int \phi^*_v(x)\phi_v(x-m)dx, \tag{9}$$

where $\phi_v \in L^2(R)$ is the unique function, constructed as in Theorem 1, that satisfies Eq. (5) and the condition $\int \phi_v = 1$,

(iii) a subset

$$W_N = \{v = \{C_0,...,C_{2N-1}\} \in V_N \mid DP_v(1) = 0\}, \tag{10}$$

where DP_v denotes the derivative of the characteristic polynomial P_v of S_v. The following result shows that "most," but for $N \geqslant 2$ not all, of the tight frames constructed above are orthonormal bases.

Theorem 3: For all $v \in V_N$, $S_v(\delta_{0m}) = \delta_{0m}$ and $S_v(A_v) = A_v$. The subset of W_N of V_N is a proper algebraic subset of V_N. For $N \geqslant 2$, $v = \{1,0,...,0,1\} \in V_N$, $v \in W_N$, and the set $\{\psi_{jk} \mid j,k \in Z\}$ is not orthonormal. If $v \in V_N$ and $v \notin W_N$ then $\{\psi_{jk} \mid j,k \in Z\}$ is an orthonormal basis for $L^2(R)$.

III. DERIVATIONS

Several intermediate results are required for the proofs.

Lemma 1: The sequence $\{\phi^j\}$ in Theorem 1 converges to a distribution ϕ whose support is $[0,2N-1]$.

Proof: The Fourier transforms Φ^j of ϕ^j satisfy $\Phi^{j+1}(y) = \Phi^j(y/2)\ m_0(y/2)$ for $j \geqslant 1$ and $\Phi^0(y) = \exp(\pi i y)\mathrm{sinc}(\pi y)$, where $m_0(y) = \frac{1}{2}\Sigma_n C_n \exp(2\pi i n y)$. Equation (4) implies $m_0(y) = 1$, therefore the sequence Φ^j converges uniformly to the function $\Phi(y) = \Pi_{j>1} m_0(2^{-j}y)$ on compact subsets of R. Furthermore, for $|y| \geqslant 1$, each Φ^j satisfies the inequality $|\Phi^j(y)| \leqslant CB^{1+\log_2(y)}$, where $C = \max\{|\Phi(y)|, |y| \leqslant 1\}$ and $B = \max\{|m_0(y)|\}$. Thus Φ^j and Φ are bounded by polynomial growth and $\{\phi^j\}$ converges to a distribution ϕ such that $\hat{\phi} = \Phi$. Clearly, $\int_\phi = \Phi(0) = 1$. Also, $\text{support}(\phi) = \lim \text{support}(\phi^j) = \lim[0,(N-1)(1-2^{-j})] = [0,N-1]$.

Lemma 2: For all $j \geqslant 0$, $\langle \phi^j \mid \phi^j \rangle = 1$. The distribution ϕ is in $L^2(R)$.

Proof: Clearly for any integer n, $\langle \phi^0(x) \mid \phi^0(x-n) \rangle = \delta_{0n}$. By (7), (3) and induction

$$\langle \phi^{j+1}(x) \mid \phi^{j+1}(x-n) \rangle$$

$$= \langle \sum_m C_m \phi^j(2x-m) \mid \sum_k C_k \phi^j(2x-2n-k) \rangle$$

$$= \sum_m \sum_k C^*_m C_k \langle \phi^j(2x-m) \mid \phi^{j+1}(2x-2n-k) \rangle$$

$$= \frac{1}{2} \sum_m C^*_m C_{m-2k} = \delta_{0n}.$$

Therefore, since the unit ball in $L^2(R)$ is weakly compact, any subsequence $\{g_k\}$ of $\{\phi^j\}$ has at least one accumulation

point, say g, in the weak topology. Then there exists a subsequence of $\{g_k\}$ that converges weakly to g. Since the weak topology on $L^2(R)$ is stronger than the distribution topology, $g=\phi$. Therefore, $\{\phi^j\}$ converges weakly to $\phi \in L^2(R)$.

Proof of Theorem 1: It suffices, from Lemmas 1 and 2, to demonstrate the uniqueness of ϕ. Assume ϕ is any distribution satisfying the scaling recursion (7) and the condition $\int\phi = 1$. The Fourier transform satisfies $\hat{\phi}(y) = \Phi(y)\hat{\phi}^{(0)} = \Phi(y)$ where Φ is the function represented by the infinite product expansion in Lemma 2. This concludes the proof.

Lemma 3: The set $\{\phi(x-m)\,|\,m\in Z\}$ forms a partition of unity.

Proof: Let $\{\phi^j\}$ be the sequence in Theorem 1. Since $\phi = \lim \phi^j$ it suffices to prove $\{\phi^j(x-m)\,|\,m\in Z\}$ forms a partition of unity for all $j=1$. For $j=1$ this holds since ϕ^0 = characteristic function of $[0,1)$. The result for $j\geqslant 1$ follows by induction.

For any vectors g and h in a Hilbert space H let $|g\rangle\langle h|$ denote the operator, called the *outer product* of g and h, defined by $|g\rangle\langle h|(f) = \langle h|f\rangle g$. Therefore, a subset $\{h_\alpha\}$ of H is a tight frame if and only if $\Sigma_\alpha |h_\alpha\rangle\langle h_\alpha| = I$, where I denotes the identity operator on H and the limit is in the weak sense.

Lemma 4: Let ϕ be a scaling function with support $[0,2N-1]$ and for any integer J let I_J denote the operator $\Sigma_m |2^{J/2}\phi(2^Jx-m)\rangle\langle 2^{J/2}\phi(2^Jx-m)|$. Then limit $I_J = I$.

Proof: The set of operators $\{I_J\}$ is uniformly bounded since for any $f\in L^2(R)$, the orthonormality of $\{2^{J/2}\phi(2^Jy-n(2N-1))/\|\phi\|\,|\,n\in Z\}$ and Bessel's inequality[7] imply

$$\|I_J(f)\| = \Big\|\sum_{n=0}^{n=2N-2}\sum_{m=n \bmod(2N-1)}\langle 2^{J/2}\phi(2^Jy-m)|f(y)\rangle \times 2^{J/2}\phi(2^Jx-m)\Big\| \leqslant (2N-1)\|\phi\|^2\|f\|.$$

Therefore, it suffices to prove that limit $I_J(f) = f$ for all f in a dense subset of $L^2(R)$. Assume f is continuous and has compact support. Then if $m\to\infty$ and $J\to\infty$ such that $2^{-J}m\to x$, $\lim 2^J\langle\phi(2^Jy-m)|f(y)\rangle = f(x)$ uniformly in x. The result now follows since, by Lemma 3, $\{\phi(2^Jy-m)\,|\,m\in Z\}$ forms a partition of unity.

Lemma 5: Let ϕ and ψ denote the scaling function and wavelet corresponding to a $\{C_0,\ldots,C_{2N-1}\}\in V_N$. For any integer J define the operators I_J as in Lemma 4 and $F_J = \Sigma_m |2^{J/2}\psi(2^Jx-m)\rangle\langle 2^{J/2}\psi(2^Jx-m)|$. Then for all J, $I_J + F_J = I_{J+1}$.

Proof: Substitute the expressions for ϕ and ψ in Eqs. (5) and (6) to obtain

$$\begin{aligned} I_J + F_J &= \frac{1}{2}\sum_p C_p^* \sum_q C_q \sum_m |2^{(J+1)/2}\phi(2^{J+1}x-2m-p)\rangle\langle 2^{(J+1)/2}\phi(2^{J+1}x-2m-q)| \\ &\quad + \frac{1}{2}\sum_p (-1)^p C_{1-p}\sum_q (-1)^q C_{1-q}^* \sum_m |2^{(J+1)/2}\phi(2^{J+1}x-2m-p)\rangle\langle 2^{(J+1)/2}\phi(2^{J+1}x-2m-q)| \\ &= \sum_i\sum_k C(i,k)|2^{(J+1)/2}\phi(2^{J+1}x-i)\rangle\langle 2^{(J+1)/2}\phi(2^{J+1}x-k)|, \end{aligned}$$

where

$$C(i,k) = \frac{1}{2}\sum_m [C_{2m+i}^* C_{2m+k} + (-1)^{2m+i} \times C_{1-2m-i}(-1)^{2m+k}C_{1-2m-k}^*].$$

It suffices to prove that $C(i,k) = \delta_{ik}$. Assume i is even and k is odd. Then replace m by $m - i/2$ in the first expression and replace m by $-m-(1-k)/2$ in the second expression (within brackets) above to obtain

$$C(i,k) = \frac{1}{2}\sum_m [C_{2m}^* C_{2m-i+k} - C_{2m-i+k}C_{2m}^*] = 0.$$

If i is odd and k is even it follows similarly that $C(i,k)=0$. Now assume i is even and k is even. Then replace m by $m-k/2$, respectively $-m-i/2$ in the first and second expression above to obtain

$$C(i,k) = \frac{1}{2}\sum_m [C_{2m-k+i}^* C_{2m} + C_{1+2m}C_{1+2m-k+i}^*].$$

Since $2m$ ranges over all even integers and $1+2m$ ranges over all odd integers, this expression is equivalent to $\frac{1}{2}\Sigma_m C_{m-k+i}^* C_{2m}$. Equation (3) implies this expression equals δ_{ik} because $-k+1$ is even. Finally, if both i and k are odd, replace m by $-m-(i+k)/2$ in the second expression to obtain

$$\frac{1}{2}\sum_m [C_{2m+i}^* C_{2m+k} + C_{1+2m+k}C_{1+2m+i}^*] = \delta_{ik}.$$

Proof of Theorem 2: The first conclusion follows from Lemma 3. The second conclusion follows from Lemma 4 and the repeated application of Lemma 5 to obtain

$$I = \lim_{J\to\infty} I_J = \lim_{J\to\infty} F_{J-1} + F_{J-2} + \cdots + F_{-J} + I_{-J} = \sum_j F_j.$$

Proof of Theorem 3: Let $N\geqslant 1$ be an integer, let $v = \{C_0,\ldots,C_{2N-1}\}\in V_N$, and let S_v denote the operator defined by (8). Then $S_v(\delta_{0m}) = \delta_{0m}$ follows from (3) and

$S_v(A_v) = A_v$ is obtained by substituting the scaling relation (5) into the integrand in Eq. (9). Clearly, $DP_v(1)$ is a polynomial in terms of the scaling parameters $\{C_m\}$ and their complex conjugates. Therefore, W_N is an algebraic subset of V_N. Also, $W_N \neq V_N$ because the scaling parameters constructed by Daubechies are in V_N and not in W_N, therefore W_N is a proper algebraic subset of V_N. Clearly, for $N \geqslant 2$, $v = \{1,0,...,0,1\} \in W_N$ and also the set $\{\psi_{jk} | j,k \in Z\}$ is not orthonormal. Assume $\{\psi_{jk} | j,k \in Z\}$ corresponding to $v \in V_N$ is not an orthonormal basis for $L^2(R)$. It suffices to show $v \in W_N$. Consider the sequence A_v defined by Eq. (9). If A_v is a linear multiple of the sequence δ_{0m} then the functions $\{\phi(x-m) | m \in Z\}$ are orthogonal. However, by construction $\int_\phi = 1$, and by Lemma 3, $\{\phi(x-m) | m \in Z\}$ forms a partition of unity, therefore,

$$A_v(0) = \int \phi^*(x)\phi(x)dx = \int \phi^*(x) \sum_k \phi(x-k)dx$$

$$= \int \phi^*(x)dx = 1.$$

Then Eq. (8), describing the construction of the wavelet ψ in terms of ϕ, implies the functions $\{\psi_{jk} | j,k \in Z\}$ are orthogonal. Since by Theorem 2, this set is a also a quasiorthonormal basis for $L^2(R)$,

$$\psi_{jk} = \sum_{m,n} \langle \psi_{mn} | \psi_{jk} \rangle \psi_{mn} = \langle \psi_{jk} | \psi_{jk} \rangle \psi_{jk},$$

for all j, $k \in Z$, hence $\{\Psi_{jk} | j,k \in Z\}$ is an orthonormal basis. Since this contradicts the original assumption, it follows that A_v is not a linear multiple of δ_{0m}. Therefore, A_v and δ_{0m} are linearly independent eigenvectors in L_N, having eigenvalue 1, for the linear operator S_v defined by Eq. (8). Therefore the characteristic polynomial P_v of S_v has the root 1 occur with multiplicity $\geqslant 2$. Hence its derivative DP_v satisfies $DP_v(1) = 0$ and $v \in W_N$. The proof is complete.

Remark: It follows from Pollen's factorization theorem[8] for unitary matrices over the ring of Laurent polynomials that V_N is an irreducible algebraic set. Therefore, the subset W_N of V_N has zero measure.

Open Problem: Does every $v \in W_N$ give rise to a nonorthonormal basis?

ACKNOWLEDGMENTS

The author is indebted to Ingrid Daubechies for permission to include her proof of Theorem 1, communicated to me in a letter of April 1989, in this paper and for encouragement to publish these results. Special thanks are given to my colleages at AWARE, Inc., in particular to Ramesh Gopinath, who uncovered several errors in earlier versions of the manuscript and provided helpful editorial comments; to David Linden, who discovered the tight frame properties of wavelet bases through numerical experiments prior to theoretical investigations; to David Pollen, who elucidated the detailed structure of the parameter set for wavelets; and to Howard Resnikoff, who first suggested the examination of the eigenvalues of linear operators derived from the scaling relation as used in the proof of Theorem 2.

This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F49620-89-C-0125. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

[1] A. Grossman and J. Morlet, SIAM J. Math. Anal. **15**, 723 (1984).
[2] A. Grossman and J. Morlet, J. Math. Phys. **26**, 2473 (1985).
[3] J. R. Klauder and B. S. Skagerstam, *Coherent States. Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1984).
[4] I. Daubechies, A. Grossman, and Y. Meyer, J. Math. Phys. **27**, 1271 (1985).
[5] Alfred Haar, Math. Ann. **69**, 336 (1910).
[6] I. Daubechies, Commun. Pure Appl. Math. **41**, 909 (1988).
[7] G. Dahlquist and A. Bjorck, *Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1979).
[8] D. Pollen, "$SU_I(2,F[z,1/z])$ for F a subfield of C," J. Am. Math. Soc. (to be published).

Exact and steady-state solutions to sinusoidally excited, half-infinite chains of harmonic oscillators with one isotopic defect

E. L. Mokole
Naval Research Laboratory, Radar Division, Washington, DC 20375

A. L. Mullikin
Georgia Tech Research Institute, Atlanta, Georgia 30332

M. B. Sledd[a)]
Stone Mountain, Georgia 30083

(Received 14 November 1989; accepted for publication 28 March 1990)

A half-infinite chain of spring-mass oscillators with nearest-neighbor coupling is excited by a sinusoidal force applied to the mass at the accessible end of the chain. The identical linear springs are massless. Each mass has measure $m(m>0)$ except one, which has measure $\mu m(\mu>0)$. An exact solution is given for the initial-value problem in which all initial velocities and displacements are zero. Behavior of the solution for large t (time) is examined. When the concept of average power supplied by the source to the chain in steady state is meaningful, expressions for average power are deduced.

I. INTRODUCTION

A half-infinite chain of harmonic oscillators is considered in which each mass is coupled to its nearest neighbors by massless linear springs (Fig. 1). The chain is called half-infinite because it extends indefinitely to the right only. The leftmost mass, which is acted upon by an applied sinusoidal force $F\sin(\omega t)$, is not connected to the inertial frame. The supporting plane is horizontal and frictionless. All the springs are alike with linear spring constant $k(k>0)$; and each mass has measure $m(m>0)$ except one, which has measure $\mu m(\mu>0)$. The masses are numbered 0,1,2,... from the left; and the anomolous mass, which is thought of as an isotopic defect,[1,2] is the $(N-1)$th mass. The displacement of the nth mass from its equilibrium position is represented by x_n. Note that any configuration of the masses in which each spring has its natural length is *an* equilibrium position; and from all such configurations, an arbitrary one may be selected and called *the* equilibrium configuration.

In Sec. II the general case of a single isotopic defect in an arbitrary position is discussed. An initial-value problem is considered in which all velocities and displacements are zero at time $t=0$. A solution of this problem is defined to be a sequence $\{x_n\}_{n=0}^{\infty}$ of twice-differentiable, real-valued functions of t that satisfy the prescribed initial conditions and reduce the differential equations to identities in t for $t>0$. Such a solution is unique if each x_n $(n\geqslant 0)$ is required to be analytic (that is, expansible in a power series). The initial-value problem is solved by giving an integral representation[3] of x_n in terms of a suitable set of orthogonal polynomials and a generalized weight function (distribution). The behavior of the displacement x_n is analyzed for large t, and the results are summarized in Tables II and III. When the concept of average power supplied to the chain by the source is meaningful, the average power is calculated.

For given m, k, and F, the other parameters of the problem are the angular frequency (ω) of the exciting force and the location (N) and seriousness (μ) of the defect. In analyzing the solution, it is convenient to consider ten cases, which are determined essentially by the relation of ω to the natural frequency $\sqrt{4k/m}$ and of μ to a critical defect factor μ_c that will be defined subsequently. The various possibilities are enumerated in Table I.

The remaining sections consider special cases or extensions of the work undertaken in Sec. II. In Sec. III the case of a single isotopic defect in the leftmost position is discussed in greater detail by using the results of Sec. II for $N=1$. Results for the uniform half-infinite chain with no defect are obtained (Sec. IV) by setting $\mu=1$ in the appropriate locations in Sec. II. Section V examines the effects of the change in phase when the exciting force $F\sin(\omega t)$ is replaced by $F\sin(\omega t-\phi)$.

II. THE UNIFORM CHAIN WITH A SINGLE ISOTOPIC DEFECT IN AN ARBITRARY POSITION

The equations of motion of the half-infinite uniform chain with a single isotopic defect in an arbitrary position are

$$\mu m\ddot{x}_0 = k(x_1 - x_0) + F\sin(\omega t), \tag{1a}$$

$$m\ddot{x}_n = k(x_{n+1} - 2x_n + x_{n-1}), \quad n\geqslant 1, \tag{1b}$$

when $N=1$, or when $N\geqslant 2$,

$$m\ddot{x}_0 = k(x_1 - x_0) + F\sin(\omega t), \tag{2a}$$

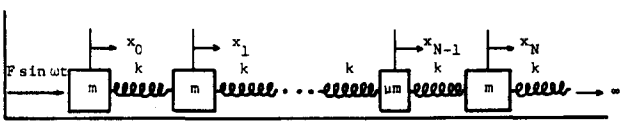


FIG. 1. A uniform half-infinite chain with a single isotopic defect.

[a)] Dr. Marvin B. Sledd passed away on 30 October 1988.

 0022-2488/90/081902-12$03.00

$$\mu m\ddot{x}_{N-1} = k(x_N - 2x_{N-1} + x_{N-2}), \tag{2b}$$

$$m\ddot{x}_n = k(x_{n+1} - 2x_n + x_{n-1}),$$

$$1 \leqslant n \leqslant N-2 \text{ and } n \geqslant N, \tag{2c}$$

where $x_n(N,t)$ represents the displacement of the nth mass from its equilibrium position and the dots above a variable signify differentiation with respect to time t. To Eqs. (1) and (2) the initial conditions

$$x_n(N,0) = 0, \quad \dot{x}_n(N,0) = 0, \quad n \geqslant 0, \tag{3}$$

are adjoined for arbitrary $N \geqslant 1$.

It has been shown[4] that, for fixed $N \geqslant 1$, each component $x_n(N,t)$ of the solution of the systems (1),(3) and (2),(3) can be written in the form

$$x_n(N,t) = \frac{1}{m_0} \int_{-\infty}^{+\infty} W(y,t;0) P_n(N,y) d\alpha(N,y), \quad n \geqslant 0, \tag{4}$$

where m_0 is the measure of the leftmost mass ($m_0 = \mu m$ when $N = 1$ and $m_0 = m$ when $N \geqslant 2$). Then, $W(y,t;0)$ is the solution of the differential system

$$W_{tt}(y,t;0) + yW(y,t;0) = F\sin(\omega t),$$

$$W(y,0;0) = 0, \quad W_t(y,0;0) = 0. \tag{5}$$

Thus

$$W(y,t;0) = \begin{cases} \dfrac{F[\omega \sin(\sqrt{y}t) - \sqrt{y}\sin(\omega t)]}{\sqrt{y}(\omega^2 - y)}, & \text{if } y \neq \omega^2, \\ \dfrac{F[\sin(\omega t) - \omega t \cos(\omega t)]}{2\omega^2}, & \text{if } y = \omega^2 \end{cases} \tag{6}$$

Notice that $W(y,t;0)$ is a continuous function of y for $y \geqslant 0$ provided $W(0,t;0)$ is defined to be $F[\omega t - \sin(\omega t)]/\omega^2$. The sequence $\{P_n(N,y)\}_{n=0}^{\infty}$ consists of polynomials that satisfy the recurrence

$$P_0(1,y) = 1,$$

$$P_1(1,y) = 1 - m\mu y k^{-1}, \tag{7}$$

$$P_{n+1}(1,y) = (2 - myk^{-1})P_n(1,y) - P_{n-1}(1,y), \quad n \geqslant 1,$$

when $N = 1$ or the recurrence

$$P_0(N,y) = 1,$$

$$P_1(N,y) = 1 - myk^{-1},$$

$$P_N(N,y) = (2 - m\mu y k^{-1})P_{N-1}(N,y) - P_{N-2}(N,y),$$

$$P_n(N,y) = (2 - myk^{-1})P_{n-1}(N,y) - P_{n-2}(N,y),$$

$$1 \leqslant n \leqslant N-2 \text{ and } n \geqslant N, \tag{8}$$

when $N \geqslant 2$. Each system may be solved using standard techniques for difference equations.[5,6] A convenient form for this solution is obtained by making the change of variable $2\cos\beta = -2 + myk^{-1}$ when $0 < y < 4k/m$ or $2\cosh\psi = -2 + myk^{-1}$ when $y > 4k/m$. This yields

$$P_n(N,y) = \begin{cases} (-1)^n \dfrac{\sin(n+\frac{1}{2})\beta}{\sin(\frac{1}{2}\beta)}, & \text{if } 0 \leqslant n \leqslant N-1, \\ (-1)^n \dfrac{\sin(n+\frac{1}{2})\beta}{\sin(\frac{1}{2}\beta)} - (-1)^n 2(1-\mu)(1+\cos\beta) \dfrac{\sin(n-N+1)\beta}{\sin\beta} \dfrac{\sin(N-\frac{1}{2})\beta}{\sin(\frac{1}{2}\beta)}, & \text{if } N \leqslant n, \end{cases} \tag{9}$$

for $0 < y < 4k/m$, or

$$P_n(N,y) = \begin{cases} (-1)^n \dfrac{\sinh(n+\frac{1}{2})\psi}{\sinh(\frac{1}{2}\psi)}, & \text{if } 0 \leqslant n \leqslant N-1, \\ (-1)^n \dfrac{\sinh(n+\frac{1}{2})\psi}{\sinh(\frac{1}{2}\psi)} - (-1)^n 2(1-\mu)(1+\cosh\psi) \dfrac{\sinh(n-N+1)\psi}{\sinh\psi} \dfrac{\sinh(N-\frac{1}{2})\psi}{\sinh(\frac{1}{2}\psi)}, & \text{if } N \leqslant n, \end{cases} \tag{10}$$

for $y > 4k/m$. By applying a known test,[7] it can be verified that these polynomials are orthogonal with respect to some weight function (distribution). It can also be shown that these polynomials are a special case of a class of polynomials for which McKibben[8] has described a method for constructing the appropriate distribution. In addition to an absolutely continuous part, the distribution may contain a singular part represented by a *Dirac delta*.[9] The location and weight of this distribution are determined by certain properties of a polynomial $P(N,z)$ and by the distribution corresponding to a set of orthogonal polynomials associated with the polynomials $P_n(N,y)$. As a practical matter, it is not always possible to carry out McKibben's procedure. However, in the present case, the distribution is given by

$$d\alpha(N,y)=\begin{cases}\left[\dfrac{m_0\sqrt{(4k-my)/my}\,I_{[0,4k/m]}(y)}{2\pi kT_1(N,1-my/(2k))}+(1/m)J(x_n)\delta(y-y_N)\right]dy, & \text{if } 0<\mu<(4N-3)/(4N-2),\\ \left[\dfrac{m_0\sqrt{(4k-my)/my}\,I_{[0,4k/m]}(y)}{2\pi kT_1(N,1-my/(2k))}\right]dy, & \text{if } (4N-3)/(4N-2)\leqslant\mu,\end{cases} \tag{11}$$

where T_1 is defined by

$$T_1(N,x)\triangleq\left[(\mu-1)^2\left|z^{2N-1}-2\sum_{l=1}^{2N-2}(-z)^l+\frac{\mu}{1-\mu}\right|^2\right]\Bigg|_{z=e^{i\theta}}$$

with $x=\cos\theta=1-(my)/(2k)$. Here, $I_{[0,4k/m]}(y)$ is the indicator function of the interval $0\leqslant y\leqslant 4k/m$. The numbers y_N and x_N needed to determine the location and strength of the Dirac delta in (11) are given by the relation

$$y_N=\frac{2k}{m}(1-x_N)=-\frac{k}{m}\frac{(z_N-1)^2}{z_n}, \tag{12}$$

where z_N is the single zero in the interval $(-1,0)$ of the $(2N)$th-degree polynomial

$$P(N,z)=(z-1)(\mu-1)\times\left[z^{2N-1}-2\sum_{l=1}^{2N-2}(-z)^l+\frac{\mu}{1-\mu}\right], \tag{13}$$

the summation being defined to be zero when $N=1$. The strength of the Dirac delta is

$$\frac{J(x_N)}{m}=\frac{(z_N-z_N^{-1})^2}{mz_N(dP(N,z_N)/dz)P(N,z_N^{-1})}. \tag{14}$$

More detail on the preceding calculations can be found in Mokole;[10] in particular, the function T_1 is the function S_1 on page 97 of this reference.

Observe that the presence or absence of a Dirac delta in the distribution $d\alpha(N,y)$ depends on the relation of the parameter μ (seriousness of the single isotopic defect) to the critical value $\mu_c(N)=(4N-3)/(4N-2)$. This special value of μ is obtained from (13) by setting $P(N,-1)$ to zero. That the Dirac delta is present when $\mu<\mu_c(N)$ and absent when $\mu\geqslant\mu_c(N)$ increases the number of cases that must be analyzed.

A major objective of the analysis is to derive useful information about $x_n(N,t)$ for large t. Since Eq. (4) is not very well adapted to this end when all frequency ranges are considered, the expression for $x_n(N,t)$ is now reformulated.

The form in which $x_n(N,t)$ is to be recast is suggested by past experience with similarly excited finite frictionless chains. Except when resonance is involved, the displacements in such systems are sinusoids with the possible addition of an affine term. So a representation of $x_n(N,t)$, $n\geqslant 0$, of the form

$$x_n(N,t)=\begin{cases}A_n(N,t)+C_n(N,t)\cos(\omega t)+S_n(N,t)\sin(\omega t)\\ \quad+\dfrac{1}{m}J(x_N)P_n(N,y_N)W(y_N,t;0), & \text{if } 0<\mu<\mu_c(N)\\ A_n(N,t)+C_n(N,t)\cos(\omega t)+S_n(N,t)\sin(\omega t), & \text{if } \mu_c(N)\leqslant\mu\end{cases} \tag{15}$$

is chosen, where $A_n(N,t)$, $C_n(N,t)$, and $S_n(N,t)$ have finite limits as t approaches positive infinity. The expressions,

$$m^{-1}J(x_N)P_n(N,y_N)W(y_N,t;0) \text{ and } A_n(N,t)+C_n(N,t)\cos(\omega t)+S_n(N,t)\sin(\omega t),$$

are obtained respectively from the Dirac delta and the absolutely continuous parts of the generalized weight function appearing in (11). One means for arriving at (15) from (4) is first to rewrite (6) as

$$W(y,t;0)=\frac{F\sin(\sqrt{y}t)}{\sqrt{y}(\omega+\sqrt{y})}-\frac{F\sin[\frac{1}{2}(\sqrt{y}-\omega)t]}{\frac{1}{2}(\sqrt{y}-\omega)}\frac{\cos[\frac{1}{2}(\sqrt{y}+\omega)t]}{(\sqrt{y}+\omega)}, \tag{16}$$

with the understanding that $W(\omega^2,t;0)=\lim_{y\to\omega^2}W(y,t;0)$. The substitution of (16) and (11) into (4), followed by the changes of variable $v=\sqrt{y}$ and $u=v-\omega$, leads to Eq. (15) with

$$A_n(N,t)=\frac{F}{\pi}\int_0^{\sqrt{4k/m}}\frac{P_n(N,v^2)\sqrt{(4k/m)-v^2}}{(\omega+v)kT_1(N,1-mv^2 2^{-1}k^{-1})}\frac{\sin(vt)}{v}\,dv, \tag{17a}$$

$$C_n(N,t) = -\frac{F}{\pi}\int_{-\omega}^{\sqrt{4k/m}-\omega} \frac{P_n(N,(u+\omega)^2)\sqrt{(4k/m)-(u+\omega)^2}}{(u+2\omega)kT_1(N,1-m(u+\omega)^2 2^{-1}k^{-1})}\frac{\sin(ut)}{u}\,du, \tag{17b}$$

$$S_n(N,t) = \frac{F}{\pi}\int_{-\omega}^{\sqrt{4k/m}-\omega} \frac{P_n(N,(u+\omega)^2)\sqrt{(4k/m)-(u+\omega)^2}}{(u+2\omega)kT_1(N,1-m(u+\omega)^2 2^{-1}k^{-1})}\frac{1-\cos(ut)}{u}\,du. \tag{17c}$$

With the following definitions:

$$f(N,v) = \frac{F\sqrt{(4k/m)-v^2}}{\pi k(v+\omega)T_1(N,1-mv^2 2^{-1}k^{-1})}, \tag{18a}$$

$$g(N,u) = f(N,u+\omega), \tag{18b}$$

$$f_n(N,v) = P_n(N,v^2)f(N,v), \tag{18c}$$

$$g_n(N,u) = f_n(N,u+\omega) = P_n(N,(u+\omega)^2)g(N,u), \tag{18d}$$

Eqs. (17) become

$$A_n(N,t) = \int_0^{\sqrt{4k/m}} f_n(N,v)\frac{\sin(vt)}{v}\,dv, \tag{19a}$$

$$C_n(N,t) = \int_{-\omega}^{\sqrt{4k/m}-\omega} -g_n(N,u)\frac{\sin(ut)}{u}\,du, \tag{19b}$$

$$S_n(N,t) = \int_{-\omega}^{\sqrt{4k/m}-\omega} g_n(N,u)\frac{1-\cos(ut)}{u}\,du. \tag{19c}$$

Observe that in the integrals for $C_n(N,t)$ and $S_n(N,t)$ in (19) the upper limits of integration are positive, zero, or negative depending on whether the angular frequency ω is less than, equal to, or greater than $\sqrt{4k/m}$. Also note that the form that $x_n(N,t)$ takes in (15) depends on whether the physical parameter μ is less than, equal to, or greater than the critical value $\mu_c(N) = (4N-3)/(4N-2)$. Thus, the ten cases listed in Table I are considered.

Because case 5 involves both a special frequency and a special value of μ, it is handled separately. The remaining nine cases are now considered collectively. By using the Riemann–Lebesgue lemma and similar concepts, it can be shown[11] that for each $n \geqslant 0$ the functions $A_n(N,t)$, $C_n(N,t)$, and $S_n(N,t)$ in (19) have finite limits as t approaches positive infinity. Call these limits $a_n(N)$, $c_n(N)$, and $s_n(N)$, respectively. Then, for each fixed $N \geqslant 1$ and for each $n \geqslant 0$, the function

$$x_{n,\mathrm{ss}}(N,t) = \begin{cases} a_n(N) + c_n(N)\cos(\omega t) + s_n(N)\sin(\omega t) + (1/m)J(x_N)P_n(N,y_N)W(y_N,t;0), & \text{if } 0<\mu<\mu_c(N), \\ a_n(N) + c_n(N)\cos(\omega t) + s_n(N)\sin(\omega t), & \text{if } \mu_c(N)\leqslant\mu, \end{cases} \tag{20}$$

represents $x_n(N,t)$ for large t. In particular, $\lim_{t\to+\infty}[x_n(N,t) - x_{n,\mathrm{ss}}(N,t)] = 0$ for each $n \geqslant 0$. The portion of $x_n(N,t)$ that arises from the absolutely continuous part of the distribution $d\alpha$ is written in the form $A_n(N,t) + C_n(N,t)\cos(\omega t) + S_n(N,t)\sin(\omega t)$, where $A_n(N,t)$, $C_n(N,t)$, and $S_n(N,t)$ have finite limits as t approaches infinity. Since trigonometric identities may be used to arrive at different representations for the coefficients A_n, C_n, and S_n, it is unclear whether $a_n(N)$, $c_n(N)$, and $s_n(N)$ are unique. This issue is resolved by the following theorem.

Theorem: Suppose that $\omega > 0$ and that a real-valued function f of a real variable can be represented for all real t as

$$f(t) = A(t) + C(t)\cos(\omega t) + S(t)\sin(\omega t),$$

and as

$$f(t) = \bar{A}(t) + \bar{C}(t)\cos(\omega t) + \bar{S}(t)\sin(\omega t).$$

Also suppose the real-valued functions $A(t)$, $C(t)$, $S(t)$, $\bar{A}(t)$, $\bar{C}(t)$, and $\bar{S}(t)$ have the properties

$$\lim_{t\to+\infty}A(t) = a, \quad \lim_{t\to+\infty}C(t) = c,$$
$$\lim_{t\to+\infty}S(t) = s,$$
$$\lim_{t\to+\infty}\bar{A}(t) = \bar{a}, \quad \lim_{t\to+\infty}\bar{C}(t) = \bar{c},$$
$$\lim_{t\to+\infty}\bar{S}(t) = \bar{s},$$

for the finite real numbers a, c, s, $\bar{a}$, $\bar{c}$, $\bar{s}$. Then $a = \bar{a}$, $c = \bar{c}$, and $s = \bar{s}$.

Argument: It is sufficient to prove that if $\lim_{t\to+\infty}F(t) = f$, $\lim_{t\to+\infty}G(t) = g$, and $\lim_{t\to+\infty}H(t) = h$ and if $F(t) + G(t)\cos(\omega t) + H(t)\sin(\omega t) = 0$ for $t > 0$ and $\omega > 0$, then $f = g = h = 0$. Let $t_n = 2n\pi/\omega$ for $n = 1,2,\ldots,$

TABLE I. Enumeration of the ten cases to be considered by the ranges of the parameters μ and ω.

	$0<\omega^2<4k/m$	$\omega^2 = 4k/m$	$4k/m<\omega^2$
$\mu_c(N)<\mu$	Case(1)	Case(2)	Case(3)
$\mu=\mu_c(N)$	Case(4)	Case(5)	Case(6)
$0<\mu<\mu_c(N)$	Case(7)	Case(8)	Case(9)
			$\omega^2 = y_N$
			Case(10)
			$\omega^2 \neq y_N$

then $\lim_{n\to\infty}[F(t_n)+G(t_n)\cos(\omega t_n)+H(t_n)\sin(\omega t_n)]=f+g=0$. Now let $\tau_n=(2n+1)\pi/\omega$ for $n=1,2,\ldots$, then $\lim_{n\to\infty}[F(\tau_n)+G(\tau_n)\cos(\omega\tau_n)+H(\tau_n)\sin(\omega\tau_n)]=f-g=0$. Hence $f=g=0$. The proof of $h=0$ is similar.

Thus, although the choice of $A_n(N,t)$, $C_n(N,t)$, and $S_n(N,t)$ in (15) is not unique, it follows by the preceding theorem that every representation of the exact solution $\{x_n(N,t)\}_{n=0}^{\infty}$ in the form (15) leads to the unique result (20) provided only that $A_n(N,t)$, $C_n(N,t)$, and $S_n(N,t)$ have finite limits as t approaches positive infinity.

To evaluate $\lim_{t\to+\infty}A_n(N,t)=a_n(N)$, consider (17a). It can be shown that $T_1(N,[2k-mv^2]/[2k])$ is nonzero on the interval of integration except when $\mu=\mu_c(N)$. In this case the ratio

$$\sqrt{[4k-mv^2]/m}/T_1(N,[2k-mv^2]/[2k])$$

has an absolutely (improperly) integrable singularity at $v=\sqrt{4k/m}$. So for all choices of μ, the limit is given by $a_n(N)=[FP_n(N,0)]/[\omega T_1(N,1)\sqrt{mk}]$, $n\geqslant 0$. Since $P_n(N,0)=\mu$ and $T_1(N,1)=1$, $a_n(N)=F/[\omega\sqrt{mk}]$ for $n\geqslant 0$.

By using similar techniques, intricate algebraic arguments, and various limit theorems like the Riemann–Lebesgue lemma, the limits $c_n(N)$ and $s_n(N)$, $n\geqslant 0$ and $N\geqslant 1$, can be calculated.[12] An indirect approach which provides the same results and is considerably easier to execute consists of the following steps.

(i) First the sequence $\{x_{n,\mathrm{ss}}(N,t)\}_{n=0}^{\infty}$ is shown to satisfy the differential equations (1) when $N=1$ or the differential equations (2) when $N\geqslant 2$.[13]

(ii) It is observed that if an expression for $x_{0,\mathrm{ss}}(N,t)$ is known, the remaining $x_{n,\mathrm{ss}}(N,t)$, $n\geqslant 1$, are uniquely determined by the differential equations.

(iii) By using the differential equations and standard techniques of difference equations, the coefficients $c_n(N)$ and $s_n(N)$, $n\geqslant 1$, can be expressed in terms of $c_0(N)$ and $s_0(N)$, which are recorded in Table II for all cases except

TABLE II. Coefficients $c_0(N)$ and $s_0(N)$ of Eq. (20). A horizontal bar through an integral denotes the Cauchy principal value.

Case	$c_0(N)$	$s_0(N)$
(1)	$\frac{-F\sqrt{(4k/m)-\omega^2}}{2k\omega T_1(N,1-m\omega^2 2^{-1}k^{-1})}$	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$
(2)	0	$\int_0^{\omega}\frac{-F}{\pi k\sqrt{\omega^2-v^2}T_1(N,1-2v^2\omega^{-2})}dv$
(3)	0	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$
(4)	$\frac{-F\sqrt{(4k/m)-\omega^2}}{2k\omega T_1(N,1-m\omega^2 2^{-1}k^{-1})}$	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$
(6)	0	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$
(7)	$\frac{-F\sqrt{(4k/m)-\omega^2}}{2k\omega T_1(N,1-m\omega^2 2^{-1}k^{-1})}$	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$
(8)	0	$\int_0^{\omega}\frac{-F}{\pi k\sqrt{\omega^2-v^2}T_1(N,1-2v^2\omega^{-2})}dv$
(9)	0	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$
(10)	0	$\int_0^{\sqrt{4k/m}}\frac{F\sqrt{(4k/m)-v^2}}{\pi k(v^2-\omega^2)T_1(N,1-mv^2 2^{-1}k^{-1})}dv$

case (5). In addition, the expressions for $c_n(N)$ and $s_n(N)$ are listed in Table III.

For given values of N, μ, and ω^2, evaluation of $c_0(N)$ involves only algebraic operations, and $s_0(N)$ is represented by an integral that can be evaluated exactly for $N=1$. For $N>1$, the integral expression for $s_0(N)$ has not been determined explicitly, but various numerical techniques can be used to approximate its value.

In case (5) [where $\mu=\mu_c(N)$ and $\omega^2=4k/m$], the displacement $x_n(N,t)$ for large t is treated somewhat differently because $C_n(N,t)$ and $S_n(N,t)$ do not have finite limits as t approaches positive infinity. However, since $t^{-1/2}C_n(N,t)$ and $t^{-1/2}S_n(N,t)$ have finite *nonzero* limits as t approaches infinity,[14] rewrite $x_n(N,t)$, $n\geqslant 0$, in the form

$$x_n(N,t)=A_n(N,t)+t^{1/2}\overline{C}_n(N,t)\cos(\omega t)+t^{1/2}\overline{S}_n(N,t)\sin(\omega t), \tag{21}$$

where

$$\overline{C}_n(N,t)=t^{-1/2}C_n(N,t) \text{ and } \overline{S}_n(N,t)=t^{-1/2}S_n(N,t). \tag{22}$$

In this instance the behavior of $x_n(N,t)$ for large t is represented by

$$x_{n,\mathrm{ss}}(N,t)=a_n(N)+\bar{c}_n(N)t^{1/2}\cos(\omega t)+\bar{s}_n(N)t^{1/2}\sin(\omega t), \tag{23}$$

where

$$a_n(N)=\lim_{t\to+\infty}A_n(N,t),\quad \bar{c}_n(N)=\lim_{t\to+\infty}\overline{C}_n(N,t),\quad \bar{s}_n(N)=\lim_{t\to+\infty}\overline{S}_n(N,t). \tag{24}$$

By an argument not very different from the one used to prove the previous theorem, it can be shown that although $A_n(N,t)$, $\overline{C}_n(N,t)$, and $\overline{S}_n(N,t)$ are not unique, the result $x_{n,\mathrm{ss}}(N,t)$ is unique provided that $A_n(N,t)$, $\overline{C}_n(N,t)$, and $\overline{S}_n(N,t)$ have finite limits as t approaches infinity. Calculating the appropriate limits yields

$$x_{n,\mathrm{ss}}(N,t)=\begin{cases}\dfrac{F}{2k}-\dfrac{F(-1)^n}{2k}\sqrt{\dfrac{\omega}{\pi}}\dfrac{(2n+1)}{(2N-1)^2}\sqrt{t}\cos(\omega t)-\dfrac{F(-1)^n}{2k}\sqrt{\dfrac{\omega}{\pi}}\dfrac{(2n+1)}{(2N-1)^2}\sqrt{t}\sin(\omega t), & \text{if } 0\leqslant n\leqslant N-1,\\ \dfrac{F}{2k}-\dfrac{F(-1)^n}{2k}\sqrt{\dfrac{\omega}{\pi}}\dfrac{1}{2N-1}\sqrt{t}\cos(\omega t)-\dfrac{F(-1)^n}{2k}\sqrt{\dfrac{\omega}{\pi}}\dfrac{1}{2N-1}\sqrt{t}\sin(\omega t), & \text{if } n\geqslant N.\end{cases} \tag{25}$$

TABLE III. Coefficients $a_n(N)$, $c_n(N)$, and $s_n(N)$ of Eq. (20) in terms of $a_0(N)$, $c_0(N)$, and $s_0(N)$. $a_n(N)=a_0(N)=F/[\omega\sqrt{mk}]$ for all cases, $\theta=\arcsin\sqrt{m\omega^2/4k}$, $2\cosh\psi=-2+m\omega^2/k$, and $2\cosh\psi_N=-2+my_N/k$.

Case	$c_n(N)$ [If $c_0(N)=0$, $c_n(N)=0$ for all n]	$s_n(N)$
(1), $0\leqslant n\leqslant N-1$	$c_0(N)\dfrac{\cos(2n+1)\theta}{\cos\theta}$	$s_0(N)\dfrac{\cos(2n+1)\theta}{\cos\theta}-\dfrac{F}{k}\dfrac{\sin 2n\theta}{\sin 2\theta}$
(1), $n\geqslant N$	$c_0(N)\left[\dfrac{\cos(2n+1)\theta}{\cos\theta}+2(1-\mu)(1-\cos 2\theta)\dfrac{\cos(2N-1)\theta}{\cos\theta}\dfrac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right]$	$s_0(N)\left[\dfrac{\cos(2n+1)\theta}{\cos\theta}+2(1-\mu)(1-\cos 2\theta)\dfrac{\cos(2N-1)\theta}{\cos\theta}\dfrac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right]-\dfrac{F}{k}\left[\dfrac{\sin 2n\theta}{\sin 2\theta}+2(1-\mu)(1-\cos 2\theta)\dfrac{\sin(2N-2)\theta}{\sin 2\theta}\dfrac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right]$
(2), $0\leqslant n\leqslant N-1$	0	$(-1)^n\left[s_0(N)(2n+1)+\dfrac{F}{k}n\right]$
(2), $n\geqslant N$	0	$(-1)^n\left[s_0(N)(2n+1)+\dfrac{F}{k}n\right]-4(1-\mu)(n-N+1)(-1)^n\left[s_0(N)(2N-1)+\dfrac{F}{k}(N-1)\right]$
(3), $0\leqslant n\leqslant N-1$	0	$(-1)^n\left[s_0(N)\dfrac{\sinh(n+1/2)\psi}{\sinh\psi/2}+\dfrac{F}{k}\dfrac{\sinh n\psi}{\sinh\psi}\right]$
(3), $n\geqslant N$	0	$(-1)^n\left[s_0(N)\dfrac{\sinh(n+1/2)\psi}{\sinh\psi/2}+\dfrac{F}{k}\dfrac{\sinh n\psi}{\sinh\psi}\right]-2(-1)^n(1-\mu)(1+\cosh\psi)\dfrac{\sinh(n-N+1)\psi}{\sinh\psi}\cdot\left[s_0(N)\dfrac{\sinh(N-1/2)\psi}{\sinh\psi/2}+\dfrac{F}{k}\dfrac{\sinh(N-1)\psi}{\sinh\psi}\right]$
(4), $0\leqslant n\leqslant N-1$	$c_0(N)\dfrac{\cos(2n+1)\theta}{\cos\theta}$	$s_0(N)\dfrac{\cos(2n+1)\theta}{\cos\theta}-\dfrac{F}{k}\dfrac{\sin 2n\theta}{\sin 2\theta}$

TABLE III. (*Continued.*)

Case	$c_n(N)$ [If $c_0(N)=0$, $c_n(N)=0$ for all n]	$s_n(N)$
(4), $n>N$	$c_0(N)\left[\frac{\cos(2n+1)\theta}{\cos\theta} + \frac{2(1-\cos 2\theta)}{4N-2}\,\frac{\cos(2N-1)\theta}{\cos\theta}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right]$	$s_0(N)\left[\frac{\cos(2n+1)\theta}{\cos\theta} + \frac{2(1-\cos 2\theta)}{4N-2}\,\frac{\cos(2N-1)\theta}{\cos\theta}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right] - \frac{F}{k}\left[\frac{\sin 2n\theta}{\sin 2\theta} + \frac{2(1-\cos 2\theta)}{4N-2}\,\frac{\sin(2N-2)\theta}{\sin 2\theta}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right]$
(6), $0\le n\le N-1$	0	$(-1)^n\left[s_0(N)\,\frac{\sinh(n+1/2)\psi}{\sinh\psi/2} + \frac{F}{k}\,\frac{\sinh n\psi}{\sinh\psi}\right]$
(6), $n>N$	0	$(-1)^n\left[s_0(N)\,\frac{\sinh(n+1/2)\psi}{\sinh\psi/2} + \frac{F}{k}\,\frac{\sinh n\psi}{\sinh\psi}\right] - \frac{2(-1)^n(1+\cosh\psi)}{4N-2}\,\frac{\sinh(n-N+1)\psi}{\sinh\psi} \cdot\left[s_0(N)\,\frac{\sinh(N-1/2)\psi}{\sinh\psi/2} + \frac{F}{k}\,\frac{\sinh(N-1)\psi}{\sinh\psi}\right]$
(7), $0\le n\le N-1$	$c_0(N)\,\frac{\cos(2n+1)\theta}{\cos\theta}$	$\left[s_0(N)\,\frac{\cos(2n+1)\theta}{\cos\theta} - \frac{F}{k}\,\frac{\sin 2n\theta}{\sin 2\theta}\right] + \frac{FJ(x_N)}{k}\,\frac{\left[\frac{\cos(2n+1)\theta}{\cos\theta} + (-1)^{n+1}\,\frac{\sinh(n+1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N+\cos 2\theta)}$
(7), $n>N$	$c_0(N)\left[\frac{\cos(2n+1)\theta}{\cos\theta} + 2(1-\mu)(1-\cos 2\theta)\,\frac{\cos(2N-1)}{\cos\theta}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right]$	$s_0(N)\left[\frac{\cos(2n+1)\theta}{\cos\theta} + 2(1-\mu)(1-\cos 2\theta)\,\frac{\cos(2N-1)\theta}{\cos\theta}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right] - \frac{F}{k}\left[\frac{\sin 2n\theta}{\sin 2\theta} + 2(1-\mu)(1-\cos 2\theta)\,\frac{\sin(2N-2)\theta}{\sin 2\theta}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\right] + \frac{FJ(x_N)}{k}\,\frac{\left[\frac{\cos(2n+1)\theta}{\cos\theta} + (-1)^{n+1}\,\frac{\sinh(n+1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N+\cos 2\theta)} + \frac{FJ(x_N)}{k}\,\frac{2(1-\mu)(1-\cos 2\theta)}{2(\cosh\psi_N+\cos 2\theta)}\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta} \cdot\left[\frac{\cos(2N-1)\theta}{\cos\theta} + (-1)^N\,\frac{\sinh(N-1/2)\psi_N}{\sinh\psi_N/2}\right] + \frac{FJ(x_N)}{k}\,(1-\mu)(-1)^N\,\frac{\sin(2n-2N+2)\theta}{\sin 2\theta}\,\frac{\sinh(N-1/2)\psi_N}{\sinh\psi_N/2}$

Case	$c_n(N)$ [If $c_0(N)=0$, $c_n(N)=0$ for all n]	$s_n(N)$
(8), $0\le n\le N-1$	0	$(-1)^n\left[s_0(N)(2n+1)+\frac{F}{k}n\right] + \frac{FJ(x_N)}{k}(-1)^n\,\frac{\left[(2n+1) - \frac{\sinh(n+1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N-1)}$
(8), $n>N$	0	$(-1)^n\left[s_0(N)(2n+1)+\frac{F}{k}n\right] - 4(1-\mu)(n-N+1)(-1)^n\left[s_0(N)(2N-1)+\frac{F}{k}(N-1)\right] + \frac{FJ(x_N)}{k}(-1)^n\,\frac{\left[(2n+1) - \frac{\sinh(n+1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N-1)} - \frac{FJ(x_N)}{k}\,4(1-\mu)(n-N+1)(-1)^n\,\frac{\left[(2N-1) - \frac{\sinh(N-1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N-1)} + \frac{FJ(x_N)}{k}(1-\mu)(n-N+1)(-1)^n\,\frac{\sinh(N-1/2)\psi_N}{\sinh\psi_N/2}$
(9), $0\le n\le N-1$	0	$(-1)^n\left[s_0(N)\,\frac{\sinh(n+1/2)\psi}{\sinh\psi/2} + \frac{F}{k}\,\frac{\sinh n\psi}{\sinh\psi}\right] + \frac{FJ(x_N)}{k}(-1)^n\,\frac{\left[\frac{\sinh(n+1/2)\psi}{\sinh\psi/2} - \frac{\sinh(n+1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N-\cosh\psi)}$
(9), $n>N$	0	$(-1)^n\left[s_0(N)\,\frac{\sinh(n+1/2)\psi}{\sinh\psi/2} + \frac{F}{k}\,\frac{\sinh n\psi}{\sinh\psi}\right] - 2(-1)^n(1-\mu)(1+\cosh\psi)\,\frac{\sinh(n-N+1)\psi}{\sinh\psi} \cdot\left[s_0(N)\,\frac{\sinh(N-1/2)\psi}{\sinh\psi/2} + \frac{F}{k}\,\frac{\sinh(N-1)\psi}{\sinh\psi}\right] + \frac{FJ(x_N)}{k}(-1)^n\,\frac{\left[\frac{\sinh(n+1/2)\psi}{\sinh\psi/2} - \frac{\sinh(n+1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N-\cosh\psi)} - \frac{FJ(x_N)}{k}(-1)^n 2(1-\mu)(1+\cosh\psi)\,\frac{\sinh(n-N+1)\psi}{\sinh\psi} \cdot\frac{\left[\frac{\sinh(N-1/2)\psi}{\sinh\psi/2} - \frac{\sinh(N-1/2)\psi_N}{\sinh\psi_N/2}\right]}{2(\cosh\psi_N-\cosh\psi)} + \frac{FJ(x_N)}{k}(-1)^n(1-\mu)\,\frac{\sinh(n-N+1)\psi}{\sinh\psi}\,\frac{\sinh(N-1/2)\psi_N}{\sinh\psi_N/2}$

TABLE III. (*Continued.*)

Case	$c_n(N)$ [If $c_0(N)=0$, $c_n(N)=0$ for all n]	$s_n(N)$
(10), $0 \leqslant n \leqslant N-1$	0	$(-1)^n\left[s_0(N)\frac{\sinh(n+1/2)\psi}{\sinh\psi/2}+\frac{F}{k}\frac{\sinh n\psi}{\sinh\psi}\right]$ $+\frac{FJ(x_N)}{k}(-1)^n\frac{\left[\frac{\sinh(n+1/2)\psi}{\sinh\psi/2}-(2n+1)\frac{\cosh(n+1/2)\psi}{\cosh\psi/2}\right]}{4(\cosh\psi-1)}$
(10), $n \geqslant N$	0	$(-1)^n\left[s_0(N)\frac{\sinh(n+1/2)\psi}{\sinh\psi/2}+\frac{F}{k}\frac{\sinh n\psi}{\sinh\psi}\right]$ $-2(-1)^n(1-\mu)(1+\cosh\psi)\frac{\sinh(n-N+1)\psi}{\sinh\psi}$ $\cdot\left[s_0(N)\frac{\sinh(N-1/2)\psi}{\sinh\psi/2}+\frac{F}{k}\frac{\sinh(N-1)\psi}{\sinh\psi}\right]$ $+\frac{FJ(x_N)}{k}(-1)^n\frac{\left[\frac{\sinh(n+1/2)\psi}{\sinh\psi/2}-(2n+1)\frac{\cosh(n+1/2)\psi}{\cosh\psi/2}\right]}{4(\cosh\psi-1)}$ $-\frac{FJ(x_N)}{k}2(-1)^n(1-\mu)(1+\cosh\psi)\frac{\sinh(n-N+1)\psi}{\sinh\psi}$ $\cdot\frac{\left[\frac{\sinh(N-1/2)\psi}{\sinh\psi/2}-(2N-1)\frac{\cosh(N-1/2)\psi}{\cosh\psi/2}\right]}{4(\cosh\psi-1)}$ $+\frac{FJ(x_N)}{k}(-1)^n(1-\mu)\frac{\sinh(n-N+1)\psi}{\sinh\psi}\frac{\sinh(N-1/2)\psi}{\sinh\psi/2}$

In determining the behavior of $x_n(N,t)$ for large t in all cases except case (5), the fact that $\{x_{n,\mathrm{ss}}(N,t)\}_{n=0}^{\infty}$ satisfies the differential equations (1) or (2) is used. For case (5), $x_{n,\mathrm{ss}}(N,t)$, which has a $\sqrt{t}$ dependence [see Eq. (25)], fails to satisfy Eqs. (1a) and (2a) by a term of order t^0 and fails to satisfy the remaining differential equations by a term of order $1/\sqrt{t}$. So in case (5), the expression for $x_{n,\mathrm{ss}}(N,t)$ is *not* obtained by beginning with $x_{0,\mathrm{ss}}(N,t)$ and using the successive differential equations; rather it is obtained for each n by calculating directly the limits of $A_n(N,t)$, $\bar{C}_n(N,t)$, and $\bar{S}_n(N,t)$ as t approaches infinity. Another feature which distinguishes case (5) from the other nine cases is that the behavior of $x_n(N,t)$ for large t implied by (25) does not agree with the usual concept of *steady state* but is more like undamped resonance.

The remaining question to be considered in this section is: For large values of t, what is the *steady-state* average power supplied by the source to the chain? The concept of steady-state average power is meaningful if the expression (the integral is the energy transmitted to the system over one cycle of the applied force),

$$\frac{\omega}{2\pi}\int_T^{T+2\pi/\omega}\dot{x}_0(N,t)F\sin(\omega t)dt, \tag{26}$$

has a finite limit as the positive number T approaches infinity. When this limit exists, it is by definition the steady-state average power $P_{\mathrm{av,ss}}(N)$. Thus,

$$P_{\mathrm{av,ss}}(N)=\lim_{T\to\infty}\frac{\omega}{2\pi}\int_T^{T+2\pi/\omega}\dot{x}_0(N,t)F\sin(\omega t)dt. \tag{27}$$

Intuitively, one might expect the expressions

$$\tilde{P}_{\mathrm{av,ss}}(N)=\frac{\omega}{2\pi}\int_T^{T+2\pi/\omega}\frac{d}{dt}\left[x_{0,\mathrm{ss}}(N,t)\right]F\sin(\omega t)dt \tag{28}$$

and

$$\hat{P}_{\mathrm{av,ss}}(N)=\frac{\omega}{2\pi}\int_T^{T+2\pi/\omega}\left[\dot{x}_0(N,t)\right]_{\mathrm{ss}}F\sin(\omega t)dt \tag{29}$$

to be related, at least in some cases, to $P_{\mathrm{av,ss}}(N)$. The steady-state velocity (in brackets) is obtained in the same way that the steady-state displacements of Eq. (20) are derived. It can be shown[15] that whenever $P_{\mathrm{av,ss}}(N)$ is meaningful,

$$\tilde{P}_{\mathrm{av,ss}}(N)=\hat{P}_{\mathrm{av,ss}}(N)=P_{\mathrm{av,ss}}(N).$$

The expressions for $\tilde{P}_{\mathrm{av,ss}}(N)$ and $\hat{P}_{\mathrm{av,ss}}(N)$ are frequently more amenable to evaluation than the expression for $P_{\mathrm{av,ss}}(N)$.

If $\mu>\mu_c(N)$, the steady-state average power supplied by the source to the chain is a meaningful concept for all ranges of the angular frequency ω. Its value is

$$P_{\mathrm{av,ss}}(N)=\begin{cases}\dfrac{F^2\sqrt{(4k/m)-\omega^2}}{4kT_1(N,1-m\omega^2 2^{-1}k^{-1})}, & \text{if } 0<\omega^2<4k/m,\\ 0, & \text{if } 4k/m\leqslant\omega^2.\end{cases} \tag{30}$$

In this case $x_{0,\mathrm{ss}}(N,t)=a_0(N)+c_0(N)\cos(\omega t)+s_0(N)\sin(\omega t)$. When $c_0(N)\neq 0$, the *dephasing term* $c_0(N)\cos(\omega t)$ introduces a phase shift in the oscillatory part of $x_{0,\mathrm{ss}}(N,t)$ relative to the applied force $F\sin(\omega t)$. Recall from cases (2) and (3) of Table II that the *dephasing term* vanishes for $\omega^2\geqslant 4k/m$; and as is indicated by (30), no power is absorbed by the chain for high frequencies. Therefore one might think of the half-infinite chain as acting like a perfect filter for angular frequencies greater than $\sqrt{4k/m}$. But for each frequency less than $\sqrt{4k/m}$ [see case (1) of

Table II] the chain receives a positive[16] steady-state average power from the source; and consequently energy is absorbed from the source even though the chain is frictionless and in steady-state motion. Any finite frictionless chain receiving a positive average power could not display steady-state motion. The difference in the properties of the half-infinite and the finite frictionless chains is that in the half-infinite chain the same steady-state average power that is supplied by the source to the leftmost mass is also transmitted from any mass to the mass to its right; and thus, in a sense, energy is transmitted down the half-infinite chain and never returns, which is an impossibility for the finite frictionless chain.

If $\mu = \mu_c(N)$, the concept of steady-state average power supplied by the source to the chain is meaningful for all values of the frequency except $\omega = \sqrt{4k/m}$. It is not meaningful in this special case because the limit in (27) is positive infinity. So, for $\mu = \mu_c(N)$ and $\omega \neq \sqrt{4k/m}$, $P_{\text{av,ss}}(N)$ is also given by Eq. (30).

If $0 < \mu < \mu_c(N)$, the steady-state average power supplied by the source to the chain is meaningful if and only if $\omega = \omega_n \triangleq \sqrt{y_N}/n (n = 2,3,4,\ldots)$; that is, ω is a subharmonic of the frequency $\sqrt{y_N}$. The average power for these discrete values of ω is also given by (30) with ω replaced by ω_n.

III. THE UNIFORM CHAIN WITH A SINGLE ISOTOPIC DEFECT IN THE LEFTMOST POSITION

The special case $N = 1$ is now considered. This example is chosen because it illustrates the results of the previous section and because some of the computational difficulties encountered for large N are avoided. In this case,

(i) $T_1(1,x)$ is a first-degree polynomial in x, and the integral present in $s_0(1)$ can be evaluated exactly in each of the ten cases except case (5);

(ii) the expression $y_1 = k/[m\mu(1-\mu)]$ determines y_1 exactly [see Eqs. (12) and (13)];

(iii) it is relatively easy to derive information about $P_{\text{av,ss}}(1)$ [graphs of $P_{\text{av,ss}}(1)$ as a function of ω for several values of μ are shown].

First, the behavior of $x_n(1,t)$ for large t as represented by $x_{n,\text{ss}}(1,t)$ for all ten cases is summarized in Table IV.

To obtain expressions for the *steady-state* average power supplied by the source to the chain for the three ranges of the physical parameter μ, set $N = 1$ in (30). Since

$$T_1(1, 1 - m\omega^2 2^{-1} k^{-1}) = 1 + \mu(\mu - 1) m\omega^2 k^{-1},$$

the expressions that describe the behavior of $P_{\text{av,ss}}(1)$ as a function of the angular frequency ω for fixed values of μ are simple.

If $\frac{1}{2} \triangleq \mu_c(1) < \mu$,

$$P_{\text{av,ss}}(1) = \begin{cases} \dfrac{F^2\sqrt{(4k/m) - \omega^2}}{4[k + \mu(\mu-1)m\omega^2]}, & \text{if } 0 < \omega^2 < 4k/m, \\ 0, & \text{if } 4k/m \leqslant \omega^2. \end{cases} \tag{31}$$

Graphs of $P_{\text{av,ss}}(1)$ as a function of ω for several values of $\mu (\mu > 1/2)$ are shown in Fig. 2. All of the graphs for

TABLE IV. Summary of the expressions for $x_{n,\text{ss}}(1,t)$, $n \geq 0$. $x_{n,\text{ss}}(1,t) = a_n(1) + q_n(1,t)\cos(\omega t) + r_n(1,t)\sin(\omega t) + k^{-1}(1-2\mu)y_1 \times [\mu/(\mu-1)]^n W(y_1,t;0)$, $a_n(1) = F/[\omega\sqrt{mk}]$ for all cases, and $\theta = \arcsin\sqrt{m\omega^2/4k}$.

Case	$q_n(1,t)$	$r_n(1,t)$	$\frac{1-2\mu}{k} y_1 \left(\frac{\mu}{\mu-1}\right)^n W(y_1,t;0)$
(1)	$\frac{-F}{\omega\sqrt{mk}} \frac{\mu\cos(2n+1)\theta + (1-\mu)\cos(2n-1)\theta}{1+\mu(\mu-1)\frac{m\omega^2}{k}}$	$\frac{-F}{\omega\sqrt{mk}} \frac{\mu\sin(2n+1)\theta + (1-\mu)\sin(2n-1)\theta}{1+\mu(\mu-1)\frac{m\omega^2}{k}}$	(No Such Term Appears)
(2)	0	$\frac{-F}{2k} \frac{(-1)^n}{2\mu-1}$	(No Such Term Appears)
(3)	0	$\frac{-F(-1)^n}{\omega\sqrt{mk}} \left[\frac{(4k/m)^{1/2}}{\omega + (\omega^2 - \frac{4k}{m})^{1/2}}\right]^{2n+1} \frac{m[\omega^2(\mu-1) + \frac{2k}{m} + \omega(\mu-1)(\omega^2 - \frac{4k}{m})^{1/2}]}{2[k + \mu(\mu-1)m\omega^2]}$	(No Such Term Appears)
(4)	$\frac{-F}{\omega\sqrt{mk}} \frac{\cos(2n+1)\theta + \cos(2n-1)\theta}{2(1 - \frac{m\omega^2}{4k})}$	$\frac{-F}{\omega\sqrt{mk}} \frac{\sin(2n+1)\theta + \sin(2n-1)\theta}{2(1 - m\omega^2/4k)}$	(No Such Term Appears)
(5)	$\frac{-F}{2k}(-1)^n(\omega/\pi)^{1/2} t^{1/2}$	$\frac{-F}{2k}(-1)^n(\omega/\pi)^{1/2} t^{1/2}$	(No Such Term Appears)
(6)	0	$\frac{-F(-1)^n}{\omega\sqrt{mk}} \left[\frac{(4k/m)^{1/2}}{\omega + (\omega^2 - \frac{4k}{m})^{1/2}}\right]^{2n+1} \frac{\omega + (\omega^2 - \frac{4k}{m})^{1/2}}{(\omega^2 - \frac{4k}{m})^{1/2}}$	(No Such Term Appears)
(7)	$\frac{-Fy_1}{\omega\sqrt{mk}} \frac{\mu\cos(2n+1)\theta + (1-\mu)\cos(2n-1)\theta}{y_1 - \omega^2}$	$\frac{-Fy_1}{\omega\sqrt{mk}} \frac{\mu\sin(2n+1)\theta + (1-\mu)\sin(2n-1)\theta}{y_1 - \omega^2} - \left(\frac{\mu}{\mu-1}\right)^n \frac{Fy_1(1-2\mu)}{k(y_1 - \omega^2)}$	$\left(\frac{\mu}{\mu-1}\right)^n \frac{F\sqrt{y_1}(1-2\mu)}{k(y_1-\omega^2)} (\sqrt{y_1}\sin\omega t - \omega\sin\sqrt{y_1}\,t)$
(8)	0	$\frac{F}{2k(1-2\mu)} \left[(-1)^n - 2\left(\frac{\mu}{\mu-1}\right)^n\right]$	$\left(\frac{\mu}{\mu-1}\right)^n \frac{F}{k\sqrt{y_1}(1-2\mu)} (\sqrt{y_1}\sin\omega t - \omega\sin\sqrt{y_1}\,t)$
(9)	0	$\frac{-F(-1)^n y_1[\omega(\mu-1)(\omega + (\omega^2 - \frac{4k}{m})^{1/2}) + \frac{2k}{m}]}{\omega(\frac{4k}{m})^{1/2}k(y_1 - \omega^2)} \left[\frac{(4k/m)^{1/2}}{\omega + (\omega^2 - \frac{4k}{m})^{1/2}}\right]^{2n+1} - \left(\frac{\mu}{\mu-1}\right)^n \frac{Fy_1(1-2\mu)}{k(y_1-\omega^2)}$	$\left(\frac{\mu}{\mu-1}\right)^n \frac{F\sqrt{y_1}(1-2\mu)}{k(y_1-\omega^2)} (\sqrt{y_1}\sin\omega t - \omega\sin\sqrt{y_1}\,t)$
(10)	0	$\frac{-F}{k}\left(\frac{\mu}{\mu-1}\right)^n \left[\frac{\mu(1-\mu)}{1-2\mu} - n\right]$	$\left(\frac{\mu}{\mu-1}\right)^n \frac{F(1-2\mu)}{2k} (\sin\omega t - \omega t\cos\omega t)$

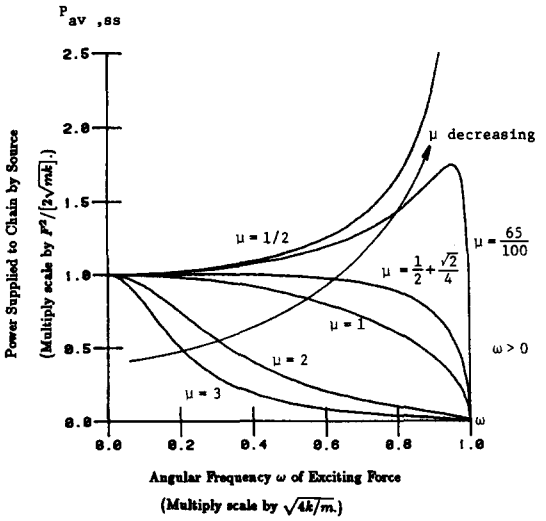


FIG. 2. "Steady-state" average power as a function of the angular frequency of the applied force for a half-infinite chain with a single isotopic defect in the leftmost position ($\mu \geq 1/2$).

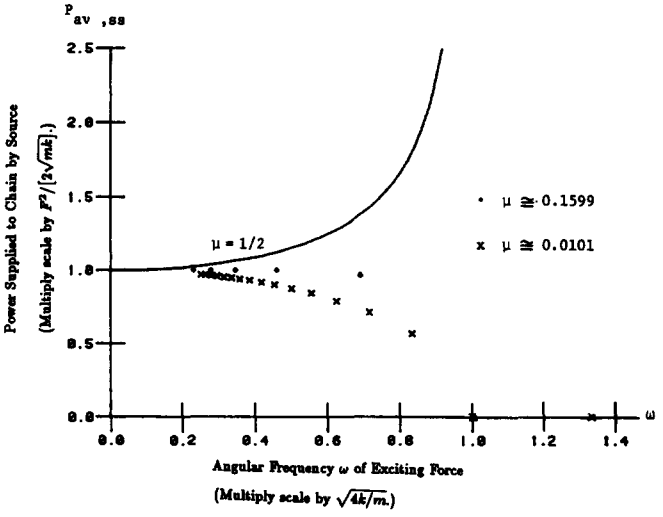


FIG. 3. "Steady-state" average power as a function of the angular frequency of the applied force for a half-infinite chain with a single isotopic defect in the leftmost position ($0 < \mu \leq 1/2$).

$\frac{1}{2} < \mu < \frac{1}{2} + \sqrt{2}/4$ have a single point of inflection and a maximum ordinate exceeding $F^2/[2\sqrt{mk}]$. For $\mu > \mu_0$ ($\cong 1.25771$), the graphs have two points of inflection and are strictly decreasing. For $\frac{1}{2} + \sqrt{2}/4 \leq \mu \leq \mu_0$, the graphs have no inflection points and are strictly decreasing.

When $\mu = \frac{1}{2}$,

$$P_{av,ss}(1) = \begin{cases} \dfrac{F^2}{m\sqrt{(4k/m) - \omega^2}}, & \text{if } 0 < \omega^2 < 4k/m, \\ 0, & \text{if } 4k/m < \omega^2. \end{cases} \tag{32}$$

The graph of $P_{av,ss}(1)$ for $\mu = \frac{1}{2}$ appears as the uppermost curve in Fig. 2. Notice that as μ approaches $\frac{1}{2}$ from above, the curves for $\mu > \frac{1}{2}$ more closely resemble the curve for $\mu = \frac{1}{2}$.

Lastly, if $0 < \mu < \frac{1}{2}$ and $\omega^2 = \omega_n^2 = n^{-2} y_1$ ($n = 2,3,4,\ldots$),

$$P_{av,ss}(1) = \begin{cases} \dfrac{F^2 n \sqrt{y_1}}{4k(n^2-1)} \sqrt{4\mu(1-\mu)n^2 - 1}, \\ \quad \text{if } 0 < \omega_n^2 < 4k/m, \\ 0, \quad \text{if } 4k/m \leq \omega_n^2. \end{cases} \tag{33}$$

Graphs of $P_{av,ss}(1)$ as a function of the discrete variable ω_n appear in Fig. 3 for two values of μ ($0 < \mu < \frac{1}{2}$). Observe that there are at most finitely many values of ω_n that are equal to or greater than $\sqrt{4k/m}$; so $P_{av,ss}(1)$ is zero at finitely many points.

IV. THE UNIFORM CHAIN WITH NO DEFECT

The equations of motion of the half-infinite chain with no defect (Fig. 1 with $\mu = 1$) are obtained from (1) or (2) by setting $\mu = 1$. The behavior of the displacement $x_n(t)$ for large t can be obtained by letting $\mu = 1$ in cases (1), (2), and (3) of Tables II and III. Thus

$$x_{n,ss}(t) = \begin{cases} \dfrac{F}{\omega\sqrt{mk}} \{1 - \cos[\omega t - (2n+1)\arcsin(\sqrt{[m\omega^2]/[4k]})]\}, & \text{if } 0 < \omega^2 < 4k/m, \\ \dfrac{F}{2k} \{1 - (-1)^n \sin(\omega t)\}, & \text{if } \omega^2 = 4k/m, \\ \dfrac{F}{\omega\sqrt{mk}} \left\{1 - (-1)^n \left[\dfrac{\sqrt{4k/m}}{\omega + \sqrt{\omega^2 - (4k/m)}}\right]^{2n+1} \sin(\omega t)\right\}, & \text{if } 4k/m < \omega^2, \end{cases} \tag{34}$$

where the principle value of the arcsine is used.

The steady-state average power supplied by the source to the chain is a meaningful concept for all values of the frequency ω, and the graph is the curve designated $\mu = 1$ in Fig. 2.

V. THE EFFECT OF THE PHASE OF THE EXCITING FORCE

This section discusses the effect of replacing the force $F\sin(\omega t)$ applied to the leftmost mass by $F\sin(\omega t - \phi)$,

where ϕ is real. In particular, it describes the effect of such a change of phase on the behavior of $x_n(N,t)$, $n \geqslant 0$, for large t.

The unique analytic solution of the slightly modified differential system is

$$x_n(N,t,\phi) = \frac{1}{m_0} \int_{-\infty}^{+\infty} P_n(N,y) W(y,t,\phi;0) d\alpha(N,y), \quad n \geqslant 0, \tag{35}$$

where $d\alpha(N,y)$ and $P_n(N,y)$ are given, as before, by Eqs. (9)–(11), but $W(y,t,\phi;0)$ is the solution of the differential system [compare to Eq. (5)]

$$W_{tt}(y,t,\phi;0) + yW(y,t,\phi;0) = F\sin(\omega t - \phi),$$
$$W(y,0,\phi;0) = W_t(y,0,\phi;0) = 0. \tag{36}$$

Solving (36) leads to

$$W(y,t,\phi;0) = \cos\phi \frac{F\sin[\sqrt{y}t]}{\sqrt{y}(\omega+\sqrt{y})} - \cos\phi \frac{F\sin[(\sqrt{y}-\omega)t/2]}{(\sqrt{y}-\omega)/2} \times \frac{\cos[(\sqrt{y}+\omega)t/2]}{(\sqrt{y}+\omega)} - \sin\phi \frac{F\sin[(\sqrt{y}-\omega)t/2]}{(\sqrt{y}-\omega)/2} \frac{\sin[(\sqrt{y}+\omega)t/2]}{(\sqrt{y}+\omega)}, \tag{37}$$

with the understanding that $W(\omega^2, t, \phi; 0) = \lim_{y \to \omega^2} W(y, t, \phi; 0)$.

The analysis of the phase for $N \geqslant 1$ is divided into three possibilities:

(i) $\mu \geqslant \mu_c(N)$ except when $\mu = \mu_c(N)$ and $\omega^2 = 4k/m$;

(ii) $\mu = \mu_c(N)$ and $\omega^2 = 4k/m$;

(iii) $0 < \mu < \mu_c(N)$.

In the first case, substitute (37) into (35) and recast $x_n(N,t,\phi)$ in the form

$$x_n(N,t,\phi) = A_n(N,t)\cos\phi + C_n(N,t)\cos(\omega t - \phi) + S_n(N,t)\sin(\omega t - \phi), \quad n \geqslant 0, \tag{38}$$

where $A_n(N,t)$, $C_n(N,t)$, and $S_n(N,t)$ are specified by Eqs. (17). Hence

$$x_{n,\mathrm{ss}}(N,t,\phi) = a_n(N)\cos\phi + c_n(N)\cos(\omega t - \phi) + s_n(N)\sin(\omega t - \phi), \quad n \geqslant 0. \tag{39}$$

When $\mu = \mu_c(N)$ and $\omega^2 = 4k/m$, the reformulation of (35) is similar to the preceding case. Since the limits of $C_n(N,t)$ and $S_n(N,t)$ as t approaches infinity through positive values are infinite, the substitutions $\bar{C}_n(N,t) = C_n(N,t)/\sqrt{t}$ and $\bar{S}_n(N,t) = S_n(N,t)/\sqrt{t}$ are made so that [compare to Eq. (21)]

$$x_n(N,t,\phi) = A_n(N,t)\cos\phi + t^{1/2}\bar{C}_n(N,t)\cos(\omega t - \phi) + t^{1/2}\bar{S}_n(N,t)\sin(\omega t - \phi), \quad n \geqslant 0, \tag{40}$$

from which it follows that

$$x_{n,\mathrm{ss}}(N,t,\phi) = a_n(N)\cos\phi + t^{1/2}\bar{c}_n(N)\cos(\omega t - \phi) + t^{1/2}\bar{s}_n(N)\sin(\omega t - \phi), \quad n \geqslant 0. \tag{41}$$

If $0 < \mu < \mu_c(N)$, recall from (11) that the distribution $d\alpha(N,y)$ is the sum of an absolutely continuous component and a Dirac delta component. In the integrand of Eq. (35), $W(y,t,\phi;0)$ is multiplied by this sum. The parts of $x_n(N,t,\phi)$ and $x_{n,\mathrm{ss}}(N,t,\phi)$ associated with the absolutely continuous component of $d\alpha(N,y)$ are given by (38) and (39), respectively. The Dirac delta component yields the term $J(x_N)P_n(N,y_N)W(y_N,t,\phi;0)/m$, where (37) is used to evaluate $W(y,t,\phi;0)$ at $y = y_n$. Thus, when $0 < \mu < \mu_c(N)$,

$$x_n(N,t,\phi) = A_n(N,t)\cos\phi + C_n(N,t)\cos(\omega t - \phi) + S_n(N,t)\sin(\omega t - \phi) + \frac{FJ(x_N)P_n(N,y_N)}{m\sqrt{y_N}(\omega^2 - y_N)} \times \left[-\sqrt{y_N}\sin(\omega t - \phi) + \omega\sin(\sqrt{y_N}t)\cos\phi - \sqrt{y_N}\cos(\sqrt{y_N}t)\sin\phi \right], \quad n \geqslant 0, \tag{42}$$

and

$$x_{n,\mathrm{ss}}(N,t,\phi) = a_n(N)\cos\phi + c_n(N)\cos(\omega t - \phi) + s_n(N)\sin(\omega t - \phi) + \frac{FJ(x_N)P_n(N,y_N)}{m\sqrt{y_N}(\omega^2 - y_N)} \times \left[-\sqrt{y_N}\sin(\omega t - \phi) + \omega\sin(\sqrt{y_N}t)\cos\phi - \sqrt{y_N}\cos(\sqrt{y_N}t)\sin\phi \right], \quad n \geqslant 0. \tag{43}$$

VI. COMMENTS

In essence, the main results of the preceding analysis are: (a) finding the exact solution of the initial-value problem associated with the half-infinite chain of harmonic oscillators with one isotopic defect; (b) determining the behavior of the displacement x_n of the nth mass for large t (steady state); (c) calculating the steady-state average power supplied to the chain by the source over one period of the source; and (d) determining the effect of the phase of the source upon the behavior of x_n for large t. In this effort, various derivations are truncated or omitted entirely for the sake of brevity. Further details are found in Ref. 10.

Although the mathematics of this investigation is presented in terms of a mechanical system of harmonic oscillators (this physical system is chosen because it allows for good intuition), the system of ordinary differential equations representing the corresponding equations of motion is reasonably general and satisfies the second-order vector matrix differential equation, $\ddot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{F}(t)$, where $\mathbf{X}$ is an infinite column vector of the displacements of the masses from equilibrium, $\mathbf{F}$ is an infinite column vector representing the forcing function, and $\mathbf{A}$ is an infinite tridiagonal matrix. Since the mathematical structure has some generality, the approach contained herein may lend itself to the understanding and solution of problems in other areas such as those involving large electrical networks, defects in crystals, vibrational properties of disordered solids,[17] and dynamics of atomic lattices.[18] In fact, by making the usual identification between mechanical and electrical quantities, the displacements of this paper are the charges across the capaci-

tors in a half-infinite ladder network of LC circuits, where all the capacitances are the same and one of the inductances, the one corresponding to the defective mass, is different from the rest.

[1] E. W. Montroll and R. B. Potts, Phys. Rev. **100**, 525 (1955).
[2] W. P. McKibben, doctoral dissertation, Georgia Institute of Technology, 1973, pp. 6–7.
[3] The integral representation of this solution is given by Theorem 2 in W. G. Christian, A. G. Law, W. F. Martens, A. L. Mullikin, and M. B. Sledd, J. Math. Phys. **17**, 146 (1976).
[4] See Ref. 3, p. 147, Eq. (8).
[5] L. Brand, *Differential and Difference Equations* (Wiley, New York, 1966).
[6] One particular technique that is used is found on pp. 106ff of H. Levy and F. Lessman, *Finite Difference Equations* (Macmillan, New York, 1961).
[7] See Ref. 3, p. 149.
[8] See Ref. 2.
[9] This is referred to as the saltus by Ya. L. Geronimus, "On Some Finite Difference Equations and Corresponding Systems of Orthogonal Polynomials," *Memoirs of the Mathematical Section of the Faculty of Mathematics and Physics at Kharkov State University and the Kharkov Mathematical Society* (1957), Vol. 25 pp. 87–100.
[10] E. L. Mokole, doctoral dissertation, Georgia Institute of Technology, 1982.
[11] See Ref. 10, pp. 106–118.
[12] See Ref. 10, pp. 106–118.
[13] See Ref. 10, Appendix III.
[14] See Ref. 10, p. 126.
[15] The verification that $\tilde{P}_{\mathrm{av,ss}} = \hat{P}_{\mathrm{av,ss}} = P_{\mathrm{av,ss}}$ for $N = 1$ is found on pp. 64–77 of Ref. 10.
[16] The expression $T_1(N, 1 - m\omega^2 2^{-1} k^{-1})$ is positive for $0 < \omega^2 < 4k/m$ as a result of the properties of $S_1(N,x)$ on pp. 98–99 of Ref. 10.
[17] A. S. Barker, Jr. and A. J. Sievers, Rev. Mod. Phys. **47**, Suppl. No. 2 (Fall 1975).
[18] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation*, 2nd ed. (Academic, New York, 1971), pp. 514–519.

Solution of Laplace's equation in plane single-connected regions bounded by arbitrary single curves

F. Minotti[a] and C. Moreno[a]
Laboratorio de Física del Plasma, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pab. I, 1428 Buenos Aires, Argentina

(Received 15 September 1989; accepted for publication 21 February 1990)

A method is developed to solve Laplace's equation with Dirichlet's or Neumann's conditions in plane, single-connected regions bounded by arbitrary single curves. It is based on the existence of a conformal transformation that reduces the original problem to another whose solution is known. The main advantage of the method is that it does not require the knowledge of the transformation itself, so it is applicable even when no transformation is available. The solution and its higher-order derivatives are expressed in terms of explicit quadratures easy to evaluate numerically or even analytically.

I. INTRODUCTION

Laplace's equation appears in many physical problems: gravitational or electrostatic problems; the steady, irrotational flow of incompressible fluids; the steady-state flow of heat; the steady diffusion of a solute, of neutrons, and generally, in steady diffusive processes. Consequently, several methods have been developed to solve it. There are essentially two different approaches: variable separation and Green functions (or a combination of both).[1-3] In addition, bidimensional problems can be treated by means of conformal mapping techniques that provide a powerful alternative method. None of these methods, of course, is always applicable and its success depends on the problem at hand. In fact, the boundary conditions and the contour itself may prevent the variable separation; moreover, Green's function or, in bidimensional problems, an appropriate conformal transformation, is generally not easy to find. On the other hand, in very complicated cases one must resort to numerical calculations.[4-6]

In this work we propose a method that allows the obtention of the general solution in bidimensional single-connected regions bounded by single closed curves. In spite of the fact that the method is based on the conformal mapping technique, the explicit knowledge of the conformal transformation is avoided. The solution is then expressed by explicit quadratures that are easy to evaluate numerically or even analytically.

In addition, the same formalism allows the direct calculation of certain magnitudes of interest as, for instance, the normal derivatives at the boundary in Dirichlet's problem, and higher-order derivatives of the unknown potential function.

II. METHOD

A. Dirichlet's conditions

In order to develop the method we use the following holomorphic functions:

$$f = \psi + i\phi, \tag{1a}$$

$$g = F + iG, \tag{1b}$$

where ψ and F are the conjugated functions of the potential ϕ, and Green's function G, corresponding to Dirichlet's conditions in the region of interest, respectively. With these definitions two equivalent expressions of the formal solution can be given as follows:

$$\phi(x',y') = \int_{\partial(R)} \phi(x,y)\,\frac{dg}{dZ}\,dZ, \tag{2}$$

from Green's formula, and

$$\phi(x',y') = \mathrm{Im}\left\{\frac{1}{2\pi i}\int_{\partial(R)} \frac{f(Z)}{Z-Z'}\,dZ\right\}, \tag{3}$$

from Cauchy's formula, where $\partial(R)$ denotes the boundary of the domain R (see Fig. 1).

We can transform the region R onto the upper half-plane R' by means of a conformal transformation $\zeta(Z)$ (see Fig 1). In the transformed domain R' the function g is given by

$$g(\zeta,\zeta',\zeta'^{*}) = \frac{i}{2\pi}\ln\left(\frac{\zeta-\zeta'^{*}}{\zeta-\zeta'}\right), \tag{4}$$

where the asterisk denotes the complex conjugate.

From Eqs. (2) and (4) the formal solution of the problem is

$$\phi[x'(\xi',\eta'),y'(\xi',\eta')] = \frac{i}{2\pi}\int_{\partial(R')}\left(\frac{1}{\zeta-\zeta'^{*}}-\frac{1}{\zeta-\zeta'}\right)\phi[Z(\zeta)]\,d\zeta. \tag{5}$$

On the other hand, by developing Eq. (3), we get,

$$\phi(x',y') = \frac{1}{2\pi}\int_{\partial(R)}\frac{\phi\cos\alpha-\psi\sin\alpha}{r}\,dl, \tag{6}$$

where α and r are those denoted in Fig. 1.

To obtain the explicit solution from Eq. (5) or (6), the explicit knowledge of $\zeta(Z)$ or ψ, respectively, is required. In fact, both problems reduce to the knowledge of $\zeta(Z)$, because the determination of ψ requires it. Although many conformal transformations that solve the problem are known,[7] this is not possible for an arbitrary region R. The goal of the method presented here is to show that (and how)

[a] Fellow of the Consejo Nacional de Investigaciones Científicas y Técnicas.

 0022-2488/90/081914-05$03.00

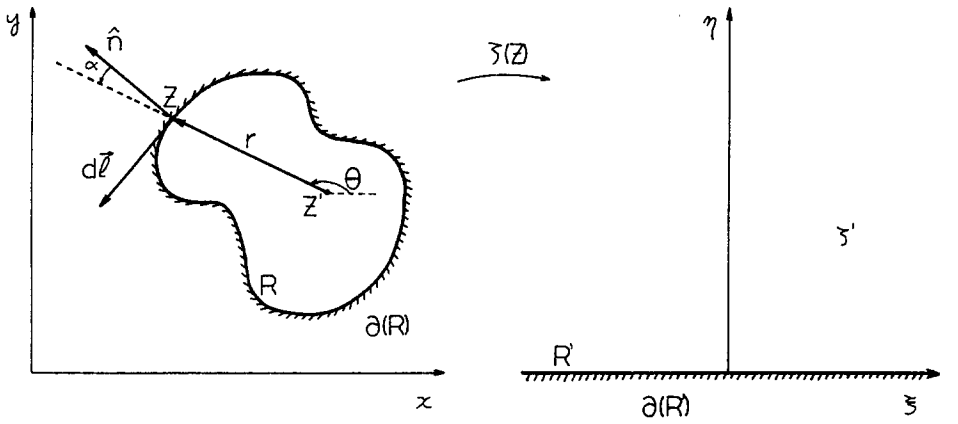


FIG. 1. The original region in Z plane and the transformated region in ζ plane.

ψ can be determined without knowing the explicit form of $\zeta(Z)$.

Noting that

$$\frac{\partial\psi}{\partial\xi'}=\frac{\partial\phi}{\partial\eta'}, \tag{7}$$

and using Eq. (5), we get,

$$\frac{\partial\psi}{\partial\xi'}=\int_{-\infty}^{+\infty} h\,[(\xi-\xi'),\eta']\phi(\xi)d\xi, \tag{8}$$

where

$$h(u,v)=(1/\pi)(u^2-v^2)/[u^2+v^2]^2. \tag{9}$$

From this definition it follows that

$$\int_{-\infty}^{+\infty} h\,[(\xi-\xi'),\eta']d\xi=0, \tag{10}$$

which allows us to evaluate Eq. (8) on the boundary $\partial(R')$ as

$$\left(\frac{\partial\psi}{\partial\xi'}\right)_{\partial(R')} = \lim_{\eta'\to 0^+}\int_{-\infty}^{+\infty} h\,[(\xi-\xi'),\eta'][\phi(\xi)-\phi(\xi')]d\xi. \tag{11}$$

Assuming that $\phi(\xi)$ can be developed in Taylor's series near ξ', it can be easily proved that this limit indeed exists and is given by

$$\left(\frac{\partial\psi}{\partial\xi'}\right)_{\partial(R')}=\frac{1}{\pi}\,\mathrm{pV}\int_{-\infty}^{+\infty}\frac{\phi(\xi)-\phi(\xi')}{(\xi-\xi')^2}d\xi, \tag{12}$$

where pV refers to the principal Cauchy value.

Equation (12) allows the calculation of the derivative of ψ along the boundary of the transformed region that on integration yields the function $\psi(\xi')$. Once ψ is known, the use of Eq. (6) leads to the determination of the potential ϕ at any point inside R.

As was pointed out by the referee, in many cases of interest one can avoid the intermediate step of evaluating the derivative $\partial\psi/\partial\xi'$ in order to obtain ψ. This is possible by virtue of the result derived by Sneddon[8] that states that if ϕ is a harmonic function in the half-plane $\eta>0$, and tends to zero as $(\xi^2+\eta^2)^{1/2}\to\infty$ then the normal derivative of ϕ on the boundary $\eta=0$ is the Hilbert transform of the tangential derivative, i.e.,

$$\left(\frac{\partial\phi}{\partial\eta'}\right)_{\partial(R')}=\frac{1}{\pi}\,\mathrm{pV}\int_{-\infty}^{+\infty}\frac{1}{(\xi-\xi')}\frac{\partial\phi}{\partial\xi}d\xi, \tag{13}$$

which, since $\partial\phi/\partial\eta=\partial\psi/\partial\xi$, and taking into account the properties of the Hilbert transform results in

$$\psi(\xi')=\frac{1}{\pi}\,\mathrm{pV}\int_{-\infty}^{+\infty}\frac{\phi(\xi)}{(\xi-\xi')}d\xi. \tag{14}$$

For this formula to hold, however, ϕ and ψ must tend to zero as $(\xi^2+\eta^2)^{1/2}\to\infty$. This is possible, for instance, if the potential is constant (which without loss of generality can be taken as zero) on some finite piece of the original boundary. In this case, by choosing the parametrization in such a way that $s\to\pm\infty$ corresponds to points inside the piece at constant potential, the potential at the transformed boundary will be different from zero only on a finite interval, assuring that ϕ and ψ tend to zero as $(\xi^2+\eta^2)^{1/2}\to\infty$. Formula (14) allows then the obtention of ψ on the boundary without going through the integration of expression (12).

In order to evaluate Eq. (12) or, when possible, Eq. (14), only the values of the potential ϕ on the boundary are required, so that all we need to know is the transformation $\zeta(Z)$ restricted to that boundary. This can be easily done if the boundary of the original region R is given in parametric form. In fact, if the boundary is given by

$$x=X(s), \tag{15a}$$

$$y=Y(s), \tag{15b}$$

where s is a real parameter which, when ranging from $-\infty$ to ∞ counterclockwise, describes the full contour $\partial(R)$, the restricted transformation is

$$\zeta(\xi)=X(\xi)+iY(\xi). \tag{16}$$

In this manner, the potential on the boundary is given as a function of ξ by

$$\phi(\xi)=\phi[X(\xi),Y(\xi)]. \tag{17}$$

In order to evaluate Eq. (6), the explicit form of the terms entering in the integrand can be expressed in terms of known functions as

$$\frac{\phi \cos\alpha\, dl}{r} = \phi[X(\xi),Y(\xi)] \times \frac{[X(\xi)-x'](dY/d\xi)-[Y(\xi)-y'](dX/d\xi)}{[X(\xi)-x']^2+[Y(\xi)-y']^2} \times d\xi, \tag{18a}$$

$$\frac{\psi \sin\alpha\, dl}{r} = \psi[X(\xi),Y(\xi)] \times \frac{[Y(\xi)-y'](dY/d\xi)+[X(\xi)-x'](dX/d\xi)}{[X(\xi)-x']^2+[Y(\xi)-y']^2} \times d\xi. \tag{18b}$$

As can be seen, we have avoided the use of the explicit conformal transformation, which is the major difficulty in actual calculations.

B. Neumann's conditions

Using the definitions given by Eqs. (1), G being now Green's function for Neumann's conditions, the formal solution is given by

$$\phi(x',y') = \mathrm{Im}\left\{\int_{\partial(R)} g\frac{df}{dZ}\,dZ\right\}. \tag{19}$$

We can transform the original region R onto the upper half-plane R' as in Dirichlet's case by means of a conformal transformation $\zeta(Z)$. In R' the function g is now given by

$$g(\zeta,\zeta',\zeta'^*) = -K\zeta + (i/2\pi)\ln[(\zeta-\zeta')(\zeta-\zeta'^*)], \tag{20}$$

where K is a real constant.

We can then obtain the derivative of the potential ϕ along the contour $\partial(R')$ from Eq. (19). A straightforward calculation leads to

$$\left(\frac{\partial\phi}{\partial\xi'}\right)_{\partial(R')} = -\frac{1}{\pi}\,\mathrm{pV}\int_{-\infty}^{+\infty}\frac{1}{\xi-\xi'}\frac{\partial\phi}{\partial\eta}\,d\xi, \tag{21}$$

whose integrand is known through the relationship

$$\left(\frac{\partial\phi}{\partial\eta}\right)_{\partial(R')} d\xi = -\left(\frac{\partial\phi}{\partial n}\right)_{\partial(R)} dl. \tag{22}$$

On the other hand, since f is holomorphic,

$$\frac{\partial\psi}{\partial\xi} = \frac{\partial\phi}{\partial\eta}, \tag{23}$$

we can obtain the values of ϕ and ψ on the boundary $\partial(R')$ by integration of Eqs. (21) and (23). The potential $\phi(x',y')$ can then be obtained from Eqs. (6) and (18).

It is interesting to note that the "energy" U associated to the potential ϕ:

$$U = \int\int_R |\nabla\phi|^2 dx\,dy, \tag{24}$$

and expressed in terms of the magnitudes evaluated at the boundary as

$$U = \int_{\partial(R)} \phi\frac{\partial\phi}{\partial n}\,dl = -\int_{\partial(R')} \phi\frac{\partial\phi}{\partial\eta}\,d\xi, \tag{25}$$

can be easily evaluated from Eq. (12) [or Eq. (14) if possible] in Dirichlet's case, and from Eqs. (21) and (22) in Neumann's case.

C. Determination of higher-order derivatives of ϕ

We have developed a method that allows the evaluation of ϕ from the knowledge of the function f at the boundary in Secs. II A and II B. From this function we can also evaluate the higher-order derivatives of the potential ϕ. In fact, by using the integral Cauchy formulas and taking real and imaginary parts, the order-p derivatives of ϕ can be written in terms of

$$\mathrm{Re}\left\{\frac{d^p f}{dz^p}\right\} = \mathrm{Re}\left\{\frac{p!}{2\pi i}\int_{\partial(R)}\frac{f(Z)}{(Z-Z')^{p+1}}\,dZ\right\} = \frac{p!}{2\pi}\int_{\partial(R)}\frac{\phi\sin\chi_p+\psi\cos\chi_p}{r^{p+1}}\,dl, \tag{26a}$$

$$\mathrm{Im}\left\{\frac{d^p f}{dz^p}\right\} = \mathrm{Im}\left\{\frac{p!}{2\pi i}\int_{\partial(R)}\frac{f(Z)}{(Z-Z')^{p+1}}\,dZ\right\} = \frac{p!}{2\pi}\int_{\partial(R)}\frac{\phi\cos\chi_p-\psi\sin\chi_p}{r^{p+1}}\,dl, \tag{26b}$$

where χ_p is given in terms of the angles denoted in Fig. 1 by $\chi_p = \alpha + p\theta$.

As an example, developing the integrands of Eqs. (26) as in Sec. II A we get the analogous of Eqs. (18) for the case $p=1$:

$$\cos\chi_1 = \frac{[(X(\xi)-x')^2-(Y(\xi)-y')^2](dY/d\xi)-2(X(\xi)-x')(Y(\xi)-y')(dX/d\xi)}{r^2}, \tag{27a}$$

$$\sin\chi_1 = \frac{2(X(\xi)-x')(Y(\xi)-y')(dY/d\xi)+[(X(\xi)-x')^2-(Y(\xi)-y')^2](dX/d\zeta)}{r^2}. \tag{27b}$$

The higher-order derivatives can be easily obtained in the same way. The advantage of this method is that derivatives of any order can be obtained from the same data as the required for the potential.

III. APPLICATIONS

The method outlined in Sec. II A is specially suited for situations in which the potential ϕ is piecewise constant, because in those cases the evaluation of Eq. (12) and its subse-

quent integration to obtain ψ can be performed analytically. So, the general results are

$$\left(\frac{\partial\psi}{\partial\xi}\right)_{\partial(R')} = \frac{1}{\pi}\sum_{j=0}^{N}(\phi_{j+1}-\phi_{i+1})\left[\frac{1-\delta_{j,0}}{\xi_j-\xi}-\frac{1-\delta_{j,N}}{\xi_{j+1}-\xi}\right], \tag{28}$$

$$(\psi)_{\partial(R')} = \frac{1}{\pi}\sum_{j=0}^{N}\{(\phi_{j+1}-\phi_{i+1})[(1-\delta_{j,N}) \times\ln|\xi_{j+1}-\xi|-(1-\delta_{j,0})\ln|\xi_j-\xi|]\}, \tag{29}$$

where the notations of Fig. 2 was used and $\xi_i<\xi<\xi_{i+1}$. We mention in passing that formula (29) could be obtained from the direct evaluation of Eq. (14).

A case of practical interest is given by the electrostatic potential generated by conducting boundaries. In this situation the inducted surface charge density per unit length, $\sigma=\partial\phi/\partial n$, is given by

$$\sigma[X(\xi),Y(\xi)] = -\left(\frac{\partial\psi}{\partial\xi}\right)_{\partial(R')}\Big/\sqrt{\left(\frac{dX}{d\xi}\right)^2+\left(\frac{dY}{d\xi}\right)^2}. \tag{30}$$

Let us consider as an example the problem of a square with one side at a potential assumed to be one, and the rest of the boundary at zero potential. The length of each side is normalized to one. We immediately get by using Eqs. (28) and (30):

$$\sigma = -\frac{1}{\pi}\frac{1/t^2+1/(4-t)^2}{1/(4-t)-1/t+\frac{2}{3}}. \tag{31}$$

To obtain this formula we have used the following parametric representation of the boundary:

$$\begin{aligned} &X(t)=t, && Y(t)=0, && \text{if } 0\leqslant t\leqslant 1,\\ &X(t)=1, && Y(t)=t-1, && \text{if } 1\leqslant t\leqslant 2,\\ &X(t)=3-t, && Y(t)=1, && \text{if } 2\leqslant t\leqslant 3,\\ &X(t)=0, && Y(t)=4-t, && \text{if } 3\leqslant t\leqslant 4, \end{aligned}$$

with the potential: 1 if $0\leqslant t\leqslant 1$; 0 if $1<t<4$; and with the relation $s=1/(4-t)-1/t$ between t and the s parameter appearing in Eqs. (15).

It is important to point out that, even in this simple case, the obtention of σ from the expression of ϕ calculated by means of the usual method of variable separation, is not possible due to the fact that the arising series is not derivable, a "blemish" commonly encountered in these methods when the potential on the boundary is discontinuous.

As a second example let us consider a circle of radius one with boundary conditions of the Neumann type given by: $\partial\phi/\partial\rho=\cos(\theta)$ on $\rho=1$; where (ρ,θ) are polar coordinates. Using Eq. (23), the expression of ψ on the boundary is immediately found to be: $\psi=-\sin(\theta)+\text{const}$. With the transformation $\xi=-\cot(\theta/2)$, Eq. (21) leads to

$$\left[\frac{\partial\phi}{\partial\theta'}\right]_{\rho=1} = -\frac{1}{2\pi}\csc^2(\theta'/2)\mathrm{pV}\times\int_0^{2\pi}\frac{\cos(\theta)}{\cot(\theta/2)-\cot(\theta'/2)}\,d\theta. \tag{32}$$

The integration can be analytically performed and the result is simply: $\partial\phi/\partial\theta'=-\sin(\theta')$ on the boundary that on integration yields: $\phi=\cos(\theta')+\text{const}$. With these expressions of ϕ and ψ, Eq. (6) leads on integration to: $\phi=\rho'\times\cos(\theta')+\text{const}$, which agrees with the result obtained by classical methods.

These simple examples show how the proposed method can be used. Analytical solutions, however, are not always possible and in more involved cases one must resort to numerical evaluation of the integrals appearing in Eqs. (6) and (11) or (21). We wish to point out here that this numerical procedure is fast and easy to perform even in very complicated problems.

IV. FINAL REMARKS AND CONCLUSIONS

We have developed a method that allows to obtain the potential satisfying Laplace's equation with Dirichlet's or Neumann's conditions on a plane region bounded by an arbitrary curve.

The potential and its derivatives of any order are expressed by explicit quadratures which allows further developments. The great advantage of the method presented here is that the explicit form of the conformal transformation is not required. Therefore, the potential in very complicated regions, for which no conformal transformations are available, can be directly computed from the boundary conditions alone.

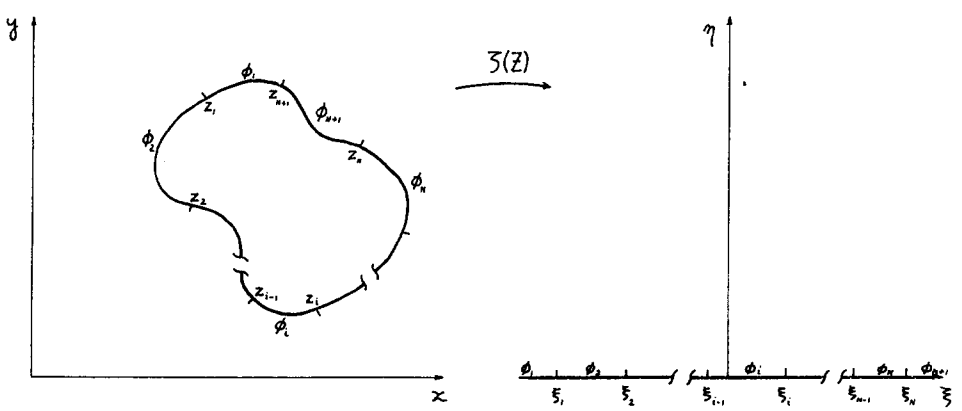


FIG. 2. Sketch of a Dirichlet's problem with a piecewise constant potential at the boundary (Z and ζ planes).

From the numerical point of view, the method is extremely low memory consuming and it allows the evaluation of the potential function ϕ with any precision avoiding the use of grids or numerical transformations.

Moreover, special cases of practical interest as, for instance, the electrostatic potential generated by conducting boundaries or the temperature distribution in regions whose boundary is composed by several constant temperature sections, etc.; admits solutions given by only a quadrature easy to evaluate even in very complicated regions. As a bonus, some magnitudes of interest (viz. the induced charge or the heat flow through the boundary) are obtained analytically.

Finally, for both Neumann's and Dirichlet's conditions, the total energy associated to the potential field can be straightforwardly computed from the formalism itself.

ACKNOWLEDGMENTS

We would like to thank Professor Constantino Ferro Fontán for his assistance and Professor H. J. Kelly and Professor A. R. Piriz for a critical reading of the manuscript. It is a pleasure to thank the anonymous referee who suggested Eqs. (13) and (14).

This work was supported by grants of the Organization of American States, the Consejo Nacional de Investigaciones Científicas y Técnicas, and the Universidad de Buenos Aires.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[2] H. F. Weinberger, *Introduction to Partial Differential Equations with Methods of Complex Variables and Integral Transforms* (Blaisdell, New York, 1965).
[3] J. H. Jeans, *Mathematical Theory of Electricity and Magnetism* (Cambridge U.P., Cambridge, 1958), 5th ed.
[4] R. Menikoff and C. Zemach, J. Comput. Phys. **36**, 366 (1980).
[5] D. I. Meiron, S. A. Orszag, and M. Israeli, J. Comput. Phys. **40**, 345 (1981).
[6] R. Hostens and G. De Mey, Comput. Phys. Commun. **16**, 5 (1978).
[7] H. Kober, *Dictionary of Conformal Representations* (Dover, New York, 1957).
[8] I. N. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).

Integrability condition and finite-periodic Toda lattice

Susumu Okubo
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 6 February 1990; accepted for publication 21 March 1990)

A further generalization to construct integrable dynamical systems with hierarchy has been developed, based upon a generalization of the zero Nijenhuis tensor condition in a symplectic manifold with double symplectic structures. As an example, the periodic Toda lattice Hamiltonian has been shown to satisfy these conditions. The case of the three-dimensional Coulomb potential problem has also been analyzed in a similar fashion.

I. INTRODUCTION

There are many works[1-6] on integrable systems, both finite and infinite dimensional. To be definite, we will consider only the finite case. Let M be a $2N$-dimensional symplectic manifold with local coordinate $x^{\mu}(\mu = 1,2,...,2N)$. Let $f_{\mu\nu} = -f_{\nu\mu}$ be the symplectic tensor field so that it satisfies

$$\partial_{\lambda} f_{\mu\nu} + \partial_{\mu} f_{\nu\lambda} + \partial_{\nu} f_{\lambda\mu} = 0. \tag{1.1}$$

The Poisson bracket of two smooth functions $h(x)$ and $g(x)$ in M is then defined as usual[7,8] by

$$\{h,g\} = f^{\mu\nu}\,\partial_{\mu} h\,\partial_{\nu} g, \tag{1.2}$$

where $f^{\mu\nu}$ is the inverse tensor of $f_{\mu\nu}$ and repeated Greek indices imply automatical summations on $2N$ values $1,2,...,2N$. According to the well-known Darboux theorem,[9] there exists a local coordinate frame that is called the canonical frame, where the tensor $f_{\mu\nu}$ is constant with values

$$f_{jk} = f_{j+N,k+N} = 0,$$
$$f_{j,k+N} = -\delta_{jk}, \tag{1.3}$$

for $j,k = 1,2,...,N$. In that case, the coordinate x^{μ} may be explicitly written as

$$x^{\mu} = \{q_1,q_2,...,q_N,p_1,p_2,...,p_N\}, \tag{1.4}$$

in terms of N-configuration variables q_j and their canonical conjugate momenta $p_j\,(j = 1,2,...,N)$.

Although many finite integrable models[2,3] are now known, there exists no systematical method of discovering all of them. The most commonly used procedure is first to find N algebraically independent functions $K_1,K_2,...,K_N$ of the coordinate x^{μ}'s, such that they are constants of motion with respect to a given Hamiltonian H. Ordinarily, this may be accomplished by finding a suitable Lax pair.[10] Then, we verify the involution property $\{K_j,K_k\} = 0$ for $j,k = 1,2,...,N$, so that the system is integrable by Liouville's theorem.[11,12] We now note that we can reverse the above procedure as follows. In contrast, we do not assume any particular Hamiltonian, and let I be an index set that ordinarily consists of all integer values $0,1,2,...$. Suppose that we can find a set of functions K_n $(n \in I)$ that are in involution, i.e.,

$$\{K_n,K_m\} = 0, \tag{1.5}$$

for $n, m \in I$. Moreover, we assume that we can find N algebraically independent terms among them, which we identify with $K_1,K_2,...,K_N$ for simplicity. Now, let $p \in I$ be an arbitrary but fixed integer and we choose K_p to be the Hamiltonian of the system, i.e.,

$$H = K_p. \tag{1.6}$$

For this choice of H, we designate the corresponding time variable as t_p with the Hamilton's equation of motion

$$\frac{d}{dt_p} h(x) = \{H,h\} = f^{\mu\nu}\,\partial_{\mu} K_p\,\partial_{\nu} h, \tag{1.7}$$

for any smooth function $h(x)$. Then, in view of the involution property Eq. (1.5) we see readily the validity of

$$\frac{d}{dt_p} K_n = 0, \tag{1.8}$$

for any $n \in I$, so that K_n's are constants of motion with respect to the Hamiltonian $H = K_p$. Because of our assumption of $K_1,K_2,...,K_N$ being algebraically independent, then the Liouville theorem assures the complete integrability of the system. Note that Eq. (1.7) is a prototype of the hierarchy equation[6] familiar in the K dV system.

In summary, a problem of constructing integrable systems is reduced in principle to that of finding a set of functions $\{K_n \mid n \in I\}$ satisfying the involution property Eq. (1.5), provided that N terms among them are algebraically independent. Presently, there exist two semisystematic approaches towards this goal. The first one is algebraic[13] in spirit, related to Cartan matrix of simple Lie algebras. The second method[14-20] that is the topic of this paper is geometrical in nature, involving the notion of the Nijenhuis tensor[21] (or its generalization). Let the original symplectic manifold M possess the second symplectic (or more generally presymplectic) tensor field $F_{\mu\nu} = -F_{\nu\mu}$, so that we have

$$\partial_{\lambda} F_{\mu\nu} + \partial_{\mu} F_{\nu\lambda} + \partial_{\nu} F_{\lambda\mu} = 0. \tag{1.9}$$

We can then construct the (1,1) tensor S_{μ}^{ν} by

$$S_{\mu}^{\nu} = F_{\mu\lambda} f^{\lambda\nu}, \tag{1.10}$$

and the Nijenhuis tensor $N_{\mu\nu}^{\lambda}$ by

$$N_{\mu\nu}^{\lambda} = S_{\mu}^{\alpha}\partial_{\alpha} S_{\nu}^{\lambda} - S_{\nu}^{\alpha}\partial_{\alpha} S_{\mu}^{\lambda} - S_{\alpha}^{\lambda}(\partial_{\mu} S_{\nu}^{\alpha} - \partial_{\nu} S_{\mu}^{\alpha}). \tag{1.11}$$

We remark that $N_{\mu\nu}^{\lambda}$ is a genuine (1,2) tensor under general coordinate transformations. Suppose now that S_{μ}^{ν} satisfies the zero Nijenhuis tensor condition

$$N_{\mu\nu}^{\lambda} = 0. \tag{1.12}$$

 0022-2488/90/081919-10$03.00

In previous publications,[18-28] we have presented the following facts. First, defining K_n's for any integer n by

$$K_n = \begin{cases} (1/2n)\mathrm{Tr}\, S^n = (1/2n)(S^n)^\mu_\mu & (n \neq 0) \\ \frac{1}{2}\log\det S & (n=0) \end{cases} \tag{1.13}$$

they satisfy the recursion relation

$$S^\nu_\mu\, \partial_\nu K_n = \partial_\mu K_{n+1}, \tag{1.14}$$

from which the involution property Eq. (1.5) can be shown to automatically follow. Second, we introduce the nth power antisymmetric tensor $(F^n)_{\mu\nu}$ inductively by

$$(F^{n+1})_{\mu\nu} = (F^n)_{\mu\alpha} f^{\alpha\beta} F_{\beta\nu},$$
$$(F^0)_{\mu\nu} = f_{\mu\nu}. \tag{1.15}$$

Then, it has been shown in Refs. 19 and 20 that they are all symplectic, i.e.,

$$\partial_\lambda (F^n)_{\mu\nu} + \partial_\mu (F^n)_{\nu\lambda} + \partial_\nu (F^n)_{\lambda\mu} = 0, \tag{1.16}$$

for any integer n. Moreover, the condition $N^\lambda_{\mu\nu} = 0$ is equivalent to the validity of Eq. (1.16) for the special case $n = 2$. In other words, the general result Eq. (1.16) for any integer n follows from its three special cases $n = 0,1$, and 2.

One notable example obeying the zero Nijenhuis tensor condition Eq. (1.12) is the Toda lattice solution[18,19] where we have

$$F_{jk} = f_j \delta_{j+1,k} e^{q_j - q_k} - f_k \delta_{k+1,j} e^{q_k - q_j},$$
$$F_{j,k+N} = -p_j \delta_{jk}, \tag{1.17}$$
$$F_{j+N,k+N} = -\epsilon(j-k),$$

in the canonical coordinate frame Eq. (1.3), where $\epsilon(j-k)$ is the sign function

$$\epsilon(j-k) = \begin{cases} 1, & j > k, \\ 0, & j = k, \\ -1, & j < k, \end{cases} \tag{1.18}$$

and where $f_j (j = 1,2,\ldots,N)$ with $f_N = 0$ are arbitrary real constants. The standard Toda Hamiltonian is identified by $H = K_2$, i.e., $p = 2$ in Eq. (1.6) with

$$H = \frac{1}{4}\mathrm{Tr}\, S^2 = \frac{1}{2}\sum_{j=1}^N (p_j)^2 + \sum_{j=1}^N f_j e^{q_j - q_{j+1}}. \tag{1.19}$$

However, the case of the periodic Toda lattice with $f_N \neq 0$ and with identification $q_{N+1} = q_1$ in Eq. (1.19) cannot be covered in this way. The reason is that the sign function $\epsilon(j-k)$ is not invariant under the cyclic transformation $1 \to 2 \to 3 \to \cdots \to N \to 1$. The main purpose of this note is to demonstrate the fact that we must generalize our method in order to accomodate the case of the periodic Toda lattice. For the sake of simplicity, let us set

$$\Delta_{\mu\nu\lambda}(G) = \partial_\lambda G_{\mu\nu} + \partial_\mu G_{\nu\lambda} + \partial_\nu G_{\lambda\mu}, \tag{1.20}$$

for any antisymmetric tensor $G_{\mu\nu} = -G_{\nu\mu}$, and we modify the zero Nijenhuis tensor condition $\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(F^2) = 0$ by

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(\tilde{F}^2) = 0, \tag{1.21}$$

where $(\tilde{F}^2)_{\mu\nu}$ is defined by

$$(\tilde{F}^2)_{\mu\nu} \equiv (F^2)_{\mu\nu} + f_{\mu\nu}\left[\tfrac{1}{2}C_1 (K_1)^2 + C_2 K_2\right], \tag{1.22}$$

for some constants C_1 and C_2. In the next section, we will show that a weaker analog of the recursion relation Eq. (1.14) holds nevertheless (see Proposition 1 in Sec. II) and that the involution law Eq. (1.5) for positive integers n and m still follows from it in general except for some special values of the constant C_2 (see Proposition 2 in Sec. II). Using these results, we will demonstrate in Sec. III and in the Appendix that we can accommodate the periodic Toda lattice and it is still integrable with some interesting properties, where $F_{\mu\nu}$ are now given by

$$F_{jk} = f_j \delta_{k,j+1} e^{q_j - q_k} - f_k \delta_{j,k+1} e^{q_k - q_j},$$
$$F_{j,k+N} = -\delta_{jk}\left[p_j - \frac{1}{N}\sum_{\ell=1}^N p_\ell\right] + \frac{1}{N}p_j, \tag{1.23}$$
$$F_{j+N,k+N} = -\bar{\epsilon}(j-k) \equiv -\epsilon(j-k) + (2/N)(j-k),$$

for $j,k = 1,2,\ldots,N$ with understanding $q_{N+1} = q_1$ and $\delta_{1,N+1} = 1$ instead of the usual value zero as in Eq. (1.17). Note that in spite of its form $\bar{\epsilon}(j-k)$ is invariant now under cyclic permutation $1 \to 2 \to \cdots \to N \to 1$. Also, the Hamiltonian H for the periodic Toda lattice is not exactly K_2 now but has a form of $aK_2 + b(K_1)^2$ for some constants a and b. In Sec. IV, we will give some examples of systems that satisfy the new conditions but are not necessarily integrable, since K_n for $n \geqslant 3$ for these models turn out to be polynomials of K_1 and K_2.

II. MODIFIED ZERO NIJENHUIS TENSOR CONDITION

Here, we assume only that the antisymmetric tensor $F_{\mu\nu} = -F_{\nu\mu}$ is presymplectic and satisfies condition Eq. (1.21), i.e.,

$$\Delta_{\lambda\mu\nu}(F) = 0, \tag{2.1}$$
$$\Delta_{\lambda\mu\nu}(\tilde{F}^2) = 0, \tag{2.2}$$

where $(\tilde{F}^2)_{\mu\nu}$ is defined by

$$(\tilde{F}^2)_{\mu\nu} = (F^2)_{\mu\nu} + f_{\mu\nu}\left[\tfrac{1}{2}C_1 (K_1)^2 + C_2 K_2\right], \tag{2.3}$$

for some constants C_1 and C_2. For the special case of $C_1 = C_2 = 0$, these conditions reduce to that of the zero Nijenhuis tensor conditions as we have noted in Sec. I. Next, we first note that for $n \geqslant 1$ we have

$$K_n = (1/2n)\mathrm{Tr}\, S^n = (1/2n)(F^n)_{\mu\nu} f^{\nu\mu}. \tag{2.4}$$

Also, for simplicity, we will work hereafter in the canonical frame where $f_{\mu\nu}$ and $f^{\mu\nu}$ are constants. Multiplying $(F^n)_{\alpha\beta} f^{\alpha\lambda} f^{\beta\nu}$ to Eq. (2.1), we find that

$$2(S^n)^\beta_\alpha\, \partial_\beta S^\alpha_\mu = (S^n)^\alpha_\beta\, \partial_\mu S^\beta_\alpha = 2\,\partial_\mu K_{n+1},$$

which can be moreover rewritten as

$$\partial_\lambda (S^{n+1})^\lambda_\mu - S^\alpha_\mu\, \partial_\lambda (S^n)^\lambda_\alpha = \partial_\mu K_{n+1}, \tag{2.5}$$

after some calculations. Similarly, we multiply $(F^{n-1})_{\alpha\beta} f^{\alpha\lambda} f^{\beta\nu}$ to Eq. (2.2) to find now

$$2(S^{n-1})^\lambda_\alpha\, \partial_\lambda (S^2)^\alpha_\mu - (S^{n-1})^\alpha_\beta\, \partial_\mu (S^2)^\beta_\alpha$$
$$= \{2(n-1)K_{n-1} + 2N\delta_{n,1}\}\partial_\mu R - 2(S^{n-1})^\lambda_\mu \partial_\lambda R, \tag{2.6}$$

where we have set for simplicity

$$R = \tfrac{1}{2}C_1 (K_1)^2 + K_2, \tag{2.7}$$

and noted

$$\mathrm{Tr}\, S^n = (S^n)^\mu_\mu = 2nK_n + 2N\delta_{n,0}. \tag{2.8}$$

We next calculate

$$(S^{n-1})^\alpha_\beta \, \partial_\mu (S^2)^\beta_\alpha = 2(S^n)^\alpha_\beta \, \partial_\mu S^\beta_\alpha = 4 \, \partial_\mu K_{n+1},$$

as well as

$$\begin{aligned}(S^{n-1})^\lambda_\alpha \, \partial_\lambda (S^2)^\alpha_\mu &= \partial_\lambda [(S^{n+1})^\lambda_\mu] - (S^2)^\alpha_\mu \, \partial_\lambda (S^{n-1})^\lambda_\alpha \\ &= \partial_\mu K_{n+1} + S^\beta_\mu \{\partial_\lambda (S^n)^\lambda_\beta \\ &\quad - S^\alpha_\beta \, \partial_\lambda (S^{n-1})^\lambda_\alpha\} \\ &= \partial_\mu K_{n+1} + S^\beta_\mu \, \partial_\beta K_n,\end{aligned}$$

when we utilize Eq. (2.5). Inserting these relations to Eq. (2.6), we obtain

$$\begin{aligned}&S^\lambda_\mu \, \partial_\lambda K_n - \partial_\mu K_{n+1} \\ &\quad = \{(n-1)K_{n-1} + N\delta_{n,1}\}\partial_\mu R - (S^{n-1})^\lambda_\mu \, \partial_\lambda R.\end{aligned} \tag{2.9}$$

If $C_1 = C_2 = 0$ and hence $R = 0$, this reproduces, of course, the recursion relation Eq. (1.14).

Setting $n = 1$ in Eq. (2.9), we find

$$S^\lambda_\mu \, \partial_\lambda K_1 = \{1 + (N-1)C_2\}\partial_\mu K_2 + (N-1)C_1 K_1 \, \partial_\mu K_1, \tag{2.10}$$

while for $n \geqslant 2$, it gives

$$\begin{aligned}&S^\lambda_\mu \, \partial_\lambda K_n + (S^{n-1})^\lambda_\mu [C_1 K_1 \, \partial_\lambda K_1 + C_2 \, \partial_\lambda K_2] - \partial_\mu K_{n+1} \\ &\quad = (n-1)K_{n-1}\{C_1 K_1 \, \partial_\mu K_1 + C_2 \, \partial_\mu K_2\} \quad (n \geqslant 2).\end{aligned} \tag{2.11}$$

Now, we want to prove the following recursion relation that generalizes Eq. (1.14).

Proposition 1: Let us set

$$\theta_n = \begin{cases} 1 + (n-1)C_2 & (n \geqslant 1), \\ 1 + (N-1)C_2 = \theta_N & (n = 0). \end{cases} \tag{2.12}$$

Moreover, suppose that none of $\theta_1, \theta_2, \ldots, \theta_{\ell-1}$ for a positive integer ℓ vanishes. Then, we have

$$\theta_n S^\lambda_\mu \, \partial_\lambda K_n - \theta_{n-1} \, \partial_\mu K_{n+1} = \sum_{m=1}^{n} P_{nm} \, \partial_\mu K_m, \tag{2.13}$$

for any $1 \leqslant n \leqslant \ell$, where P_{nm} are polynomials of $K_1, K_2, \ldots, K_n$ of the form

$$P_{nm} = \sum_{n_1, n_2, \ldots} C_{n_1 n_2 \ldots} (K_1)^{n_1} (K_2)^{n_2} (K_3)^{n_3} \cdots, \tag{2.14a}$$

$$n_1 + 2n_2 + 3n_3 + \cdots = n - m + 1, \tag{2.14b}$$

for some constants $C_{n_1, n_2, \ldots}$ that may depend upon n and m.

Proof of Proposition 1: We shall prove a slightly more general formula;

$$\theta_n S^\lambda_\mu \, \partial_\lambda K_n - \theta_{n-1} \, \partial_\mu K_{n+1} = \sum_{m=1}^{n} P_{nm} \, \partial_\mu K_m, \tag{2.15a}$$

$$\theta_n (S^{n-1})^\lambda_\mu \, \partial_\lambda K_2 = \partial_\mu K_{n+1} + \sum_{m=1}^{n} Q_{nm} \, \partial_\mu K_m, \tag{2.15b}$$

$$\theta_n (S^n)^\lambda_\mu \, \partial_\lambda K_1 = \theta_0 \, \partial_\mu K_{n+1} + \sum_{m=1}^{n} R_{nm} \, \partial_\mu K_m, \tag{2.15c}$$

for $n \geqslant 1$, provided that none of $\theta_1, \theta_2, \ldots, \theta_{n-1}$ does not vanish. Here, Q_{nm} and R_{nm} are polynomials of $K_1, K_2, \ldots, K_n$ with the same structure as in Eq. (2.14), but with different coefficients in general. We shall now prove Eqs. (2.15) inductively. For $n = 1$, they are easily verified on the basis of Eq. (2.10) and (2.11) with $n = 2$. Suppose that they are valid for any n with $1 \leqslant n \leqslant \ell$ for some integer ℓ. We will prove then that they will hold also for $n = \ell + 1$, provided that $\theta_1, \theta_2, \ldots, \theta_\ell$ are nonzero. Setting $n = \ell + 1 \geqslant 2$ in Eq. (2.11), and multiplying by θ_ℓ, we find that

$$\begin{aligned}&\theta_\ell S^\lambda_\mu \, \partial_\lambda K_{\ell+1} - \theta_\ell \, \partial_\mu K_{\ell+2} + \theta_\ell (S^\ell)^\lambda_\mu \{C_1 K_1 \, \partial_\lambda K_1 \\ &\quad + C_2 \, \partial_\lambda K_2\} \\ &\quad = \ell\theta_\ell K_\ell \{C_1 K_1 \, \partial_\mu K_1 + C_2 \, \partial_\mu K_2\}.\end{aligned} \tag{2.16}$$

By the induction hypothesis, we have

$$\begin{aligned}&C_1 K_1 \theta_\ell (S^\ell)^\lambda_\mu \, \partial_\lambda K_1 \\ &\quad = C_1 K_1 \left\{\theta_0 \, \partial_\mu K_{\ell+1} + \sum_{m=1}^{\ell} R_{\ell m} \, \partial_\mu K_m\right\}.\end{aligned}$$

Also, we calculate

$$\begin{aligned}\theta_\ell (S^\ell)^\lambda_\mu \, \partial_\lambda K_2 &= \theta_\ell S^\alpha_\mu (S^{\ell-1})^\lambda_\alpha \, \partial_\lambda K_2 \\ &= S^\alpha_\mu \left\{\partial_\alpha K_{\ell+1} + \sum_{m=1}^{\ell} Q_{\ell m} \, \partial_\alpha K_m\right\} \\ &= S^\lambda_\mu \, \partial_\lambda K_{\ell+1} + \sum_{m=1}^{\ell} \left(\frac{Q_{\ell m}}{\theta_m}\right) \\ &\quad \times \left[\theta_{m-1} \, \partial_\mu K_{m+1} + \sum_{j=1}^{m} P_{mj} \, \partial_\mu K_j\right],\end{aligned} \tag{2.17}$$

by the induction hypothesis on Eqs. (2.15a) and (2.15b). Inserting these equations into Eq. (2.16), we obtain

$$\begin{aligned}&(\theta_\ell + C_2) S^\lambda_\mu \, \partial_\lambda K_{\ell+1} - \theta_\ell \, \partial_\mu K_{\ell+2} \\ &\quad = \ell\theta_\ell \{C_1 K_1 K_\ell \, \partial_\mu K_1 + C_2 K_\ell \, \partial_\mu K_2\} \\ &\qquad - C_1 \left\{\theta_0 K_1 \, \partial_\mu K_{\ell+1} + \sum_{m=1}^{\ell} K_1 R_{\ell m} \, \partial_\mu K_m\right\} \\ &\qquad - C_2 \sum_{m=1}^{\ell} \left(\frac{Q_{\ell m}}{\theta_m}\right)\left[\theta_{m-1} \, \partial_\mu K_{m+1} \right. \\ &\qquad \left. + \sum_{j=1}^{m} P_{mj} \, \partial_\mu K_j\right].\end{aligned}$$

Noting $\theta_\ell + C_2 = \theta_{\ell+1}$, this proves then that Eq. (2.15a) is also valid for $n = \ell + 1$ with

$$\sum_{m=1}^{\ell+1} P_{\ell+1,m}\,\partial_\mu K_m$$
$$= \ell\theta_\ell\{C_1K_1K_\ell\,\partial_\mu K_1 + C_2K_\ell\,\partial_\mu K_2\}$$
$$- C_1\left\{\theta_0K_1\,\partial_\mu K_{\ell+1} + \sum_{m=1}^{\ell} K_1R_{\ell m}\,\partial_\mu K_m\right\}$$
$$- C_2\sum_{m=1}^{\ell}\left(\frac{Q_{\ell m}}{\theta_m}\right)\left[\theta_{m-1}\,\partial_\mu K_{m+1} + \sum_{j=1}^{m} P_{mj}\,\partial_\mu K_j\right], \tag{2.18}$$

which determines $P_{\ell+1,m}$ inductively. Then, Eq. (2.17) can be rewritten as

$$\theta_{\ell+1}(S^\ell)^\lambda_\mu\partial_\lambda K_2 = \left(\frac{\theta_{\ell+1}}{\theta_\ell}\right)S^\lambda_\mu\partial_\lambda K_{\ell+1} + \frac{\theta_{\ell+1}}{\theta_\ell}\sum_{m=1}^{\ell}\left(\frac{Q_{\ell m}}{\theta_m}\right)\left[\theta_{m-1}\,\partial_\mu K_{m+1} + \sum_{j=1}^{m} P_{mj}\,\partial_\mu K_j\right]$$
$$= \frac{1}{\theta_\ell}\left\{\theta_\ell\,\partial_\mu K_{\ell+2} + \sum_{m=1}^{\ell+1} P_{\ell+1,m}\,\partial_\mu K_m\right\} + \frac{\theta_{\ell+1}}{\theta_\ell}\sum_{m=1}^{\ell}\left(\frac{Q_{\ell m}}{\theta_m}\right)\left[\theta_{m-1}\,\partial_\mu K_{m+1} + \sum_{j=1}^{m} P_{mj}\,\partial_\mu K_j\right],$$

by utilizing the result we have just now proved. This shows the validity of Eq. (2.15b) for $n = \ell+1$ with

$$\sum_{m=1}^{\ell+1} Q_{\ell+1,m}\,\partial_\mu K_m$$
$$= \frac{1}{\theta_\ell}\sum_{m=1}^{\ell+1} P_{\ell+1,m}\,\partial_\mu K_m + \frac{\theta_{\ell+1}}{\theta_\ell}\sum_{m=1}^{\ell}\left(\frac{Q_{\ell m}}{\theta_m}\right)$$
$$\times\left[\theta_{m-1}\,\partial_\mu K_{m+1} + \sum_{j=1}^{m} P_{mj}\,\partial_\mu K_j\right]. \tag{2.19}$$

Finally, multiplying $\theta_{\ell+1}(S^\ell)^\mu_\nu$ to Eq. (2.10) and changing the indices suitably, it gives

$$\theta_{\ell+1}(S^{\ell+1})^\lambda_\mu\,\partial_\lambda K_1 = \theta_0\theta_{\ell+1}(S^\ell)^\lambda_\mu\,\partial_\lambda K_2$$
$$+ (N-1)C_1\theta_{\ell+1}K_1(S^\ell)^\lambda_\mu\,\partial_\lambda K_1$$
$$= \theta_0\left\{\partial_\mu K_{\ell+2} + \sum_{m=1}^{\ell+1} Q_{\ell+1,m}\,\partial_\mu K_m\right\}$$
$$+ (N-1)C_1\theta_{\ell+1}K_1\frac{1}{\theta_\ell}$$
$$\times\left[\theta_0\,\partial_\mu K_{\ell+1} + \sum_{m=1}^{\ell} R_{\ell m}\,\partial_\mu K_m\right],$$

again by the induction hypothesis. This proves the validity of Eq. (2.15c) for $n = \ell+1$ with

$$\sum_{m=1}^{\ell+1} R_{\ell+1,m}\,\partial_\mu K_m$$
$$= \theta_0\sum_{m=1}^{\ell+1} Q_{\ell+1,m}\,\partial_\mu K_m + (N-1)C_1\frac{\theta_{\ell+1}}{\theta_\ell}$$
$$\times\left[\theta_0K_1\,\partial_\mu K_{\ell+1} + \sum_{m=1}^{\ell} K_1R_{\ell m}\,\partial_\mu K_m\right], \tag{2.20}$$

which will determine $R_{\ell,m}$ inductively. This completes the proof of Proposition 1.

Proposition 2: Suppose that $\theta_0,\theta_1,\theta_2,\ldots,\theta_{\ell-1}$ are nonzero. Then, for any two positive integers n, and m satisfying $1\leqslant n\leqslant\ell$ and $1\leqslant m\leqslant\ell$ we have

$$\{K_n,K_m\} = 0. \tag{2.21}$$

Proof: We first note the following identity,

$$f^{\mu\lambda}S^\nu_\lambda = -f^{\nu\lambda}S^\mu_\lambda, \tag{2.22}$$

when we note

$$f^{\mu\lambda}S^\nu_\lambda = f^{\mu\lambda}F_{\lambda\alpha}f^{\alpha\nu}.$$

Then, we calculate

$$f^{\mu\nu}S^\lambda_\mu\,\partial_\lambda K_n\,\partial_\nu K_m = -f^{\mu\lambda}S^\nu_\mu\,\partial_\lambda K_n\,\partial_\nu K_m$$
$$= -f^{\mu\lambda}S^\nu_\mu\,\partial_\nu K_m\,\partial_\lambda K_n$$
$$= -f^{\mu\nu}S^\lambda_\mu\,\partial_\lambda K_m\,\partial_\nu K_n,$$

where we interchanged indices ν and λ in the last step. Hence, we have

$$f^{\mu\nu}S^\lambda_\mu\,\partial_\lambda K_n\,\partial_\nu K_m = -f^{\mu\nu}S^\lambda_\mu\,\partial_\lambda K_m\,\partial_\nu K_n, \tag{2.23}$$

which is antisymmetric for interchange of n and m. Under these preparations, we now proceed to the proof of the Proposition by induction. For $\ell = 1$, there is nothing to prove. For $\ell = 2$, Eq. (2.10) gives

$$f^{\mu\nu}S^\lambda_\mu\,\partial_\lambda K_1\,\partial_\nu K_1 = \theta_0\{K_2,K_1\} + (N-1)C_1K_1\{K_1,K_1\}$$
$$= \theta_0\{K_2,K_1\}.$$

In view of Eq. (2.23) for $n = m = 1$, this leads to the desired result $\{K_2,K_1\} = 0$, provided $\theta_0 \neq 0$.

Now, suppose that we have

$$\{K_n,K_m\} = 0,$$

for any positive integer n and m satisfying $1\leqslant n$, $m\leqslant\ell$, provided that none of $\theta_0,\theta_1,\ldots,\theta_{\ell-1}$ vanishes. We shall then prove the same for $1\leqslant n\leqslant\ell+1$ and $1\leqslant m\leqslant\ell+1$, assuming $\theta_\ell\neq 0$ in addition, as follows. From Eq. (2.13), we calculate for n, $m\leqslant\ell$

$$\theta_nf^{\mu\nu}S^\lambda_\mu\,\partial_\lambda K_n\,\partial_\nu K_m - \theta_{n-1}\{K_{n+1},K_m\}$$
$$= \sum_{j=1}^{n} P_{nj}\{K_j,K_m\} = 0,$$

by the induction hypothesis. Using Eq. (2.23), this gives

$$\theta_m\theta_{n-1}\{K_{n+1},K_m\} = -\theta_n\theta_{m-1}\{K_{m+1},K_n\}$$
$$= \theta_n\theta_{m-1}\{K_n,K_{m+1}\}.$$

Repeating this relation successively by letting $n\to n-1$ and $m\to m+1$, we find

$$\{K_{n+1},K_m\}=\frac{\theta_n}{\theta_{n-1}}\frac{\theta_{m-1}}{\theta_m}\{K_n,K_{m+1}\}$$
$$=\frac{\theta_n}{\theta_{n-2}}\frac{\theta_{m-1}}{\theta_{m+1}}\{K_{n-1},K_{m+2}\}$$
$$=\cdots=\frac{\theta_n}{\theta_{n-k}}\frac{\theta_{m-1}}{\theta_{m+k-1}}\times\{K_{n-k+1},K_{m+k}\},$$

for any integer k satisfying $n\geqslant k$. We may now assume $n\geqslant m$ without loss of generality, since we may otherwise interchange the role of n and m. Then, choosing $k=n-m+1$, this gives

$$\{K_{n+1},K_m\}=\frac{\theta_n}{\theta_{m-1}}\frac{\theta_{m-1}}{\theta_n}\{K_m,K_{n+1}\}$$
$$=\{K_m,K_{n+1}\}=-\{K_{n+1},K_m\}=0.$$

Also, we have $\{K_{\ell+1},K_{\ell+1}\}=0$ trivially. This proves the validity of the induction hypothesis for ℓ being replaced by $\ell+1$. This completes the proof of Proposition 2.

Remark 1: If C_2 is not equal to any of the values, -1, $-\frac{1}{2}$, $-\frac{1}{3}$,... then θ_n for $n\geqslant 0$ can never become zero. Therefore, we conclude $\{K_n,K_m\}=0$ for any positive integers n and m for such a case. However, in the next section, we will show $C_2=-1/(N-2)$ for the periodic Toda lattice ($N\geqslant 3$), so that we find $\theta_{N-1}=0$ but $\theta_n\neq 0$ for $0\leqslant n\leqslant N-2$. For such a case, Eq. (2.13) with $n=N-1$ gives

$$-\theta_{N-2}\,\partial_\mu K_N=\sum_{m=1}^{N-1}P_{N-1,m}\,\partial_\mu K_m.$$

Since $P_{N-1,m}$ is a polynomial of $K_1,K_2,...,K_{N-1}$, this implies that K_N must be a polynomial of $K_1,K_2,...,K_{N-1}$. We will show in Sec. III that this is indeed the case for the periodic Toda lattice. However $K_1,K_2,...,K_{N-1}$ as well as one more quantity $\tilde{K}_N$ are shown to be algebracially independent and in involution.

III. PERIODIC TODA LATTICE

For simplicity, we will work again in the canonical coordinate frame and set

$$\bar{j}=j+N\quad(j=1,2,...,N),\qquad \overline{jk}=\bar{j}\,\bar{k},\tag{3.1}$$

as in Ref. 19. We seek a solution satisfying Eqs. (2.1)–(2.3) with an ansatz of

$$F_{jk}=a_{jk}\exp[\lambda_{jk}(q_j-q_k)]-a_{kj}\exp[\lambda_{kj}(q_k-q_j)],\tag{3.2a}$$
$$F_{j\bar{k}}=-\delta_{jk}(Ap_j+\xi H_1)-\xi p_j,\tag{3.2b}$$
$$F_{\bar{j}\bar{k}}=h_{jk}=-h_{kj},\tag{3.2c}$$

where we have set

$$H_1=\sum_{\ell=1}^{N}p_\ell,\tag{3.3}$$

and where, A,ξ, a_{jk}, λ_{jk}, and h_{jk} are some constants to be determined. It is easy to verify the validity of

$$\Delta_{\lambda\mu\nu}(F)=\partial_\lambda F_{\mu\nu}+\partial_\mu F_{\nu\lambda}+\partial_\nu F_{\lambda\mu}=0.\tag{3.4}$$

Next, we calculate

$$(F^2)_{jk}=\{A(p_j+p_k)+2\xi H_1\}F_{jk}+\xi p_k\sum_{m=1}^{N}F_{jm}-\xi p_j\sum_{m=1}^{N}F_{km},\tag{3.5a}$$
$$(F^2)_{\bar{j}\bar{k}}=\{A(p_j+p_k)+2\xi H_1\}h_{jk}+\xi\sum_{m=1}^{N}(h_{jm}-h_{km})p_m,\tag{3.5b}$$
$$(F^2)_{j\bar{k}}=-\sum_{m=1}^{N}h_{km}F_{jm}-\delta_{jk}(Ap_j+\xi H_1)^2-\xi p_j[A(p_j+p_k)+3\xi H_1].\tag{3.5c}$$

We then compute

$$K_1=\tfrac{1}{2}\operatorname{Tr}S=[A+(N+1)\xi]H_1,\tag{3.6a}$$
$$K_2=\frac{1}{4}\operatorname{Tr}(S^2)=\frac{1}{2}\sum_{j,k=1}^{N}h_{jk}F_{jk}+\frac{1}{2}A(A+2\xi)\sum_{k=1}^{N}(p_k)^2+\tfrac{1}{2}\xi[2A+(N+3)\xi](H_1)^2.\tag{3.6b}$$

We shall assume

$$A+(N+1)\xi\neq 0,\quad A(A+2\xi)\neq 0,\tag{3.7}$$

so that K_1 and K_2 are physically nontrivial. Then, when we choose constants C_1 and C_2 to be

$$C_2=\xi/(A+2\xi),\tag{3.8a}$$
$$C_1=-\xi^2/(A+2\xi)[A+(N+1)\xi],\tag{3.8b}$$

and set

$$(\tilde{F}^2)_{\mu\nu}=(F^2)_{\mu\nu}+f_{\mu\nu}[C_2K_2+\tfrac{1}{2}C_1(K_1)^2],\tag{3.9}$$

it is straightforward to find

$$\Delta_{jk\ell}(\tilde{F}^2)=\Delta_{\bar{j}k\ell}(\tilde{F}^2)=\Delta_{\bar{j}\bar{k}\ell}(\tilde{F}^2)=0,\tag{3.10a}$$

but

$$\Delta_{jk\bar{\ell}}(\tilde{F}^2)=\delta_{\ell j}\left\{AF_{jk}+\xi\sum_{m=1}^{N}F_{mk}+C_2\sum_{m=1}^{N}h_{km}F'_{km}\right\}-\delta_{\ell k}\left\{AF_{kj}+\xi\sum_{m=1}^{N}F_{mj}+C_2\sum_{m=1}^{N}h_{jm}F'_{jm}\right\}+2\xi F_{jk}+(h_{\ell j}-h_{\ell k})F'_{jk},\tag{3.10b}$$

where we have set

$$F'_{jk}=a_{jk}\lambda_{jk}\exp[\lambda_{jk}(q_j-q_k)]+a_{kj}\lambda_{kj}\exp[\lambda_{kj}(q_k-q_j)].\tag{3.11}$$

Suppose that F_{jk} satisfies a differential equation

$$(h_{\ell j}-h_{\ell k})F'_{jk}+2\xi F_{jk}+A(\delta_{\ell j}+\delta_{\ell k})F_{jk}=0.\tag{3.12}$$

Then, summing Eq. (3.12) for k with $\ell=k$, and noting Eq. (3.8a), it is easy to find

$$\xi\sum_{m=1}^{N}F_{mj}+C_2\sum_{m=1}^{N}h_{jm}F'_{jm}=0.$$

Therefore, Eq. (3.10b) leads to $\Delta_{jk\bar{\ell}}(\tilde{F}^2)=0$ and hence we will have Eq. (2.2), i.e.,

$$\Delta_{\lambda\mu\nu}(\tilde{F}^2)=0,\tag{3.13}$$

in view of Eq. (3.10a). Using Eqs. (3.2a) and (3.11), we can

rewrite Eq. (3.12) as

$$\{(h_j - h_k)\lambda_{jk} + 2\xi + A(\delta_{lj} + \delta_{lk})\}a_{jk} = 0. \quad (3.14)$$

In conclusion, if the constants A, ξ, λ_{jk}, h_{jk}, and a_{jk} satisfy the condition Eq. (3.14), then we have the desired relation Eq. (3.13). Equation (3.14) implies that whenever we have $a_{jk} \neq 0$, we must have

$$(h_j - h_k)\lambda_{jk} + 2\xi + A(\delta_{lj} + \delta_{lk}) = 0, \quad (3.15)$$

for arbitrary values of $l = 1,2,...,N$. This condition is very strong and essentially determines all these constants as we will see in the Appendix. For this purpose, it is convenient to use the diagrammatic approach familiar in the theory of Dynkin diagrams[22] of simple Lie algebras. When we have $a_{jk} \neq 0$, then we draw the diagram, Fig. 1, for a straight line joining two points j and k with the arrow direction from the point j to the point k. As we see from Eq. (3.14), both $a_{jk} \neq 0$, and $a_{kj} \neq 0$ are not in general possible. However, we can always reverse the direction of the arrow by using the new quantities $\tilde{a}_{kj}$ and $\tilde{\lambda}_{kj}$ defined by

$$\tilde{a}_{kj} = -a_{jk}, \quad \tilde{\lambda}_{kj} = -\lambda_{jk}, \quad (3.16)$$

since the numerical expression for F_{jk} remains unchanged under this transformation. Assuming $A \neq 0$ and excluding the trivial case of all λ_{jk} being zero, we will show in the Appendix that the solution of the problem allows only two possibilities of either the straight line joining 1,2,...,N successively or the circle joining 1,2,3,...,N as in Figs. 2 and 3.

Moreover, keeping the directions of the arrows continuously in the same direction, the solution for Fig. 2 corresponding to the simple Toda lattice will be shown in the Appendix to be given by

$$\begin{aligned} F_{jk} &= f_j \delta_{kj+1} \exp\{\lambda(q_j - q_k)\} \\ &\quad - f_k \delta_{j,k+1} \exp\{\lambda(q_k - q_j)\}, \\ F_{j\bar{k}} &= -\delta_{jk}(Ap_j + \xi H_1) - \xi p_j, \\ F_{\bar{j}\bar{k}} &= (1/\lambda)\{A\epsilon(k-j) + 2\xi(k-j)\}, \end{aligned} \quad (3.17)$$

where $\epsilon(k-j)$ is the sign function Eq. (1.18), and f_j, λ, A, ξ are arbitrary constants. Since A and ξ are arbitrary, it is more advantageous to assume $\xi = 0$, and $A = \lambda = 1$. Then, this solution reduces to the standard Toda lattice solution Eq. (1.17) with $C_1 = C_2 = 0$.

However, for the circular case of Fig. 3 corresponding to the periodic Toda lattice, the situation is slightly more stringent. We must first reinterpret $\delta_{j,k+1}$ in Eq. (3.17) to imply (also with $q_{N+1} = q_1$)

$$\delta_{j,k+1} = \begin{cases} 1, & \text{if } j = k+1 \text{ for } 1 \leqslant k \leqslant N-1, \\ 1, & \text{if } j = 1 \text{ and } k = N, \\ 0, & \text{otherwise.} \end{cases} \quad (3.18)$$

Second, A and ξ are no longer arbitrary, but we must have

$$\xi = -(1/N)A, \quad (3.19)$$

assuming $A \neq 0$. If we set $\lambda = A = 1$, then it reduces to the solution Eq. (1.23).

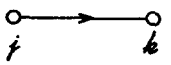


FIG. 1. The assignment of the line with the arrow when we have $a_{jk} \neq 0$. The open circles at the ends imply that the line may continue.

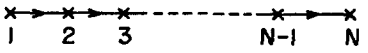


FIG. 2. The diagram corresponding to the situation when we have $a_{jj+1} \neq 0$ for all $j = 1,2,...,N-1$. The cross symbols at the end of lines imply that no other line can be joined to them.

We note that $\bar{\epsilon}(j-k)$ defined by

$$\bar{\epsilon}(j-k) = \epsilon(j-k) - (2/N)(j-k), \quad (3.20)$$

is invariant under cyclic permutation. To see it more clearly, let us define

$$\hat{j} = \begin{cases} j+1, & \text{if } 1 \leqslant j \leqslant N-1, \\ 1, & \text{if } j = N. \end{cases} \quad (3.21)$$

Then, it is not difficult to verify the identify

$$\bar{\epsilon}(\hat{j} - \hat{k}) = \bar{\epsilon}(j-k), \quad (3.22)$$

for $1 \leqslant j,k \leqslant N$, proving its cyclic invariance. We also note

$$\sum_{k=1}^{N} \bar{\epsilon}(j-k) = 0. \quad (3.23)$$

Returning to the general case, we will give explicit forms for K_1, K_2, and K_3,

$$K_1 = [A + (N+1)\xi] \sum_{j=1}^{N} p_j, \quad (3.24a)$$

$$\begin{aligned} K_2 = \frac{1}{2}(A+2\xi)\left\{A\sum_{j=1}^{N}(p_j)^2 + \frac{2}{\lambda}\sum_{j=1}^{N} f_j e^{\lambda(q_j - q_{j+1})}\right\} \\ + \tfrac{1}{2}\xi[2A + (N+3)\xi](H_1)^2, \end{aligned} \quad (3.24b)$$

$$\begin{aligned} K_3 = \frac{1}{3}(A+3\xi)\left\{A^2\sum_{j=1}^{N}(p_j)^3 + 3A\xi H_1 \sum_{j=1}^{N}(p_j)^2 \right. \\ \left. + \frac{3}{\lambda}\sum_{j=1}^{N}[A(p_j + p_{j+1}) + 2\xi H_1] f_j e^{\lambda(q_j - q_{j+1})}\right\} \\ + \tfrac{1}{3}\xi^2[3A + (N+7)\xi](H_1)^3. \end{aligned} \quad (3.24c)$$

Hereafter, we will concentrate our attention only to the case of the periodic Toda lattice solution ($N \geqslant 3$). Because of Eqs. (3.8) and (3.19), we have

$$-C_1 = C_2 = -1/(N-2), \quad (3.25)$$

so that θ_n defined by Eq. (2.12) has the value

$$\theta_n = (N-n-1)/(N-2). \quad (3.26)$$

Especially, we see that

$$\theta_{N-1} = 0, \quad (3.27)$$

while all other θ_n ($n \geqslant 0$) are nonzero. Because of the Proposition 2 of Sec. II, this implies then

$$\{K_n, K_m\} = 0, \quad (3.28)$$

for any n and m satisfying $1 \leqslant n, m \leqslant N-1$. Moreover, Remark 1 of the same section suggests that K_N will be a polyno-

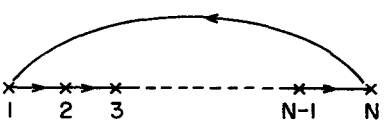


FIG. 3. The diagram corresponding to the case in which we have $a_{jj+1} \neq 0$ for $j = 1,2,...,N-1$ and $a_{N,1} \neq 0$.

mial of $K_1, K_2, \ldots, K_{N-1}$. Here, we will show first that $K_1, K_2, \ldots, K_{N-1}$ are algebracially independent, while K_N is indeed a polynomial of these $N-1$ quantities for the periodic Toda lattice. To prove the algebraic independence, it suffices to regard K_n's as a polynomial of momentum variables p_j, discarding the q_j dependent terms. By induction, we can show easily that

$$(F^n)_{j\bar{k}} = -\delta_{jk} A^n (p_j)^n - \xi A^{n-1} \sum_{m=1}^{n} (p_j)^m (p_k)^{n-m} + \cdots,$$

for $n \geqslant 1$ where terms neglected are either lower-polynomials of p_j's or terms containing $H_1 = \Sigma_{j=1}^N p_j$, as well as q_j-dependent quantities. Therefore, we calculate

$$K_n = -\frac{1}{n} \sum_{j=1}^{N} (F^n)_{j\bar{j}} = \frac{1}{n} A^{n-1} (A + n\xi) \sum_{j=1}^{N} (p_j)^n + \cdots,$$

for $n \geqslant 1$, where terms neglected are those which contain functions only of lower-order polynmial $\Sigma_{j=1}^N (p_j)^m$ with $m \leqslant n-1$. For the periodic Toda lattice, we have $A = -N\xi$, so that the coefficients of $\Sigma_{j=1}^N (p_j)^n$ is nonzero for $n = 1,2,\ldots,N-1$ but vanishes for $n = N$. This proves the algebraic independence of $K_1, K_2, \ldots, K_{N-1}$. In order to prove that K_N is a polynomial of $K_1, K_2, \ldots, K_{N-1}$, we regard S_μ^ν, $F_{\mu\nu}$, and $f_{\mu\nu}$ for $\mu, \nu = 1,2,\ldots,2N$ be $2N \times 2N$ matrices, and consider the secular equation

$$\det(S - \lambda I) = 0,$$

where I is the $2N \times 2N$ unit matrix. However, since

$$\det(S - \lambda I) = \det(Ff^{-1} - \lambda I) = \det(F - \lambda f),$$

and $2N \times 2N$ matrix $F - \lambda f$ is now antisymmetric, its determinant is a square of its phaffian. Therefore, we conclude that eigenvalue λ must always appear in the pairs. Labelling them as (λ_1, λ_1), $(\lambda_2, \lambda_2), \ldots, (\lambda_N, \lambda_N)$, then we have

$$K_n = \frac{1}{2n} \mathrm{Tr}(S^n) = \frac{1}{n} \sum_{j=1}^{N} (\lambda_j)^n, \tag{3.29}$$

for $n \geqslant 1$. Next, we shall prove that $\lambda_1 = \xi H_1 = K_1$ is an eigenvalue of S_μ^ν for the periodic Toda lattice. Indeed, consider the eigenvalue problem

$$S_\mu^\nu \phi_\nu = \xi H_1 \phi_\mu,$$

whose nontrivial solution is indeed given by

$$\phi_j = 1, \quad \phi_{\bar{j}} = 0,$$

for all $j = 1,2,\ldots,N$, when we note Eq. (3.23). Because $\lambda_1 = K_1$, and $K_1 = \Sigma_{j=1}^N \lambda_j$, we are forced to conclude that

$$\lambda_2 + \lambda_3 + \cdots + \lambda_N = 0, \tag{3.30}$$

so that K_N must be a polynomial of $K_1, K_2, \ldots, K_{N-1}$. For $N = 3$, Eq. (3.24c) with $A = -3\xi$ indeed leads to

$$K_3 = \tfrac{1}{3}\xi^3 (H_1)^3 = \tfrac{1}{3}(K_1)^3.$$

Similarly, for $N = 4$, we can verify

$$K_4 = \tfrac{1}{2}(K_2)^2 - \tfrac{1}{2}K_1^2 K_2 + \tfrac{1}{8}(K_1)^4.$$

Since we have only $N-1$ algebraically independent functions $K_1, K_2, \ldots, K_{N-1}$ in involution, we must find additional term $\bar{K}_N$ that is algebracially independent of and in involution with, all these $N-1$ terms, if the periodic Toda lattice is integrable. We can fortunately find such an extra term as follows. Let us suppose for a moment that we have $f_N = 0$. Then, since we have $F_{N,1} = 0$, it corresponds to the straight chain diagram Fig. 2 rather than Fig. 3. In that case, there is no reason to set $A = -N\xi$. However, in order to avoid confusion, let us write K_n explicitly as $K_n(\xi, A)$ as functions of two unconstrained variables ξ and A. We know then we have

$$\{K_N(\xi, A), K_n(\xi, A)\} = 0, \tag{3.31}$$

for all $n = 1,2,\ldots,N-1$. Moreover, if we set $A = -N\xi$, then $K_N(\xi, A)$ reduces to a polynomial $P(K_1, K_2, \ldots, K_{N-1}) \equiv P(\xi, A)$. Therefore, we may set

$$K_N(\xi, A) = (A + N\xi)\, \bar{K}_N(\xi, A) + P(\xi, A). \tag{3.32}$$

We note that $\bar{K}_N(\xi, A)$ contains a term proportional to $\Sigma_{j=1}^N (p_j)^N$ and does not vanish for $A = -N\xi$. Therefore, $\bar{K}_N(\xi, A)$ is algebraically independent of $K_1, K_2, \ldots, K_{N-1}$ even in the limit $A = -N\xi$. Moreover, since $P(\xi, A)$ is a polynomial of $K_1, K_2, \ldots, K_{N-1}$, we must have

$$\{P(\xi, A), K_n(\xi, A)\} = 0,$$

for all $n = 1,2,\ldots,N-1$. Hence Eqs. (3.31) and (3.32) require the validity of

$$\{\bar{K}_N(\xi, A), K_n(\xi, A)\} = 0.$$

Now setting $A = -N\xi$, and dropping the ξ-dependence of functions, this leads to

$$\{\bar{K}_N, K_n\} = 0,$$

for all $n = 1,2,\ldots,N-1$. However, as we noted in the beginning of the argument, this presupposes $f_N = 0$. Hence, if $f_N \neq 0$, we conclude that $\{\bar{K}_N, K_n\}$ must be proportional to f_N for the original periodic Toda lattice. We now note that the periodic Toda lattice is cyclic invariant. Hence, replacing $1 \to 2 \to 3 \to \cdots \to N \to 1$, and so on successively, we conclude that $\{\bar{K}_N, K_n\}$ must be proportional to the product $f_1 f_2 \cdots f_N$ of all coupling parameters. On the other hand, it is easy to see that K_n as a function of $f_1, f_2, \ldots, f_N$ is a polynomial of degree $[n/2]$, where $[n/2]$ is the largest integer contained in $n/2$. Hence, $\{\bar{K}_N, K_n\}$ for $n = 1,2,\ldots,N-1$ is a function of $f_1, f_2, \ldots, f_N$ with the degree less than N but never equal to N. Therefore, we conclude that we must have $\{\bar{K}_N, K_n\} = 0$ for all $n = 1,2,\ldots,N-1$. Hence, we have found N algebraically independent functions $K_1, K_2, \ldots, K_{N-1}$ and $\bar{K}_N$ that are in involution, and hence the periodic Toda lattice is integrable.

In ending this section, we note the following. First, we set

$$\bar{K}_2 = [1/(N-2)]\{K_2 - \tfrac{1}{2}(K_1)^2\}, \tag{3.33a}$$

$$\bar{K}_3 = [1/(N-3)]\{K_3 - \tfrac{1}{3}(K_1)^3\}, \tag{3.33b}$$

$$\bar{K}_4 = [1/(N-4)]\{K_4 + \tfrac{1}{2}K_2[(K_1)^2 - K_2] - \tfrac{3}{8}(K_1)^4\}. \tag{3.33c}$$

Then, the recursion relations Eq. (2.13) for $n = 1,2$, and 3 can be simplified to become

$$S_\mu^\lambda \partial_\lambda K_1 = \partial_\mu \{\tfrac{1}{2}(K_1)^2 - \bar{K}_2\}, \tag{3.34a}$$

$$S_\mu^\lambda \partial_\lambda \bar{K}_2 = \partial_\mu \bar{K}_3, \tag{3.34b}$$

$$S_\mu^\lambda \partial_\lambda \bar{K}_3 = \partial_\mu \{\bar{K}_4 + \tfrac{1}{2}(N-2)(\bar{K}_2)^2\}, \tag{3.34c}$$

which are reminiscent of Eq. (1.14). Note that each of $(N-2)\bar{K}_2$, $(N-3)\bar{K}_3$ and $(N-4)\bar{K}_4$ vanishes for

$N = 2,3$, and 4, respectively. The validity of Eqs. (3.34) implies that the Lax equation can be derived from these as their integrability conditions. Moreover, we can find more than two algebraically independent Lagrangians which nevertheless lead to the same equations of motion. For details, see Refs. 18 and 19.

Secondly, we can also find an analog of Eq. (1.16) for the periodic Toda lattice solution at least for $n = 1,2,3$, and 4 as follows. We set

$$(\tilde{F}^2)_{\mu\nu} = (F^2)_{\mu\nu} - \bar{K}_2 f_{\mu\nu}, \tag{3.35a}$$

$$(\tilde{F}^3)_{\mu\nu} = (F^3)_{\mu\nu} - 2\bar{K}_2 F_{\mu\nu} - \bar{K}_3 f_{\mu\nu}, \tag{3.35b}$$

$$(\tilde{F}^4)_{\mu\nu} = (F^4)_{\mu\nu} - 3\bar{K}_2 (F^2)_{\mu\nu} - 2\bar{K}_3 F_{\mu\nu} - \{\bar{K}_4 + \tfrac{1}{2}(N-5)(\bar{K}_2)^2\} f_{\mu\nu}. \tag{3.35c}$$

Then, modifying the proof of Eq. (1.16) in Refs. 19 and 20, we can prove the validity of

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(\tilde{F}^2) = \Delta_{\lambda\mu\nu}(\tilde{F}^3) = \Delta_{\lambda\mu\nu}(\tilde{F}^4) = 0, \tag{3.36}$$

which now replace Eq. (1.16). Although we suspect that Eqs. (3.34) and (3.36) will hold for more general cases, we could not prove the conjecture.

Lastly, for the special case of $N = 3$, we can verify the validity of

$$(F^3)_{\mu\nu} - K_1 (F^2)_{\mu\nu} - \left[K_2 - \tfrac{1}{2}(K_1)^2\right](F_{\mu\nu} - K_1 f_{\mu\nu}) = 0, \tag{3.37}$$

which is equivalent to a statement that the 6×6 matrix S^{ν}_{μ} is completely diagonalizeable. Presumably, we may also find analogous formulas for general N.

IV. OTHER EXAMPLES

We will now give a few examples that satisfy the condition Eq. (1.21) but are not necessarily integrable nevertheless since K_n for $n \geqslant 3$ turn out to polynomials of K_1 and K_2.

A. Model A

Let $\xi, \bar{\xi}, \eta, b_j$, and $C_{jk} = -C_{kj}$ be arbitrary constants, and set

$$F_{j\bar{k}} = b_j - b_k, \tag{4.1a}$$

$$F_{j\bar{k}} = -\delta_{jk}\xi H_1 - \xi p_j - \bar{\xi} p_k - \eta H_1, \tag{4.1b}$$

$$F_{jk} = \psi_j - \psi_k, \tag{4.1c}$$

where H_1 and ψ_j are given by

$$H_1 = \sum_{k=1}^{N} p_k, \tag{4.2a}$$

$$\psi_j = \sum_{\ell=1}^{N} C_{j\ell} \exp\left[\frac{2\xi}{b_j - b_\ell}(q_j - q_\ell)\right]. \tag{4.2b}$$

Now, define $(\tilde{F}^2)_{\mu\nu}$ by

$$(\tilde{F}^2)_{\mu\nu} = (F^2)_{\mu\nu} - (\bar{\xi} + N\eta) H_1 F_{\mu\nu} + \tfrac{1}{2} f_{\mu\nu}\{K_2 - b(H_1)^2\}, \tag{4.3a}$$

with

$$b = \tfrac{1}{2}N(N+1)\xi^2 + \eta[(N-1)\xi + \bar{\xi}N] + \tfrac{1}{2}\bar{\xi}^2 + \tfrac{1}{2}\eta^2 N^2. \tag{4.3b}$$

Then, after some calculations, we can verify the validity of

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(\tilde{F}^2) = 0. \tag{4.4}$$

Moreover, if we choose constants to satisfy $\bar{\xi} + N\eta = 0$, then $(\tilde{F}^2)_{\mu\nu}$ has the same form as Eq. (1.22). However, the present model may not necessarily be integrable for the following reason. We can verify the validity of a cubic identity

$$(F^3)_{\mu\nu} = (4\xi + \bar{\xi} + \eta N) H_1 (F^2)_{\mu\nu} + H_2 F_{\mu\nu} + H_3 f_{\mu\nu}, \tag{4.5}$$

by a straightforward but tedious calculation, where H_2 and H_3 are defined by

$$\begin{aligned} H_2 &= K_2 - \tfrac{1}{2}(H_1)^2[(N+13)\xi^2 + \bar{\xi}^2 + 8\xi\bar{\xi} \\ &\quad + 2(4\xi + \bar{\xi})\eta N + \eta^2 N^2] \\ &= N\sum_{j=1}^{N} b_j \psi_j + N\xi\bar{\xi}\sum_{j=1}^{N}(p_j)^2 \\ &\quad - \xi(5\xi + 2\eta N + 3\bar{\xi})(H_1)^2 \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} H_3 &= -\xi H_1 K_2 + \tfrac{1}{2}\xi[(N+7)\xi^2 + \bar{\xi}^2 + 6\xi\bar{\xi} \\ &\quad + 2(3\xi + \bar{\xi})\eta N + \eta^2 N^2](H_1)^3. \end{aligned} \tag{4.7}$$

Multiplying $f^{\nu\mu}$ to Eq. (4.5), we find then

$$\begin{aligned} K_3 &= \tfrac{1}{6}\operatorname{Tr} S^3 \\ &= (3\xi + \bar{\xi} + N\eta) H_1 K_2 - \tfrac{1}{6}\{(3\xi + \bar{\xi} + N\eta)^3 \\ &\quad + (N-2)\xi^2[7\xi + 3\bar{\xi} + 3N\eta]\}(H_1)^3, \end{aligned} \tag{4.8}$$

so that it is a polynomial of $K_1 = (2\xi + \bar{\xi} + \eta)H_1$ and K_2. Moreover, by repeated use of Eq. (4.5), we can similarly conclude that all $K_n (n \geqslant 3)$ are polynomials of K_1 and K_2. Especially, the involution property $\{K_n, K_m\} = 0$ guaranteed by Proposition 2 (with $C_2 = \frac{1}{2}$ and hence $\theta_n \neq 0$) is rather trivial. Especially, the present model is not necessarily integrable except for the simple case of $N = 2$.

B. Model B: Coulomb potential in three dimensions

The Hamiltonian of the three dimensional Coulomb potential is as usual given by

$$H = (1/2m)\mathbf{p}^2 + \alpha/r, \tag{4.9}$$

for mass m and coupling parameter α, where we have set

$$\mathbf{p}^2 = p_1^2 + p_2^2 + p_3^2, \tag{4.10a}$$

$$r = [q_1^2 + q_2^2 + q_3^2]^{1/2}. \tag{4.10b}$$

Now, define $F_{\mu\nu}(\mu,\nu = 1,2,\ldots,6)$ by

$$F_{jk} = m\alpha \sum_{\ell=1}^{3} \epsilon_{jk\ell}\frac{q_\ell}{r^3}, \tag{4.11a}$$

$$F_{j\bar{k}} = \sum_{\ell=1}^{3} \epsilon_{jk\ell} p_\ell, \tag{4.11b}$$

$$F_{\bar{j}\bar{k}} = -2\sum_{\ell=1}^{3} \epsilon_{jk\ell} q_\ell, \tag{4.11c}$$

where $\epsilon_{jk\ell}$ is the three-dimensional Levi-Civita symbol. Then, we find first $\Delta_{\lambda\mu\nu}(F) = 0$ when we note the following identity valid only for $N = 3$,

$$\epsilon_{jkm} q_\ell + \epsilon_{k\ell m} q_j + \epsilon_{\ell jm} q_k = \epsilon_{jk\ell} q_m.$$

Moreover, we have

$$K_1 = \tfrac{1}{2}\,\mathrm{Tr}\,S = 0, \tag{4.12a}$$

$$K_2 = \tfrac{1}{4}\,\mathrm{Tr}\,S^2 = -\underline{p}^2 - 2m\alpha/r = -2mH. \tag{4.12b}$$

Setting further

$$(\tilde{F}^2)_{\mu\nu} = (F^2)_{\mu\nu} - \tfrac{1}{2}K_2 f_{\mu\nu}, \tag{4.13}$$

we can verify the validity of

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(\tilde{F}^2) = 0, \tag{4.14}$$

with $C_2 = -\frac{1}{2}$, $C_1 = 0$, $N = 3$, and $\theta_0 = \theta_3 = 0$. Then, from the results of Sec. II, K_3 must be proportional to $K_1 K_2$ so that it must be identically zero. More generally, we find the validity of a special third-order polynomial identity

$$(F^3)_{\mu\nu} - K_2 F_{\mu\nu} = 0, \tag{4.15}$$

and hence all K_n ($n \geqslant 3$) are functions of K_2. For example, we have

$$K_3 = 0, \quad K_4 = \tfrac{1}{2}(K_2)^2,$$

from Eq. (4.15). Then, the recursion relation Eq. (2.11) for $n = 2$ becomes

$$S^\lambda_\mu\, \partial_\lambda K_2 = 0, \tag{4.16}$$

while Eq. (2.10) is trivially satisfied.

Further, if we set

$$\mathbf{L}^2 = (\mathbf{q}\times\mathbf{p})^2 = r^2\mathbf{p}^2 - (\mathbf{q}\cdot\mathbf{p})^2, \tag{4.17}$$

we find also a rather peculiar relation

$$(S^2)^\lambda_\mu\, \partial_\lambda \mathbf{L}^2 = \partial_\mu (K_2 \cdot \mathbf{L}^2), \tag{4.18}$$

whose implication is, however, obscure at the moment. This model is, nevertheless, integrable since H, $\mathbf{L}^2$, and $L_3 = q_1 p_2 - q_2 p_1$ are algebracially independent and in involution.

C. Model C: Harmonic oscillator

Returning to the original N-dimensional case, we now consider

$$F_{jk} = A(q_j p_k - q_k p_j), \tag{4.19a}$$

$$F_{j\bar{k}} = B(q_j p_k - q_k p_j), \tag{4.19b}$$

$$F_{j\bar{k}} = A q_j q_k + B p_j p_k, \tag{4.19c}$$

for $j,k = 1,2,\ldots,N$, where A and B are some constants. Then,

$$K_1 = \tfrac{1}{2}\,\mathrm{Tr}\,S = -[A\mathbf{q}^2 + B\mathbf{p}^2] \tag{4.20a}$$

$$K_2 = \tfrac{1}{4}\,\mathrm{Tr}\,S^2 = \tfrac{1}{2}(K_1)^2. \tag{4.20b}$$

Moreover, we have a simple quadratic equation

$$(F^2)_{\mu\nu} = K_1 F_{\mu\nu}, \tag{4.21}$$

so that all K_n ($n \geqslant 2$) are polynomials of K_1. We can verify the presymplectic tensor condition,

$$\Delta_{\lambda\mu\nu}(F) = 0, \tag{4.22}$$

Also, if we set

$$(\tilde{F}^2)_{\mu\nu} = (F^2)_{\mu\nu} - K_1 F_{\mu\nu}, \tag{4.23}$$

it is identically zero by Eq. (4.21) so that we have trivially

$$\Delta_{\lambda\mu\nu}(\tilde{F}^2) = 0, \tag{4.24}$$

also. Therefore, this case satisfies conditions $\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(\tilde{F}^2) = 0$. Although the explicit form of $(\tilde{F}^2)_{\mu\nu}$ given by Eq. (4.23) is slightly different from that discussed in Sec. II, we can proceed in essentially the same way. In this case, we find

$$(S^2)^\lambda_\mu\, \partial_\lambda K_1 = K_1 S^\lambda_\mu\, \partial_\lambda K_1, \tag{4.25}$$

which can be split into

$$S^\lambda_\mu\, \partial_\lambda \mathbf{L}^2 = 0, \tag{4.26a}$$

$$S^\lambda_\mu\, \partial_\lambda K_1 = \partial_\mu\{\tfrac{1}{2}(K_1)^2 - 2AB\,\mathbf{L}^2\}, \tag{4.26b}$$

when we set

$$\mathbf{L}^2 = \mathbf{q}^2\cdot\mathbf{p}^2 - (\mathbf{q}\cdot\mathbf{p})^2. \tag{4.27}$$

Again, the reason for the validity of Eqs. (4.26) is not so clear at the present. Also, instead of Eqs. (4.19), we may consider

$$F_{jk} = A(q_j p_k - q_k p_j), \tag{4.28a}$$

$$F_{j\bar{k}} = B(q_j p_k - q_k p_j), \tag{4.28b}$$

$$F_{j\bar{k}} = -A(q_j q_k + \delta_{jk}\mathbf{q}^2) - B(p_j p_k + \delta_{jk}\mathbf{p}^2), \tag{4.28c}$$

which satisfy now

$$\Delta_{\lambda\mu\nu}(F) = 0, \tag{4.29}$$

as well as a modified quadratic equation

$$(F^2)_{\mu\nu} = [3/(N+1)]K_1 F_{\mu\nu} - [2/(N+1)^2](K_1)^2 f_{\mu\nu}, \tag{4.30}$$

with

$$K_1 = \tfrac{1}{2}\,\mathrm{Tr}\,S = (N+1)[A\mathbf{q}^2 + B\mathbf{p}^2]. \tag{4.31}$$

There are a few other choices in which we can have zero Nijenhuis tensor itself.

1. Choice (i)

$$F_{jk} = 0, \quad F_{j\bar{k}} = q_j p_k - q_k p_j, \quad F_{j\bar{k}} = p_j p_k, \tag{4.32}$$

where we have

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(F^2) = 0,$$

with a quadratic equation

$$(F^2)_{\mu\nu} = -\mathbf{p}^2 F_{\mu\nu},$$

as well as

$$S^\lambda_\mu \partial_\lambda \mathbf{L}^2 = 0.$$

2. Case (ii)

$$F_{jk} = 0, \quad F_{j\bar{k}} = -2(q_j p_k - q_k p_j), \quad F_{j\bar{k}} = \delta_{jk}\mathbf{p}^2, \tag{4.33}$$

where we have

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(F^2) = 0,$$

now with a cubic equation

$$(F^3)_{\mu\nu} = (\mathbf{p}^2)^2 F_{\mu\nu}.$$

ACKNOWLEDGMENT

This work is supported in part by the U.S. Department of Energy Contract No. DE-AC02-76ER13065.

APPENDIX: SOLUTIONS OF EQ. (3.14)

We shall prove that only solutions of Eq. (3.14) are limited to those corresponding to Figs. 2 and 3. First, we

recall our convention that we draw a directed arrow from the point j to the point k as in Fig. 1, only when we have $a_{jk} \neq 0 (j \neq k)$. Now, suppose that we have $a_{12} a_{23} \neq 0$ corresponding to Fig. 4. Then, first setting $j = 1$ and $k = 2$ in Eq. (3.14), we find

$$(h_{\ell} - h_{\ell 2})\lambda_{12} + A(\delta_{\ell} + \delta_{\ell 2}) + 2\xi = 0. \tag{A1}$$

Choosing $\ell = 1$ and $\ell = 3$, this gives

$$h_{12}\lambda_{12} = A + 2\xi,$$
$$(h_{13} - h_{23})\lambda_{12} = 2\xi, \tag{A2}$$

respectively. Similarly, $a_{23} \neq 0$ in Eq. (3.14) leads to

$$(h_{\ell 2} - h_{\ell 3})\lambda_{23} + A(\delta_{\ell 2} + \delta_{\ell 3}) + 2\xi = 0,$$

which gives

$$h_{23}\lambda_{23} = A + 2\xi,$$
$$(h_{12} - h_{13})\lambda_{23} + 2\xi = 0, \tag{A3}$$

for $\ell = 2$ and $\ell = 1$, respectively. Since we are assuming $\lambda_{jk} \neq 0$ always, it is easy to find that Eqs. (A2) and (A3) lead to

$$\lambda_{12} = \lambda_{23}, \tag{A4}$$

as well as

$$h_{12} = h_{23} = (A + 2\xi)/\lambda_{12}, \quad h_{13} = (1/\lambda_{12})(A + 4\xi). \tag{A5}$$

It is also evident that we cannot have in general both $a_{12} \neq 0$ and $a_{21} \neq 0$.

We shall now show that any diagram cannot contain a branch line. Suppose that we have a diagram corresponding to Fig. 5.

Then, considering two chains $1 \to 2 \to 3$ and $1 \to 2 \to 4$, we conclude $\lambda_{12} = \lambda_{23} = \lambda_{24}$ from Eq. (A4). Next, we reverse the direction from the point 2 to 4 by setting $\tilde{\lambda}_{42} = -\lambda_{24}$ as has been explained in Eq. (3.16). Then, applying Eq. (A4) to the chain $4 \to 2 \to 3$, it gives $\tilde{\lambda}_{42} = \lambda_{23}$ so that $\lambda_{24} = -\lambda_{23}$. Comparing this to the previous result, we fine $\lambda_{12} = \lambda_{23} = \lambda_{24} = 0$ identically in contradiction to our hypothesis. Therefore, we conclude that only possible diagrams should not contain any branch and hence only Figs. 2 and 3 are possible by relabelling indices suitably.

Next, suppose that we have the Fig. 2, i.e., the chain $1 \to 2 \to 3 \to \cdots \to N$. Then, starting with our result of Eqs. (A4) and (A5), it is not difficult to show by induction on Eq. (3.14) that we have

$$\lambda_{12} = \lambda_{23} = \cdots = \lambda_{N-1,N} \equiv \lambda, \tag{A6}$$

and

$$h_{jk} = (1/\lambda)\{A\epsilon(k-j) + 2\xi(k-j)\}, \tag{A7}$$

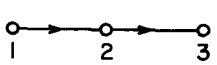


FIG. 4. The diagram when we have $a_{12} \neq 0$ an $a_{23} \neq 0$. The open circles imply that the end may be joined by other lines.

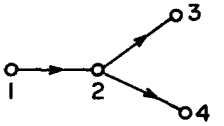


FIG. 5. The diagram when we have $a_{12} \neq 0$, $a_{23} \neq 0$, and $a_{24} \neq 0$. This diagram is not possible as is explained in the text.

for any j, and k satisfying $1 \leqslant j,k \leqslant N$, where $\epsilon(k-j)$ is the sign function defined by Eq. (1.18).

When we are considering the cyclic case of Fig. 3, then we have additional condition $a_{N1} \neq 0$. Applying Eqs. (A4) and (A5), to the chain $N \to 1 \to 2$, we obtain $\lambda_{N1} = \lambda_{12} = \lambda$ as well as

$$h_{N1} = (1/\lambda)[A + 2\xi].$$

However, from Eq. (A7), we know

$$h_{N1} = -h_{1N} = -(1/\lambda)\{A + 2\xi(N-1)\}.$$

Both are consistent only if $A = -\xi N$. We can verify that this is the only additional condition. Therefore, for the cyclic case corresponding to Fig. 3, we have

$$\lambda_{12} = \lambda_{23} = \cdots = \lambda_{N-1,N} = \lambda_{N,1} = \lambda, \tag{A8}$$
$$A = -N\xi, \tag{A9}$$
$$h_{jk} = (\xi/\lambda)\{2(k-j) - N\epsilon(k-j)\}, \tag{A10}$$

for $1 \leqslant j,k \leqslant N$.

This completes the proof of the results stated in Sec. III.

[1] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[2] M. Toda, *Theory of Nonlinear Lattices* (Springer, New York, 1981).
[3] M. A. Olshansky and A. M. Perelomov, Phys. Rep. **71**, 315 (1981).
[4] A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1988).
[5] L. D. Fadeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, New York, 1987).
[6] A. Das, *Integrable Models* (World Scientific, Singapore, 1989).
[7] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics; A Modern Approach* (Wiley, New York, 1974).
[8] P. Libermann and C-M. Marle, *Symplectic Geometry and Analytical Mechanics*, translated by P. E. Schwarzback (Reidel, Dordrecht, 1987).
[9] R. Abraham and J. E. Marsden, *Foundation of Mechanics* (Benjamin–Cummings, Reading, MA, 1978).
[10] P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968); **28**, 141 (1975).
[11] V. I. Arnold, *Mathmatical Methods of Classical Mechanics* (Springer, New York, 1978).
[12] W. D. Curtis and F. R. Miller, *Differential Manifolds and Theoretical Physics* (Academic, New York, 1974).
[13] D. Olive and N. Turok, Nucl. Phys. B **220**, 491 (1983); **265**, 469 (1986).
[14] I. M. Gel'fand and I. Y. Dorfman, Funct. Anal. Appl. **13**, 248 (1979).
[15] S. DeFilippo, G. Vilasi, G. Marmo, and M. Salerno, Nuovo Cimento B **83**, 97 (1984).
[16] P. Antonini, G. Marmo, and C. Rubano, Nuovo Cimento B **86**, 17 (1985).
[17] F. Magri, C. Morosi, and O. Ragnisco, Commun. Math. Phys. **99**, 115 (1985).
[18] A. Das and S. Okubo, Ann. Phys. (NY) **190**, 215 (1989).
[19] S. Okubo, J. Math. Phys. **30**, 834 and 1043 (1989).
[20] S. Okubo, Algebra Groups Geometr. **6**, 65 (1989).
[21] A. Nijenhuis, Indago. Math. **13**, 200 (1951).
[22] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).

Particle symmetry and antisymmetry in approximately relativistic Lagrangians[a)]

Harry W. Woodcock
Department of Physics, Philadelphia College of Textiles and Science, Philadelphia, Pennslyvania 19144 [b)]
and Department of Physics, Temple University, Philadelphia, Pennslyvania 19122

Warren N. Herman[c)]
Department of Physics, Spring Garden College, Philadelphia, Pennslyvania 19119

(Received 8 July 1988; accepted for publication 11 April 1990)

For a single-time approximation of the type discussed by Woodcock and Havas [Phys. Rev. D **6**, 3422 (1972)] applied to particle-asymmetric Poincaré-invariant variational principles (VPs) of the Fokker type, a method is presented for expressing approximately relativistic Lagrangians (ARLs) to any order in c^{-1} in a form such that coefficients of functions of the instantaneous three-separation r_{ij} are either particle symmetric or antisymmetric. These functions of r_{ij} are determined solely by the corresponding particle symmetric or antisymmetric parts of the exactly relativistic kernel of the VP describing two-body interactions of N classical point particles. While the exact kernel involving the particles' four-separations and four-velocities is particle asymmetric, the built-in static Newtonian limit is particle symmetric. Using this method to reformulate previously published ARLs to order c^{-3} makes it obvious that a sufficient condition for acceleration-free ARLs to order c^{-3} is that the kernel of the exact non-time-reversal-invariant interaction be particle symmetric.

I. INTRODUCTION

In a series of papers addressing the problem of obtaining approximately relativistic equations of motion for a system of N classical interacting point particles, the approach taken was to obtain an approximately relativistic Lagrangian (ARL) by direct expansion of the exactly relativistic variational principle (VP). In part I (herein referred to as WH),[1] a general form to order c^{-2} was established for ARLs obtained by making a single-time approximation of Poincaré-invariant[2] VPs[3] of the Fokker type[4] describing N point particles with two-body interactions depending on the particles' four-separations and four-velocities. Only interactions with a static Newtonian limit [i.e., a potential $V_{ij}(r_{ij})$ describing the instantaneous interaction of particles i and j separated by the distance r_{ij}] were considered. While the static Newtonian limit is symmetric in the particles' variables, the relativistic interactions considered were not so restricted.

Part II[5] extended these results to order c^{-3} for non-time-reversal invariant relativistic VPs and gave examples that could be easily integrated and used to compound acceleration-free ARLs and Hamiltonians with $6N$ canonical variables. However, no general conditions for acceleration-free ARLs were found.

In part III,[6] ARLs to order c^{-2} were computed by the method of WH and II for relativistic VPs involving classical isospin. However, it should be noted that isospin is not treated in the present paper.

Each of the ARLs derived in the above-mentioned papers[7] lacked a certain elegance of form in that the coefficients of most functions of r_{ij} were either particle symmetric or particle antisymmetric, but some failed to have either character. One of our objectives had been to see if particle nonsymmetry in the relativistic progenitor has observable consequences in the ARL; in order c^{-2}, as noted in the discussion of WH, this consequence turned out to be realized in the function W, which from its definition [Eq. (WH67b)] can be seen to vanish for particle-symmetric relativistic interactions. Its coefficients happened to be asymmetric. Therefore, the particle asymmetry of the coefficients of some of the functions of r_{ij} in the ARL was accepted.

In order c^{-3}, the final form[7] of the interaction terms of the ARL was characterized by the coefficients of two of the four terms being particle symmetric with the remaining ones asymmetric. However, in II there were no functions analogous to the W function of WH that allowed one to see from the final form of the ARL where particle asymmetry in the exact interaction manifested itself in order c^{-3}. Furthermore, to be able to write acceleration-free ARLs and Hamiltonians with only $6N$ canonical variables to order c^{-3}, the device used in II was to compound particular examples of ARLs so as to eliminate the acceleration-dependent terms. Indeed, it was specifically noted in the Discussion of II that: "The general form of relativistic kernels (or their linear combinations) for which the approximate Lagrangian is acceleration-free still remains to be determined" (in order c^{-3}).

We will show here that a sufficient condition on the relativistic kernel that results in the ARL being acceleration free in order c^{-3} can be inferred from a linear combination

[a)] A preliminary report on the results of this paper was given at the Spring Meeting of The American Physical Society [Bull. Am. Phys. Soc. **32**, 1022(1987)].
[b)] Permanent address.
[c)] Present address: Naval Air Development Center, Code 5052, Warminster, PA 18976.

 0022-2488/90/081929-10$03.00

ARL calculated by adding one-half of the ARL computed by integrating first on the world line of particle i to one-half of the ARL computed with the order of integrations reversed. The resulting ARL, called $\tilde{I}$, can be written (to all existing orders as shown in the Discussion) so that the coefficient of each function of r_{ij} is either particle symmetric or particle antisymmetric. Furthermore, the functions of r_{ij} having particle-symmetric (-antisymmetric) coefficients can be defined in such a way that they are determined by the corresponding particle-symmetric (-antisymmetric) part of the exactly relativistic interaction kernel. To order c^{-2}, $\tilde{I}$ is just a physically equivalent but more elegant form than that of WH. In order c^{-3}, however, $\tilde{I}$ makes it not only possible to trace the effects of asymmetry in the relativistic kernel but also easy to see that acceleration-free ARLs and their corresponding Hamiltonians follow from relativistic kernels that are particle symmetric but not time-reversal invariant; the latter condition is needed for nonvanishing contributions in that order.

In addition, the derivations of the ARLs in the first three papers of this series involved judicious use of integration by parts on the single time variable and (except for paper II) on the variable ζ, an integration variable appearing in the definitions of the functions of r_{ij}. The integrated terms then were assumed to vanish.[8] The vanishing of these terms had the redeeming value of leading to final expressions that could be connected to well-known forms of ARLs as well as to specially contrived easily integrated examples, but had no obvious motivation.

Here we supply a superior justification. Initially in this paper, we will retain all terms in the ARL except for the total time derivative terms, which make no contribution to the approximate equations of motion. Then we will show that if the exact nondependence of the results on the order of integrations is required of the order-by-order Taylor and Lagrange expansions, this leads to conditions allowing the deletion of the $d/d\zeta$ terms thereby justifying the assumptions of WH. In order c^{-3}, the conditions suggested by this requirement dramatically simplify the form of the W functions and lead to the sufficient condition for acceleration-free ARLs already noted.

It should be emphasized that this paper does not contradict the results of WH, II, and III. We present a method that gives justification for simplifications, conditions that must be satisfied by the relativistic kernel for the expansions to be meaningful, and a technique for expressing results in a form that better lends itself to interpretation. Alternative derivations verifying the final form for the ARLs had been made in the theses[9,10] on which WH and III were based; Ref. 10 showed to order c^{-2} that the final form did not depend on the order of integrations under assumptions that are justified here.

In Sec. II we give the exact relativistic VP, its relationship to ARLs, and establish the notation. The resulting terms to orders c^0 and c^{-2} are noted here. Simplification of the ARL by deleting d/dt and $d/d\zeta$ terms and the calculation of $\tilde{I}$ are carried out in detail in order c^{-3} in Sec. III, where acceleration-free ARLs (to order c^{-3}) are also considered. The results are discussed in Sec. IV.

II. POINCARÉ-INVARIANT VPS AND ARLS

Poincaré-invariant equations of motion for a system of N point particles interacting through two-body forces can be obtained from a parameter-invariant VP:[3,11]

$$\delta\mathscr{I}=0, \quad \mathscr{I}=\mathscr{I}_K+\mathscr{I}_I, \tag{1a}$$

where the kinetic term $\mathscr{I}_K$ is

$$\mathscr{I}_K\equiv -c^2\sum_i\int_{-\infty}^{\infty} dT_i\, m_i(\eta_{\mu\nu}\mathfrak{v}_i^\mu\mathfrak{v}_i^\nu)^{1/2}, \tag{1b}$$

with

$$\eta_{\mu\nu}\equiv 0 \text{ if } \mu\neq\nu, \quad \eta_{00}\equiv 1, \quad \eta_{11}=\eta_{22}=\eta_{33}=-c^{-2} \tag{1c}$$

as the metric of the four-space, and the interaction term $\mathscr{I}_I$ is

$$\mathscr{I}_I\equiv -\sum_{i<j}\sum\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dT_i\, dT_j\, \mathfrak{v}_i\mathfrak{v}_j\Lambda_{ij}\left(s_{ij}^\mu,\frac{\mathfrak{v}_i^\mu}{\mathfrak{v}_i},\frac{\mathfrak{v}_j^\mu}{\mathfrak{v}_j}\right),$$

$$s_{ij}^\mu\equiv z_i^\mu(T_i)-z_j^\mu(T_j), \quad \mathfrak{v}_i^\mu(T_i)\equiv\frac{dz_i^\mu(T_i)}{dT_i}. \tag{1d}$$

The coordinates $z_i^\mu(T_i)$ of the world line of the ith particle (of inertial mass m_i) are defined as

$$\begin{aligned} z_i^\mu(T_i)&\equiv(z_i^0,z_i^1,z_i^2,z_i^3)\\ &\equiv[t_i(T_i),x_i(T_i),y_i(T_i),z_i(T_i)], \end{aligned} \tag{2}$$

and their derivatives are $\mathfrak{v}_i^\mu$. The interaction is characterized by $N(N-1)/2$ possibly distinct functions Λ_{ij} (not necessarily symmetric in i and j). Each Λ_{ij} is assumed to be invariant under the infinitesimal transformations of the Poincaré group and to depend only on the positions and velocities of the particles. Thus it can be a function only of the two-body invariants of the group, given below.

If T_i is chosen to be a Poincaré-invariant parameter, then $\mathfrak{v}_i^\mu$ is a four velocity. The magnitude of $\mathfrak{v}_i^\mu$

$$\mathfrak{v}_i\equiv(\mathfrak{v}_i^\mu\mathfrak{v}_{i\mu})^{1/2} \tag{3}$$

is constrained to be equal to 1 if the arbitrary parameter T_i is chosen to be the proper time τ_i, i.e., such that

$$dT_i\to d\tau_i\equiv(\eta_{\mu\nu}\,dz_i^\mu\,dz_i^\nu)^{1/2}, \quad \mathfrak{v}_i\to 1. \tag{4}$$

Denoting the proper-time four-velocity by v_i^μ, it follows that

$$v_i^\mu(\tau_i)\equiv\frac{dz_i^\mu}{d\tau_i}=\frac{\mathfrak{v}_i^\mu}{\mathfrak{v}_i}, \quad \mathfrak{v}_i=\frac{d\tau_i}{dT_i}. \tag{5}$$

Choosing the parameter to be the proper time involves the constraint (4) but has the redeeming feature of a known connection to the particle's three coordinates and the time through

$$\begin{aligned} d\tau_i&=dt_i[1-\mathbf{v}_i(t_i)\cdot\mathbf{v}_i(t_i)/c^2]^{1/2},\\ \mathbf{v}_i(t_i)&\equiv\frac{d\mathbf{r}_i(t_i)}{dt_i}. \end{aligned} \tag{6}$$

Thus the four-dimensional formulation can be connected to a three-plus-one formulation either by going $T_i\to\tau_i\to t_i$ as in WH and II or by going $T_i\to t_i$ directly, as in III. [The latter is particularly useful in approximating the ten exactly con-

served quantities[3,6,9,10] that follow from the Poincaré invariance of Eqs. (1).]

Variation of Eq. (1a) results[1,3,6,9,10] in parameter-invariant Poincaré-invariant equations of motion:

$$\frac{d}{dT_i}\frac{m_i c^2 \mathfrak{v}_{i\mu}}{\mathfrak{v}_i} = \mathscr{L}_{i\mu}(\mathfrak{v}_i V_i), \tag{7a}$$

where $\mathscr{L}_{i\mu}$ is the four-dimensional Lagrangian derivative

$$\mathscr{L}_{i\mu} \equiv \frac{\partial}{\partial z_i^\mu} - \frac{d}{dT_i}\frac{\partial}{\partial \mathfrak{v}_i^\mu} \tag{7b}$$

and

$$V_i\left(z_i^\mu, \frac{\mathfrak{v}_i^\mu}{\mathfrak{v}_i}\right) = \sum_{j>i}\int_{-\infty}^{\infty} dT_j\, \mathfrak{v}_j \Lambda_{ij} + \sum_{j<i}\int_{-\infty}^{\infty} dT_j\, \mathfrak{v}_j \Lambda_{ji} \tag{7c}$$

is the generalized potential.[12] Then taking the Newtonian limit ($c^{-1}\to 0$) changes the description from one involving $4N$ coordinates and N parameters in a four-dimensional space to one involving $3N$ coordinates and one parameter in a three-dimensional space. As defined in WH, the static Newtonian limit *chosen*[13] for Eqs. (1) is

$$\delta \mathscr{I}_0 = 0,$$
$$\mathscr{I}_0 \equiv \int_{-\infty}^{\infty} dt\, L_{(0)}[\mathbf{r}_i(t), \mathbf{v}_i(t)], \quad i = 1,\ldots,N, \tag{8}$$

which is invariant up to a divergence, i.e., a total time derivative, under the infinitesimal transformations of the Galilei group. The VP (8) was found by putting the Poincaré-invariant VP (1) into the form of an integral over a single time and taking the limit $c^{-1}\to 0$. But as in WH, II, and III, our interest is in expanding Eqs. (1) in powers of c^{-1} yielding

$$\delta \mathscr{I} = 0,$$
$$\mathscr{I} \approx \int_{-\infty}^{\infty} dt\, L_{(n)}[\mathbf{r}_i(t), \mathbf{v}_i(t), \mathbf{a}_i(t), \ldots], \quad i = 1,\ldots,N, \tag{9a}$$

where

$$L_{(n)} \equiv K_{(n)} + I_{(n)},$$
$$K_{(n)} \equiv -\sum_i m_i c^2 + K_0 + c^{-2}K_2 + c^{-4}K_4 + \cdots + c^{-n}K_n,$$
$$I_{(n)} \equiv I_0 + c^{-1}I_1 + c^{-2}I_2 + c^{-3}I_3 + \cdots + c^{-n}I_n,$$
$$I_0 \equiv -\sum_{i<j}\sum V_{ij}(r_{ij}), \quad K_0 \equiv \sum_i \frac{1}{2} m_i \mathbf{v}_i^2(t). \tag{9b}$$

The case $n = 0$ is Eq. (8). It will be convenient here to refine the notation and define

$$I_2 \equiv \frac{1}{2}\sum_{i<j}\sum I_{2ij} \text{ and } I_3 \equiv -\frac{1}{2}\sum_{i<j}\sum I_{3ij}; \tag{10}$$

$K_{(n)}$ was calculated in WH [Eq. (WH35)] and involves only even powers of c^{-1}. [If n is odd, up to the $(n-1)$th term of $K_{(n)}$ is kept in the ARL.] I_1 (for isospin-independent interactions) was shown in WH [Eq. (WH61)] to be a total time derivative having no effect on the approximate equations of motion and is thus deleted from the ARL to any order.

As shown in WH [Eq. (WH32)–(WH38)], the Poincaré invariance and the choice of a static Newtonian limit for Eq. (1d) is ensured by setting $\Lambda_{ij} = cU_{ij}$ and choosing the arguments of U_{ij} to be

$$\begin{aligned}
\sigma_{ij} &\equiv c^2 s_{ij}^2 = c^2(t_i - t_j)^2 - |\mathbf{r}_i(t_i) - \mathbf{r}_j(t_j)|^2 \\
&\equiv c^2 t_{ij}^2 - r_{ij}^2(t_i,t_j), \\
\omega_{ij} &\equiv (\mathfrak{v}_i^{-1}\mathfrak{v}_j^{-1})\mathfrak{v}_i^\mu \mathfrak{v}_{j\mu} = \gamma_i\gamma_j[1 - c^{-2}\mathbf{v}_i(t_i)\cdot\mathbf{v}_j(t_j)], \\
\chi_{ij} &\equiv c\mathfrak{v}_i^{-1}\mathfrak{v}_i^\mu s_{ji\mu} = -\gamma_i[ct_{ij} - c^{-1}\mathbf{v}_i(t_i)\cdot\mathbf{r}_{ij}(t_i,t_j)], \\
\zeta_{ij} &\equiv c\mathfrak{v}_j^{-1}\mathfrak{v}_j^\mu s_{ij\mu} = \gamma_j[ct_{ij} - c^{-1}\mathbf{v}_j(t_j)\cdot\mathbf{r}_{ij}(t_i,t_j)],
\end{aligned} \tag{11}$$

where $U_{ij}(\sigma_{ij},\omega_{ij},\chi_{ij},\zeta_{ij})$ does not involve c except through the invariants noted in Eq. (11), and where

$$\gamma_i \equiv [1 - c^{-2}\mathbf{v}_i(t_i)\cdot\mathbf{v}_i(t_i)]^{-1/2}. \tag{12}$$

In this paper it is necessary to distinguish between the variable changes $(t_i,t_j)\to(\zeta_{ij},t_j)$ in the double integral of (1d) (in order to get a single time) and the variable change $(t_i,t_j)\to(t_i,\chi_{ij})$. The former variable change (used in WH) is designated

$$\begin{aligned}
\mathscr{I}_I &\equiv -\sum_{i<j}\sum\int_{-\infty}^{\infty} dt_j\left[\gamma_j^{-1}\int_{-\infty}^{\infty} dt_i\, \gamma_i^{-1} cU_{ij}\right] \\
&= -\sum_{i<j}\sum\int_{-\infty}^{\infty} dt_j\left[\gamma_j^{-1}\int_{-\infty}^{\infty} d\zeta_{ij} \times \frac{\gamma_i^{-1} cU_{ij}(\hat{\sigma}_{ij},\omega_{ij},\hat{\chi}_{ij},\zeta_{ij})}{(d\zeta_{ij}/dt_i)}\right]\Bigg|_+,
\end{aligned} \tag{13a}$$

where

$$\begin{aligned}
\hat{\sigma}_{ij} &\equiv (\zeta_{ij}\gamma_j^{-1} + c^{-1}\mathbf{v}_j\cdot\mathbf{r}_{ij})^2 - \mathbf{r}_{ij}^2, \\
\hat{\chi}_{ij} &\equiv -\gamma_i\gamma_j^{-1}\zeta_{ij} + c^{-1}\gamma_i(\mathbf{v}_i - \mathbf{v}_j)\cdot\mathbf{r}_{ij},
\end{aligned} \tag{13b}$$

and $|_+$ means the expression is evaluated with

$$t_i = t_j + c^{-1}[\zeta_{ij}\gamma_j^{-1} + c^{-1}\mathbf{v}_j\cdot\mathbf{r}_{ij}(t_i,t_j)], \tag{14}$$

which results in $\sigma_{ij}\to\hat{\sigma}_{ij}$ and $\chi_{ij}\to\hat{\chi}_{ij}$ on substitution for explicit t_i's; Eq. (14) follows from the definition of ζ_{ij} in Eq. (11). The latter variable change[10] is designated

$$\begin{aligned}
\mathscr{I}_I^* &\equiv -\sum_{i<j}\sum\int_{-\infty}^{\infty} dt_i\left[\gamma_i^{-1}\int_{-\infty}^{\infty} dt_j\, \gamma_j^{-1} cU_{ij}\right] \\
&= -\sum_{i<j}\sum\int_{-\infty}^{\infty} dt_i\left[\gamma_i^{-1}\int_{-\infty}^{\infty} d\chi_{ij} \times \frac{\gamma_j^{-1} cU_{ij}(\check{\sigma}_{ij},\omega_{ij},\chi_{ij},\check{\zeta}_{ij})}{(d\chi_{ij}/dt_j)}\right]\Bigg|_-,
\end{aligned} \tag{13a'}$$

where

$$\begin{aligned}
\check{\sigma}_{ij} &\equiv (\chi_{ij}\gamma_i^{-1} - c^{-1}\mathbf{v}_i\cdot\mathbf{r}_{ij})^2 - \mathbf{r}_{ij}^2, \\
\check{\zeta}_{ij} &= -\gamma_j\gamma_i^{-1}\chi_{ij} + c^{-1}\gamma_j(\mathbf{v}_i - \mathbf{v}_j)\cdot\mathbf{r}_{ij},
\end{aligned} \tag{13b'}$$

the $|_-$ means evaluated with

$$t_j = t_i + c^{-1}[\chi_{ij}\gamma_i^{-1} - c^{-1}\mathbf{v}_i\cdot\mathbf{r}_{ij}(t_i,t_j)], \tag{14'}$$

which came from the definition of χ_{ij} in Eqs. (11); the substitution for explicit t_j's has resulted in $\sigma_{ij}\to\check{\sigma}_{ij}$ and $\zeta_{ij}\to\check{\zeta}_{ij}$. The arbitrary parameters T_i and T_j have been chosen to be the coordinate times t_i and t_j. This choice means that

$$\mathfrak{v}_i^\mu \to \mathrm{v}_i^\mu(t_i) = \frac{dz_i^\mu(t_i)}{dt_i} = [1, \mathbf{v}_i(t_i)],$$

$$\mathfrak{v}_i \to \mathrm{v}_i = (\mathrm{v}_i^\mu \mathrm{v}_{i\mu})^{1/2} = \gamma_i^{-1}, \tag{15}$$

using Eqs. (2), (3), and (5).

The implicit-function difficulty indicated in Eqs. (14) was handled in WH, II, and III by a Lagrange expansion.[14] The details of doing the Lagrange and Taylor expansions of Eq. (13) using (14) in powers of c^{-1} were covered in WH, II, and III and will not be reiterated here. The details of Eqs. (13′) and (14′) to order c^{-2} were given in Ref. 10. However, as pointed out there, the results of the starred expansion can be obtained easily from the unstarred one.

This can be understood by noting that formally interchanging particle variables together with an interchange of the third and fourth arguments of the function U_{ij} and its derivatives turns Eqs. (13) with (14) into (13′) with (14′) and vice versa. [Note that $\zeta_{ji} = \chi_{ij}$ and $\chi_{ji} = \zeta_{ij}$, which follows from the definitions (11), that the ij subscripts on U_{ij} are not reversed because they are part of the name of the function, and that the order of the particle double sum remains the same with $i<j$.] Then, since we are interested in the results of an expansion in powers of c^{-1} and since the variable change $t_i \to \zeta$ in (13) results in the limit $c^{-1} \to 0$ in the arguments of U_{ij} being $U_{ij}(\zeta^2 - r^2{}_{ij}, 1, -\zeta, \zeta)$ whereas the $t_j \to \chi$ variable change in (13′) results (in the same limit) in $U_{ij}(\chi^2 - r^2{}_{ij}, 1, \chi, -\chi)$, it is convenient for the purposes of comparison to make the trivial transformation $\chi \to -\zeta$ on the dummy integration variable χ in Eq. (13′) and (14′). Then the expansions of Eqs. (13) and (13′) can be fairly compared, since in (13) the variable changes $(t_i, t_j) \to (\zeta_{ij}, t_j) \to (\zeta, t)$ and in (13′) the variable changes $(t_i, t_j) \to (t_i, \chi_{ij}) \to (t, \chi) \to (t, \zeta)$ result both ways in equations involving a single time variable and another integration variable ζ. Thus to obtain from the already known expansion of (13) the expansion that results from (13′) (after $\chi \to -\zeta$), we perform the operations

$$i \leftrightarrow j, \quad \zeta \to -\zeta, \quad \frac{\partial}{\partial \chi_{ij}} \leftrightarrow \frac{\partial}{\partial \zeta_{ij}}, \tag{16}$$

and interchange the third and fourth arguments of U_{ij} and its derivatives; we shall refer to this as the "star operation." The star operation can also be used to take the result of (13′) (after $\chi \to -\zeta$) into the expansion that results from (13) so that two consecutive applications of the star operation are equivalent to an identity operation. Notationally, when Eq. (13) is expanded using Eqs. (9) and (10) we get

$$\begin{aligned}
\mathscr{I}_I &\approx \sum_n c^{-n} \mathscr{I}_{In} \\
&= \int_{-\infty}^{\infty} dt\, I_{(n)} = \sum_{i<j} \sum \int_{-\infty}^{\infty} dt\, I_{(n)ij} \\
&= \sum_{i<j} \sum \int_{-\infty}^{\infty} dt \Big[-V_{ij} \\
&\quad + \frac{1}{2} c^{-2} I_{2ij} - \frac{1}{2} c^{-3} I_{3ij} + \cdots \Big],
\end{aligned} \tag{17}$$

while (13′) gives

$$\begin{aligned}
\mathscr{I}^*_I &\approx \sum_n c^{-n} \mathscr{I}^*_{In} \\
&= \int_{-\infty}^{\infty} dt\, I^*_{(n)} \\
&= \sum_{i<j} \sum \int_{-\infty}^{\infty} dt\, I^*_{(n)ij} \\
&= \sum_{i<j} \sum \int_{-\infty}^{\infty} dt \Big[-V^*_{ij} + \frac{1}{2} c^{-2} I^*_{2ij} \\
&\quad - \frac{1}{2} c^{-3} I^*_{3ij} + \cdots \Big].
\end{aligned} \tag{17′}$$

An obvious identity involving Eqs. (13) and (13′) is

$$\begin{aligned}
\mathscr{I}_I &\equiv (1/2)(\mathscr{I}_I + \mathscr{I}^*_I) + (1/2)(\mathscr{I}_I - \mathscr{I}^*_I) \\
&\equiv \tilde{\mathscr{I}}_I + (1/2)\mathscr{D}_I,
\end{aligned} \tag{18a}$$

where

$$\tilde{\mathscr{I}}_I \equiv (1/2)(\mathscr{I}_I + \mathscr{I}^*_I), \tag{18b}$$

$$\mathscr{D}_I \equiv \mathscr{I}_I - \mathscr{I}^*_I. \tag{18c}$$

Of course, Eqs. (13) and (13′) represent the *same* numerical value, a definite (double) integral; the distinction between them is the order of integration. But, in the variations leading to the exact equations of motion,[1,3,6] nondependence of the results on the order of integration was used and thus implicitly assumed, implying that the difference $\mathscr{D}_I$ of the Poincaré invariant Eqs. (13) and (13′) is

$$\mathscr{D}_I = 0, \tag{19}$$

yielding from (18a)

$$\mathscr{I}_I = \tilde{\mathscr{I}}_I. \tag{20}$$

Equation (18b) shows that

$$\tilde{\mathscr{I}}^*_I = \tilde{\mathscr{I}}_I. \tag{21}$$

However, after Lagrange and Taylor expansions, Eqs. (13) and (13′) become (17) and (17′); there is no guarantee that the order-by-order differences

$$\mathscr{D}_{In} \equiv \mathscr{I}_{In} - \mathscr{I}^*_{In} \tag{22}$$

vanish without introducing restrictions in some (or even all) orders of the expansion. Thus, for mathematical consistency, Eq. (19) yields

$$\mathscr{D}_{In} = 0 \tag{23}$$

for each n.

The case $n = 0$ from Eqs. (8) and (9) turns out to satisfy Eq. (23) automatically. The expression for the Newtonian potential (WH66) is

$$V_{ij}(r_{ij}) \equiv \int_{-\infty}^{\infty} d\zeta\, \bar{U}_{ij}(\zeta, r_{ij}), \tag{24a}$$

where the bar over a letter means that (1) the substitution (14) has been made, (2) all explicit $c^{-1} = 0$ so that $t_i = t_j = t$, and (3) e.g., [cf. Eq. (11)],

$$\begin{aligned}
\bar{U}_{ij}(\zeta, r_{ij}) &= U(\bar{\sigma}_{ij}, \bar{\omega}_{ij}, \bar{\chi}_{ij}, \bar{\zeta}_{ij}) \\
&= U(\zeta^2 - r^2{}_{ij}, 1, -\zeta, \zeta).
\end{aligned} \tag{24b}$$

Note that the ij subscripts have been deleted from U_{ij} for simplicity and from ζ_{ij} because it is now a dummy integration variable; also, it often will be convenient to delete the ij subscripts from r_{ij} to simplify the notation. It should be understood that all expressions that are coefficients of a barred quantity are now (after Lagrange and Taylor expansion around $c^{-1}=0$ in powers of c^{-1}) functions of the one time t. Use of the star operation (16) on Eq. (24) yields

$$V^*_{ij} = V_{ij}, \tag{25a}$$

so that (23) with $n=0$ gives

$$\mathscr{D}_{10} = 0. \tag{25b}$$

The case $n=2$ leads to no new physical results. Consequently, only the results will be summarized. Demanding that $\mathscr{D}_{12}=0$ yields the condition on $\bar{U}$ and its derivatives that had been assumed in WH.[15] Then using Eq. (18b) on I_{2ij} [defined in (10)] to calculate

$$\tilde{I}_{2ij} = (1/2)(I_{2ij} + I^*_{2ij}), \tag{26a}$$

where I_{2ij} is the main result of WH (WH75c), yields

$$\begin{aligned}\tilde{I}_{2ij} = {} & \mathbf{v}_i\cdot\mathbf{v}_j V_{ij} - \mathbf{v}_i\cdot\mathbf{r}\mathbf{v}_j\cdot\mathbf{r}\,\frac{1}{r}\frac{dV_{ij}}{dr} \\ & + (\mathbf{v}_i-\mathbf{v}_j)^2(V_{ij}+\tilde{X}_{ij}) \\ & + [(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^2\tilde{Y}_{ij} + (\mathbf{v}_i^2-\mathbf{v}_j^2)\tilde{W}_{ij} \\ & + [(\mathbf{v}_i\cdot\mathbf{r})^2 - (\mathbf{v}_j\cdot\mathbf{r})^2]\,\frac{1}{r}\frac{d\tilde{W}_{ij}}{dr},\end{aligned} \tag{26b}$$

where V_{ij} is given by Eq. (24) and

$$\begin{aligned}\tilde{X}_{ij} &\equiv -\int_{-\infty}^{\infty} d\zeta\left[\bar{U}_\omega + \frac{1}{2}(\zeta\bar{U}_\zeta - \zeta\bar{U}_\chi)\right], \\ \tilde{Y}_{ij} &\equiv -\frac{1}{2}\int_{-\infty}^{\infty} d\zeta(\bar{U}_{\chi\chi} + \bar{U}_{\zeta\zeta}), \\ \tilde{W}_{ij} &\equiv \frac{1}{2}W_{ij} \equiv \frac{1}{2}\int_{-\infty}^{\infty} d\zeta\,\zeta(\bar{U}_\chi + \bar{U}_\zeta).\end{aligned} \tag{26c}$$

A Poincaré invariant used as a subscript implies partial differentiation with respect to that invariant (before it is barred), e.g.,

$$\bar{U}_{\chi\chi} = \overline{\left(\frac{\partial^2 U}{\partial\chi^2}\right)}. \tag{26d}$$

The only noteworthy change in form from (WH75c) occurs in the terms involving the W functions. Their form in (26b) is just the form examined by Nordtvedt when he investigated the gravitational consequences of such terms for celestial body dynamics,[16] while (WH75c) found these terms to be

$$(\mathbf{v}_i^2 - \mathbf{v}_i\cdot\mathbf{v}_j)W_{ij}(r_{ij}) + [\mathbf{v}_i\cdot\mathbf{r}_{ij}\mathbf{v}_j\cdot\mathbf{r}_{ij} - (\mathbf{v}_j\cdot\mathbf{r}_{ij})^2]\,\frac{1}{r_{ij}}\frac{dW_{ij}}{dr_{ij}},$$

$$W_{ij}(r_{ij}) \equiv \int_{-\infty}^{\infty} d\zeta\,\zeta(\bar{U}_\chi + \bar{U}_\zeta). \tag{26e}$$

$\tilde{I}_{2ij}$ in (26b) is *not* particle symmetric, but it does have the property

$$\tilde{I}^*_{2ij} = \tilde{I}_{2ij}, \tag{27}$$

which is consistent with (21).

III. TO ORDER c^{-3}

The final result in II[5] for I_{3ij} [defined in Eq. (10)] (called I'_{3ijW} here) was given in Eq. (II25a') as[17]

$$\begin{aligned}I'_{3ijW} = {} & (\mathbf{v}_i-\mathbf{v}_j)^2(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W'_{1ij} \\ & + [(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^3 W'_{2ij} \\ & + \mathbf{a}_i\cdot(\mathbf{v}_i-\mathbf{v}_j)W'_{3ij} + \mathbf{a}_i\cdot\mathbf{r}(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W'_{4ij},\end{aligned} \tag{28a}$$

where

$$\begin{aligned}W'_{1ij} &\equiv \int_{-\infty}^{\infty} d\zeta\,[4\zeta^3\bar{U}_{\sigma\sigma} - 4\zeta^2\bar{U}_{\sigma\chi} + \zeta\bar{U}_{\chi\chi} \\ &\quad + 6\zeta\bar{U}_\sigma - 2\bar{U}_\chi - 2\zeta\bar{U}_{\sigma\omega} + \bar{U}_{\omega\chi}], \\ W'_{2ij} &\equiv \int_{-\infty}^{\infty} d\zeta\left[\frac{8}{3}\zeta^3\bar{U}_{\sigma\sigma\sigma} + 4\zeta^2\bar{U}_{\sigma\sigma\chi} - 2\zeta\bar{U}_{\sigma\chi\chi}\right. \\ &\quad \left. + \frac{1}{3}\bar{U}_{\chi\chi\chi} + 2\bar{U}_{\chi\sigma} - 4\zeta\bar{U}_{\sigma\sigma}\right], \\ W'_{3ij} &\equiv \int_{-\infty}^{\infty} d\zeta\left[2\zeta\bar{U}_\omega - 2\zeta\bar{U} - \frac{4}{3}\zeta^3\bar{U}_\sigma\right], \\ W'_{4ij} &\equiv \int_{-\infty}^{\infty} d\zeta\left[\frac{8}{3}\zeta^3\bar{U}_{\sigma\sigma} - 4\zeta^2\bar{U}_{\sigma\chi}\right. \\ &\quad \left. + 2\zeta\bar{U}_{\chi\chi} - 2\bar{U}_\chi + 4\zeta\bar{U}_\sigma\right],\end{aligned} \tag{28b}$$

and the primes on these functions, necessary in II, have been retained for the moment. A total time derivative

$$\begin{aligned}\frac{d}{dt}\Bigg[(\mathbf{v}_i\cdot\mathbf{r})^2\int_{-\infty}^{\infty} d\zeta(\bar{U}_\chi - 2\zeta\bar{U}_\sigma) \\ + \mathbf{a}_i\cdot\mathbf{r}\int_{-\infty}^{\infty} d\zeta\left(\zeta^2\bar{U}_\chi - \frac{2}{3}\zeta^3\bar{U}_\sigma\right)\Bigg],\end{aligned} \tag{29}$$

was omitted from Eq. (28) as irrelevant to a variational principle. The expressions (28b) can be rearranged to exhibit $d/d\zeta$ terms as

$$\begin{aligned}W'_{1ij} &= \int_{-\infty}^{\infty} d\zeta\left[(\bar{U}_{\omega\zeta} + \zeta\bar{U}_{\zeta\zeta})\right. \\ &\quad \left. + \frac{d}{d\zeta}(2\zeta^2\bar{U}_\sigma - \zeta\bar{U}_\chi - \zeta\bar{U}_\zeta + \bar{U} - \bar{U}_\omega)\right], \\ W'_{2ij} &= \frac{1}{3}\int_{-\infty}^{\infty} d\zeta\left[\bar{U}_{\zeta\zeta\zeta} + \frac{d}{d\zeta}(-4\zeta^2\bar{U}_{\sigma\sigma} + 4\zeta\bar{U}_{\sigma\chi}\right. \\ &\quad \left. + 2\zeta\bar{U}_{\sigma\zeta} - 2\bar{U}_\sigma - \bar{U}_{\chi\chi} - \bar{U}_{\chi\zeta} - \bar{U}_{\zeta\zeta})\right], \\ W'_{3ij} &= \frac{2}{3}\int_{-\infty}^{\infty} d\zeta\left[(3\zeta\bar{U}_\omega - \zeta\bar{U} + \zeta^2\bar{U}_\zeta - \zeta^2\bar{U}_\chi)\right. \\ &\quad \left. + \frac{d}{d\zeta}(-\zeta^2\bar{U})\right], \\ W'_{4ij} &= \frac{2}{3}\int_{-\infty}^{\infty} d\zeta\,[\zeta(\bar{U}_{\chi\chi} + \bar{U}_{\chi\zeta} + \bar{U}_{\zeta\zeta}) \\ &\quad + \frac{d}{d\zeta}(2\zeta^2\bar{U}_\sigma - 2\zeta\bar{U}_\chi - \zeta\bar{U}_\zeta + \bar{U})\Big].\end{aligned} \tag{30}$$

The route from Eq. (28b) to (30) is not obvious; what one does is systematically eliminate σ partial derivatives in favor of ζ partial derivatives using expressions analogous to the "chain-rule" differentiations

$$\frac{d}{d\zeta}\,\overline{U}(\zeta,r)=\frac{d}{d\zeta}\,U(\zeta^2-r^2,1,-\zeta,\zeta)$$
$$=2\zeta\overline{U}_\sigma-\overline{U}_\chi+\overline{U}_\zeta,$$
$$\frac{d\overline{U}}{dr}=-2r\overline{U}_\sigma. \tag{31}$$

A series of equally tedious calculations using the star operation (16) on Eqs. (28a) and (30) gives

$$I^*_{3ij}=I'^*_{3ijW}=(\mathbf{v}_i-\mathbf{v}_j)^2(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W'^*_{1ij}$$
$$+[(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^3W'^*_{2ij}-\mathbf{a}_j\cdot(\mathbf{v}_i-\mathbf{v}_j)W'^*_{3ij}$$
$$-\mathbf{a}_j\cdot\mathbf{r}(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W'^*_{4ij}, \tag{28a'}$$

where the omitted total time derivative is

$$\frac{d}{dt}\left[(\mathbf{v}_j\cdot\mathbf{r})^2\int_{-\infty}^{\infty}d\zeta(\overline{U}_\zeta+2\zeta\overline{U}_\sigma)\right.$$
$$\left.-\mathbf{a}_j\cdot\mathbf{r}\int_{-\infty}^{\infty}d\zeta(\zeta^2\overline{U}_\zeta+\frac{2}{3}\zeta^3\overline{U}_\sigma)\right], \tag{29'}$$

and where

$$W'^*_{1ij}=W'_{1ij}-\frac{1}{2r}\frac{d}{dr}W'_{3ij}-W'_{4ij}-\int_{-\infty}^{\infty}d\zeta\frac{d}{d\zeta}$$
$$\times[-\overline{U}_\omega+\overline{U}+2\zeta^2\overline{U}_\sigma+\zeta\overline{U}_\chi+\zeta\overline{U}_\zeta],$$
$$W'^*_{2ij}=W'_{2ij}-\frac{1}{2r}\frac{d}{dr}W'_{4ij} \tag{30'}$$
$$+\frac{1}{3}\int_{-\infty}^{\infty}d\zeta\frac{d}{d\zeta}(4\zeta^2\overline{U}_{\sigma\sigma}+4\zeta\overline{U}_{\sigma\zeta}$$
$$+2\zeta\overline{U}_{\sigma\chi}+2\overline{U}_\sigma+\overline{U}_{\chi\chi}+\overline{U}_{\chi\zeta}+\overline{U}_{\zeta\zeta}),$$
$$W'^*_{3ij}=-W'_{3ij},$$
$$W'^*_{4ij}=-W'_{4ij}-\int_{-\infty}^{\infty}d\zeta\frac{d}{d\zeta}(2\zeta\overline{U}_\chi+2\zeta\overline{U}_\zeta).$$

Then, using Eq. (28a) with (30) along with (28a') and (30') in (22) with $n=3$, we obtain for $\mathscr{D}_{I3}$ the expression

$$\mathscr{D}_{I3}=-\frac{1}{2c^3}\sum_{i<j}\sum\int_{-\infty}^{\infty}dt\left[\frac{d}{dt}\left\{\frac{1}{2}(\mathbf{v}_i-\mathbf{v}_j)^2W'_3+\frac{1}{2}[(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^2W'_4\right\}+(\mathbf{v}_i-\mathbf{v}_j)^2(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}\right.$$
$$\times\int_{-\infty}^{\infty}d\zeta\frac{d}{d\zeta}(-\overline{U}_\omega+\overline{U}+2\zeta^2\overline{U}_\sigma+\zeta\overline{U}_\chi+\zeta\overline{U}_\zeta)-\frac{1}{3}[(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^3\int_{-\infty}^{\infty}d\zeta\frac{d}{d\zeta}(4\zeta^2\overline{U}_{\sigma\sigma}+4\zeta U_{\sigma\zeta}$$
$$\left.+2\zeta\overline{U}_{\sigma\chi}+2\overline{U}_\sigma+\overline{U}_{\chi\chi}+\overline{U}_{\chi\zeta}+\overline{U}_{\zeta\zeta})-\mathbf{a}_j\cdot\mathbf{r}(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}\int_{-\infty}^{\infty}d\zeta\frac{d}{d\zeta}(2\zeta\overline{U}_\chi+2\zeta\overline{U}_\zeta)\right]. \tag{32}$$

The d/dt terms are of no consequence for the VP, but the $d/d\zeta$ terms in Eq. (32) are required to vanish so that $\mathscr{D}_{I3}=0$, i.e.,

$$(-\overline{U}_\omega+\overline{U}+2\zeta^2\overline{U}_\sigma+\zeta\overline{U}_\chi+\zeta\overline{U}_\zeta)|^\infty_{\zeta=-\infty}=0,$$
$$(4\zeta^2\overline{U}_{\sigma\sigma}+4\zeta\overline{U}_{\sigma\zeta}+2\zeta\overline{U}_{\sigma\chi}+2\overline{U}_\sigma+\overline{U}_{\chi\chi}$$
$$+\overline{U}_{\chi\zeta}+\overline{U}_{\zeta\zeta})|^\infty_{\zeta=-\infty}=0, \tag{33}$$
$$(\zeta\overline{U}_\zeta+\zeta\overline{U}_\chi)|^\infty_{\zeta=-\infty}=0;$$

this requirement is consistent with the assumption[15] made in WH "that U and its derivatives vanish sufficiently rapidly at infinity."

Using the above-mentioned assumption in Eqs. (28a) and (30) we find the simplified forms of I_{3ij} (deleting the prime along with the $d/d\zeta$ terms)

$$I_{3ij}=(\mathbf{v}_i-\mathbf{v}_j)^2(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W_{1ij}$$
$$+[(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^3W_{2ij}+\mathbf{a}_i\cdot(\mathbf{v}_i-\mathbf{v}_j)W_{3ij}$$
$$+\mathbf{a}_i\cdot\mathbf{r}(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W_{4ij}, \tag{34a}$$

where

$$W_{1ij}\equiv\int_{-\infty}^{\infty}d\zeta(\overline{U}_{\omega\zeta}+\zeta\overline{U}_{\zeta\zeta}),\quad W_{2ij}\equiv\frac{1}{3}\int_{-\infty}^{\infty}d\zeta\,\overline{U}_{\zeta\zeta\zeta},$$
$$W_{3ij}\equiv\frac{2}{3}\int_{-\infty}^{\infty}d\zeta(3\zeta\overline{U}_\omega-\zeta\overline{U}+\zeta^2\overline{U}_\zeta-\zeta^2\overline{U}_\chi), \tag{34b}$$
$$W_{4ij}\equiv\frac{2}{3}\int_{-\infty}^{\infty}d\zeta\,\zeta(\overline{U}_{\chi\chi}+\overline{U}_{\chi\zeta}+\overline{U}_{\zeta\zeta}).$$

Similarly, the use of the assumption turns Eq. (28a') and (30') into

$$I^*_{3ij}=(\mathbf{v}_i-\mathbf{v}_j)^2(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W^*_{1ij}+[(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^3W^*_{2ij}$$
$$-\mathbf{a}_j\cdot(\mathbf{v}_i-\mathbf{v}_j)W^*_{3ij}-\mathbf{a}_j\cdot\mathbf{r}(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}W^*_{4ij}, \tag{34a'}$$

where

$$W^*_{1ij}=W_{1ij}-\frac{1}{2r}\frac{dW_{3ij}}{dr}-W_{4ij},\quad W^*_{3ij}=-W_{3ij},$$
$$W^*_{2ij}=W_{2ij}-\frac{1}{2r}\frac{dW_{4ij}}{dr},\quad W^*_{4ij}=-W_{4ij}. \tag{34b'}$$

Then use of Eqs. (34) and (34') in (22) with $n=3$ gives only total time derivative terms, viz.,

$$\mathscr{D}_{I3} = -\frac{1}{2c^3}\sum_{i<j}\sum\int_{-\infty}^{\infty} dt\,\frac{d}{dt}\left\{\frac{1}{2}(\mathbf{v}_i-\mathbf{v}_j)^2 W_{3ij} + \frac{1}{2}[(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^2 W_{4ij}\right\}, \tag{35}$$

which are irrelevant to a VP.

Using Eqs. (34) and (34'), the consequence of Eq. (18b) in order c^{-3} is

$$\begin{aligned}\tilde{I}_{3ij} &= \tfrac{1}{2}(I_{3ij} + I^*_{3ij})\\ &= (\mathbf{v}_i-\mathbf{v}_j)^2(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}\,\tilde{W}_{1ij}\\ &\quad + [(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}]^3\tilde{W}_{2ij} + (\mathbf{a}_i+\mathbf{a}_j)\cdot(\mathbf{v}_i-\mathbf{v}_j)\tilde{W}_{3ij}\\ &\quad + (\mathbf{a}_i+\mathbf{a}_j)\cdot\mathbf{r}(\mathbf{v}_i-\mathbf{v}_j)\cdot\mathbf{r}\,\tilde{W}_{4ij},\end{aligned} \tag{36a}$$

where

$$\begin{aligned}\tilde{W}_{1ij} &\equiv W_{1ij} - \frac{1}{2}\left(W_{4ij} + \frac{1}{2r}\frac{dW_{3ij}}{dr}\right)\\ &= \frac{1}{2}\int_{-\infty}^{\infty} d\zeta\,[\zeta(\bar{U}_{\zeta\zeta} - \bar{U}_{\chi\chi}) + (\bar{U}_{\omega\chi} + \bar{U}_{\omega\zeta})],\end{aligned}$$

$$\tilde{W}_{2ij} \equiv W_{2ij} - \frac{1}{4r}\frac{dW_{4ij}}{dr} = \frac{1}{6}\int_{-\infty}^{\infty} d\zeta(\bar{U}_{\chi\chi\chi} + \bar{U}_{\zeta\zeta\zeta}), \tag{36b}$$

$$\tilde{W}_{3ij} \equiv \frac{1}{2}W_{3ij} = \frac{1}{3}\int_{-\infty}^{\infty} d\zeta(3\zeta\bar{U}_{\omega} - \zeta\bar{U} + \zeta^2\bar{U}_{\zeta} - \zeta^2\bar{U}_{\chi}),$$

$$\tilde{W}_{4ij} \equiv \frac{1}{2}W_{4ij} = \frac{1}{3}\int_{-\infty}^{\infty} d\zeta\,\zeta(\bar{U}_{\chi\chi} + \bar{U}_{\chi\zeta} + \bar{U}_{\zeta\zeta}).$$

Equations (36) likewise have the property that

$$\tilde{I}^*_{3ij} = \tilde{I}_{3ij}, \tag{37a}$$

where

$$\begin{aligned}&\tilde{W}^*_{1ij} = \tilde{W}_{1ij}, \quad \tilde{W}^*_{2ij} = \tilde{W}_{2ij},\\ &\tilde{W}^*_{3ij} = -\tilde{W}_{3ij}, \quad \tilde{W}^*_{4ij} = -\tilde{W}_{4ij}.\end{aligned} \tag{37b}$$

Each term of $\tilde{I}_{2ij}$ and $\tilde{I}_{3ij}$ has the property that if the coefficient of a function of r_{ij} is particle symmetric (antisymmetric), then that function of r_{ij} is determined solely by the particle-symmetric (-antisymmetric) part of the exactly relativistic kernel. This can be seen by defining a particle-exchange operator R_{ij} by its effect on U_{ij}:

$$\begin{aligned}&R_{ij}U_{ij}(\sigma_{ij},\omega_{ij},\chi_{ij},\zeta_{ij})\\ &\quad = U_{ij}(\sigma_{ji},\omega_{ji},\chi_{ji},\zeta_{ji})\\ &\quad = U_{ij}(\sigma_{ij},\omega_{ij},\zeta_{ij},\chi_{ij}),\end{aligned} \tag{38a}$$

where we have used

$$R_{ij}\chi_{ij} = \zeta_{ij}, R_{ij}\zeta_{ij} = \chi_{ij}, \tag{38b}$$

σ_{ij} and ω_{ij} are particle symmetric by their definitions in Eq. (11), and the ij subscripts on U_{ij}, as noted in Sec. II, are part of the name of the function and are not reversed by R_{ij}. As Eq. (38a) shows, the effect of R_{ij} on U_{ij} is to exchange χ_{ij} and ζ_{ij}. Then U_{ij} then can be written as the sum of a particle-symmetric term $U^S{}_{ij}$ and a particle-antisymmetric term $U^A{}_{ij}$ as follows:

$$\begin{aligned}&U_{ij} = U^S{}_{ij} + U^A{}_{ij},\\ &U^S{}_{ij} \equiv \tfrac{1}{2}U_{ij} + (\tfrac{1}{2})R_{ij}U_{ij},\\ &U^A{}_{ij} \equiv \tfrac{1}{2}U_{ij} - (\tfrac{1}{2})R_{ij}U_{ij},\end{aligned} \tag{39a}$$

where

$$R_{ij}U^S{}_{ij} = U^S{}_{ij}, \quad R_{ij}U^A{}_{ij} = (-1)U^A{}_{ij}, \tag{39b}$$

i.e., $U^S{}_{ij}$ is even on particle exchange while $U^A{}_{ij}$ is odd on particle exchange. When the functions of r_{ij} in $\tilde{I}_{2ij}$ and $\tilde{I}_{3ij}$ are written as integrals of a "barred U_{ij}" [recall Eqs. (24b), (26c), and (28b)], it turns out that $\bar{U}^S{}_{ij}$ is even in its argument ζ while $\bar{U}^A{}_{ij}$ is odd in the integration variable ζ. This becomes obvious from

$$\begin{aligned}\bar{U}^S{}_{ij} &= \tfrac{1}{2}U_{ij}(\zeta^2 - r^2, 1, -\zeta, \zeta)\\ &\quad + \tfrac{1}{2}U_{ij}(\zeta^2 - r^2, 1, \zeta, -\zeta),\\ \bar{U}^A{}_{ij} &= \tfrac{1}{2}U_{ij}(\zeta^2 - r^2, 1, -\zeta, \zeta)\\ &\quad - \tfrac{1}{2}U_{ij}(\zeta^2 - r^2, 1, \zeta, -\zeta),\end{aligned} \tag{40}$$

using Eqs. (39) and (24b). The only mild complication occurs when a partial derivative with respect to either χ or ζ is involved. For these cases it can be shown that

$$R_{ij}\frac{\partial}{\partial\zeta}U_{ij} = \frac{\partial}{\partial\chi}R_{ij}U_{ij}, \quad R_{ij}\frac{\partial}{\partial\chi}U_{ij} = \frac{\partial}{\partial\zeta}R_{ij}U_{ij}. \tag{41}$$

We can use these properties to show that, for example, $\tilde{W}_{3ij}$, whose coefficient in Eq. (36a) is particle antisymmetric, is determined solely by $U^A{}_{ij}$. Its integrand, from Eq. (36b), is $\frac{1}{3}[3\zeta\bar{U} - \zeta\bar{U} + \zeta^2(\bar{U}_{\zeta} - \bar{U}_{\chi})]$. Using the first of Eqs. (39a), the integrand can be separated into

$$\begin{aligned}&\tfrac{1}{3}[3\zeta\bar{U}^S_{\omega} - \zeta\bar{U}^S + \zeta^2(\bar{U}^S_{\zeta} - \bar{U}^S_{\chi})]\\ &\quad + \tfrac{1}{3}[3\zeta\bar{U}^A_{\omega} - \zeta\bar{U}^A + \zeta^2(\bar{U}^A_{\zeta} - \bar{U}^A_{\chi})].\end{aligned} \tag{42}$$

But from Eqs. (40) and (41), we have

$$\begin{aligned}R_{ij}(U^S_{\zeta} - U^S_{\chi}) &= R_{ij}\left(\frac{\partial}{\partial\zeta}U^S - \frac{\partial}{\partial\chi}U^S\right)\\ &= \frac{\partial}{\partial\chi}R_{ij}U^S - \frac{\partial}{\partial\zeta}R_{ij}U^S\\ &= \frac{\partial}{\partial\chi}U^S - \frac{\partial}{\partial\zeta}U^S = -(U^S_{\zeta} - U^S_{\chi}),\end{aligned} \tag{43}$$

which shows that the terms involving $\bar{U}^S$ in Eq. (42) are odd in ζ and consequently these terms will vanish when integrated on ζ from $-\infty$ to $+\infty$. From a similar calculation, the terms involving $\bar{U}^A$ in Eq. (42) are even in ζ and can give a nonzero contribution to the integral. Thus, we see that $\tilde{W}_{3ij}$, whose coefficient in $\tilde{I}_{3ij}$ is particle antisymmetric, is determined only by the particle-antisymmetric part of the relativistic kernel. From such considerations, it can be shown that only $U^S{}_{ij}$ contributes to V_{ij}, $\tilde{X}_{ij}$, $\tilde{Y}_{ij}$, $\tilde{W}_{1ij}$, and $\tilde{W}_{2ij}$, while only $U^A{}_{ij}$ contributes to $\tilde{W}$, $\tilde{W}_{3ij}$, and $\tilde{W}_{4ij}$. The immediate consequence is that if $U_{ij} = U^S{}_{ij}$, then $\tilde{W}_3$ and $\tilde{W}_4$ are zero, which kills the acceleration dependence of $\tilde{I}_{3ij}$. Thus, to order c^{-3}, a sufficient condition for acceleration-free ARLs and their associated Hamiltonians is that the relativistic kernel be particle symmetric and non-time-reversal invariant.[18] Of course, relativistic interactions that *are* time-reversal invariant have no terms of odd powers in c^{-1} in their ARLs making them trivially acceleration free to order c^{-3}.

IV. DISCUSSION

We have presented a method for writing ARLs that are physically equivalent to those derived in WH (to order c^{-2}) and in II (in order c^{-3}) but which have the property that coefficients of functions of r_{ij} are either particle symmetric or particle antisymmetic, and these functions of r_{ij} are determined solely by the corresponding particle-symmetric or -antisymmetric part of the relativistic kernel. To order c^{-2} this form differs only slightly from the one derived in WH and leads to no further insights. To order c^{-3}, however, it makes it obvious that a sufficient condition for acceleration-free ARLs is that the kernel of the exact non-time-reversal-invariant interaction be particle symmetric. For such interactions the functions $\tilde{W}_{3ij}$ and $\tilde{W}_{4ij}$ given in Eq. (36b) vanish since their integrands are then odd in the integration variable ζ. The prescription[19] given in paper II to calculate the third-order approximately relativistic Hamiltonian corresponding to the ARL gives, using $\mathbf{v}_i \approx \mathbf{p}_i/m_i$ in $\tilde{I}_{3ij}$,

$$\tilde{H}_3 = -\frac{1}{2}\sum_{i<j}\sum\left\{\left(\frac{\mathbf{p}_i}{m_i}-\frac{\mathbf{p}_j}{m_j}\right)^2\left(\frac{\mathbf{p}_i}{m_i}-\frac{\mathbf{p}_j}{m_j}\right)\cdot\mathbf{r}_{ij}\,\tilde{W}_{1ij}(r_{ij}) + \left[\left(\frac{\mathbf{p}_i}{m_i}-\frac{\mathbf{p}_j}{m_j}\right)\cdot\mathbf{r}_{ij}\right]^3\tilde{W}_{2ij}(r_{ij})\right\} \tag{44}$$

with the usual $6N$ canonical variables for an N particle system. The corresponding results in II involved particular examples only and far less generality.

We have also indicated that terms to order c^{-2} in the ARL that previously were deleted by assuming conditions (guided by known examples) on the relativistic kernel now necessarily vanish when we demand in each order that the results do not depend on the order of the double integration in the method used to obtain the ARL from the exact VP. This requirement in order c^{-3} suggests the same condition on the relativistic interaction; invoking it dramatically simplifies the definitions of the functions of r_{ij} appearing in the ARL and aids greatly in obtaining the form (36) of $\tilde{I}_{3ij}$.

The prescription for calculating an ARL having coefficients of functions of r_{ij} that are particle symmetric or antisymmetric was given in Eq. (18b). To order c^{-2} this gives Eq. (26) and in order c^{-3} Eq. (36). We now will demonstrate that it would work to all existing orders. It should be noted, however, that the applicability of our prescription to higher-order approximations does not imply that all such higher orders exist, nor that they are acceleration free. Martin and Sanz have shown[20] that in predictive relativistic mechanics (a subset of relativistic mechanics) a canonical approach[21] constrains a system of particles to be free particles only "from order c^{-6}" on, i.e., the no-interaction theorem[22] begins its effect in order c^{-6}. Whether the same is true for approximations following from Poincaré-invariant action principles is not as yet known, but we are exploring this alternative. With this caveat in mind, we suppose that a term T of $\mathscr{J}$ involves a coefficient c_{ij} of a function $F_{ij}(r_{ij})$ in the form

$$T = c_{ij}F_{ij}. \tag{45a}$$

Then, by Eq. (18b), $\tilde{\mathscr{J}}$ will have a term

$$\tilde{T} = \tfrac{1}{2}[c_{ij}F_{ij} + c_{ij}F^*{}_{ij}]. \tag{45b}$$

This is identical to

$$\tilde{T} = \tfrac{1}{2}[\tfrac{1}{2}(c_{ij}+c_{ji})(F_{ij}+F^*{}_{ij})] + \tfrac{1}{2}[\tfrac{1}{2}(c_{ij}-c_{ji})(F_{ij}-F^*{}_{ij})], \tag{45c}$$

which has particle symmetry in the coefficient of the first term and particle antisymmetry in the coefficient of the second. It is clear from Eq. (45c) that $\tilde{T}^* = \tilde{T}$.

An alternative way to construct $\tilde{\mathscr{J}}_I$ is to use Eq. (18a) to obtain

$$\tilde{\mathscr{J}}_I = \mathscr{J}^*{}_I + \tfrac{1}{2}\mathscr{D}_I = \mathscr{J}_I - \tfrac{1}{2}\mathscr{D}_I. \tag{46}$$

Equation (46) shows that an expansion of $\tilde{\mathscr{J}}_I$ differs from either $\mathscr{J}_I$ or $\mathscr{J}_I{}^*$ only by terms that are time derivatives, which are irrelevant to a VP, and/or ζ derivatives, which we have demanded to be zero for consistency to the exact expression. Equation (46) is useful in calculating approximately conserved quantities from exact ones. While in II the explicit forms of approximate conservation laws were not written out because the form given in II led to long and noninstructive forms for the conserved quantities, the form $\tilde{I}_{3ij}$ shortens the calculations greatly. These are not included here because they still would significantly lengthen this paper.

A third alternative method for obtaining $\tilde{I}_{nij}$ from an already calculated I_{nij} is to write any asymmetric coefficients as sums of particle-symmetric and particle-antisymmetric coefficients, combine like terms, and e.g., in order c^{-3}, perform an integration by parts on the time variable. This method has the advantage that $\tilde{I}_{nij}$ seems to follow directly from either I_{nij} (or $I^*{}_{nij}$) rather than from a linear combination. However, simplification of the forms of the definitions of the functions of r_{ij} would still require subjecting $I^*{}_{nij}$ to a similar procedure and then requiring the vanishing of the difference $I_{nij} - I^*{}_{nij}$; the vanishing of $d/d\zeta$ terms would follow from requiring, e.g., that $W_{1ij} - W^*{}_{1ij}$ be zero.

It is possible to approximate the exact equations of motion (7) and get the same post-Newtonian corrections to the equations of motion as follow from $\tilde{I}_{2ij}$ and $\tilde{I}_{3ij}$. But Eq. (7c) shows that both the variable changes (17) and (17′) would be required, and consequently this is no advantage over the alternative approaches discussed above. Furthermore, both are also required in calculating approximate conservation laws from exact ones, as noted in the second alternative above.

It should be noted that $\tilde{I}_{3ij}$ could be written so that every term is acceleration dependent. This is possible because both of the first two terms have $(\mathbf{v}_i - \mathbf{v}_j)\cdot\mathbf{r}_{ij} = r\,dr/dt$ multiplied by a function of r as a factor and thus an integration by parts on the time results in acceleration-dependent terms. However, this would seem to introduce higher derivatives artificially.

An easily integrated example of a particle-nonsymmetric interaction that is not time-reversal invariant and leads to arbitrary functions of r_{ij} in the ARL (as opposed to the example in II that led to powers of r_{ij} only) is

$$U_{ij} = 2\omega'_{ij}\delta(\sigma_{ij})f_{ij}(\zeta_{ij})\theta(\zeta_{ij}), \tag{47a}$$

where ℓ is a constant, δ is the Dirac delta function, and[23]

$$\theta(x)=\begin{cases}1, & x>0,\\ 0, & x<0.\end{cases} \tag{47b}$$

For this interaction, using Eqs. (24a), (26c), and (36b) to find the functions appearing in $\tilde{I}_{2ij}$ and $\tilde{I}_{3ij}$ results in

$$\begin{aligned}&V_{ij}(r_{ij})=f_{ij}(r_{ij})/r_{ij},\quad \tilde{X}_{ij}=-\ell V_{ij}-\tfrac{1}{2}f'_{ij}(r_{ij}),\\ &\tilde{Y}_{ij}=-\tfrac{1}{2}f''_{ij}/r_{ij},\quad \tilde{W}_{ij}=\tfrac{1}{2}f'_{ij},\\ &\tilde{W}_{1ij}=\tfrac{1}{2}(f''_{ij}+\ell f'_{ij}/r_{ij}),\quad \tilde{W}_{2ij}=\tfrac{1}{6}f'''_{ij}/r_{ij},\\ &\tilde{W}_{3ij}=\tfrac{1}{3}[(3\ell-1)f_{ij}+r_{ij}f'_{ij}],\quad \tilde{W}_{4ij}=\tfrac{1}{3}f''_{ij},\\ &f'_{ij}\equiv\frac{df_{ij}}{dr_{ij}},\end{aligned} \tag{48}$$

enabling one to trace the effects of f_{ij} in each order. The example (47) can be modified to be particle symmetric but still not time-reversal invariant in the form

$$U_{ij}=\omega'_{ij}\delta(\sigma_{ij})[f_{ij}(\zeta_{ij})\theta(\zeta_{ij})+f_{ij}(\chi_{ij})\theta(\chi_{ij})]. \tag{49}$$

Thus, Eq. (49) is an example involving an arbitrary function f_{ij} but yielding an acceleration-free ARL to order c^{-3}.

Since the form of ARL given in WH has been verified to order c^{-2} by other approaches in other papers and we have presented another form shown to be equivalent to WH, we should comment on ways in which the other approaches could obtain $\tilde{I}_{2ij}$.

Starting from a Newtonian theory characterized by interactions $V_{ij}(r_{ij})$ and by the ten constants of the motion satisfying the Lie-bracket relations of the Galilei group, Stachel and Havas[24] looked for corrections to order c^{-2} of these ten quantities such that the new quantities satisfied the Lie-bracket relations of the Poincaré group to that order. They found that requiring the corrections to be consistent with the expansion of an exact relativistic expression in powers of c^{-2} resulted in a Hamiltonian of the form obtained from the ARL in WH. By choosing a different particular solution to one of their equations and then redefining the form of some arbitrary functions, it can be shown[25] that our form (26b) just as easily could have been obtained as the original WH form.

In another approach, Coester and Havas[26] started from the exact relativistic quantum mechanical canonical formalism of Bakamjian and Thomas and expanded in powers of c^{-2} to obtain an approximately relativistic Hamiltonian in agreement with the post-Newtonian interactions of WH. Again it can be shown[27] that their derivation can lead just as naturally to the form given here by choosing alternate representations of arbitrary functions.

As noted in the Introduction, isospin is not treated in this paper. However, it easily can be shown that the isospin-dependent ARL given to order c^{-2} in III has the property that it is invariant under application of the star operation. No results on expansions to order c^{-3} of isospin-dependent relativistic interactions have been published.

Although the recent theoretical literature of particle physics is dominated by gauge field theories, an approach starting from classical direct-particle interactions is not only interesting in itself, but occasionally such approaches are found useful by those working in QCD.[28]

ACKNOWLEDGMENTS

We thank Professor Peter Havas for many helpful suggestions on both this work and on the manuscript.

This research was supported in part by a sabbatical leave granted to H. W. Woodcock by the Philadelphia College of Textiles and Science that was spent in residence at Temple University.

ERRATA TO II

Equation (II36c) should have $f'dt'$ on the right-hand side. In Eq. (II46), replace 8/3 by 4/3. In Eq. (II59), $A/2$ should read $-A/2$ and the coefficient of $(7/2+\ell)$ should be $(\mathbf{p}_i/m_i-\mathbf{p}_j/m_j)^2$. The * on the author's first address indicates his permanent address.

[1] H. W. Woodcock and P. Havas, Phys. Rev. D **6**, 3422 (1972). Referred to here both as part I and as WH, the latter to avoid confusion with interaction terms denoted by I.

[2] Invariance under the proper orthochronous subgroup of the full inhomogeneous Lorentz group (called the Poincaré group) is the physically important invariance group and it is this subgroup which is meant here.

[3] P. Havas, in *Problems in the Foundations of Physics*, edited by M. Bunge (Springer, Berlin, 1971), p. 31; errata are in footnote 37 of WH.

[4] A. D. Fokker, Z. Physik **58**, 386 (1929).

[5] H. W. Woodcock, Phys. Rev. D **17**, 1539 (1978), referred to as II; errata to I are in II. Errata to II precede these footnotes.

[6] W. N. Herman and P. Havas, Phys. Rev. D **17**, 1985 (1978), referred to as III.

[7] Equations (WH75c), (II25a'), and (III49d).

[8] Reference 1, immediately below Eq. (WH62). In II, it adds: "at most a constant to the action principle," immediately below Eq. (II24), while in III, it is: "omitting the irrelevant integrated term," below Eq. (IIIA14).

[9] H. W. Woodcock, Temple University thesis, 1972 (unpublished), Appendix B. This alternative calculation of the ARL was done not by $t_j\to\chi$ but by $t_j\to\zeta$, which is a monotonic variable change only to order c^{-2}.

[10] W. N. Herman, Temple University thesis, 1976 (unpublished), Appendix B. 1.

[11] The notations of WH, II, and III were slightly different, adapted to their purposes. Here we use mainly the notations of II, where, e.g., the coupling constants g_ig_j used in WH have been absorbed into U_{ij}.

[12] The notation of III is chosen for V_i; it is slightly different in WH and II.

[13] Nonstatic limits are also possible and were discussed in WH.

[14] K. G. Dedrick and E. L. Chu, Arch. Ration. Mech. Anal. **16**, 385 (1964).

[15] Reference 1, p. 3431, immediately below Eq. (WH62).

[16] K. Nordtvedt, Astrophys. J. **297**, 390 (1985).

[17] Equation (25a') is found in Ref. 5 in the "Notes added in proof"; the d/dt term there is omitted here and the names of the functions have been simplified, e.g., W'_{1ij} replaces $(W_{1ij}+W_{5ij}-2F_{ij})$, etc.

[18] In four-dimensional form, $z'^{\mu}_i(\tau'_i)=z^{\mu}_i(\tau_i)$ and $\tau'_i=k-\tau_i$, or $d\tau'_i=-d\tau_i$ expresses reversal of time. Since only χ_{ij} or ζ_{ij} of the four Poincaré invariants are odd in $d\tau_i$ and $d\tau_j$ [Eqs. (11)], only U_{ij}'s that are odd in the pair of variables χ_{ij} and ζ_{ij} are non-time-reversal invariant.

[19] Equations (II56) and (II57).

[20] J. Martin and J. L. Sanz, J. Math. Phys. **19**, 780 (1978).

[21] For an alternative canonical approach, see M. Pauri and G. M. Prosperi, J. Math. Phys. **17**, 1468 (1976).

[22] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).

[23] See, e.g., I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1, p. 21.

[24] J. Stachel and P. Havas, Phys. Rev. D **12**, 1598 (1976).

[25] The method of Ref. 24 yields our form when instead of the choice of particular solution given in their Eq. (A17), one chooses (in their notation with $\pi_{RS}=\mathbf{p}_R/m_R-\mathbf{p}_S/m_S$)

$$\bar{B}_{RS} = -\frac{1}{4}\left[\boldsymbol{\pi}_{RS}\cdot\left(\frac{\mathbf{p}_R}{m_R}+\frac{\mathbf{p}_S}{m_S}\right)W_{RS} + \boldsymbol{\pi}_{RS}\cdot\mathbf{r}_{RS}\left(\frac{\mathbf{p}_R}{m_R}+\frac{\mathbf{p}_S}{m_S}\right)\cdot\mathbf{r}_{RS}\frac{1}{r}\frac{dW_{RS}}{dr}\right] + B^{\text{hom}}_{RS},$$

together with

$$-2B^{\text{hom}}_{IK} = \left(\frac{\mathbf{p}_I}{m_I}-\frac{\mathbf{p}_K}{m_K}\right)^2\tilde{X}_{IK} + \left[\left(\frac{\mathbf{p}_I}{m_I}-\frac{\mathbf{p}_K}{m_K}\right)\cdot\mathbf{r}_{IK}\right]^2\tilde{Y}_{IK}$$

in place of their Eq. (69). [Note that the factor of -2 in their Eq. (69) should be a $-1/2$]. Finally, letting $W_{IJ} = 2\tilde{W}_{IJ}$ yields our result.

[26] F. Coester and P. Havas, Phys. Rev. D **14**, 2556 (1976).

[27] To obtain our form via the derivation in Ref. 26, simply replace their definitions (106) of the arbitrary functions X and Y by

$$\tilde{X} = A - \frac{m}{M}V + Z - \frac{m_2-m_1}{M}\tilde{W},$$

$$\tilde{Y} = B - \frac{m}{M}\frac{1}{r}\frac{dV}{dr} + \frac{1}{r}\frac{dZ}{dr} - \frac{m_2-m_1}{M}\frac{1}{r}\frac{d\tilde{W}}{dr},$$

and use these in their Eqs. (85), (87), and (104) together with $W = 2\tilde{W}$.

[28] R. W. Childers, Phys. Rev. D **36**, 606, 3813 (1987).

The Goursat problem for the homogeneous wave equation

Pierre Hillion
Institut Henri Poincaré, 75231 Paris, France

(Received 18 October 1989; accepted for publication 14 March 1990)

The Goursat problem for the 3-D homogeneous wave equation is presented and some methods to solve it are discussed.

I. INTRODUCTION

The Goursat problem[1] is a boundary value problem with data given on the characteristics $\xi = z - x_0$, $\eta = z + x_0$ ($x_0 = ct$) of the 1-D wave equation $\partial_z^2\psi - \partial_{x_0}^2\psi = 0$. Some authors speak of boundary value problems of the first kind[2,3] or of characteristic initial value problems.[4]

The Goursat problem may be generalized to the 3-D wave equation:

$$(\Delta - \partial_{x_0}^2)\psi = 0, \quad \Delta = \partial_x^2 + \partial_y^2 + \partial_z^2, \tag{1}$$

which becomes with the variables ξ, η:

$$4\partial_\xi\partial_\eta\psi + \Delta_\perp\psi = 0, \quad \Delta_\perp = \partial_x^2 + \partial_y^2, \tag{2}$$

by considering the transverse variables x, y, as some parameters.

Then the Goursat problem is defined by Eq. (2) together with the boundary conditions on $\xi = 0$ and $\eta = 0$:

$$\psi(x,y,\xi,0) = \varphi_1(x,y,\xi), \quad \psi(x,y,0,\eta) = \varphi_2(x,y,\eta), \tag{3}$$

satisfying the compatibility condition:

$$\varphi_1(x,y,0) = \varphi_2(x,y,0) = \varphi(x,y). \tag{3'}$$

We do not discuss here the existence and the uniqueness of solutions for the Goursat problem. We first present the modal waves that seem to be most natural solutions and we discuss some methods to solve the Goursat problem when solutions exist.

II. MODAL WAVES

Introducing the variables $\zeta = x - iy$, $\bar{\zeta} = x + iy (i = \sqrt{-1})$, Eq. (2) becomes:

$$\partial_\xi\partial_\eta\psi + \partial_\xi\partial_{\bar{\xi}}\psi = 0, \tag{4}$$

which is invariant under the transformations:

$$\xi \leftrightarrow \eta, \quad \zeta \leftrightarrow \bar{\zeta}, \quad (\xi,\eta) \leftrightarrow (\zeta,\bar{\zeta}). \tag{4'}$$

This suggests that we introduce the variable u with some similar symmetry:

$$u = (a+\xi)(b+\eta) + (c+\zeta)(d+\bar{\zeta}), \tag{5}$$

where a, b, c, d are arbitrary constants. We note τ_μ, $\mu = 1,2,3,4$, the monomials

$$\tau_1 = a + \xi, \quad \tau_2 = b + \eta, \quad \tau_3 = c + \zeta, \quad \tau_4 = d + \bar{\zeta}. \tag{5'}$$

Let $\mathcal{F}$ be an arbitrary function with sufficient derivatives. One checks easily by a direct calculation that for any τ_μ

$$\psi^0 = (1/\tau_\mu)\mathcal{F}((\omega/\tau_\mu)u), \tag{6}$$

where ω is some parameter is the solution of Eq. (4) that we call the fundamental mode since there exists higher-order modes obtained by successive derivatives of (6) with respect to ξ, η, ζ, $\bar{\zeta}$:

$$\psi^0_{i,j,k,l} = \partial_\xi^i\partial_\eta^i\partial_\zeta^k\partial_{\bar{\zeta}}^l\psi^0, \quad i,j,k,l = 0,1,2,\ldots. \tag{6'}$$

The solutions $\{\psi^0_{i,j,k,l}\}$ constitute the first set of modal solutions of Eq. (4) the most famous being the focus wave modes[5] obtained when $\mathcal{F}$ is an exponential function and $\tau_\mu = \tau_1$ or τ_2.

To obtain a second set, we start with the particular solutions

$$\psi^0 = \frac{1}{u}, \quad \psi^0_{i,j,k,l} = \frac{\tau_1^i\tau_2^j\tau_3^k\tau_4^l}{u^{i+j+k+l+1}}. \tag{7}$$

Now let the function f be analytical, then using the Maclaurin series for f, one checks easily that

$$\psi^1 = (1/\tau_\mu)f(\tau_\mu/\omega u) \tag{8}$$

is a solution of Eq. (4) since each term in the expansion of (8) is of the form (7). The analyticity condition imposed on f is unduly restrictive, only sufficient derivatives are required. Of course one can also check by a direct calculation that ψ^1 is a solution of Eq. (4).

The higher modes $\psi^1_{i,j,k,l}$ are defined as in (6'). So the solutions $\{\psi^1_{i,j,k,l}\}$ obtained by changing $\omega u/\tau_\mu$ into $\tau_\mu/\omega u$ in ψ^0 constitute the second set of modal waves.

One verifies easily that both sets of modal solutions satisfy boundary conditions of the type (3), (3') so that they may be considered as solutions of a Goursat problem.

III. SEPARABLE GOURSAT PROBLEMS

The Goursat problem is called separable if the boundary conditions (3), (3') are of the form

$$\varphi_1(x,y,\xi) = f_1(\xi)\varphi(x,y),$$
$$\varphi_2(x,y,\eta) = f_2(\eta)\varphi(x,y), \quad f_1(0) = f_2(0). \tag{9}$$

We start with the particular case where the data (9) do not depend upon ξ and η, that is,

$$\varphi_2(x,y,\xi) = \varphi_2(x,y,\eta) = \varphi(x,y). \tag{10}$$

This suggests that we look for the solution of Eq. (2) in the form

$$\psi(x,y,\xi,\eta) = h(v)g(x,y), \quad v = (\xi\eta)^{1/2}. \tag{11}$$

Substituting (11) into (2) leads to the system of equations:

$$\partial_v^2 h + (1/v)\partial_v h - k^2 h = 0, \tag{12a}$$

$$\Delta_\perp g + k^2 g = 0. \tag{12b}$$

The solution of Eq. (12a) is

$$h = I_0(k\sqrt{\xi\eta}), \tag{13}$$

 0022-2488/90/081939-03$03.00

where I_0 is the modified Bessel function of the first kind of order zero. For $\Delta_\perp = \partial_x^2 + \partial_y^2$ or $\Delta_\perp = \partial_r^2 + (1/r)\partial_r$ the solution of (12b) is

$$g(x,y) = e^{i(lx+my)}, \quad k^2 = l^2 + m^2, \tag{14a}$$

$$g(r) = J_0(kr), \qquad r^2 = x^2 + y^2, \tag{14b}$$

where J_0 is the Bessel function of the first kind of order zero. Consequently, if $\varphi(x,y)$ in (10) has the Fourier expansion

$$\varphi(x,y) = \sum_{l,m=-\infty}^{+\infty} a_{lm} e^{i(lx+my)}, \tag{15a}$$

or the Fourier–Bessel expansion[6]

$$\varphi(r) = \sum_{m=1}^{\infty} a_m J_0(kj_m r), \tag{15b}$$

where $j_1, j_2, \ldots$, denote the positive zeros of $J_0(\xi)$ arranged in ascending order of amplitude, the solutions of the Goursat problem are

$$\psi(x,y,\xi,\eta) = \sum_{l,m=-\infty}^{+\infty} a_{em} e^{i(lx+my)} I_0(\sqrt{(l^2+m^2)\xi\eta}), \tag{16a}$$

$$\psi(r,\xi,\eta) = \sum_{m=1}^{\infty} a_m I_0(kj_m\sqrt{\xi\eta}) J_0(kj_m r). \tag{16b}$$

For instance, for the boundary condition $\varphi(r) = J_0(kr)$, one has

$$\psi(r,\xi,\eta) = I_0(k\sqrt{\xi\eta}) J_0(kr). \tag{17}$$

We may now consider the general separable boundary data (9) leading us to look for the solution of Eq. (2) in the form

$$\psi(x,y,\xi,\eta) = h(\xi,\eta) g(x,y). \tag{18}$$

Substituting (18) into (2) supplies (12b) and the equation

$$4\partial_\xi \partial_\eta h - k^2 h = 0. \tag{19}$$

The Green's function of Eq. (19) is $I_0(k(\xi-\xi_0)^{1/2}(\eta-\eta_0)^{1/2})$, and using the Riemann method[2-4] we get a solution of the Goursat problem for Eq. (19) with boundary conditions $f_1(\xi)$ and $f_2(\eta)$:

$$h(\xi,\eta) = \int_0^\xi I_0(k\sqrt{\eta(\xi-s)})\partial_s f_1(s)\, ds + \int_0^\eta I_0(k\sqrt{\xi(\eta-s)})\partial_s f_2(s)\, ds. \tag{20}$$

Consequently, if $\varphi(x,y)$ has the expansion (15a) or (15b), the solution of the Goursat problem (9) is given by (16a) or (16b) with the Bessel function I_0 replaced by (20). If $\partial_s f_1(s) = \partial_s f_2(s) = \delta(s)$, where $\delta(s)$ is the Dirac distribution we get exactly (16a) and (16b) (see the Appendix).

We may generalize the previous results to partially separable Goursat problems defined by the boundary conditions:

$$\psi(x,y,\xi,0) = \varphi_1(x,y,\xi),$$
$$\psi(x,y,0,\eta) = f(\eta)\varphi(x,y), \quad \varphi_1(x,y,0) = \varphi(x,y). \tag{21}$$

Then we look for the solution of Eq. (2) in the form:

$$\psi(x,y,\xi,\eta) = h(\eta) g(x,y,\xi). \tag{22}$$

Substituting (22) into Eq. (2) supplies the following two equations:

$$\partial_\eta h = (\lambda/4)h, \tag{23a}$$

$$\lambda\partial_\xi g + \Delta_\perp g = 0. \tag{23b}$$

From (23a), we get $h = e^{(\lambda/4)\eta}$ while the solution of the diffusionlike equation (23b) satisfying the boundary condition $\varphi(x,y)$ for $\xi = 0$ is with $\mathrm{Re}\,\lambda < 0$ and $\xi > 0$:

$$g(x,y,\xi) = \frac{\lambda}{4\pi\xi} \times \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left(\frac{\lambda}{4\xi}[(x-x_0)^2 + (y-y_0)^2]\right) \times \varphi(x_0,y_0)\, dx_0\, dy_0. \tag{24}$$

Consequently, if $\varphi_1(x,y,\xi) \neq g(x,y,\xi)$ the Goursat problem (2) has no solution. But if $\varphi_1(x,y,\xi) = g(x,y,\xi)$ and if $f(\eta)$ has the Fourier expansion

$$f(\eta) = \sum_\lambda a_\lambda e^{(\lambda/4)\eta},$$

the solution of the Goursat problem (21) is

$$\psi(x,y,\xi,\eta) = \sum_\lambda \frac{\lambda a_\lambda}{4\pi\xi} \exp\left(\frac{\lambda}{4}\eta\right) \times \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left(\frac{\lambda}{4\xi}[(x-x_0)^2 + (y-y_0)^2]\right) \varphi(x_0,y_0)\, dx_0\, dy_0. \tag{25}$$

For instance if $f(\eta) = e^{(\lambda/4)\eta}$ and $\varphi(x,y) = \delta(x)\delta(y)$ we get, according to (25),

$$\psi(x,y,\xi,\eta) = \frac{\lambda}{4\pi}\frac{e^{(\lambda/4)\eta}}{\xi} e^{(\lambda/4\xi)(x^2+y^2)}, \tag{26}$$

which is a focus wave mode.

As a final generalization of the separable problems, we consider the case where the boundary conditions (3) have the form (24), that is

$$\varphi_1(x,y,\xi) = \frac{\lambda}{4\pi\xi}\int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} \exp\left(\frac{\lambda}{4\xi}[(x-x_0)^2 + (y-y_0)^2]\right) \times \varphi(x_0,y_0)\, dx_0\, dy_0, \tag{27a}$$

$$\varphi_2(x,y,\eta) = \frac{\mu}{4\pi\eta}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left(\frac{\mu}{4\eta}[(x-x_0)^2 + (y-y_0)^2]\right) \times \varphi(x_0,y_0)\, dx_0\, dy_0, \tag{27b}$$

then one checks easily that the solution of the Goursat problem is, provided that $4\Delta_\perp\varphi + \lambda\mu\varphi = 0$,

$$\psi(x,y,\xi,\eta) = e^{\lambda\eta/4}\varphi_1(x,y,\xi) + e^{\mu\xi/4}\varphi_2(x,y,\eta) - e^{[(\lambda/4)\eta + (\mu/4)\xi]}\varphi(x,y). \tag{28}$$

This seems to be the most general situation where the solutions of the Goursat problem may be obtained in a closed

form. For arbitrary boundary conditions (3), one probably has to resort to numerical methods.

IV. CONCLUSION

The discussion of the Goursat problem makes clear that the homogeneous 3-D wave equation is rich in unexpected solutions. Are some of these solutions interesting from a physical point of view? We think that they could intervene to develop the relativistic front form of dynamics initiated by Dirac[7] many years ago. It is, in particular, gratifying to see that the focus wave modes appear as solutions of a boundary value problem. For instance, the solution (26) may be interpreted as a wave generated by a point harmonic source in the hyperplane $\xi = 0$. This result, shown as the relation (22) to obtain the focus wave modes, is well known to people working in this field.[8]

APPENDIX

Let us remark that Eq. (19) has the solutions

$$h(\xi,\eta) = e^{\mu\xi + \nu\eta}, \quad 4\mu\nu = k^2,$$

consequently when the boundary conditions are of the type

$$\varphi_1(x,y,\xi) = e^{\mu\xi}\varphi(x,y), \quad \varphi_2(x,y,\eta) = e^{\nu\eta}\varphi(x,y).$$

One may use (16a) and (16b) together with relations (4') rather than Eq. (20) to obtain the solution of the Goursat problem.

For instance, this leads to

$$\psi(x,y,\xi,\eta) = \sum_{\mu,l,m} a_{lm\zeta} e^{\mu\xi + \nu\eta} e^{i(lx + mg)}, \quad 4\mu\nu = l^2 + m^2,$$

$$\psi(x,y,\xi,\eta) = \sum_{\mu\nu} a_{\mu\nu} e^{\mu\xi + \nu\eta} J_0(kr), \qquad k^2 = 4\mu\nu.$$

[1] A. N. Tychonov and A. A. Samarski, *Partial Differential Equations of Mathematical Physics* (Holden-Day, San Francisco, 1964).
[2] S. L. Sobolev, *Partial Differential Equations of Mathematical Physics* (Pergamon, New York, 1964).
[3] I. N. Sneddon, *Elements of Partial Differential Equations* (MacGraw-Hill, New York, 1967).
[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics—Vol. II* (Interscience, New York, 1962).
[5] P. A. Bélanger, J. Opt. Soc. Am. A **1**, 723 (1984).
[6] G. N. Watson, *Theory of Bessel Functions* (Cambridge U. P., Cambridge, England, 1962).
[7] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[8] J. N. Brittingham, J. Appl. Phys. **54**, 1179 (1983); A. P. Kiselev, Radiophys. Quant. Elec. **26**, 1014 (1983); T. T. Wu and R. W. P. King, J. Appl. Phys. **56**, 2587 (1984); A. Sezginer, J. Appl. Phys. **57**, 678 (1985); R. W. Ziolwoski, J. Math. Phys. **26**, 861 (1985); A. Wünsche, J. Opt. Soc. Am. A **6**, 1320, 1661 (1989).

Berry's phase for photons and topology in Maxwell's theory

Y. Q. Cai, G. Papini, and W. R. Wood
Department of Physics, University of Regina, Regina, Saskatchewan, S4S 0A2, Canada

(Received 13 April 1989; accepted for publication 4 April 1990)

The classical nature of Berry's phase for photons is shown to arise from the intrinsic topological structure of Maxwell's theory. The phase is developed in the context of fiber-bundle theory and is discussed in some detail.

I. INTRODUCTION

The surprising discovery of Berry's phase[1] has recently received considerable attention and has shown the importance of geometrical structures in Hilbert space. In the case of Berry's phase, which is a manifestation of the anholonomy in a Hermitian line bundle,[2] the structure can be described by a gauge potential in the parameter space of the quantum system. While the original derivation of Berry's phase relied on an adiabatic approximation, this condition can be removed, as shown by Aharonov and Anandan,[3] by observing that the projective Hilbert space of a quantum system can itself act as a parameter space. Thus, Berry's phase originates from the nontrivial topology of the complex projective Hilbert space.[4]

In their studies of Berry's phase for photons, Chiao and Wu[5] took the momentum **k** space as the parameter space and predicted a rotation of the polarization vector for a linearly polarized laser beam travelling through a single, helically wound optical fiber. Their result was soon experimentally confirmed.[6] Although Chiao and Wu regarded the manifestation of Berry's phase for photons (sometimes referred to in the literature as the Chiao–Wu phase) as "topological features of classical Maxwell theory that originate at the quantum level, but survive the correspondence-principle limit ($\hbar \to 0$) into the classical level," several authors disagree and have argued that the Chiao–Wu phase can be obtained without any reference to quantum mechanics. For instance, by adopting the notion of torsion used by Ross,[7] Haldane[8] attributes the Chiao–Wu phase to the geometrical structure of the optical fibers. Segert's discussion of the Chiao–Wu phase,[9] which is based on the S^2 topology of the unit sphere defined by $|\mathbf{t}|^2 = 1$ with **t** being the unit tangent vector of an optical fiber, is also classical in nature. He constructs a tangent bundle over S^2 and finds a local expression for the relative connection one-form. This latter description of the Chiao–Wu phase bears some similarities to the original formalism of Berry,[1] except for its classical nature, as one may consider the **t** space to be the parameter space and the one-form defined on it to give the corresponding gauge structure. The elegance of these two approaches is, however, somewhat blemished by the explicit introduction of geometrical objects directly related to the medium. This makes the intrinsic relationship between the Chiao–Wu phase and Maxwell's theory less obvious. An approach, aimed at showing this interdependence more clearly, should follow the work of Iwo and Zofia Bialynicki-Birula,[10] wherein the Chiao–Wu phase appears as a result of the existence of a connection one-form in the reciprocal space of Maxwell's theory.[11] While the latter approach, which motivates this work, is more important in studying the nature of the Chiao–Wu phase, the methods of Haldane and Segert are conceptually simpler and more suitable for experimental applications.

The emphasis here is on the origin of the Chiao–Wu phase via an S^2 topology intrinsic to Maxwell's theory and on the corresponding fiber-bundle formalism. As such, the method of Ref. 10 is closely followed and is realized in a simpler and more illustrative form. A directly visible geometric picture is provided to simplify the discussion. This picture is also helpful in understanding the abstract mathematical concepts of fiber bundles and topology used in studies of Berry's phase.

For the sake of completeness, the main results of Maxwell's theory as given in Ref. 11 are briefly reviewed in Sec. II. The S^2 topology inherent in the theory is discussed in some detail and the tangent bundle over S^2 is also constructed. The fiber bundle and its connection one-form are introduced in Sec. III. The gauge structure for the Chiao–Wu phase is then derived and a simple geometrical picture is presented. In Sec. IV, Haldane's method is rephrased in terms of Maxwell's theory to make the discussion complete. The conclusions are contained in Sec. V.

II. MAXWELL THEORY AND THE INTRINSIC S^2 TOPOLOGY

The Lagrangian density of Maxwell's theory is

$$\mathcal{L} = -\tfrac{1}{4} f_{\mu\nu} f^{\mu\nu}, \tag{2.1}$$

and the Maxwell equations *in vacuo* are

$$\partial^\mu f_{\mu\nu} = 0, \quad \partial^\mu \tilde{f}_{\mu\nu} = 0, \tag{2.2}$$

where $\tilde{f}^{\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\lambda\rho} f_{\lambda\rho}$ is the dual of $f_{\mu\nu}$ and $\epsilon^{\mu\nu\lambda\rho}$ is the completely antisymmetric tensor with $\epsilon^{0123} = 1$. Following Ref. 11, one can introduce an auxiliary complex antisymmetric tensor $F_{\mu\nu}$;

$$F_{\mu\nu}(x) \equiv \tfrac{1}{2}(f_{\mu\nu}(x) + i\tilde{f}_{\mu\nu}(x)), \tag{2.3}$$

and it is easy to verify that $F_{\mu\nu}$ is self-dual, i.e.,

$$\tilde{F}_{\mu\nu}(x) = -iF_{\mu\nu}(x). \tag{2.4}$$

With this new tensor field, the Maxwell equations (2.2) can be rewritten as

$$\partial^\mu F_{\mu\nu}(x) = 0. \tag{2.5}$$

A consequence of Eqs. (2.4) and (2.5) is that $F_{\mu\nu}(x)$ satisfies the d'Alembert equation

$$\Box F_{\mu\nu}(x) = 0. \tag{2.6}$$

It is well known that Maxwell's theory is a linear theory

 0022-2488/90/081942-05$03.00

that admits plane-wave solutions. These features are implicit in Eq. (2.6). A general solution of the d'Alembert equation can be written as

$$F_{\mu\nu}(x) = \int d\Gamma \left[a^{+}_{\mu\nu}(\mathbf{k}) e^{-ik\cdot x} + a^{-}_{\mu\nu}(\mathbf{k}) e^{ik\cdot x} \right], \quad (2.7)$$

with $d\Gamma \equiv d^3\mathbf{k}/2\omega(\mathbf{k})(2\pi)^3$ and $\omega(\mathbf{k}) \equiv |\mathbf{k}|$. The symbol $k\cdot x$ denotes the scalar product of the four-vectors k_μ and x^μ with a Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$. The antisymmetric tensors $a^{+}_{\mu\nu}$ and $a^{-}_{\mu\nu}$ are coefficients in the decomposition of the field $F_{\mu\nu}$ into plane waves; due to Eqs. (2.4) and (2.5), they can be expressed as

$$a^{+}_{\mu\nu}(\mathbf{k}) = e_{\mu\nu}(\mathbf{k}) f(\mathbf{k}, +1), \quad a^{-}_{\mu\nu}(\mathbf{k}) = e_{\mu\nu}(\mathbf{k}) f^*(\mathbf{k}, -1), \quad (2.8)$$

in terms of a complex antisymmetric tensor $e_{\mu\nu}(\mathbf{k})$ and a complex function f. The functions $f(\mathbf{k}, \pm 1)$ describe the amplitudes of the plane waves and are determined by the sources generating the waves. In contrast, $e_{\mu\nu}(\mathbf{k})$ is an intrinsic quantity of the theory that provides a description of the polarization. The polarization tensor satisfies the transverse condition

$$k^\mu e_{\mu\nu}(\mathbf{k}) = 0, \quad (2.9)$$

the self-dual condition

$$\tilde{e}_{\mu\nu}(\mathbf{k}) = -ie_{\mu\nu}(\mathbf{k}), \quad (2.10)$$

and the normalization condition

$$e_{\mu\nu}(\mathbf{k}) e^{*\nu}{}_{\lambda}(\mathbf{k}) = k_\mu k_\lambda. \quad (2.11)$$

As a result, $e_{\mu\nu}(\mathbf{k})$ is also constrained by the following relation:

$$e_{\mu\nu}(\mathbf{k}) e^{*}_{\lambda\rho}(\mathbf{k}) = \tfrac{1}{2} k_{[\mu} g_{\nu][\lambda} k_{\rho]} + \tfrac{1}{4} i k^\alpha \epsilon_{\alpha\mu\nu[\lambda} k_{\rho]} - \tfrac{1}{4} i k^\alpha \epsilon_{\alpha\lambda\rho[\mu} k_{\nu]}. \quad (2.12)$$

Equations (2.9)–(2.12) are an alternative expression of the Maxwell equations. By introducing a complex three-vector $\mathbf{e}(\mathbf{k})$ to represent the polarization and relating it to the components of the tensor $e_{\mu\nu}(\mathbf{k})$ by

$$e_{0i}(\mathbf{k}) = i\omega(\mathbf{k}) e_i(\mathbf{k}) = -\tfrac{1}{2} i \epsilon_{ijk} e_{jk}(\mathbf{k}), \quad (2.13)$$

the Maxwell equations reduce to the following vectorial relations:

$$\mathbf{k}\cdot\mathbf{e}(\mathbf{k}) = 0, \quad (2.14)$$

$$\mathbf{k}\times\mathbf{e}(\mathbf{k}) = -i\omega(\mathbf{k})\mathbf{e}(\mathbf{k}), \quad (2.15)$$

$$\mathbf{e}^*(\mathbf{k})\cdot\mathbf{e}(\mathbf{k}) = 1, \quad (2.16)$$

$$\mathbf{e}(\mathbf{k})\cdot\mathbf{e}(\mathbf{k}) = 0 = \mathbf{e}^*(-\mathbf{k})\cdot\mathbf{e}(\mathbf{k}), \quad (2.17)$$

$$\mathbf{e}^*(\mathbf{k})\times\mathbf{e}(\mathbf{k}) = i\mathbf{n}, \quad (2.18)$$

$$\mathbf{e}^*(-\mathbf{k})\times\mathbf{e}(\mathbf{k}) = 0, \quad (2.19)$$

where $\mathbf{n}$ is the unit vector in the $\mathbf{k}$ direction;

$$\mathbf{n} = \mathbf{k}/\omega(\mathbf{k}). \quad (2.20)$$

The following discussion will be based on this representation of Maxwell's theory.

As shown by Eqs. (2.14)–(2.19), the intrinsic properties of an electromagnetic wave are its wave vector $\mathbf{k}$ and polarization $\mathbf{e}(\mathbf{k})$. Since Maxwell's theory can be formulated as a representation of Poincaré symmetry,[11] all possible wave vectors form a vector space, named $\mathbf{k}$ space, with a flat Euclidean metric structure. The on-mass-shell condition

$$|\mathbf{k}|^2 = \omega^2(\mathbf{k}), \quad (2.21)$$

defines a constraint submanifold in $\mathbf{k}$ space. In fact, $\omega(\mathbf{k})$ is just a convenient notation for $\omega/|\mathbf{v}(\mathbf{k})|$, where ω is a constant frequency determined by the source and $\mathbf{v}(\mathbf{k})$ is the phase velocity of the wave with magnitude c *in vacuo*. Therefore, the submanifold is the surface of a sphere (or S^2 surface in short). The vector $\mathbf{n}$ gives the propagation direction of an electromagnetic wave and, in the case of light propagating in an optical fiber, coincides with the tangent vector of the fiber. Hence, the S^2 used by Segert[9] originates from the S^2 topology implicit in Maxwell's theory.

In order to describe the polarization of an electromagnetic wave, one can decompose the complex polarization vector $\mathbf{e}(\mathbf{k})$ into[11]

$$\mathbf{e}(\mathbf{k}) = (1/\sqrt{2})(\mathbf{I}_1(\mathbf{k}) + e^{i\delta}\mathbf{I}_2(\mathbf{k})), \quad (2.22)$$

with $\mathbf{I}_1(\mathbf{k})$ and $\mathbf{I}_2(\mathbf{k})$ two independent real unit vectors. The presence of the arbitrary phase δ is due to the fact that the Maxwell equations can only determine $\mathbf{e}(\mathbf{k})$ up to a phase factor.[11] It is always possible to choose $\mathbf{I}_1$ and $\mathbf{I}_2$ such that $(\mathbf{n}, \mathbf{I}_1, \mathbf{I}_2)$ forms an orthonormal basis, i.e.,

$$\mathbf{I}_1\times\mathbf{I}_2 = \mathbf{n}, \quad \mathbf{I}_2\times\mathbf{n} = \mathbf{I}_1, \quad \mathbf{n}\times\mathbf{I}_1 = \mathbf{I}_2. \quad (2.23)$$

Moreover, Eq. (2.23) shows that $\mathbf{I}_1$ and $\mathbf{I}_2$ actually form a tangent bundle over the S^2 surface and, consequently, the S^2 geometry induces a connection in the tangent bundle. The Chiao–Wu phase, therefore, results as a manifestation of the connection one-form.

Alternatively, one could introduce a vector space of all possible linear polarizations, hereafter called $\mathbf{p}$ space, and express an arbitrary $\mathbf{e}(\mathbf{k})$ in terms of two orthonormal vectors in $\mathbf{p}$ space with an appropriate phase δ. The orthonormal condition, which is due to Eq. (2.16), defines an S^2 surface in $\mathbf{p}$ space and the two vectors can be represented by two points on this surface. It is not convenient to describe the changes of polarization in $\mathbf{p}$ space, or more precisely, on the S^2 surface in $\mathbf{p}$ space. Nevertheless, since the polarization states are well described on the Poincaré sphere,[12] there must exist a mapping that will map the S^2 surface in $\mathbf{p}$ space onto the Poincaré sphere, in the sense that any two points together with a phase δ will be mapped to a point on the Poincaré sphere. One can thus choose the three-dimensional space, in which the Poincaré sphere is defined, as the parameter space. A cyclic evolution of a polarization vector will trace out a closed curve on the Poincaré sphere and generate a Pancharatnam phase.[13,14] A recent study has shown both theoretically and experimentally that the Chiao–Wu and Pancharatnam phases are additive.[15] However, this paper will concentrate on the Chiao–Wu phase only.

III. THE CONNECTION ONE-FORM AND FIBER-BUNDLE FORMALISM

The approach presented in the previous section is very similar to that of Berry[1] for quantum mechanics. If one takes the Maxwell equations (2.5) to be the corresponding Schrödinger equation, then the d'Alembert equation (2.6),

or equivalently the self-dual condition (2.4), plays the role of the instantaneous stationary Schrödinger equation. The important difference is that no adiabatic condition is needed here. The plane-wave solutions with all possible $\mathbf{k}$ and $\mathbf{e}(\mathbf{k})$ constitute a Hilbert space $\mathscr{H}$ and a general solution (2.7) is then expressed as a vector in $\mathscr{H}$. Therefore, a polarized plane wave in Maxwell's theory resembles the energy eigenfunction $\psi_n(x)$ in quantum mechanics. Furthermore, as the Maxwell equations can only determine the polarization tensor $e_{\mu\nu}(\mathbf{k})$ or $\mathbf{e}(\mathbf{k})$ up to an arbitrary phase factor,[11] one may treat $e_{\mu\nu}(\mathbf{k})$ or $\mathbf{e}(\mathbf{k})$ as the analog of the phase factor in $\psi_n(x)$. As a result, the complex projective Hilbert space can be parameterized by $\mathbf{k}$ only, and may be mapped isomorphically onto $\mathbf{k}$ space. However, one should be reminded that there are fundamental differences between the phase factor of a wave function and the polarization vector of a plane wave: The former is not an observable, while the latter can be measured with a polarizer. The formal analogy of the two is used only for the sake of constructing a fiber-bundle formalism in Maxwell's theory.

Before going further into the details of the construction, it is helpful to make use of the simple geometrical triad given in Sec. II to derive the gauge structure in the $\mathbf{k}$ space and to see how it gives rise to the Chiao–Wu phase. This will provide the reader with a visual picture for the fiber-bundle formalism.

As an electromagnetic wave propagates, its wave vector $\mathbf{k}$ traces out a curve on the S^2 surface in $\mathbf{k}$ space. Along it the polarization vector $\mathbf{e}(\mathbf{k})$, or equivalently an element of the typical fiber, is transported according to the Maxwell equations (2.14)–(2.19). Because of the geometry of the surface, $\mathbf{e}(\mathbf{k})$ will in general change, the change being given by a connection one-form in the tangent bundle. Since this tangent bundle is two dimensional with $\mathbf{I}_1(\mathbf{k})$ and $\mathbf{I}_2(\mathbf{k})$ forming a local basis for each fiber, the connection one-form may be given by

$$\alpha(\mathbf{k}) = \mathbf{I}_2(\mathbf{k}) \cdot d\,\mathbf{I}_1(\mathbf{k}), \tag{3.1}$$

for circular polarization when $\delta = \pi/2$. For an arbitrary δ in general, $\alpha(\mathbf{k})$ should be rewritten as[11]

$$\alpha(\mathbf{k}) = i\mathbf{e}^*(\mathbf{k}) \cdot d\mathbf{e}(\mathbf{k}), \tag{3.2}$$

or equivalently, with $\alpha(\mathbf{k}) = \alpha_i(\mathbf{k}) dk^i$,

$$\alpha_i(\mathbf{k}) = i\mathbf{e}^*(\mathbf{k}) \cdot (\nabla_i \mathbf{e}(\mathbf{k})). \tag{3.3}$$

It provides a gauge structure in $\mathbf{k}$ space and generates a curvature[11]

$$\nabla \times \alpha(\mathbf{k}) = -\frac{\mathbf{n}(\mathbf{k})}{\mathbf{k}^2}. \tag{3.4}$$

The rotation of the polarization vector $\mathbf{e}(\mathbf{k})$, when it is parallel transported along the curve by the connection $\alpha(\mathbf{k})$, is given by the integral of the one-form along a curve in $\mathbf{k}$ space. When the curve is a closed path C on the S^2 surface the rotation angle resulting from the cyclic evolution is given by

$$\oint_C \alpha(\mathbf{k}) = \int\int_{\Sigma_C} (\nabla \times \alpha(\mathbf{k})) \cdot d\,\mathbf{S}_n = -\Omega(C), \tag{3.5}$$

where $\Omega(C)$ is the solid angle of the loop C with respect to the origin in $\mathbf{k}$ space.

The above discussion can be illustrated by the following simple example. Assuming that at $t = 0$, $(\mathbf{n},\mathbf{I}_1,\mathbf{I}_2)$ are chosen to be at (1,0,0) in the $\mathbf{k}$ space and directed along $(\mathbf{x},\mathbf{y},\mathbf{z})$, respectively. First, by carrying the triad along the equator to the point (0,1,0) and keeping $\mathbf{I}_2$ parallel to $\mathbf{z}$, the triad at (0,1,0) is in the direction of $(\mathbf{y}, -\mathbf{x},\mathbf{z})$. Next, the triad is transported along the meridian to (0,0,1) with the direction of $\mathbf{I}_1$ unchanged, the triad at (0,0,1) is then in the direction of $(\mathbf{z}, -\mathbf{x}, -\mathbf{y})$. Finally, upon returning to the starting point (1,0,0) along the meridian with $\mathbf{I}_2$ keeping its direction unaltered, the triad ends with the orientation $(\mathbf{x},\mathbf{z}, -\mathbf{y})$ and both $\mathbf{I}_1$ and $\mathbf{I}_2$ differ from their original orientation by a rotation about the $\mathbf{x}$ axis by $\pi/2$. In fact, this rotation angle equals the solid angle of the closed path traced out by $\mathbf{k}$ on the S^2 surface in $\mathbf{k}$ space, which is $\frac{1}{8}$ of the total solid angle of the S^2 surface. The result of Eq. (3.5) is thus verified.

In Simon's beautiful reformulation of Berry's phase[2] in terms of fiber-bundle theory, the solution of the Schrödinger equation forms a Hermitian line bundle over the parameter space and Berry's phase results as an anholonomic effect in the line bundle. The similarity between the approach presented above and that of Berry suggests the possibility of using the language of fiber-bundle theory in the case of photons as well. In fact, by eliminating the degrees of freedom related to the polarization, the Hilbert space $\mathscr{H}$ reduces to its complex projective space $\mathscr{P}$. The restriction to a particular choice of $\omega(\mathbf{k})$ defines a submanifold $\mathfrak{F}$ in $\mathscr{P}$. Since $\mathscr{P}$ can be parameterized by $\mathbf{k}$, this parameterization defines a map so that $\mathscr{P}$ is isomorphic to $\mathbf{k}$ space and $\mathfrak{F}$ to the S^2 surface defined by Eq. (2.21). One can now treat the polarization as a Hermitian line bundle $\mathfrak{N}$ with the structure of a principal fiber bundle $P(\mathscr{P},U(1))$, which is isomorphic to $\mathscr{H}$. The structure group $U(1)$ stems from the fact that Maxwell's theory cannot determine the phase of the polarization $e_{\mu\nu}(\mathbf{k})$.[11] It is the arbitrariness in this phase that leaves room for the Chiao–Wu phase.

A cyclic evolution corresponds to a closed curve C on $\mathfrak{F}$, thus on the S^2 surface in $\mathbf{k}$ space as well, and generates a transport of $\mathbf{e}(\mathbf{k})$ in the line bundle. Using the arguments in Ref. 2, there is a natural Hermitian connection in the line bundle given by the one-form $\alpha(\mathbf{k}) = i\mathbf{e}^*(\mathbf{k}) \cdot d\mathbf{e}(\mathbf{k})$ of Eq. (3.2). It defines the transport of a fiber, $\mathbf{e}(\mathbf{k})$ for some $\mathbf{k} \in S^2$, in the Hermitian line bundle. The Chiao–Wu phase $\gamma(C)$ is obtained in the form

$$\gamma(C) = \oint_C \alpha(\mathbf{k}) = \int\int_{\Sigma_c} V(\mathbf{k}), \tag{3.6}$$

where $V(\mathbf{k}) \equiv d\alpha(\mathbf{k}) = id\mathbf{e}^*(\mathbf{k}) \cdot d\mathbf{e}(\mathbf{k})$ is the curvature two-form and Σ_C is the oriented surface on the S^2 surface with $\partial S = C$ its boundary.

Since the Chiao–Wu phase results from the facts that the Hilbert space $\mathscr{H}$ for Maxwell's theory is topologically equivalent to a Hermitian line bundle $\mathfrak{N}$ over $\mathscr{P}$ and that, for a given ω, the constrained complex projective space $\mathfrak{F}$ has the S^2 topology,[9] one can study the Chiao–Wu phase in general by employing a particular S^2 geometrical structure, e.g. the S^2 sphere used in Secs. II and III.

IV. THE CHIAO–WU PHASE AND TORSION IN OPTICAL FIBERS

Soon after the experimental verification of the existence of the Chiao–Wu phase,[6] Haldane[8] recognized that this phase was the same as the one studied and observed by Ross[7] and explained it in terms of the geometry of the optical fiber. He further beautifully demonstrated the nature of the Chiao–Wu phase with the help of an S^2 surface spanned by the unit tangent vector. However, due to the explicit use of torsion, Haldane's approach seems to have overemphasized the role played by the geometrical structure of the optical fiber. The aim of this section is to show the compatibility of Haldane's approach and the present one. In order that Eqs. (2.14)–(2.19) be applicable in the present discussion, one may idealize the optical fibers as vacuum tubes with perfectly reflecting interior walls.

Following Haldane,[8] one may assume that the fiber axis is described by a curve $\mathbf{r}(s)$, where s is the distance along the fiber counted from some starting point. The fiber has unit length, i.e., $s=1$ at the end. Thus $\mathbf{t}(s)\equiv(\partial\mathbf{r}/\partial s)/|\partial\mathbf{r}/\partial s|$ gives the unit tangent vector at s. The local curvature $\kappa(s)$ and the unit normal vector $\mathbf{a}(s)$ are then defined by $\partial\mathbf{t}/\partial s=\kappa\mathbf{a}$ and $\mathbf{t}$, $\mathbf{a}$ and the unit binormal vector $\mathbf{b}\equiv\mathbf{t}\times\mathbf{a}$ provide a local orthonormal basis. By considering the propagation of a light beam along an optical fiber as a series of perfect reflections inside the fiber, one must assume an even number N of reflections, because for every reflection there is a phase change by π[12,16] and no such change is observed experimentally[7] for a smoothly bent fiber. Actually, when a fiber is smoothly curved the change in direction of the $\mathbf{k}$ may be considered to be continuous and N is, therefore, effectively infinity. In this case, N and $N+1$ are essentially the same and the number of reflections N cannot be assigned to be even or odd. Since the experimental results do not show the presence of an extra phase,[7] the even-N assumption can be considered as a good hypothesis in the following discussions. It has been shown[17] that, according to Maxwell's theory, if the angle between the normal vectors of two consecutive mirrors is β, the polarization vector $\mathbf{e}(\mathbf{k})$ of the beam reflected from the second mirror will further rotate by $-\beta$ in addition to the π phase change. If the two mirrors are infinitesimally close to each other, the angle β can effectively be written as $\tau\,ds$, with $\tau(s)$ a local torsion of the fiber given by $\partial\mathbf{b}/\partial s=-\tau\mathbf{a}$. Therefore, the rotation of the polarization vector of a linearly polarized light beam inside the fiber is given by the relation

$$\frac{\partial\theta(s)}{\partial s}=-\tau(s), \tag{4.1}$$

where $\theta(s)$ is the angle between the polarization vector and the $\mathbf{a}$ axis. In other words, Eq. (4.1) defines a parallel transport of the polarization vector along the fiber. Therefore, after a linearly polarized light beam travels along the fiber from s_0 to s, its polarization vector is rotated by an angle

$$\Delta\theta=\theta(s)-\theta(s_0)=-\int_{s_0}^{s}\tau(s')ds'. \tag{4.2}$$

Consider now a second optical fiber with the same $\mathbf{t}(0)$ and $\mathbf{t}(1)$ as the first one. One can construct a curved surface containing both fibers and then extend it further to form a closed two-dimensional surface. By labeling the first fiber as $\mathbf{r}(s,0)$ and the second one as $\mathbf{r}(s,1)$, one can imagine that there are many other fibers $\mathbf{r}(s,u)$, with the same $\mathbf{t}(0)$ and $\mathbf{t}(1)$, between them and that the end points of the interpolated fibers lay on the two geodesic curves $\mathbf{r}(0,u)$ and $\mathbf{r}(1,u)$ on the surface. Thus the area between the two fibers can be parameterized with s and u. The rotation angle of the polarization vector with respect to the $\mathbf{a}$ axis for each beam is given by Eq. (4.2) with the integration taken along the corresponding fiber. Since the $\mathbf{a}$ axes of the fibers have in general different orientations, one must take this into account in calculating the relative rotation of the polarization vectors when two linearly polarized beams, generated by coherent splitting at $s=0$, travel to the ends of the fibers. Using[8]

$$\tau=-\mathbf{a}\cdot\frac{\partial\mathbf{b}}{\partial s}=\mathbf{b}\cdot\frac{\partial\mathbf{a}}{\partial s},$$

and

$$\mathbf{t}\cdot\frac{\partial\mathbf{a}}{\partial s}=-\mathbf{a}\cdot\frac{\partial\mathbf{t}}{\partial s}=0,\quad\text{at } s=0,1,$$

the relative rotation angle is

$$\Delta\phi=\left[\int_0^1 du\,\mathbf{b}\cdot\frac{\partial\mathbf{a}}{\partial u}\right]\Bigg|_{s=0}^{s=1}-\left[\int_0^1 ds\,\mathbf{b}\cdot\frac{\partial\mathbf{a}}{\partial s}\right]\Bigg|_{u=0}^{u=1}, \tag{4.3}$$

where the first term describes the differences in $\mathbf{a}$ directions of the two fibers and the second one is due to the difference of the rotation angles with respect to the corresponding $\mathbf{a}$ axes. However, one can map the closed surface to the unit sphere defined by the unit vector $\mathbf{t}$ and rewrite Eq. (4.3) as[8]

$$\Delta\phi=-\int_0^1 ds\int_0^1 du\left[\mathbf{t}\cdot\frac{\partial\mathbf{t}}{\partial s}\times\frac{\partial\mathbf{t}}{\partial u}\right]. \tag{4.4}$$

The geometrical meaning of this $\Delta\phi$ is immediately obvious, provided one makes the identification $\mathbf{I}_1\equiv\partial\mathbf{t}/\partial s$, $\mathbf{I}_2\equiv\partial\mathbf{t}/\partial u$ and $\mathbf{n}\equiv\mathbf{t}$ on the unit sphere and uses Eqs. (2.23), (3.4), and (3.5).

In ending the section, one should also be reminded that the even-N assumption is really not necessary in deriving Eq. (4.1) from Maxwell's theory, since Berry[18] has pointed out, without giving the details of the argument, that Eq. (4.1) can be obtained from the Maxwell equations after some analysis.

V. CONCLUSIONS

The Chiao–Wu phase is classical in origin and stems directly from the intrinsic topology of Maxwell's theory. The rich topological structures of the theory may manifest themselves in many ways. In $\mathbf{k}$ space, they give rise to a gauge potential $\alpha(\mathbf{k})$ and generate the Chiao–Wu phase for any cyclic evolution carried out in the space. On the other hand, Pancharatnam phase implies the corresponding gauge structure in a particular space of polarization vectors in which a Poincaré sphere can be defined. These two topological structures of Maxwell's theory are independent and thus the two phases are additive.[15] Because different topological structures correspond to different constructions of a fiber

bundle in the theory, additional interesting topological structures, besides the two mentioned, might exist for Maxwell's theory. Finally, it should be pointed out that the gauge structure in **k** space is quite general and exists for any linear field theory, massless or massive and with or without spin.[10]

Although the above results, which refer only to the vacuum, show an intimate relationship between the Chiao–Wu phase and the classical Maxwell theory, they are unsatisfactory when applied to real experimental situations. Berry has recently suggested[19] a way to solve the problem by describing the electromagnetic field in terms of a six-component spinor and recasting the Maxwell equations in the medium into a form similar to a time-dependent Schrödinger equation with a Hermitian operator as Hamiltonian.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank Professor M. V. Berry for his comments and the referee for his constructive remarks.

[1] M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984).
[2] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).
[3] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).
[4] E. Kiritsis, Commun. Math. Phys. **111**, 417 (1987).
[5] R. Y. Chiao and Y.-S. Wu, Phys. Rev. Lett. **51**, 933 (1986).
[6] A. Tomita and R. Y. Chiao, Phys. Rev. Lett. **51**, 937 (1986).
[7] J. N. Ross, Opt. Quantum Electron. **16**, 455 (1984).
[8] F. D. M. Haldane, Phys. Rev. Lett. **59**, 1788 (1987); Opt. Lett. **11**, 730 (1986).
[9] J. Segert, Phys. Rev. A **36**, 10 (1987).
[10] I. Bialynicki-Birula and Z. Bialynicka-Birula, Phys. Rev. D **35**, 2383 (1987).
[11] I. Bialynicki-Birula and Z. Bialynicka-Birula, *Quantum Electrodynamics* (Pergamon, Oxford, 1975).
[12] M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1970).
[13] S. Pancharatnam, Proc. Indian Acad. Sci. Sect. A **44**, 247 (1956).
[14] M. V. Berry, J. Mod. Opt. **34**, 1401 (1987).
[15] H. Jiao, S. R. Wilkinson, R. Y. Chiao, and H. Nathel, Phys. Rev. A **39**, 3475 (1989).
[16] M. Kitano, T. Yabuzaki, and T. Ogawa, Phys. Rev. Lett. **58**, 523 (1987).
[17] W. W. Chow, J. Gea-Banacloche, L. M. Pedrotti, V. E. Sanders, W. Schleich, and M. O. Scully, Rev. Mod. Phys. **57**, 61 (1985).
[18] M. V. Berry, Nature **326**, 277 (1987).
[19] M. V. Berry, Lectures given at the Ferrara School of Theoretical Physics on Anomalies, Defects, and Phase 1989 (in press).

Nth (even)-order minimum uncertainty products

Robert Lynch and Harry A. Mavromatis
Physics Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

(Received 11 October 1989; accepted for publication 14 March 1990)

This paper considers the problem of finding the quantum states that minimize the products of the (even) Nth-order fluctuation of two canonically conjugate operators. The problem is first attacked in an abstract form and an equation derived for the desired states. A consideration of the $N=2$ case then leads to a connection with the new concept of "squeezed states" of the electromagnetic field, and the usual exact solution involving the coherent state. The concept of "higher-order squeezing" is touched on to motivate the further discussion. The cases $N=4$, 6, and 8 are then taken up, and approximate solutions to them found via a new numerical technique. During these calculations, it is noted that only the first few terms in the expansion in number states figure in the solution; this observation is then exploited to find an approximate closed solution to the problem valid for all N.

I. INTRODUCTION

In this paper, we will consider the problem of finding the quantum states that minimize the products of the "Nth-order fluctuation" (to be defined below) of two canonically conjugate operators. In Sec. II, we state the problem in abstract, formal terms, and derive the equation obeyed by the desired states. In Sec. III, we consider the case $N=2$, for which we are able to write down the well-known harmonic oscillator solution. This solution is then used as a brief introduction to "squeezed states" in Sec. IV. Having described "ordinary" ($N=2$) squeezing, we then move on to the definition of higher-order ($N>2$) squeezing in Sec. V, and also describe how this relates to the states sought in the paper. After this physical motivation, the tenor of the paper reverts to an abstract discussion as the approximate solutions to the cases $N=4$, 6, and 8 are found in Sec. VI, VII, and VIII, respectively. In Sec. IX, an approximate solution to the general N case is given, and its properties explored. Section X then concludes the paper with a brief summary of the results obtained.

II. ABSTRACT STATEMENT OF PROBLEM

Consider two operators X,P satisfying

$$[X,P]=i. \tag{1}$$

The problem considered in this paper is to find the states $|\psi\rangle$ and the corresponding values $(U_N)_{\min}$, which minimize the Nth-order uncertainty product:

$$U_N\equiv\langle\psi|(\Delta X)^N|\psi\rangle\langle\psi|(\Delta P)^N|\psi\rangle, \tag{2}$$

subject to $\langle\psi|\psi\rangle=1$. Here,

$$\Delta X\equiv X-\langle X\rangle,\quad \Delta P\equiv P-\langle P\rangle. \tag{3}$$

Before pursuing the solution, first define new operators x, p that also satisfy $[x,p]=i$, by means of

$$x\equiv X-\langle X\rangle,\quad p\equiv P-\langle P\rangle. \tag{4}$$

By definition $\langle x\rangle=\langle p\rangle=0$, and the uncertainty product takes the form

$$U_N=\langle\psi|x^N|\psi\rangle\langle\psi|p^N|\psi\rangle. \tag{5}$$

Next, follow a generalization of a procedure of Jackiw.[1] Introduce a Lagrange multiplier, m, to incorporate the condition $\langle\psi|\psi\rangle=1$; then a necessary condition that U_N be a minimum is

$$\frac{\delta U_N}{\delta\langle\psi|}=m|\psi\rangle. \tag{6}$$

To further simplify the problem, let us restrict our attention to only those states for which $\langle x^N\rangle=\langle p^N\rangle$; carrying through the derivative in Eq. (6) using the expression for U_N in Eq. (5), and determining m, one finds

$$\tfrac{1}{2}(x^N+p^N)|\psi\rangle=\lambda|\psi\rangle \tag{7}$$

where $\lambda\equiv\langle x^N\rangle=\langle p^N\rangle$. Equation (7) is an eigenvalue equation for the desired[2] state, $|\psi\rangle$, whose corresponding eigenvalue, λ, gives the value of $(U_N)_{\min}^{1/2}$ for that state.

As pointed out in the Introduction, our main interest in pursuing such states lies in their role in "higher-order squeezing," to be discussed in Sec. V. In this connection odd-N states have a number of undesirable properties,[3] so we will restrict our attention to *even* $N=2$, 4, 6, ... In the next section, we consider the lowest order such, $N=2$.

III. CASE $N=2$

For $N=2$, Eq. (7) becomes

$$\tfrac{1}{2}(x^2+p^2)|\psi\rangle=\lambda|\psi\rangle. \tag{8}$$

This is the well-known harmonic oscillator problem,[4] with solutions

$$|\psi\rangle=|n\rangle,\quad \lambda=(U_2)_{\min}^{1/2}=(n+\tfrac{1}{2}), \tag{9}$$

where $|n\rangle$ are the usual number states. Here, $(U_2)_{\min}$ reaches its absolute minimum for $|\psi\rangle=|0\rangle$, $(U_2)_{\min}=\lambda^2=\frac{1}{4}$.

If we return to the variables X,P, the more general solution is found to be[5] the *coherent state*, $|\alpha\rangle$, with $\alpha=\langle X\rangle+i\langle P\rangle$, and the same value of $(U_2)_{\min}=\frac{1}{4}$. Furthermore, for this state one can show that $\langle(\Delta X)^2\rangle=\langle(\Delta P)^2\rangle=\frac{1}{2}$.

 0022-2488/90/081947-04$03.00

IV. "ORDINARY" ($N=2$) SQUEEZED STATES

So far we have been considering the restricted class of states for which $\langle x^2\rangle = \langle p^2\rangle$. If one relaxes this condition, so that the variances need no longer be equal, $\langle x^2\rangle \neq \langle p^2\rangle$, this leads to what have become known as "squeezed states.[6]" In particular, in a certain squeezed state the variance of x, for example, might be chosen to have a value $\langle x^2\rangle < \frac{1}{2}$, while the corresponding variance of p is $\langle p^2\rangle > \frac{1}{2}$, such that one still has $(U_2)_{\min} = \frac{1}{4}$. We then have "squeezed" or reduced the uncertainty in x, while increasing that of p; one may[7] also squeeze p at the expense of x.

This playful game of squeezing takes on a practical aspect when, for example, we consider quantum states of a single-mode electromagnetic field, E, at optical frequencies. Suppose we decompose this field into quadrature phase components, E_C, E_S:

$$E = E_C \cos \Omega t + E_S \sin \Omega t, \tag{10}$$

where Ω is the frequency of the field, and t is the time. In quantizing the field, E, E_C, and E_S become operators, and with a suitable choice of units, after some calculations a result is found that very much resembles Eq. (1)

$$[E_C, E_S] = i. \tag{11}$$

The electromagnetic field of a common optical source, the laser, can be modeled quite well as a coherent state, $|\alpha\rangle$. As we have seen above, and in accordance with Eq. (11), this state has its fluctuations—which show up in a detector as *noise*—distributed equally between the two quadrature components, E_C, E_S. Recently, a number of groups[8] have managed to generate squeezed light fields, in which one quadrature component has decreased noise, at the expense of increased noise in the other quadrature. In a coherent state the fluctuations are independent of the amplitude, α. In particular, as $\alpha \to 0$, we approach the vacuum state; squeezed states are then seen to be such that one quadrature of the field has less noise than the vacuum, which accounts for some of the interest shown in them.

The techniques needed to generate squeezed light sources are quite complex, and will not be gone into here. Applications of squeezed light include low-noise optical communications, detection of gravity waves by optical interferometry, and, no doubt, other uses yet unvisualized. In the last few years work in this field has undergone an explosive increase.[9]

V. HIGHER-ORDER SQUEEZING

For "ordinary" second-order squeezing the coherent state is the benchmark, as it minimizes the second-order uncertainty product. Hong and Mandel[3] have sought to use this fact to generalize the concept of squeezing to higher-order (even) moments of the field. In a field state $|\psi\rangle$ a quadrature component E_C or E_S is said to exhibit Nth-order squeezing if

$$\langle\psi|(\Delta E_i)^N|\psi\rangle < \langle\alpha|(\Delta E_i)^N|\alpha\rangle \quad (i = C \text{ or } S), \tag{12}$$

i.e., the Nth-order fluctuation in the given state is less than that which obtains in the coherent state $|\alpha\rangle$. In fact, given the commutation relation between E_C and E_S, it is possible to show[3] that $\langle\alpha|(\Delta E_i)^N|\alpha\rangle = (N-1)!!/2^{N/2}$, so that the criterion for higher-order squeezing, Eq. (12), can be written more explicitly

$$\langle\psi|(\Delta E_i)^N|\psi\rangle < (N-1)!!/2^{N/2} \quad (i = C \text{ or } S). \tag{13}$$

Here $(N-1)!! \equiv 1\cdot 3\cdot 5\cdots(N-1)$. (For N = odd, the moment vanishes[3] in a coherent state, one of the difficulties with defining odd N higher-order squeezing, alluded to above.)

However, while the coherent state does minimize the *second-order* uncertainty product, this does not necessarily follow for $N > 2$. In fact, in a comment[10] on Hong and Mandel's papers,[3] one of us gave an example of a state infinitesimally different from the coherent state, the "Jackiw state[1]"

$$|\psi_1\rangle = \Omega[\,|\alpha\rangle + \epsilon \exp(-|\alpha|^2/2)|0\rangle\,], \tag{14}$$

in which *both* quadrature components E_C and E_S are squeezed in fourth order; from this it follows that the fourth-order uncertainty product in this state is also less than that for the coherent state. Here, Ω is a normalization factor, ϵ is a small positive parameter, and $|\alpha|$ is large.

However, in the Jackiw state, both the squeezing and the lowering of the uncertainty product are infinitesimal. Thus, although this result was of some formal value for examining the properties of higher-order squeezing, it did not have any practical usefulness. The question then became a general one, framed above, to find the states that minimize the (even) Nth-order minimum uncertainty products. Besides the possible utility of such states in discussions of higher-order squeezing, the problem is an interesting one of a general mathematical and quantum-mechanical nature. Hence, in the next three sections, which deal with the cases $N = 4$, 6, and 8, we revert to an abstract discussion of the solution of this problem.

VI. CASE $N=4$

One now needs to solve

$$\tfrac{1}{2}(x^4 + p^4)|\psi\rangle = \lambda\,|\psi\rangle, \tag{15}$$

a difficult problem that does not appear to have an exact solution.

A technique for the solution of Eq. (15) is suggested by a paper of Partovi and Blankenbecler.[11] Define

$$H_0 \equiv \tfrac{1}{4}(x^2 + p^2)^2, \tag{16}$$

$$H_1 \equiv \tfrac{1}{4}(x^2 - p^2)^2. \tag{17}$$

Equation (15) then becomes

$$(H_0 + H_1)|\psi\rangle = \lambda\,|\psi\rangle, \tag{18}$$

and one can attempt to solve this by perturbation theory.[12] The zeroth order wave functions are the eigenfunctions of H_0 that are just the number states, $|n\rangle$, with eigenvalues $(n + \frac{1}{2})^2$.

The detailed calculation to second order in H_1 has already been published,[13] so the steps of the derivation will not be repeated here. In further work, by a straightforward and laborious perturbation approach, this calculation has been extended to fifth order. The exact values for the contributions of the zeroth, first, second, third, fourth, and fifth or-

ders are, respectively, 0.25, 0.5, -0.075, 0.0375, -0.0239, and 0.01715. The value for λ to fifth order is thus found to be 0.7057, or 94.09% of the coherent state value of 0.75 as predicted by Eq. (13) above. The calculations are very tedious to perform by hand, and a numerical method was sought to expedite the calculations for this and higher values of N.

The new approach is as follows: First, define the usual raising and lowering operators a, a^+ where

$$x=(a+a^+)/(2)^{1/2},\quad p=(a-a^+)/(2)^{1/2}i. \tag{19}$$

By straightforward calculation one finds that

$$H_1=\tfrac{1}{4}(a^4+a^2a^{+2}+a^{+2}a^2+a^{+4}). \tag{20}$$

The problem then boils down to diagonalizing (H_0+H_1), with the smallest eigenvalue giving the desired value of $(U_4)^{1/2}_{\min}$.

Next, write the solution $|\psi\rangle$ in the form

$$|\psi\rangle=\sum_{n=0}^{\infty}c_n|4n\rangle. \tag{21}$$

One notes the form of $|\psi\rangle$, i.e., only containing the states $n=0, 4, 8, \ldots$. It can easily be seen that states of odd parity 1, 3, 5, ... are automatically excluded since H_0 and H_1 as defined in Eqs. (16) and (17) both commute with the parity operator. The ground state of the system being of even parity, H_1 only mixes in even parity components. That the states 2, 6, 10, etc. are not involved is a feature of H_1 that is obvious from the form[14] of Eq. (20), i.e., that H_1 only connects states n, n' that differ by multiples of 4.

The new numerical approach mentioned above consists of cutting off the sum in Eq. (21) at a finite value, K. This amounts to reducing the infinite-dimensional space of the exact problem to a more tractable, finite $(K+1)$-dimensional space. The computer program[15] MathCAD 2.0 is then utilized to iteratively determine the best numerical values of $(U_4)^{1/2}_{\min}$ for $K=0, 1, 2, \ldots$. If we denote the total Hamiltonian for the case we are working on $(N=4)$ as $H_4\equiv H_0+H_1$, then in the basis $|4n\rangle$ one finds that H_4 has the following diagonal and off-diagonal forms:

$$(H_4)^d_{m,n}=\tfrac{1}{4}[(8n+1)^2+(4n+1)(4n+2)+4n(4n-1)]\delta_{m,n}, \tag{22}$$

and

$$(H_4)^0_{m,n}=\tfrac{1}{2}[n(4n-1)(4n-2)(4n-3)]^{1/2}\delta_{m,n-1}+\tfrac{1}{2}[(4n+1)(4n+2)(4n+3)(n+1)]^{1/2}\times\delta_{m,n+1}. \tag{23}$$

[The reader may find it useful at this point to recall that $(H_4)_{m,n}=\langle 4m|H_4|4n\rangle$.]

The "root" function of MathCAD 2.0 is now utilized to iteratively determine the smallest eigenvalue, λ, that satisfies the characteristic equation, $\det(H_4-\lambda I)=0$, in the $(K+1)$-dimensional space. Carrying out the calculations, it is found that the most significant improvement toward the "true" value of λ occurs between $K=0$, which gives $\lambda=3/4=0.75$ [the coherent state value of Eq. (13)], and $K=1$, yielding $\lambda=0.7000$, just a trifle less than a 7% reduction from the coherent-state value. Successive iterations up to $K=6$ only improve results a little, with $K=6$ yielding $\lambda=0.6984$. For $K>6$ MathCAD 2.0 begins to yield erratic results, indicating that it has reached the limits of its precision. It is interesting to recall, as mentioned above, that fifth-order perturbation theory gives $\lambda=0.7057$, so it would appear that the new numerical technique is a significant improvement over conventional perturbation theory, at least for this problem.

VII. CASE $N=6$

For this case, the Hamiltonian for Eq. (7) is $H_6\equiv\frac{1}{2}(p^6+x^6)$. We rewrite this in the form $H_6=H_0+H_1$, where

$$H_0\equiv\tfrac{1}{8}(p^2+x^2)^3, \tag{24}$$

$$H_1\equiv\tfrac{1}{8}[(p^2-x^2)(p^2+x^2)(p^2-x^2)+(p^2+x^2)(p^2-x^2)^2+(p^2-x^2)^2(p^2+x^2)]. \tag{25}$$

As before, H_0 is diagonal in the $|4n\rangle$ basis, with eigenvalues $(8n+1)^3/8$. After some calculation the on- and off-diagonal terms of H_6 are found to be

$$(H_6)^d_{m,n}=\tfrac{1}{8}[(8n+1)^3+4(8n+1)(16n^2+4n+1)+(8n+5)(4n+1)(4n+2)+4n(8n-3)(4n-1)]\delta_{m,n} \tag{26}$$

and

$$(H_6)^0_{m,n}=\tfrac{3}{4}(4n+5/2)[(4n+1)(4n+2)(4n+3)\times(4n+4)]^{1/2}\delta_{m,n+1}+\tfrac{3}{4}(4n-3/2)\times[4n(4n-1)(4n-2)(4n-3)]^{1/2}\delta_{m,n-1}. \tag{27}$$

The $K=0$ value, i.e., that for the coherent state [Eq. (13)], is $(U_6)^{1/2}_{\min}=15/8=1.875$. Invoking MathCAD 2.0, this value is reduced at $K=1$ by about 19%, to 1.5240. As with $N=4$, there is again the same pattern of only small improvements of a few percent with increasing K, the value 1.4765 being found for $K=6$; for $K>6$ MathCAD 2.0 loses precision and returns erratic values.

VIII. CASE $N=8$

For $H_8=\frac{1}{2}(p^8+x^8)$ one finds the following expressions for on- and off-diagonal terms:

$$(H_8)^d_{m,n}=\{\tfrac{1}{16}[(6(4n+\tfrac{1}{2})^2+\tfrac{3}{2})^2+(4n+1)(4n+2)\times(4n+3)(4n+4)+4n(4n-1)(4n-2)\times(4n-3)]+(4n+3/2)^2(4n+1)(4n+2)+(4n-\tfrac{1}{2})^2(4n-1)(4n)\}\delta_{m,n} \tag{28}$$

and

$$(H_8)^0_{m,n}=(A+A^T+B+B^T)_{m,n}, \tag{29}$$

where

$$(A)_{m,n}\equiv[(4n+1)(4n+2)(4n+3)(4n+4)]^{1/2}\times\{(4n+\tfrac{3}{2})(4n+\tfrac{7}{2})+\tfrac{1}{16}[12(16n^2+20n+\tfrac{41}{4})+3]\}\delta_{m,n+1}, \tag{30}$$

$$(B)_{m,n} \equiv \tfrac{1}{16}[(4n+8)!/(4n)!]^{1/2}\delta_{m,n+2}, \quad (31)$$

and "T" denotes matrix transpose.

The $K=0$ or coherent-state value is now $105/16 = 6.5625$. The first iteration yields about a 30% reduction below this, with $(U_8)^{1/2}_{\min} = 4.5956$ for $K=1$. Upon higher-dimensional iteration, an approximately further 7% reduction is found, and the $K=6$ value is found to be 4.1447. Again, beyond this value of K MathCAD 2.0 becomes erratic, indicating loss of precision.

IX. APPROXIMATE SOLUTION FOR ANY N

We have seen that, in the cases $N=4$, 6, and 8, the solution in the 2×2-dimensional space yields a close approximation to that found by higher-dimensional iteration. This observation may be exploited to find an approximate solution to the general problem, good for any N.

Write

$$|\psi\rangle = \beta|0\rangle + (1-\beta^2)^{1/2}|4\rangle. \quad (32)$$

It is not too difficult to compute the 2×2 dimensional matrix $\langle\psi|H_N|\psi\rangle$ and diagonalize it exactly, whereupon one finds the result in this space of

$$(U_N)^{1/2}_{\min} = [(N-1)!!/2^{N/2}][(N^4+4N^3+20N^2+32N+48) - \{(N^4+4N^3+20N^2+32N+48)^2 - 768(N^3+2N^2+4N+3)\}^{1/2}]/48. \quad (33)$$

This result naturally agrees with the $K=1$ values quoted above for $N=4$, 6, and 8. For $N=10$, the equation yields a value of 17.9570, compared to the coherent-state result of $945/32 = 29.53125$, an almost 40% reduction below this latter value. As $N\to\infty$, Eq. (33) predicts that $(U_N)_{\min}\to 0$.

X. SUMMARY AND CONCLUSIONS

In summary, the problem of finding the even Nth order minimum uncertainty product has been studied, and the fundamental equation (7) for its value has been derived. The $N=2$ case was considered, and the usual coherent-state solution was rediscovered. For the cases $N=4$, 6, and 8, a new numerical technique was applied to obtain approximate solutions for the minimum uncertainty product. Finally, an observation made in the course of solving these cases allowed an approximate solution of the problem to be given for any N. For convenience, the results obtained are drawn together in Table I.

From an abstract point of view, the problem considered is an interesting one, and the results obtained may prove useful in a number of investigations. In particular, the numerical approach to the approximate solution of the problem, i.e., diagonalization of successively higher-dimensional approximations to the true Hamiltonian by a commercial software program (MathCAD 2.0), may indicate a useful new approach to quantum-mechanical problems usually attacked via perturbation theory of the standard variety.[12]

TABLE I. Summary of results obtained for $(U_N)^{1/2}_{\min}$.

N	Coherent-state value	Approximate solution	% of coherent state value
2	$\frac{1}{2}=0.5$	...	...
4	$\frac{3}{4}=0.75$	0.6984	93.12
6	$\frac{15}{8}=1.875$	1.4765	78.75
8	$\frac{105}{16}=6.5625$	4.1447	63.16
⋮	⋮	⋮	⋮
N	$(N-1)!!/2^{N/2}$	See Eq. (33)	...

Looking at the problem from a physical angle, it appears that states of the electromagnetic field as simple in form as Eq. (32) yield substantial reductions in the higher-order fluctuations of the quadrature field components, below their coherent-state (which includes the vacuum) values, in *both* quadratures simultaneously. While it cannot be denied that at the present time, production of just ordinary ($N=2$) squeezing is quite difficult, future techniques may evolve that allow higher-order squeezing to be produced, and its properties to be studied.

ACKNOWLEDGMENT

One of us (R. L.) would like to gratefully acknowledge the useful input provided by Dr. Robert Mills of Ohio State University in the initial stages of investigation of the general problem.

[1] R. Jackiw, J. Math. Phys. **9**, 339 (1968).

[2] Technically, these are not necessarily minimum, but extremum states.

[3] C. K. Hong and L. Mandel, Phys. Rev. Lett. **54**, 323 (1985); Phys. Rev. A **32**, 974 (1985).

[4] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), 2nd ed., Sec. 15.8.

[5] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).

[6] For an excellent recent tutorial introduction to squeezed states, see G. Leuchs, Contemp. Phys. **29**, 299 (1988).

[7] Squeezed states are actually slightly more general than this: It is possible to "squeeze" the variance along a direction that lies neither along x or p. The x,p uncertainty product then takes on a greater value than the minimum of $\frac{1}{4}$.

[8] A recent review of the field is found in R. Loudon and P. L. Knight, J. Mod. Opt. **34**, 709 (1987).

[9] A good and recent snapshot of the field may be found in the papers contained in the special issue, "Squeezed States of the Electromagnetic Field," J. Opt. Soc. Am. B **4**, (10) (1987).

[10] R. Lynch, Phys. Rev. A **33**, 4431 (1986).

[11] M. H. Partovi and R. Blankenbecler, Phys. Rev. Lett. **57**, 2887 (1986).

[12] Reference 4, Chap. 17.

[13] R. Lynch, Phys. Rev. A **36**, 4501 (1987).

[14] This also follows from the fact that $|\psi\rangle$ "is equal to its own Fourier transform (because of x–p symmetry)." (R. Mills, Ohio State University, private communication.)

[15] Product of MathSoft, Inc., One Kendall Square, Cambridge, MA 02139.

Algebraic calculation of the Green's function for the Hartmann potential

Arvind Narayan Vaidya and Henrique Boschi Filho
Instituto de Física, Universidade Federal do Rio de Janeiro, Cidade Universtitária, Ilha do Fundão, 21941, Rio de Janeiro, RJ, Brazil

(Received 1 February 1989; accepted for publication 11 April 1990)

Using the so(2,1) Lie algebra and the Baker–Campbell–Hausdorff formulas, the Green's function for the Hartmann potential is constructed and its bound state energy spectrum is found. Also, this Green's function is constructed in a coherent state basis and the equivalence of the two descriptions is shown.

I. INTRODUCTION

The potential

$$V(r,\theta) = \gamma\sigma^2(2a_0/r - \gamma a_0^2/r^2 \sin^2\theta)\,\epsilon_0, \tag{1}$$

was proposed by Hartmann[1] in 1972 to describe organic molecules like benzene. Here, a_0 is the Böhr radius, ϵ_0 the ground state energy of the hydrogen atom, and γ, σ are positive real parameters. The Schrödinger equation for this potential is exactly solvable[1] and several others solutions have appeared, among which, we mention those of Gerry[2] and Kibler and Winternitz[3] that discuss the dynamical symmetry group. A path integral calculation of the Green's function for this potential has been done by Chetouani *et al.*[4]

This paper does not give a new result for the quantization of the Hartmann potential, but introduces a new prescription for the calculation of Green's functions. The technique consists of writing the resolvent operator in the exponential representation of Schwinger.[5] This exponential, which involves the so(2,1) ⊗ so(2,1) generators, related to the dynamical symmetry of the Hartmann potential,[2] can be disentangled by using Baker–Campbell–Hausdorff (BCH) formulas. Finally, the explicit action of the exponentials of the generators on functions of coordinates is determined by using a variant of the technique of Mil'shtein and Strakhovenko,[6] which was originally applied to the Coulomb potential. This gives the explicit form of the Green's function from which one can determine the bound state energy spectrum and the corresponding wave functions. It is also possible to consider the Green's function using the so(2,1) coherent states of Perelomov[7] and apply the technique described above to obtain the final results in an equivalent manner.

This paper is organized as follows. In Sec. II, we review the so(2,1) Lie algebra and two BCH formulas for disentangling exponentials of the generators, which appear in the following sections. The explicit construction of the Green's function for the Hartmann potential and the determination of its bound state spectrum are done in Sec. III. The coherent state Green's function and its trace are obtained in Sec. IV and the conclusions are given in Sec. V.

II. THE BCH FORMULAS FOR THE so(2,1) LIE ALGEBRA

We choose the basis such that the generators $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{T}_3$ of the so(2,1) Lie algebra obey the commutation relations,

$$[\mathbf{T}_1, \mathbf{T}_2] = -i\mathbf{T}_1, \tag{2a}$$

$$[\mathbf{T}_1, \mathbf{T}_3] = -i\mathbf{T}_2, \tag{2b}$$

$$[\mathbf{T}_2, \mathbf{T}_3] = -i\mathbf{T}_3. \tag{2c}$$

A simple matrix representation of this algebra is given by

$$\mathbf{T}_1 = (\sigma_1 - i\sigma_2)/2\sqrt{2}, \tag{3a}$$

$$\mathbf{T}_2 = -i\sigma_3/2, \tag{3b}$$

$$\mathbf{T}_3 = (\sigma_1 + i\sigma_2)/2\sqrt{2}, \tag{3c}$$

where σ_i are the Pauli spin matrices.

In the following sections, we will need to use BCH formulas that allow us to disentangle the exponentials of the generators $\mathbf{T}_i$, in the case when they are realized as differential operators. The explicit forms are given later. It may be noted, however, that the disentanglement formulas that depend only on the commutation relations between the generators may be derived by using any convenient faithful realization of these. Further, since the algebra is finite dimensional, the exponential of a linear combination of the generators can be expressed as the product of exponentials of the generators.[8] The particular cases of this result are the formulas,

$$\exp\{-is(2\mathbf{T}_1 + k^2\mathbf{T}_3)\} = \exp\{-ia\mathbf{T}_3\}\exp\{-ib\mathbf{T}_2\}\exp\{-ic\mathbf{T}_1\}, \tag{4}$$

where

$$a = k\tan ks, \tag{5a}$$

$$b = 2\ln(\cos ks), \tag{5b}$$

$$c = 2k^{-1}\tan ks, \tag{5c}$$

and

$$\exp\{-ic\mathbf{T}_1\}\exp\{-i\sigma\mathbf{T}_3\} = \exp\{-i\rho\mathbf{T}_3\}\exp\{-i\beta\mathbf{T}_2\}\exp\{-i\tau\mathbf{T}_1\}, \tag{6}$$

where

$$\rho = i\sigma/(1 - ic\sigma/2), \tag{7a}$$

$$\beta = 2\ln(1 - ic\sigma/2), \tag{7b}$$

$$\tau = c/(1 - ic\sigma/2), \tag{7c}$$

with s, k, and σ real.

Using the realization given in Eqs. (3a)–(3c), it is very easy to verify that Eqs. (4) and (6) hold.

An alternative method is to differentiate both sides of Eqs. (4) and (6) with respect to variable s and equate the coefficients of $\mathbf{T}_i$ on both sides. This leads to differential equations:

 0022-2488/90/081951-04$03.00

$$\dot{c}\, e^b = 2, \tag{8a}$$

$$\dot{b} + \dot{c}\, a e^b = 0, \tag{8b}$$

$$\dot{a} + \dot{b}\, c + \dot{c} a^2 e^b/2 = k^2, \tag{8c}$$

in the first case and

$$\dot{c} + i c^2 \dot{\sigma}/2 = \dot{\tau} e^{\beta}, \tag{9a}$$

$$-i\dot{\sigma} c = \dot{\beta} + \rho\, \dot{\tau} e^{\beta}, \tag{9b}$$

$$i\dot{\sigma} = \dot{\rho} + \dot{\beta}\rho + \dot{\tau} e^{\beta} \rho^2/2, \tag{9c}$$

in the second case. Here, the overdot denotes derivative with respect to s. Solution of Eqs. (8) and (9) leads to Eqs. (5) and (7).

III. THE GREEN'S FUNCTION IN SQUARED PARABOLIC COORDINATES

In this section, we use the "squared" parabolic coordinates[2] ξ, η, and ϕ, that are related to Cartesian coordinates by

$$x = \xi\eta \cos\phi, \tag{10a}$$

$$y = \xi\eta \sin\phi, \tag{10b}$$

$$z = (\eta^2 - \xi^2)/2, \tag{10c}$$

where the former are restricted by $0 \leqslant \xi < \infty$, $0 \leqslant \eta < \infty$, $0 \leqslant \phi \leqslant 2\pi$.

The relation of ξ, η with the spherical polar coordinates r, θ is given by

$$r = (\xi^2 + \eta^2)/2, \tag{11a}$$

$$\sin\theta = 2\xi\eta/(\xi^2 + \eta^2). \tag{11b}$$

In the place of the Hartmann potential, it is convenient to use the potential

$$V(r,\theta) = -\gamma\sigma^2(1/r - q\gamma/2r^2 \sin^2\theta), \tag{12}$$

where γ, σ, and q are dimensionless parameters and atomic units $a_0 = 1$, $\epsilon_0 = -1/2$ are used. The case $q = 1$ corresponds to the Hartmann potential while $q = 0$ gives the Coulomb potential.

The time-independent Green's function $G_E(\mathbf{r}, \mathbf{r}')$, for a given energy E, satisfies the equation

$$(H - E) G_E(\mathbf{r},\mathbf{r}') = \delta^3(\mathbf{r} - \mathbf{r}'), \tag{13}$$

where the Hamiltonian H is written as

$$H = \mathbf{p}^2/2 + V(r,\theta). \tag{14}$$

Since $V(r,\theta)$ is independent of ϕ, one may expand $G_E(\mathbf{r},\mathbf{r}')$ as

$$G_E(\mathbf{r},\mathbf{r}') = \sum_{m=-\infty}^{\infty} \frac{\exp\{im(\phi - \phi')\}}{2\pi} G_{Em}(\xi,\eta;\xi',\eta'), \tag{15}$$

so Eq. (13) becomes

$$(H_m - E) G_{Em}(\xi,\eta;\xi',\eta') = \delta(\xi - \xi')\delta(\eta - \eta')/\xi\eta(\xi^2 + \eta^2), \tag{16}$$

where

$$(\xi^2 + \eta^2) H_m = -\frac{1}{2}\left(\frac{\partial^2}{\partial\xi^2} + \frac{1}{\xi}\frac{\partial}{\partial\xi} - \frac{M^2}{\xi^2}\right) - \frac{1}{2}\left(\frac{\partial^2}{\partial\eta^2} + \frac{1}{\eta}\frac{\partial}{\partial\eta} - \frac{M^2}{\eta^2}\right) - 2\sigma^2\gamma, \tag{17}$$

and the parameter M is defined by

$$M^2 = m^2 + q\gamma^2\sigma^2. \tag{18}$$

Henceforth, for convenience, we take M to be positive. It may be verified that the operators,

$$\mathbf{T}_1(\xi) = -\frac{1}{4}\left(\frac{\partial^2}{\partial\xi^2} + \frac{1}{\xi}\frac{\partial}{\partial\xi} - \frac{M^2}{\xi^2}\right), \tag{19a}$$

$$\mathbf{T}_2(\xi) = -\frac{i}{2}\left(\xi\frac{\partial}{\partial\xi} + 1\right), \tag{19b}$$

$$\mathbf{T}_3(\xi) = \tfrac{1}{2}\xi^2, \tag{19c}$$

obey the commutation relations of the so(2,1) Lie algebra as given in Eqs. (2a)–(2c). Using similar operators $\mathbf{T}_i(\eta)$ as defined for $\mathbf{T}_i(\xi)$ above, one may rewrite Eq. (16) as

$$[2(\mathbf{T}_1(\xi) + \mathbf{T}_1(\eta)) + k^2(\mathbf{T}_3(\xi) + \mathbf{T}_3(\eta)) - \sigma^2\gamma] \times G_{Em}(\xi,\eta;\xi',\eta') = \delta(\xi - \xi')\delta(\eta - \eta')/\xi'\eta', \tag{20}$$

where

$$E = -k^2/2. \tag{21}$$

Next, using the Schwinger representation,[5] we can write $G_{Em}(\xi,\eta;\xi'\eta')$ in the form,

$$G_{Em}(\xi,\eta;\xi',\eta') = \frac{i}{\xi'\eta'}\int_0^{\infty} ds \exp\{2is\sigma^2\gamma\} \times \exp\{-is[2\mathbf{T}_1(\xi) + k^2\mathbf{T}_3(\xi)]\}\delta(\xi - \xi') \times \exp\{is[2\mathbf{T}_1(\eta) + k^2\mathbf{T}_3(\eta)]\}\delta(\eta - \eta'). \tag{22}$$

Due to the symmetry in ξ, η we consider only

$$\exp\{-is[2\mathbf{T}_1 + k^2\mathbf{T}_3]\}\delta(\xi - \xi'),$$

where $\mathbf{T}_i$ stand for $\mathbf{T}_i(\xi)$. The operator $\mathbf{T}_3$ is realized here as $\xi^2/2$. The form of $\mathbf{T}_2$, given in Eq. (19b), gives

$$\exp\{-ib\,\mathbf{T}_2\} f(\xi) = \exp\{-b/2\} f(\xi \exp\{-b/2\}). \tag{23}$$

Using a variant of a technique of Mil'shtein and Strakhovenko,[6] one may find the effect of $\exp\{-ic\mathbf{T}_1\}$, on a function $f(\xi)$, which posses the inverse Laplace transform:

$$f(\xi) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} d\sigma\, F(\sigma) \exp\left\{\frac{\sigma\xi^2}{2}\right\} \xi^M, \tag{24}$$

where the usual transformation is given by

$$F(\sigma) = \int_0^{\infty} d\xi\, f(\xi) \exp\left\{\frac{-\sigma\xi^2}{2}\right\} \xi^{1-M}. \tag{25}$$

We note first that

$$\mathbf{T}_1 \xi^M = 0. \tag{26}$$

Using Eqs. (6), (7), (19), (23), and (24) we get

$$\exp\{-ic\mathbf{T}_1\}\exp\{\sigma\xi^2/2\}\,\xi^M = \xi^M \frac{\exp\{\sigma\xi^2/(2-i\sigma c)\}}{(1-i\sigma c/2)^{1+M}}. \tag{27}$$

Substituting it in Eq. (24) and using Eq. (25), we arrive at

$$\exp\{-ic\mathbf{T}_1\} f(\xi) = \int_0^\infty x\,dx\, f(x)\left(\frac{\xi}{x}\right)^M \frac{1}{2\pi i} \times \int_{-i\infty}^{i\infty} d\sigma \frac{\exp\{-\sigma x^2/2 + \sigma\xi^2/(2-i\sigma c)\}}{(1-i\sigma c/2)^{1+M}}. \tag{28}$$

Expanding the term with ξ^2 in the exponential, one can evaluate the integral in $d\sigma$, by residue calculus, getting

$$\exp\{-ic\mathbf{T}_1\} f(\xi) = -\frac{i}{c}\int_0^\infty x\,dx\, f(x)\exp\left\{i\frac{(\xi^2+x^2)}{c}\right\} \times J_M\left(2\frac{\xi x}{c}\right)\exp\{-i\pi M\}, \tag{29}$$

where $J_M(z)$ is the Bessel function[9] of order M. Making $f(\xi) = \delta(\xi - \xi')$, Eqs. (3), (16), (21), and (27) give

$$\exp\{-is[2\mathbf{T}_1 + k^2\mathbf{T}_3]\}\delta(\xi-\xi') = -\frac{ik\xi'}{\sin ks}\exp\left\{\frac{i(\xi^2+\xi'^2)}{c}\right\} \times J_M\frac{k\xi'\xi}{\sin ks}\exp\{-i\pi M\}, \tag{30}$$

where we have used the relations

$$\exp\{-b/2\}/c = k/2\sin ks \tag{31}$$

and

$$\exp\{-b\} - ac/2 = 1, \tag{32}$$

which follow from Eqs. (5a)–(5c). Going back to Eq. (20), we get

$$G_{Em}(\xi,\eta;\xi',\eta') = -i\int_0^\infty ds \frac{k^2\exp\{2is\sigma^2\gamma\}}{\sin^2 ks} \times\exp\{(ik/2)(\xi^2+\eta^2+\xi'^2+\eta'^2)\cot ks\} \times J_M(k\xi'\xi/\sin ks)\, J_M(k\eta'\eta/\sin ks) \times\exp\{-2i\pi M\}. \tag{33}$$

This result agrees, up to a phase factor, with that of the path integral calculation of Chetouani *et al.*[4] Following their subsequent manipulations of this result, one can deduce the bound state energy spectrum

$$E_N = -\gamma^2\sigma^4/2N^2, \tag{34}$$

where

$$N = n_1 + n_2 + M + 1, \tag{35}$$

and n_1, n_2 are integers. The wave functions $\Psi_{n_1,n_2,m}(\xi,\eta,\phi)$ may be similarly found.

IV. COHERENT STATES

It is convenient to now use the basis

$$\mathbf{K}_0 = \mathbf{T}_1 + \mathbf{T}_3/2, \tag{36a}$$

$$\mathbf{K}_1 = \mathbf{T}_1 - \mathbf{T}_3/2, \tag{36b}$$

$$\mathbf{K}_2 = \mathbf{T}_2, \tag{36c}$$

and put $\mathbf{K}_\pm = \mathbf{K}_1 \pm i\mathbf{K}_2$. Then Eqs. (2a)–(2c) lead to

$$[\mathbf{K}_0,\mathbf{K}_\pm] = \pm\mathbf{K}_\pm, \tag{37a}$$

$$[\mathbf{K}_+,\mathbf{K}_-] = -2\mathbf{K}_0. \tag{37b}$$

The Casimir operator $\mathbf{C}$, written in the representation given by Eqs. (19a)–(19c) is

$$\mathbf{C} = \mathbf{K}_0^2 - \mathbf{K}_1^2 - \mathbf{K}_2^2 = (M^2-1)/4. \tag{38}$$

Consider the unitary irreducible representation $\mathfrak{D}^+[(M+1)/2]$ of so(2,1) Lie algebra.[10] The basis states can be chosen to be $|p,M\rangle$, such that

$$\mathbf{K}_0|p,M\rangle = (p+M)|p,M\rangle \quad (p=0,1,2\ldots) \tag{39}$$

and

$$(\mathbf{K}_0^2 - \mathbf{K}_1^2 - \mathbf{K}_2^2)|p,M\rangle = \frac{M^2-1}{4}|p,M\rangle. \tag{40}$$

We define the Perelomov[7] coherent states by

$$|\lambda,M\rangle = \exp\{\alpha\mathbf{K}_+ - \alpha^*\mathbf{K}_-\}|0,M\rangle, \tag{41}$$

where

$$\lambda = -(\alpha/|\alpha|)\tanh|\alpha|. \tag{42}$$

Then

$$|\lambda,M\rangle = (1-|\lambda|^2)^{(M+1)/2} \times\sum_{p=0}^\infty\left(\frac{\Gamma(p+2k)}{p!\Gamma(2k)}\right)^{1/2}\lambda^p|p,M\rangle, \tag{43}$$

with

$$\langle\lambda,M|\lambda',M\rangle = [(1-|\lambda|^2)(1-|\lambda'|^2)]^{(M+1)/2} \times(1-\lambda^*\lambda')^{-(M+1)} \tag{44}$$

and the resolution of unity is given by

$$\mathbf{1} = \int d\mu_M(\lambda)|\lambda,M\rangle\langle\lambda,M|, \tag{45}$$

where

$$d\mu_M(\lambda) = M\,d^2\lambda/\pi(1-|\lambda|^2)^2. \tag{46}$$

One may introduce physical coherent states $|\lambda,M\rangle$, by a pseudo-rotation related to the parameter θ:

$$\widetilde{|\lambda,M\rangle} = \exp\{i\theta\mathbf{K}_2\}|\lambda,M\rangle, \tag{47}$$

which still obey Eqs. (44) and (45), since $\mathbf{K}_2$ is Hermitian.

The physical coherent states relevant for the Hartmann potential are

$$\widetilde{|\lambda_1,\lambda_2,M\rangle} = \widetilde{|\lambda_1,M\rangle}\otimes\widetilde{|\lambda_2,M\rangle}, \tag{48}$$

where $\widetilde{|\lambda_1,M\rangle}$, $\widetilde{|\lambda_2,M\rangle}$ are the physical coherent states associated with the representations given by $\mathbf{T}_i(\xi)$ and $\mathbf{T}_i(\eta)$, respectively.

We consider now Eq. (22) in operator form:

$$G_{Em} = i\int_0^\infty ds\exp\{2is\gamma\sigma^2\} \times\exp\{-is[2\mathbf{T}_1(\xi) + k^2\mathbf{T}_3(\xi)]\} \times\exp\{-is[2\mathbf{T}_1(\eta) + k^2\mathbf{T}_3(\eta)]\} \tag{49}$$

and the matrix elements

$$p_{Em}(\lambda_1,\lambda_2;\lambda_1',\lambda_2') = \langle\widetilde{\lambda_1,\lambda_2,M}|G_{Em}|\widetilde{\lambda_1',\lambda_2',M}\rangle. \tag{50}$$

Next we calculate

$$\exp\{-i\theta \mathbf{T}_2\}(2\mathbf{T}_1 + k^2\mathbf{T}_3)\exp\{i\theta \mathbf{T}_2\}$$
$$= e^{\theta}(2\mathbf{T}_1 + k^2 \exp\{-2\theta\}\mathbf{T}_3)$$
$$= 2k\,\mathbf{K}_0\,, \tag{51}$$

with the choice $e^{\theta} = k$. Thus

$$p_{Em}(\lambda_1,\lambda_2;\lambda_1',\lambda_2') = i\int_0^{\infty} ds \exp\{2is\gamma\sigma^2\}$$
$$\times\langle\lambda_1,\lambda_2,M|\exp\{-2iks\mathbf{K}_0(\xi)\}$$
$$\times\exp\{-2iks\mathbf{K}_0(\eta)\}|\lambda_1',\lambda_2',M\rangle\,. \tag{52}$$

Using Eq. (44), we rewrite Eq. (52) as

$$p_{Em}(\lambda_1,\lambda_2;\lambda_1',\lambda_2') = i\int_0^{\infty} ds \exp\{2is[\gamma\sigma^2 - k(M+1)]\}$$
$$\times\left[\frac{(1-|\lambda_1'|^2)(1-|\lambda_1|^2)(1-|\lambda_2'|^2)(1-|\lambda_2|^2)}{(1-\lambda_1'\lambda_1^{*}\exp\{-2iks\})^2(1-\lambda_2'\lambda_2^{*}\exp\{-2iks\})^2}\right]^{(M+1)/2}, \tag{53}$$

which leads to the Green's function for the Hartmann potential over the $so(2,1)\otimes so(2,1)$ coherent states. To find the spectrum of this system, we may take the trace of the Green's function:

$$\sum_m \mathrm{Tr}\,p_{Em}(\lambda_1,\lambda_2;\lambda_1',\lambda_2') = i\sum_m\int_0^{\infty} ds \exp\{2is[\gamma\sigma^2 - k(M+1)]\}$$
$$\times\int d\mu_M(\lambda_1)\,d\mu_M(\lambda_2)\left[\frac{(1-|\lambda_1|^2)(1-|\lambda_2|^2)}{(1-|\lambda_1|^2\exp\{-2iks\})(1-|\lambda_2|^2\exp\{-2iks\})}\right]^{M+1}. \tag{54}$$

Substituting Eq. (46) in Eq. (54) and performing the integrations, we arrive at

$$\sum_m \mathrm{Tr}\,p_{Em}(\lambda_1',\lambda_2';\lambda_1,\lambda_2)$$
$$= -\sum_{m,n_1,n_2}\frac{1}{2[\gamma\sigma^2 - k(n_1+n_2+M+1)]}, \tag{55}$$

which has poles given by Eq. (34), as was expected.

V. CONCLUSIONS

In this paper we have calculated the Green's function and the bound state spectrum for the Hartmann potential by using the BCH formulas for the exponentials of the generators of the so(2,1) Lie algebra. We have also calculated the Green's function in the $so(2,1)\otimes so(2,1)$ coherent state basis and derived equivalent results.

In connection with the method we have used, we make the following comments.

(1) The case $q = 0$ corresponds to the Coulomb potential and then $M = m =$ integer. In that case, one can use Graf's addition theorem[11] to deduce the familiar form of the Coulomb Green's function in spherical polar coordinates.[6]

(2) In spherical polar coordinates one can use separation of variables and a generalization of spherical harmonics $Y_{\lambda m}$ with λ nonintegral. Then the radial part of the Green's function can be calculated using the same technique as Mil'shtein and Strakhovenko.[6]

(3) It is known that by change of variables one can relate the Green's functions for the Coulomb, harmonic oscillator, and one-dimensional Morse potentials.[12] These coordinate transformations can be applied to the case we have considered to generate Green's functions for complicated noncentral potentials which, however, do not seem to have physical significance.

(4) Guha and Mukherjee[13] have considered a somewhat more general noncentral potential than the Hartmann one. Our technique can still be used to derive the Green's function for the case they considered.

ACKNOWLEDGMENTS

This research was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, Financiadora de Estudos e Projetos e Universidade Federal do Rio de Janeiro.

[1] H. Hartmann, Theor. Chim. Acta **24**, 201 (1972).
[2] C. C. Gerry, Phys. Lett. A **118**, 445 (1986).
[3] M. Kibler and P. Winternitz, J. Phys. A **20**, 4097 (1987).
[4] L. Chetouani, L. Guechi, and T. F. Hammann, Phys. Lett. A **125**, 277 (1987).
[5] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[6] A. I. Mil'shtein and V. M. Strakhovenko, Phys. Lett. A **90**, 447 (1982).
[7] A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[8] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).
[9] G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1952), 2nd ed.
[10] A. O. Barut and C. Fronsdal, Proc. R. Soc. London Ser. A **287**, 532 (1965).
[11] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), p. 363.
[12] N. K. Pak and I. Sökmen, Phys. Rev. A **30**, 1629 (1984).
[13] A. Guha and S. Mukherjee, J. Math. Phys. **28**, 840 (1987).

The generalized continued fractions and potentials of the Lennard-Jones type

Miloslav Znojil
Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež, Czechoslovakia

(Received 3 August 1989; accepted for publication 14 March 1990)

For a broad class of the strongly singular potentials $V(r)$, which are defined as superpositions of separate power-law components, the general solution of the corresponding Schrödinger differential equation is constructed as an analog of Mathieu functions. The analogy is supported by the use of the (generalized) continued fractions. The questions of convergence are analyzed in detail.

I. INTRODUCTION

In accord with Newton,[1] singular potentials have received a considerable amount of attention as a methodical laboratory in high-energy physics. At the same time, the strongly singular repulsive core and Schrödinger equations

$$-\frac{d^2}{dr^2}\psi(r)+(g_{2q}r^{2q}+g_{2q-1}r^{2q-1}+\cdots+g_{-2p}r^{-2p})\times\psi(r)=0,\quad g_{-2p}>0,p>1, \tag{1.1}$$

also find applications in the low-energy phenomenology. In particular, the Lennard-Jones forces [say, with $q=0$ and $p=6$ in (1.1)] play an important role in the scattering of molecules.[2]

In its simplest nontrivial special case, our Eq. (1.1) (with $q=0, p=2$, and with the zero odd couplings g_{-1} and g_1) degenerates to the exactly solvable Mathieu differential equation.[3] In our preceding papers,[4,5] slight generalizations of the Mathieu special functions have been introduced and shown to lead to the similar "exact" solvability of (1.1) even for $q=1$ and $p=2$ and 3 or, alternatively, in the $q=0$ and $p=2$ case with the arbitrary nonzero g_{-1} and g_1. In the present paper, we intend to extend these constructions and results to all the integers p and q and arbitrary couplings. In essence, we shall introduce a further generalization of Mathieu functions and describe some of its properties, especially in terms of the (generalized) convergent continued fractions.

In Sec. II, we shall start from an appropriate Laurent-series ansatz, the form of which is inspired by its above-mentioned special cases of Refs. 4 and 5. The basic problem of convergence of the Laurent series will be clarified in Sec. III. The underlying recurrences will then be studied in more detail. First of all, their maximal simplification and approximative finite-matrix reinterpretation will be described in Sec. IV. From an alternative point of view, Sec. V refines these results and converts them into our final, generalized-continued-fraction quasinonnumerical formulas. Their further rearrangement (leading to the significant acceleration of their convergence) is finally added in the Appendix, and Sec. VI gives a short summary.

II. THE GENERALIZED MATHIEU ANSATZ $\psi(r,\gamma)$

In the manner already used in Refs. 4 and 5, we may try to define all the possible solutions of our differential Schrödinger equation (1.1) by the Laurent-series formula or ansatz:

$$\psi(r)=\psi(r,\gamma)=\chi(r)\sum_{n=-\infty}^{\infty}h_n r^{n+\gamma}, \tag{2.1}$$

with a suitable Floquet[6] parameter γ and with some auxiliary function

$$\chi(r)=\exp\left(-\sum_{j=1}^{p-1}\gamma_{j+1}r^{-j}/j-\sum_{i=1}^{q+1}\beta_i r^i/i\right). \tag{2.2}$$

The latter function depends on some $p+q$ independent parameters γ_{j+1} and β_i in general.

An insertion of (2.2) and (2.1) in (1.1) converts our differential Schrödinger equation into an infinite set of the linear necessary conditions:

$$\sum_{m=-2q}^{2p}H_{nm}h_{n+m}=0,\quad n=\ldots,-1,0,1,\ldots. \tag{2.3}$$

Here, each row of the doubly infinite matrix H contains $q+p$ abbreviations or redefined couplings:

$$\begin{aligned}H_{n,-q}&=-G_q=\beta_0\beta_q+\beta_1\beta_{q-1}+\cdots+\beta_q\beta_0-g_q,\\ H_{n,-q-1}&=-G_{q+1}=\beta_1\beta_q+\beta_2\beta_{q-1}\\ &\quad+\cdots+\beta_q\beta_1-g_{q+1},\\ &\cdots\\ H_{n,-2q}&=-G_{2q}=\beta_q\beta_q-g_{2q},\end{aligned} \tag{2.4}$$

and

$$\begin{aligned}H_{n,p+2}&=-G_{-(p+2)}=\gamma_2\gamma_p+\gamma_3\gamma_{p-1}\\ &\quad+\cdots+\gamma_p\gamma_2-g_{-(p+2)},\\ &\cdots\\ H_{n,2p}&=-G_{-2p}=\gamma_p\gamma_p-g_{-2p}.\end{aligned} \tag{2.5}$$

The further matrix elements are given by the formulas

$$H_{n,m}=-G_{-m}+(\gamma_{m-1}-\beta_{1-m})(2n+2\gamma+m+1),\quad m=-q+1,-q+2,\ldots,p,p+1;\quad m\neq 2, \tag{2.6}$$

which are linear in n. We must set $\gamma_j=0$ for $j<2$ or $j>p$, and $\beta_i=0$ for $i<0$ or $i>q$. The abbreviation

$$-G_{-m}=\sum_{j=2}^{p}\gamma_j\gamma_{m-j}-2\sum_{k=0}^{q}\gamma_{m+k}\beta_k+\sum_{i=0}^{q}\beta_i\beta_{-m-i}-g_{-m} \tag{2.7}$$

 0022-2488/90/081955-07$03.00

is also introduced here. The last matrix element becomes quadratic in n,

$$H_{n2} = (n+\gamma+1)(n+\gamma+2) - g_{-2} \tag{2.8}$$

and, therefore, dominates the matrix H in the $n \gg 1$ asymptotic region.

In its fully unrestricted form, our Eq. (2.3) is a difference equation[7] of the order $2p+2q$. Of course, its solutions still depend on the Floquet parameter γ so that, in accord with the general theory of differential equations,[6] the general solutions of our radial Schrödinger equation (1.1) read

$$\psi(r) = c_a \psi(r,\gamma_a) + c_b \psi(r,\gamma_b), \tag{2.9}$$

and will still admit an imposition of the arbitrary boundary conditions.

In the above context, it is interesting to recall some studies of the $p<1$ regular potentials[8–10] where the difference (or algebraic reformulation of) Schrödinger equation of the type (2.3) may even happen to guarantee the proper bound-state asymptotics of $\psi(r,\gamma)$ [at a single particular value of γ in the present language, i.e., say, with $c_b = 0$ in (2.9)]. In contrast to this, the physical states remain the genuine two-term superpositions (2.9) for the singular forces with $p>1$ in general. Their $p+q=2$ numerical construction[4] and the $p+q>2$ examples of exceptions[5,11] may be found elsewhere.

III. THE GUARANTEE OF CONVERGENCE OF THE LAURENT SERIES

In the spirit of our preceding papers,[4,5,11] we may restrict the complete freedom in our choice of parameters in (2.2) and demand that the exponential factor $\chi(r)$ also reproduces the leading-order WKB asymptotics:

$$\beta_q = \sqrt{g_{2q}} > 0, \quad \gamma_p = \sqrt{g_{-2p}} > 0. \tag{3.1}$$

This makes the quasicouplings G_{2q} and G_{-2p} equal to zero, and leads to the new specification of the first nonzero values G_{s+p-1} and G_{-p-t-1} in the new form of our difference Schrödinger equation:

$$\sum_{m=-q-s+1}^{p+t+1} H_{nm} h_{n+m} = 0, \quad n = \ldots,-1,0,1,\ldots . \tag{3.2}$$

of the (lower) order $p+q+t+s$ with $0 \leqslant s \leqslant q$ and $0 \leqslant t \leqslant p-2$.

In the $|n| \gg 1$ asymptotic region, our understanding of Eq. (3.2) may be improved via the simple assumption

$$h_n \approx \rho^n, \quad |n| \gg 1. \tag{3.3}$$

This specifies the possible quotients ρ from the algebraic equation

$$\begin{aligned} &-G_{s+q-1} + \cdots + [-2\beta_q N + O(1)]\rho^s \\ &\quad + \cdots + [N^2 + O(N)] \\ &\quad \times \rho^{s+q+1} + \cdots + [2\gamma_p N + O(1)]\rho^{s+q+p} \\ &\quad + \cdots + (-G_{-p-t-1})\rho^{s+q+p+t} = 0, \\ &\quad |N| \gg 1 \end{aligned} \tag{3.4}$$

as its $p+q+t+s$ independent (complex) roots:

$$\begin{aligned} &\rho \approx \rho_a^{(i)} = (-G_{s+q-1}/2\beta_q N)^{1/s}, \quad i = 1,2,\ldots,s, \\ &\rho \approx \rho_b^{(j)} = (2\beta_q/N)^{1/(q+1)}, \quad j = 1,2,\ldots,q+1, \\ &\rho \approx \rho_c^{(k)} = (2\gamma_p N/G_{-p-t-1})^{1/t}, \quad k = 1,2,\ldots,t, \\ &\rho \approx \rho_d^{(l)} = (-N/2\gamma_p)^{1/(p-1)}, \quad l = 1,2,\ldots,p-1. \end{aligned} \tag{3.5}$$

The separate quotients (3.5) have to be confronted with the (pair of) standard criteria of nondivergence of the Laurent series (2.1),

$$|h_{N+1}/h_N| \leqslant 1, \quad N \gg 1, \quad |h_{-M}/h_{-M+1}| \leqslant 1, \quad M \gg 1. \tag{3.6}$$

These inequalities contradict some of formulas (3.5). Vice versa, all the "admissible" asymptotics may be given by the explicit formulas

$$h_{N+m} \approx h_{N+m}^{(+)} = \sum_{i=1}^{s} c_i^{(+)} \rho_a^{(i)m} + \sum_{j=1}^{q+1} d_j^{(+)} \rho_b^{(j)m}, \tag{3.7}$$

$$m > 0, \quad N \gg 1$$

and

$$h_{-M-m} \approx h_{-M-m}^{(-)} = \sum_{k=1}^{t} c_k^{(-)}/\rho_c^{(k)m} + \sum_{l=1}^{p-1} d_l^{(-)}/\rho_d^{(l)m}, \tag{3.8}$$

$$m > 0, \quad M \gg 1.$$

They and only they guarantee the convergence of (2.1) for all $r \in (0,\infty)$.

IV. THE MATRIX SCHRÖDINGER EQUATION

Our doubly infinite Laurent-series ansatz (2.1) makes sense if and only if we complement it by the convergence conditions (3.7) and (3.8). They must also be added to our difference Schrödinger equation as boundary conditions. In the recurrent manner, the two sets $h_N^{(+)}$, $h_{N-1}^{(+)}$,... and $h_{-M}^{(-)}$, $h_{-M+1}^{(-)}$,... of the "Jost" solutions may be then defined. The "matching" $h_n^{(+)} = h_n^{(-)}$ of these two sequences at some $s+t+p+q$ values of $n \in (-M, N)$ has to fix all the $s+t+p+q-1$ free (unnormalized) parameters $c_i^{(\pm)}$ and $d_j^{\pm}$ as well as the pair of Floquet energy-dependent parameters γ_a and γ_b in (2.9).

A. The WKB choice of $\chi(r)$

In the purely numerical context, we may mention that for $s>0$ or $t>0$, the above recurrent use of (3.2) may lead to a slight loss of precision in a way which has thoroughly been discussed in Ref. 12. This phenomenon reflects an asymptotic dominance of $c_i^{(\pm)}$'s, causing a step-by-step loss of information about the initial "subdominant" coefficients $d_j^{(\pm)}$. Hence, in what follows, we shall prefer the "extreme" WKB choice of $\chi(r)$ with $t=s=0$ in (3.2). Such a choice will simply eliminate all the $c_i^{(\pm)}$'s from our considerations.

From the constructive point of view, the requirement $t=s=0$ is equivalent to the vanishing of all the quasicouplings G in (2.4) and (2.5). This implies the trivial recurrent construction of the WKB values of the following parameters:

$$\begin{aligned} &\beta_q = \sqrt{g_{2q}}, \quad \beta_{q-1} = g_{2q-1}/2\beta_q, \\ &\cdots \\ &\beta_0 = (g_q - \beta_1\beta_{q-1} - \cdots - \beta_{q-1}\beta_1)/2\beta_q, \\ &\gamma_p = \sqrt{g_{-2p}}, \quad \gamma_{p-1} = g_{-2p+1}/2\gamma_p, \end{aligned}$$

$$\cdots$$

$$\gamma_2 = (g_{-p-2} - \gamma_3\gamma_{p-1} - \cdots - \gamma_{p-1}\gamma_3)/2\gamma_p, \quad (4.1)$$

and leads to a unique, completely WKB specification of exponentials (2.2).

The recurrent simplification and stability of the Jost solutions is an unexpected merit of the WKB choice of parameters (4.1). Another one lies in a partial preservation of analogy with the regular forces[10]—we may reduce the $n \gg 1$ and $n \ll -1$ asymptotical difference Schrödinger equation (3.4) to its respective two-term forms:

$$2\beta_q N h^{(+)}_{N+s+j} = N^2 h^{(+)}_{N+s+q+j+1} + \text{corrections},$$

$$N \gg 1, \quad j = 0,1,\ldots,q \quad (4.2a)$$

and

$$2\gamma_p M h^{(-)}_{-M+s+q+p+k} = M^2 h^{(-)}_{-M+s+q+k+1} + \text{corrections}$$

$$M \gg 1, \quad k = 0,1,\ldots,p-2. \quad (4.2b)$$

This induces a natural reparametrization of the WKB special case of (3.7),

$$h^{(+)}_{N-m} = \bar{h}_{N-m} \sum_{j=1}^{q+1} d_j^{(+)}/\rho_b^{(j)m}, \quad m = 0,1,\ldots \quad (4.3a)$$

and (3.8),

$$h^{(-)}_{-M+m} = \bar{h}_{-M+m} \sum_{i=1}^{p-1} d_i^{(-)}\rho_d^{(i)m}, \quad m = 0,1,\ldots . \quad (4.3b)$$

For $m \ll \min(M,N)$, the residual m dependence of the "barred" h's remains weak.

B. The partitioning and truncation of recurrences

In the $|n| \gg 1$ asymptotic domain of indices, the decoupled two-term asymptotic character of our recurrences (4.2) may most easily be visualized via the following partitioning of the Jost solutions:

$$Z_\beta = \begin{vmatrix} h^{(+)}_{N-q} \\ \cdots \\ h^{(+)}_{N} \end{vmatrix}, \quad Z_{\beta-1} = \begin{vmatrix} h^{(+)}_{N-2q-1} \\ \cdots \\ h^{(+)}_{N-q-1} \end{vmatrix}, \ldots \quad (4.4a)$$

and

$$z_{-\alpha} = \begin{vmatrix} h^{(-)}_{-M} \\ \cdots \\ h^{(-)}_{-M+p-2} \end{vmatrix}, \quad z_{-\alpha+1} = \begin{vmatrix} h^{(-)}_{-M+p-1} \\ \cdots \\ h^{(-)}_{-M+2p-3} \end{vmatrix}, \ldots . \quad (4.4b)$$

In such a notation, we may rewrite (4.2a) as a recurrence for vectors,

$$B_\beta Z_{\beta-1} = A_\beta Z_\beta + \text{corrections}, \quad (4.5a)$$

with the diagonally dominated $(q+1)$-dimensional matrix coefficients A and B. Similarly, an introduction of the $(p-1)$-dimensional lower-case matrices a and b enables us to rewrite (4.2b) in the analogous form,

$$b_{-\alpha} z_{-\alpha+1} = a_{-\alpha} z_{-\alpha} + \text{corrections}. \quad (4.5b)$$

In the partitioned notation, our boundary conditions (3.7) and (3.8) only reflect our freedom of choosing d's, i.e., an arbitrary initial pair of vectors $z_\beta \neq 0$ and $z_{-\alpha} \neq 0$ as a normalization. This is equivalent to the truncation of recurrences (4.2a), (4.2b)—they may now be treated as the respective matrix-looking relations:

$$\begin{vmatrix} & & \cdots & & \\ -B_{\beta-2} & A_{\beta-2} & C^{(1)}_{\beta-2} & C^{(2)}_{\beta-2} & \\ & -B_{\beta-1} & A_{\beta-1} & C^{(1)}_{\beta-1} & \\ & & -B_\beta & A_\beta & \end{vmatrix} \begin{vmatrix} \cdots \\ Z_{\beta-2} \\ Z_{\beta-1} \\ Z_\beta \end{vmatrix} = 0 \quad (4.6a)$$

and

$$\begin{vmatrix} a_{-\alpha} & -b_{-\alpha} & \\ c^{(1)}_{-\alpha+1} & a_{-\alpha+1} & -b_{-\alpha+1} \\ & \cdots & \end{vmatrix} \begin{vmatrix} z_{-\alpha} \\ z_{-\alpha+1} \\ \cdots \end{vmatrix} = 0. \quad (4.6b)$$

Moreover, an incorporation of the above-mentioned matching conditions leads us precisely to the matching of the two partitioned sets (4.6) into a single (finite-dimensional!) matrix equation. Its detailed analysis will be given in the next section.

V. THE GENERALIZED CONTINUED FRACTIONS

A. The factorization of the Laurent coefficients

For a simplification of the forthcoming discussion, let us replace the Jost solutions h_k by their smooth representants $\bar{h}_k$ as introduced in Eq. (4.3) above. With the bars omitted, our new partitioned-matrix Schrödinger equation will still preserve its old form (4.6). As an important consequence of the renormalization, the corrections C or c (which are not necessarily diagonally dominated) will remain manifestly small in comparison with the renormalized, diagonally dominated submatrices A and B or a and b:

$$\begin{aligned} &A_\xi = O(1) = B_\xi, \quad a_{-\xi} = O(1) = b_{-\xi}, \\ &|\xi| \approx M, N \gg 1, \\ &C^{(i)}_\xi = O(1/N^{i+1/(q+1)}), \\ &c^{(j)}_{-\xi} = O(1/M^{j+1/(p-1)}), \\ &i = 1,2,\ldots,S, \quad j = 1,2,\ldots,T, \\ &S = 1 + [(p-2)/(q+1)], \quad T = 1 + [q/(p-2)]. \end{aligned} \quad (5.1)$$

Obviously, we may only get $S > 1$ (and $T = 1$) for $p - 1 > q + 1$ or, alternatively, $T > 1$ (and $S = 1$) for $q + 1 > p - 1$.

As a consequence of the renormalization and (5.1), the leading-order coefficients in the two-term recurrences (4.5) remain diagonal. Hence, we may postulate

$$Z_\xi = Y_\xi Y_{\xi-1} \cdots Y_{\eta+1} Z_\eta, \quad \xi > \eta \quad (5.2a)$$

and

$$z_{-\xi} = y_{-\xi} y_{-\xi+1} \cdots y_{-\eta-1} z_{-\eta}, \quad -\xi < -\eta, \quad (5.2b)$$

and conclude that the factors Y and y are also asymptotically diagonal. Thus the simplicity of our scalar results[4,5] survives transition to the present matrix case.

Without any loss of clarity, we may recall now the $n \to -n$ symmetry of our formulas (cf. the Appendix for details) and restrict our attention to Eq. (5.2a), i.e., to the capital-letter (lower) corner of the banded matrix H in

(4.6a). For the sake of generality, we shall assume that $S>1$ and notice that relation (4.6a) may in fact be read as a definition of the auxiliary quantities Y_ξ. In this sense, it is equivalent to the nonlinear relations

$$Y_\xi = \left[(A_\xi + \sum_{i=1}^{S} C_\xi^{(i)} Y_{\xi+i} Y_{\xi+i-1} \cdots Y_{\xi+1})^{-1} \right] B_\xi, \tag{5.3}$$

$$\xi = \beta, \beta-1, \beta-2, \ldots .$$

Of course, their truncation

$$Y_{\beta+1} = Y_{\beta+2} = \cdots = 0 \tag{5.4}$$

is equivalent to the truncation of H. In the limit $N \to \infty$, the combined prescription (5.3) + (5.4) may now be interpreted as a new generalization of the standard[13] analytic continued fractions.

Due to the validity of the asymptotic estimates (5.1), the generalized continued fractions (GCF) Y_ξ may be replaced by their very good zero-order approximants

$$Y_\xi \approx A_\xi^{-1} B_\xi \tag{5.5}$$

in the $N \gg 1$ asymptotic region. An insertion of these almost diagonal matrices in the explicit product of matrices (5.2a) degenerates to the closed (scalar) gamma-function approximation formulas of Ref. 4 and 5 for all the simplest or "next-to-solvable" forces in (1.1). If needed, the higher-order corrections to (5.5) may be generated in the manner described in more detail in the Appendix below.

Our above derivation of the possible finite-dimensional truncation of the matrix H in (4.6) implies that the values of the Floquet parameters γ are not arbitrary. At each energy E, we may search for them via the standard determinantal condition of solvability:

$$\det \begin{vmatrix} a_{-\alpha} & -b_{-\alpha} & & & \\ c^{(1)}_{-\alpha+1} & a_{-\alpha+1} & -b_{-\alpha+1} & & \\ & \cdots & & & \\ & -B_{\beta-1} & A_{\beta-1} & C^{(1)}_{\beta-1} & \\ & & -B_\beta & A_\beta & \end{vmatrix} = 0. \tag{5.6}$$

In the extensive $p=2$ and $p=3$ tests as presented in our preceding papers[4,5] for the $q=1$ bound states, Eq. (5.6) has been used to fix the Floquet exponents $\gamma_{a,b}(E)$ numerically. Of course, the related determination of the physical value of the energy E was performed in the standard manner, via the requirement of vanishing of $\psi(r)$ at the boundaries $r \to 0$ and $r \to \infty$. The precision was verified by a comparison with the Runge–Kutta method.[14]

The rate of convergence of the above-mentioned numerical results proved satisfactory. The main reason may be seen in the continued fractional character of the algorithms. In the present generalizations, the numerical tests become redundant—we must only succeed in extending the recurrent, GCF approach to the secular or Hill[15] determinants of the type (5.6).

B. The GCF factorization of the Hill determinants

Without any significant loss of generality, let us pick up now $T=2$ and $S=1$ in our determinant $\det H$ in (5.6). With any such choice we have only to match the respective $(p-1)$-dimensional diagonal submatrices $a_{-\alpha}, \ldots, a_{-1}$ and their $(q+1)$-dimensional counterparts $A_1, A_2, \ldots, A_\beta$ at the zeroth, $(L \times L)$-dimensional block $\hat{a}_0 = a_0 = A_0 = A_0$. Thus, provided that we choose the zeroth dimension L sufficiently large, $L \geqslant \max(p-1, q+1)$, we may denote

$$H = \begin{vmatrix} a_{-\alpha} & -b_{-\alpha} & & & & & & & \\ c^{(1)}_{-\alpha+1} & a_{-\alpha+1} & -b_{-\alpha+1} & & & & & & \\ & \cdots & & & & & & & \\ & c^{(2)}_{-1} & c^{(1)}_{-1} & a_{-1} & -\hat{b}_{-1} & & & & \\ & & \hat{c}^{(2)}_0 & \hat{c}^{(1)}_0 & \hat{a}_0 & \hat{C}_0 & & & \\ & & & & -\hat{B}_1 & A_1 & C_1 & & \\ & & & & & -B_2 & A_2 & C_2 & \\ & & & & & & \cdots & & \\ & & & & & & -B_{\beta-1} & A_{\beta-1} & C_{\beta-1} \\ & & & & & & & -B_\beta & A_\beta \end{vmatrix} . \tag{5.7}$$

The shapes and dimensions of the new five ($= T+S+2$) "capped" submatrices follow already from our specification of the diagonal blocks.

In order to preserve an analogy with the continued-fractional treatment of the Mathieu functions[15] and their recent generalizations,[4,5] let us introduce the pair of auxiliary matrices:

$$H^{(1)} = \begin{vmatrix} 1/f_{-\alpha} & & & & & & & & \\ g_{-\alpha+1} & 1/f_{-\alpha+1} & & & & & & & \\ & \cdots & & & & & & & \\ & c^{(2)}_{-2} & g_{-2} & 1/f_{-2} & & & & & \\ & & c^{(2)}_{-1} & g_{-1} & 1/f_{-1} & & & & \\ & & & \hat{c}^{(2)}_0 & \hat{g}_0 & 1/\hat{f}_0 & \hat{C}_0 & & \\ & & & & & 1/F_1 & C_1 & & \\ & & & & & & \cdots & & \\ & & & & & & & 1/F_{\beta-1} & C_{\beta-1} \\ & & & & & & & & 1/F_\beta \end{vmatrix} , \tag{5.8}$$

$$H^{(2)} = \begin{vmatrix} 1 & -f_{-\alpha}b_{-\alpha} & & & & & \\ & \cdots & & & & & \\ & 1 & -f_{-1}\hat{b}_{-1} & & & & \\ & & 1 & & & & \\ & & -F_1\hat{B}_1 & 1 & & & \\ & & & & \cdots & & \\ & & & & -F_{\beta-1}B_{\beta-1} & 1 & \\ & & & & & -F_\beta B_\beta & 1 \end{vmatrix} . \tag{5.9}$$

Their partitioned structure copies that of H. We shall postulate that our matrix H equals the product

$$H = H^{(1)}H^{(2)}. \tag{5.10}$$

This relation has to become an algebraic identity.

In the first step in the indicated direction, we define $F_{\beta+1} = 0$ and arrive merely at the definition of the matrix continued fractions,

$$F_k = 1/(A_k + C_kF_{k+1}B_{k+1}), \quad k = \beta,\beta-1,...,1. \tag{5.11}$$

This concept appears often in the literature.[16]

In the next step, we complement (5.11) by its $T=2$ counterpart:

$$f_{-m} = 1/(a_{-m} + g_{-m}f_{-m-1}b_{-m-1}),$$
$$m = \alpha,\alpha-1,...,1, \tag{5.12}$$

with the initial choice of matrices $f_{-(\alpha+1)} = f_{-(\alpha+2)} = 0$ and with the abbreviation

$$g_{-n} = c^{(1)}_{-n} + c^{(2)}_{-n}f_{-n-2}b_{-n-2},$$
$$n = \alpha-1,\alpha-2,...,1. \tag{5.13}$$

We may immediately notice a close similarity of the new formulas to our preceding GCF construction—in general, the $T\geqslant 1$ quantities f and y or the $S\geqslant 1$ quantities F and Y may be shown to differ just by an inessential matrix multiplication factor.

In the final step, our recurrences have to be complemented by the missing capped definitions

$$\hat{g}_0 = \hat{c}^{(1)}_0 + \hat{c}^{(2)}_0 f_{-2}b_{-2},$$
$$1/\hat{f}_0 = \hat{a}_0 + \hat{g}_0 f_{-1}\hat{b}_{-1} + \hat{C}_0F_1\hat{B}_1. \tag{5.14}$$

Our construction becomes complete. In a full analogy with the analytic treatment of the Mathieu and Hill equations,[6,15] it is based on an assumption of existence of the underlying continued fractions. In particular, this means a feasibility of all the inversions in Eqs. (5.11) and (5.12): In the present notation, all the subdeterminants $\det 1/f_k$ and $\det 1/F_k$ of $H^{(1)}$ with nonzero subscripts k must differ from zero. Vice versa, all the zeros of our secular determinant $\det H = \det H^{(1)}\cdot\det H^{(2)} = \det H^{(1)}$ will then coincide with zeros of the transcendental GCF equation $\det 1/\hat{f}_0 = 0$, i.e., our condition (5.6) may be given the explicit GCF form:

$$\det(\hat{a}_0 + \hat{c}^{(1)}_0 f_{-1}\hat{b}_{-1} + \hat{c}^{(2)}_0 f_{-2}b_{-2}f_{-1}\hat{b}_{-1} + \hat{C}_0F_1\hat{B}_1) = 0. \tag{5.15}$$

This is our final formula—once again, we would like to emphasize its numerical efficiency as verified, for its first two nontrivial special cases, in Refs. 4 and 5.

VI. CONCLUDING REMARKS

(1) A number of interactions between composite systems (ranging from nucleons to complex molecules) may be described by a phenomenological or effective central potential $V(r)$ with a repulsive core:

$$V(r) = \sum_{m=1}^{D} a_m r^{-mxZ}. \tag{6.1}$$

Via the Fivel transformation[17] $r \to r^{\text{const}}$, the corresponding differential Schrödinger equation may often be given our "canonical" form (1.1). This extends the applicability of our present results—in all the modified (6.1) (and some other) cases, one of the parameters g remains proportional to the energy ($E \approx -g_a, a > -2$), the value of g_{-2} may incorporate the angular momentum $l = 0,1,...$ [i.e., $g_{-2} \approx g'_{-2} + l(l+1)$], etc. The values of integers p and q depend[11] on the detailed form of $V(r)$: In the most frequently used Lennard–Jones model with $Z=2$ and $D=6$, we have $q=0$ and $p=6$.

(2) The Lennard–Jones force exemplifies the multiplicities $T=1$ and $S=5$ of corrections in our matrix H. In the upper corner of $H^{(1)}$, the auxiliary GCF quantities f degenerate again to the above-mentioned matrix fractions. At the same time, their "extended" $S>1$ counterparts F remain scalars: Such a generalization has already been studied, e.g., in Ref. 8. Thus the genuine "matrix" and "extended" new type of GCFs will only be met for the values p and q that satisfy the inequality $q > p-2 > 0$ or, alternatively, $p-2 > q > 0$. After the Fivel transformation, the corresponding simplest "generalized Lennard-Jones" examples (6.1) of this type will be characterized by the parameters $D=10$ and $Z=1/2$ or $Z=1/3$.

(3) We may notice that a direct start of our recurrences (3.2) from the non-GCF and "more natural" trivial initial values

$$h_n = 0, h_m = 0, \quad n > N_0 \gg 1, \quad m < -M_0 \ll -1, \tag{6.2}$$

would lead immediately to severe oscillations caused by the "one-term" character of the first few iterations of (4.2). Similar observations have already been made and tested in the various slightly different contexts.[18] Always, a smooth truncation was able to weaken or even remove the oscillations. In this spirit, the present GCF quantities offer one of

the new suitable computational means: In fact, they just combine the "matrix" and "extended" idea of generalization of the continued fractions. The numerical consequences of such an idea were already analyzed in detail. In particular, an efficiency of their asymptotic acceleration of convergence [which is rather similar to our present technique and/or formula (5.5)] has been numerically illustrated in Ref. 19 and in the first item of Ref. 8, respectively.

(4) The smooth shape of our forces implies an easy numerical "brute-force" tractability of the related bound-state or scattering boundary conditions. Indeed, the threshold and asymptotic physical (WKB) logarithmic derivatives

$$\frac{\psi'(r)}{\psi(r)} \approx \sum_{j=1}^{p} \gamma_j r^{-j} - \sum_{i=0}^{q} \beta_i r^i, \quad \gamma_p > 0, \beta_q > 0 \tag{6.3}$$

may be matched to our generalized Mathieu functions (2.6) at any suitable pair of coordinates $r = r_0 \ll 1$ and $r = r_\infty \gg 1$ near the boundaries (end points). In a non-numerical setting, we may see our factor $\chi(r)$, (2.2), as obtained by an integration of this "simplest possible" formula (6.3). The Floquet factor γ appears then naturally as an integration constant.

(5) An interesting formal feature of the strongly singular interactions emerges also in connection with their so-called quasiexact solvability (i.e., with an existence of the particular elementary solutions at certain particular values of the couplings) as discovered recently in some regular (or, more precisely, weakly singular, $p \leqslant 1$) systems[20] and existing also in some[11] strongly singular equations (1.1). In the latter $p \geqslant 2$ cases, however, a limiting transition to the infinite series fails to provide exact solutions in general. In the framework of the so-called Hill-determinant method (see, e.g., Ref. 21 or the papers listed in Ref. 9), such an asymmetry between the regular and singular interactions is a little bit puzzling.

APPENDIX: AN ACCELERATION OF CONVERGENCE

For the sake of simplicity, let us contemplate the change of variables:[17]

$$r \to r' = 1/r, \quad \psi(r) \to \psi'(r') = \psi(r)/r, \tag{A1}$$

and notice that it converts (1.1) into the "primed" form of the same equation with the modified parameters:

$$p' - 1 = q + 1, \quad q' + 1 = p - 1,$$
$$g'_{-2p'} = g_{2q}, \quad g'_{-2p'+1} = g_{2q-1}, \ldots, \quad g'_{2q'} = g_{-2p}. \tag{A2}$$

Obviously, such a formal $n \to -n$ symmetry enables us to restrict our attention to, say, capital-letter quantities. Without any loss of clarity, we may also consider our recurrences [say, Eq. (4.5a)] in their first nontrivial $S = 2$ form:

$$B_\xi Z_{\xi-1} = A_\xi Z_\xi + C_\xi^{(1)} Z_{\xi+1} + C_\xi^{(2)} Z_{\xi+2}. \tag{A3}$$

A priori, the above-made asymptotic estimates (4.3) indicate a slow convergence of GCFs or recurrences (A3)—a shortcoming that has already been noticed in the older $p < 2$ studies.[8] The reason is obvious; the relevant measure of smallness (in fact, the parameter $|\rho| = N^{-1/(q+1)}$) only becomes noticeably small for the extremely large dimensions N. Here, we intend to show that the related scalar techniques of the suitable acceleration of convergence (cf. Ref. 5) may easily be extended to the present GCF case.

In general, our idea lies in a systematic and exact elimination of the dominant corrections from the two-term relations (4.5). Thus, in particular, we must eliminate the components $Z_{\xi+1}$ from (A3). After a shift of subscripts,

$$B_{\xi+1} Z_\xi = A_{\xi+1} Z_{\xi+1} + C_{\xi+1}^{(1)} Z_{\xi+2} + C_{\xi+1}^{(2)} Z_{\xi+3}, \tag{A4}$$

in the latter formula (this is quite easy) we get

$$B_\xi^{[2]} Z_{\xi-1} = A_\xi^{[2]} Z_\xi + C_\xi^{[2](2)} Z_{\xi+2} + C_\xi^{[2](3)} Z_{\xi+3}, \tag{A5}$$

with the superscripted "second-order" matrix coefficients:

$$B_\xi^{[2]} = B_\xi, \quad A_\xi^{[2]} = A_\xi + C_\xi^{(2)} A_{\xi+1}^{-1} B_{\xi+1}, \quad C_\xi^{[2](1)} = 0,$$
$$C_\xi^{[2](2)} = C_\xi^{(2)} - C_\xi^{(1)} A_{\xi+1}^{-1} C_{\xi+1}^{(1)},$$
$$C_\xi^{[2](3)} = -C_\xi^{(1)} A_{\xi+1}^{-1} C_{\xi+1}^{(2)}. \tag{A6}$$

In principle, the elimination of the dominant corrections (or of $C_\xi^{[k](k-1)}$) may be iterated: For $k = 2$, the elimination of $Z_{\xi+2}$ must be made in two steps. First, we reintroduce temporarily $Z_{\xi+1}$ by the shift of ξ in (A4),

$$B_{\xi+2} Z_{\xi+1} = A_{\xi+2} Z_{\xi+2} + C_{\xi+2}^{(1)} Z_{\xi+3} + C_{\xi+2}^{(2)} Z_{\xi+4} \tag{A7}$$

and, second, we eliminate it again via the shifted formula (A5),

$$B_{\xi+1}^{[2]} Z_\xi = A_{\xi+1}^{[2]} Z_{\xi+1} + C_{\xi+1}^{[2](2)} Z_{\xi+3} + C_{\xi+1}^{[2](3)} Z_{\xi+4}. \tag{A8}$$

In the resulting prescription

$$B_\xi^{[3]} Z_{\xi-1} = A_\xi^{[3]} Z_\xi + C_\xi^{[3](3)} Z_{\xi+3} + C_\xi^{[3](4)} Z_{\xi+4}, \tag{A9}$$

as well as in all its further (k th) generalizations, the matrices A and C may again be given an explicit form similar to (A6).

In the practical computations, linear formulas of the type (A3) or (A9) may easily be rearranged as the GCF definitions [cf. (5.3)]. In general, the k th GCF quantity $Y_\xi^{[k]}$ with S replaced by $S^{[k]} = S + k$) converges as $1/N^k$ per iteration. Thus, in particular, the formula,

$$Y_\xi = (A_\xi^{[k]})^{-1} B_\xi^{[k]} + o(N^{-k}), \tag{A10}$$

may be employed as an improvement of accuracy of our $k = 1$ approximant (5.5).

[1] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982), p. 398.
[2] S. Fluegge, *Practical Quantum Mechanics I* (Springer-Verlag, Berlin, 1971), p. 274.
[3] E. Vogt and G. H. Wanier, Phys. Rev. **95**, 1190 (1954), and further papers quoted in Ref. 1.
[4] M. Znojil, J. Math. Phys. **30**, 23 (1989).
[5] M. Znojil, J. Math. Phys. **31**, 108 (1990).
[6] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[7] N. E. Noerlund, *Vorlesungen ueber Differenzenrechnung* (Springer, Kopenhagen, 1923).
[8] M. Znojil, Lett. Math. Phys. **5**, 405 (1981); J. Math. Phys. **24**, 1136 (1983).
[9] A. Hautot, Phys. Rev. D **33**, 437 (1986).
[10] M. Znojil, Phys. Rev. D **34**, 1224 (1986); J. Math. Phys. **29**, 1433 (1988).
[11] M. Znojil, J. Phys. A **15**, 2111 (1982).
[12] M. Znojil, K. Sandler, and M. Tater, J. Phys. A **18**, 2541 (1985).
[13] O. Perron, *Die Lehre von der Kettenbruechen* (Teubner, Stuttgart, 1957);

W. B. Jones and W. J. Thron, *Continued Fractions: Analytic Theory and Applications*, Vol. 11 of Enc. of Math. and its Appl., edited by G.-C. Rota (Addison-Wesley, Reading, MA, 1980).
[14] F. S. Acton, *Numerical Methods That Work* (Harper & Row, New York, 1970).
[15] H. Bateman and A. Erdelyi, *Higher Transcendental Functions III* (McGraw-Hill, New York, 1955).
[16] S. Graffi and V. Grecchi, Lett. Nuovo Cimento **12**, 425 (1975).
[17] D. C. Fivel, Phys. Rev. **142**, 1219 (1966).
[18] J. L. Richardson and R. Blankenbecler, Phys. Rev. D **19**, 496 (1979); M. Tater, J. Phys. A: Math. Gen. **20**, 2483 (1987).
[19] M. Znojil, Phys. Rev. D **24**, 903 (1981).
[20] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978); E. Magyari, Phys. Lett. A **81**, 116 (1981); G. P. Flessas, *ibid.* **83**, 121 (1981); A. V. Turbiner and A. G. Ushveridze, *ibid.* **126**, 181 (1987); M. A. Shifman, preprint CERN-TH.5265/88 (in press).
[21] S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, Phys. Rev. D **4**, 3617 (1971).

On redshift and parallaxes in general relativistic kinematical world models

V. Perlick
Institut für Theoretische Physik der TU Berlin, Sekr. PN 7-1, Hardenbergstrasse 36, 1000 Berlin 12, Federal Republic of Germany

(Received 11 October 1989; accepted for publication 14 March 1990)

A Lorentzian space-time (M,g), along with an observer field (= timelike vector field) V, are considered. Several theorems are presented that are concerned with the redshift and parallaxes in such a model. Here the term "parallaxes" stands for the relative motion of two observers at the celestial sphere of a third observer. As an important mathematical tool, the index form along a lightlike geodesic is used.

I. INTRODUCTION

In an earlier paper[1] we presented some criteria for a space-time (M,g), with observer field V to be parallax-free in the following sense: If any two observers are seen from any third observer in the same direction at the celestial sphere (i.e., one behind the other) at one instant of time, then this will be the case for all time. In the present paper, some new theorems are presented that are concerned with this notion of parallaxes and with its relationship to redshift.

The paper is organized as follows. After introducing our general assumptions and notations in Sec. II, Sec. III reminds the reader of the index form along a lightlike geodesic and of its main properties. The index form will serve as a valuable tool to prove some of the theorems following later in the article. Section IV deals with various versions of a redshift formula, with special regard to the problems due to focusing effects. Section V is devoted to parallaxes. In order to obtain more detailed insight into this concept, a slight change of terminology in comparison to Ref. 1 is necessary: Whereas in the earlier article the property of being parallax-free was assigned to the model (M,g,V) as a whole, in the present paper it is assigned to an individual light signal. Finally, Secs. VI and VII present some theorems on redshift and parallaxes in models with special symmetry properties of the redshift.

II. GENERAL ASSUMPTIONS AND NOTATIONS

In all that follows, let (M,g,V) be a space-time with an observer field, i.e.. (i) M denotes a four-dimensional C^∞ manifold, the topology of which is Hausdorff and second countable; (ii) g denotes a C^∞ Lorentzian metric on M, where the signature is chosen to be $(+\ +\ +\ -)$; (iii) V denotes a timelike C^∞ vector field on M satisfying $g(V,V) = -1$. By assuming the global existence of V, we require (M,g) to be time orientable. The integral curves of V are interpreted as observers. The normalization condition $g(V,V) = -1$ ensures that these observers use standard clocks (but large parts of the article will be independent of this normalization condition). The integral curves of V (= observers) will usually be denoted by γ, $\tilde{\gamma}$, etc., whereas γ', $\tilde{\gamma}'$, etc. denote the corresponding tangent fields.

Let us agree to parametrize light signals (= lightlike geodesics with initial point and endpoint) always affinely, future pointing with respect to V, and such that the parameter runs through the interval $[0,1]$. Every light signal admits exactly one such parametrization. We denote the set of all these light signals by $\mathscr{L}$:

$$\mathscr{L} := \{\lambda:[0,1]\to M \mid \lambda \text{ is a } C^\infty \text{ map satisfying } \nabla_{\lambda'}\lambda' = 0, g(\lambda',\lambda') = 0, g(\lambda',V) < 0\}. \quad (2.1)$$

The symbol ∇ here and in the following always denotes the Levi-Cività connection of g. The vector fields along any $\lambda \in \mathscr{L}$ will be noted by $\mathscr{R}_\lambda$:

$$\mathscr{R}_\lambda := \{A:[0,1]\to TM \mid A \text{ is a } C^\infty \text{ map with } \pi\circ A = \lambda\}, \quad (2.2)$$

where $\pi:TM\to M$ denotes the tangent bundle over M. Furthermore, we write

$$\mathscr{R}_\lambda^\perp := \{A \in \mathscr{R}_\lambda \mid g(\lambda',A) = 0\}. \quad (2.3)$$

Note that the tangent field λ' is itself contained in $\mathscr{R}_\lambda^\perp$. For any $A \in \mathscr{R}_\lambda$, the covariant derivative with respect to the tangent field will be denoted by a prime:

$$A'(s) = \nabla_{\lambda'(s)}A \text{ for all } s \in [0,1]. \quad (2.4)$$

III. THE INDEX FORM ALONG LIGHTLIKE GEODESICS

In this section, we are going to remind the reader of the index form along lightlike geodesics and of some of its properties. For details, the reader is referred to Beem and Ehrlich.[2]

Definition 1: For $\lambda \in \mathscr{L}$ the index form

$$\mathscr{I}_\lambda:\mathscr{R}_\lambda^\perp \times \mathscr{R}_\lambda^\perp \to \mathbf{R}, \quad (A,B)\to \mathscr{I}_\lambda(A,B)$$

is defined by

$$\mathscr{I}_\lambda(A,B) := \int_0^1 (g(A',B') + g(A,R(\lambda',B,\lambda')))|_s\, ds,$$

with R denoting the curvature tensor of g.

More generally, the index form is often defined on piecewise (!) smooth vector fields. However, it will be sufficient for our purposes to have $\mathscr{I}_\lambda$ on the smooth vector fields $\mathscr{R}_\lambda^\perp$. Integration by parts results in the following representation of the index form:

$$\mathscr{I}_\lambda(A,B) = g(A(1),B'(1)) - g(A(0),B'(0)) + \int_0^1 g(A,R(\lambda',B,\lambda') - B'')|_s\, ds. \quad (3.1)$$

 0022-2488/90/081962-10$03.00

The integral in (3.1) vanishes if B happens to be a Jacobi field (i.e., a vector field satisfying the Jacobi equation $R(\lambda',B,\lambda') = B''$). Here $\mathcal{J}_\lambda$ is a symmetric bilinear form, but contrary to the nonlightlike case it has a nontrivial kernel since obviously, $\mathcal{J}_\lambda(\lambda',.) = 0$. The most important property of $\mathcal{J}_\lambda$, which will be used repeatedly below, is given by the following fundamental lemma.

Lemma 1: Assume $\lambda \in \mathcal{L}$ to be free of conjugate points. Then any $A \in \mathcal{X}_\lambda^\perp$ with $A(0) = A(1) = 0$ satisfies $\mathcal{J}_\lambda(A,A) = 0$ if and only if A is a multiple of λ'.

Whereas the "if" part of this lemma is obvious, the proof of the highly nontrivial "only if" part can be found in Beem and Ehrlich.[2]

Since $\mathcal{J}_\lambda$ degenerates on the multiples of λ', it is often recommended to factor out this subspace from $\mathcal{X}_\lambda$. The following terminology will be used.

Definition 2: Define an equivalence relation on $\mathcal{X}_\lambda$ by $A \sim B :\Leftrightarrow A - B$ is a multiple of λ'. An equivalence class $[A] \in \mathcal{X}_\lambda/_\sim$ is called a "Jacobi class" if it contains a Jacobi field.

We conclude this preparatory paragraph by proving a simple lemma, which will often be used below.

Lemma 2: Let J be a Jacobi field along any $\lambda \in \mathcal{L}$. Define a function $k:[0,1] \to \mathbf{R}$ by $k(s) := g(\lambda'(s), J(s))$; then its derivative $k' = g(\lambda',J')$ is a constant function.

Proof: Here $k'' = g(\lambda',J'') = g(\lambda',R(\lambda',J,\lambda')) = 0$ as a result of the curvature identity $g(X,R(U,V,Y)) + g(Y,R(U,V,X)) = 0$. ■

IV. REDSHIFT

This section will be concerned with the redshift, under which any observer in V is seen by any other observer. We begin with a rigorous definition of the redshift function which is developed from a purely kinematical consideration. The main idea of this consideration dates back to Weyl.[3] Let us fix two observers γ and $\tilde{\gamma}$ belonging to our observer field V and let us assume that $\tilde{\gamma}$ receives light signals from γ over a time. This situation is formalized by the following definition.

Definition 3: Let $\gamma:I \to M$ and $\tilde{\gamma}:\tilde{I} \to M$ be two integral curves of V. (Here I and $\tilde{I}$ denote some real intervals.) A C^∞ map

$$\beta:[0,1] \times I \to M, \quad (s,t) \mapsto \beta(s,t)$$

is called a "message" from γ to $\tilde{\gamma}$ iff

(i) $\beta(.,t) \in \mathcal{L}$ for all $t \in I$,

(ii) $\beta(0,.) = \gamma$,

(iii) there is some diffeomorphism $h:I \to \tilde{I}$ such that $\beta(1,.) = \tilde{\gamma} \circ h$: Here h is called the "message function" of β.

Hence, a message is a lightlike geodesic variation, the endpoints of which vary along integral curves of V; see Fig. 1. For an arbitrary pair $(\gamma,\tilde{\gamma})$ of integral curves of V, neither the existence nor the uniqueness of a message from γ to $\tilde{\gamma}$ is guaranteed. We shall return to this problem as far as the well-definedness of the redshift function is concerned. The message function h appearing in (iii) of Definition 3 is necessarily a monotonously increasing (!) diffeomorphism. (To give a rigorous proof of this fact, apply Lemma 2 to the generator J of the variation β.) The message function h compares the reading of $\tilde{\gamma}$'s clock to that of γ's, as seen from $\tilde{\gamma}$. A process near the world line of γ, which starts at t and ends at $t + \Delta t$, seems to start at $h(t)$ and end at $h(t + \Delta t)$ if it is observed from $\tilde{\gamma}$. Therefore, the redshift under which γ is seen by $\tilde{\gamma}$ is given by

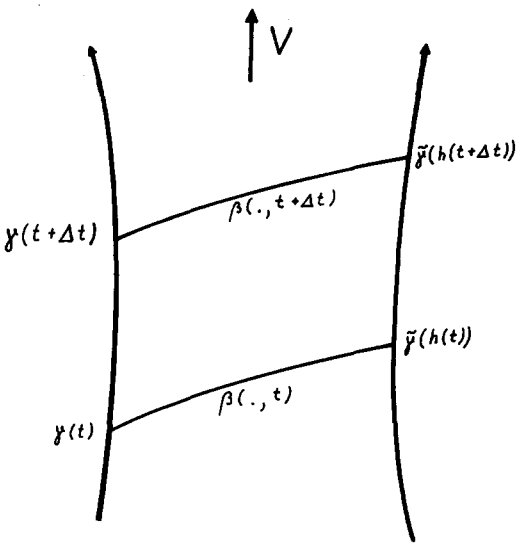


FIG. 1. A message β from γ to $\tilde{\gamma}$.

$$h'(t) = \lim_{\Delta t \to 0} \frac{h(t+\Delta t) - h(t)}{\Delta t} = \frac{\text{emitted frequency}}{\text{received frequency}}. \tag{4.1}$$

The transition $\Delta t \to 0$ to the linear approximation is necessary since a periodic process near γ will be judged as periodic from $\tilde{\gamma}$, too, if and only if h is linear. Stated in different words: Our notion of redshift makes sense only for such large emitted frequencies that h may be replaced by its linear approximation over some number of periods.

Since the limit $\Delta t \to 0$ has to be taken in (4.1), it is convenient to turn from the variation β to the corresponding variational vector field (= Jacobi field). In analogy with Definition 3, we will call this object an "infinitesimal message," where the attribute "infinitesimal" refers to the temporal duration (and not to the spatial distance).

Definition 4: Let $\lambda \in \mathcal{L}$ be any light signal. A vector field $J \in \mathcal{X}_\lambda$ is called an "infinitesimal message" along λ iff

(i) $R(\lambda',J,\lambda') = J''$,

(ii) $g(\lambda',J') = 0$,

(iii) $J(0) = V_{\lambda(0)}$,

(iv) There is some real number u such that $J(1) = e^u V_{\lambda(1)}$: Here u is called the "redshift" of the infinitesimal message J.

In Definition 4 (i) expresses the fact that J describes a second geodesic which is infinitesimally close to λ; (ii) ensures that this geodesic to be lightlike, too, whereas (iii) and (iv) express the fact that it interconnects the same two observers as λ does. [Note that the redshift u is defined by Definition 4 (iv) in such a way that the frequency ratio (4.1) corresponds to e^u.]

The redshift of an infinitesimal message can be expressed by geometric quantities. Although the resulting redshift formula is well known,[4–7] a proof will be presented since it is extremely simple and instructive: It is essentially the Brill[6] proof translated into our notation.

Theorem 1: Let J be an infinitesimal message along any light signal $\lambda \in \mathcal{L}$: Then its redshift u is given by

$$e^u = g(\lambda'(0), V_{\lambda(0)})/g(\lambda'(1), V_{\lambda(1)}).$$

Proof: By assumption, J satisfies $g(\lambda', J') = 0$: Integrating this equation over λ results in $g(\lambda'(0), J(0)) = g(\lambda'(1), J(1))$. This yields the desired equation if the boundary values $J(0) = V_{\lambda(0)}$ and $J(1) = e^u V_{\lambda(1)}$ are inserted. ∎

Please note that the Jacobi field J does not occur in the redshift formula given by Theorem 1. This proves the following corollary.

Corollary 1: If J_1 and J_2 are two infinitesimal messages along one and the same light signal λ, then J_1 and J_2 have the same redshifts.

As a consequence, the redshift may be viewed as a function on the set of all light signals. We use the letter r instead of u to denote this function.

Definition 5: The function $r: \mathcal{L} \to \mathbb{R}$ defined by

$$e^{r(\lambda)} := g(\lambda'(0), V_{\lambda(0)})/g(\lambda'(1), V_{\lambda(1)})$$

is called the "redshift function" of (M,g,V).

Obviously, $r(\lambda)$ is well defined for all $\lambda \in \mathcal{L}$. However, the interpretation of $r(\lambda)$ as redshift on the basis of the foregoing considerations requires the existence of at least one infinitesimal message J along λ. (It should be emphasized that the problem is in the existence and not the uniqueness of a Jacobi field J with the desired properties, contrary to a remark made by Herrman,[7] p. II-18.) In order to characterize the set of pathological light signals, for which this assumption is not justified, we introduce the following terminology.

Definition 6: A light signal $\lambda \in \mathcal{L}$ is called "regular" iff there is exactly one infinitesimal message J along λ. Here λ is called "singular of the first kind" iff there is more than one such J and it is called "singular of the second kind" iff there is none at all.

Hence, $r(\lambda)$ cannot be interpreted as redshift if λ is singular of the second kind. Fortunately, the following theorem clarifies that this is an exceptional case.

Theorem 2: For any light signal $\lambda \in \mathcal{L}$, the following properties are equivalent:

(i) λ is regular,

(ii) $\lambda(1)$ is not conjugate to $\lambda(0)$ along λ.

Proof: For "(i) $\Rightarrow$ (ii)" of Theorem 2: This implication will be shown indirectly, i.e., we prove "¬ (ii) $\Rightarrow$ ¬ (i)." Assume $\lambda(1)$ to be conjugate to $\lambda(0)$. This means by definition that there is a Jacobi field $J \neq 0$ satisfying $J(0) = J(1) = 0$. Hence, the function k defined in Lemma 2 satisfies $k(0) = k(1) = 0$ in this case. Since k' is constant, this requires k to vanish identically. The resulting equation $k' = g(\lambda', J') = 0$ shows that λ cannot be regular: If we are given any infinitesimal message J_1 along λ, then $J_2 = J_1 + J$ will be a second infinitesimal message different from J_1.

For "(ii) $\Rightarrow$ (i)" of Theorem 2: Since $\lambda(1)$ is assumed to be not conjugate to $\lambda(0)$, every choice of the boundary values $J(0)$ and $J(1)$ determines a unique solution of the Jacobi equation $R(\lambda', J, \lambda') = J''$. Let us consider the solution J with

$$J(0) = V_{\lambda(0)}, \quad J(1) = e^{r(\lambda)} V_{\lambda(1)}. \tag{4.2}$$

Let us apply Lemma 2 to this special J. From (4.2) we find $k(0) = k(1) = g(\lambda'(0), V_{\lambda(0)})$, which requires k to be constant as a consequence of Lemma 2. The resulting equation $k' = g(\lambda', J') = 0$ shows that J is indeed an infinitesimal message along λ. Since J is the only Jacobi field satisfying (4.2), there cannot be any further infinitesimal message along λ. ∎

In particular, we can read from Theorem 2 that every singular light signal can be made into a regular light signal just by cutting it a little bit shorter. Furthermore, we learn that the regularity of a light signal is a property of the spacetime (M,g) alone, i.e., independent of the observer field V. On the other hand, it does depend on V, whether a singular light signal is of the first or second kind. We illustrate this by example. Consider the Einstein cylinder universe $M = S^3 \times \mathbb{R}$, $g = \pi^* g^3 - dt^2$, where $t: M \to \mathbb{R}$ and $\pi: M \to S^3$ denote the natural projections and g^3 stands for the canonical Riemannian metric on the three-sphere S^3. First let V be the standard observer field $V = \partial/\partial t$. In this case, any light signal λ connecting the antipodal points p and q (see Fig. 2) turns out to be singular of the first kind. However, if we switch over to an observer field $\tilde{V}$ by deforming V inside some small neighborhood of q in such a way that $\tilde{V}_q$ is not a linear combination of V_q and $\lambda'(1)$, then the same light signal λ becomes singular of the second kind.

With respect to the physical interpretation of the above, it should be kept in mind that the appearance of conjugate points characterizes a focusing effect and that near focal points the ray optical approximation is not admissible. We have to switch over to a wave optical consideration and leave our concept of a pointlike observer. From this point of view, the foregoing considerations can be summed up in the following way: In every case that allows for a ray optical treatment, $r(\lambda)$ can be interpreted as redshift.

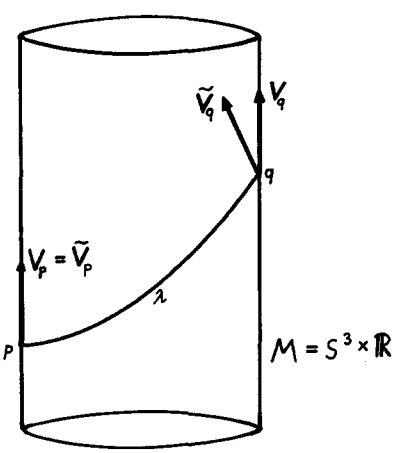


FIG. 2. A singular light ray λ in the Einstein cylinder universe.

We conclude this section by giving an alternative representation of the redshift function, which will be used repeatedly below.

Theorem 3: The redshift function r of (M,g,V) satisfies

$$r(\lambda) = -\frac{1}{2}\int_0^1 \frac{(L_V g)(\lambda'(s),\lambda'(s))}{g(\lambda'(s),V_{\lambda(s)})}\,ds, \quad \text{for all } \lambda \in \mathscr{L}.$$

Here $L_V g$ denotes the Lie derivative of the metric g.

Proof: Let X denote a vector field having λ as an integral curve. (This X is only introduced for notational convenience.) Then we find

$$\begin{aligned} r(\lambda) &= \ln g(V,X)|_{\lambda(0)} - \ln g(V,X)|_{\lambda(1)} \\ &= -\int_0^1 \frac{Xg(V,X)}{g(V,X)}\bigg|_{\lambda(s)} ds = -\int_0^1 \frac{g(\nabla_X V,X)}{g(V,X)}\bigg|_{\lambda(s)} ds \\ &= \int_0^1 \frac{g(L_V X,X)}{g(V,X)}\bigg|_{\lambda(s)} ds \\ &= -\frac{1}{2}\int_0^1 \frac{(L_V g)(X,X)}{g(V,X)}\bigg|_{\lambda(s)} ds. \end{aligned}$$ ■

This representation of $r(\lambda)$ has some advantages over the representation given in Definition 5: In particular, it preserves its form under reparametrization of the light signal. This shows immediately that the redshift has to be additive along every light signal. By the way, this is the main advantage of our function r over

$$\begin{aligned} z &= \frac{\text{emitted frequency} - \text{received frequency}}{\text{received frequency}} \\ &= e^r - 1, \end{aligned}$$

which is used by astronomers. Furthermore, Theorem 3 demonstrates the well-known fact that r vanishes identically iff V is a conformal Killing field. [The renormalization condition $g(V,V) = -1$, which has not been used up to now, then requires V even to be a Killing field.] Finally, Theorem 3 allows for relating the redshift to kinematical quantities by the well-known formula

$$\begin{aligned} \tfrac{1}{2}L_V g &= (\theta/3)g + (\theta/3)g(V,.) \vee g(V,.) \\ &\quad + \sigma - g(\nabla_V V,.) \vee g(V,.). \end{aligned} \tag{4.3}$$

Here θ and σ denote the volume expansion and shear of V, respectively, and $\vee$ stands for the symmetrized tensor product. If we introduce the one-form

$$\rho := g(\nabla_V V,.) - (\theta/3)g(V,.) \tag{4.4}$$

into (4.3) and insert this result into the representation of the redshift function given by Theorem 3, we arrive at

$$r(\lambda) = \int_\lambda \rho - \int_0^1 \frac{\sigma(\lambda'(s),\lambda'(s))}{g(\lambda'(s),V_{\lambda(s)})}\,ds. \tag{4.5}$$

For infinitesimally short λ, the redshift (4.5) formula has already been given by Ehlers.[8]

V. PARALLAXES

In order to make precise what we will mean by the term "parallaxes" here and in the following, let us recall Definition 3 of a message β from γ to $\tilde{\gamma}$. In a typical situation, the image of such a map β is a two-surface $\mathscr{F}$ bounded by γ and $\tilde{\gamma}$ and spanned by the light signals $\beta(.,t)$. In the general case, the observer field V will not be tangential to $\mathscr{F}$, with the exception of the boundary curves γ and $\tilde{\gamma}$. If we fix a light signal $\lambda = \beta(.,t_0)$ and consider all those integral curves of V that intersect λ, we obtain a two-surface $\hat{\mathscr{F}}$ which does not coincide with $\mathscr{F}$ generically. In a typical situation, $\mathscr{F}$ and $\hat{\mathscr{F}}$ form a baglike shape, as shown in Fig. 3. The aperture of this bag, i.e., the fact that $\mathscr{F}$ and $\hat{\mathscr{F}}$ do not coincide, expresses the "parallax effect": Those observers who are seen from $\tilde{\gamma}$ exactly before γ (i.e., covering γ) at one instant are not seen exactly before γ at a later instant. Therefore, we introduce the following terminology.

Definition 7: A message β is called "parallax-free" iff $\beta(s,.)'(t)$ is a linear combination of $\beta(.,t)'(s)$ and $V_\beta(s,t)$ for all $(s,t) \in [0,1]\times I$.

Turning over to an infinitesimal message, the analogous definition reads as in the following.

Definition 8: An infinitesimal message J along any light signal $\lambda \in \mathscr{L}$ is called "parallax-free" iff $J(s)$ is a linear combination of $\lambda'(s)$ and $V_{\lambda(s)}$ for all $s \in [0,1]$.

Furthermore, the property of being parallax-free can be assigned to the light signal λ itself. However, we have to be a little bit careful to include the case of a light signal that is singular in the sense of Definition 6.

Definition 9: A light signal $\lambda \in \mathscr{L}$ is called "parallax-free" iff for every regular $\tilde{\lambda} \in \mathscr{L}$ with image $\tilde{\lambda} \subseteq$ image λ the uniquely existing infinitesimal message $\tilde{J}$ along $\tilde{\lambda}$ is parallax-free in the sense of Definition 8.

Before we are ready to present several criteria for a light signal to be parallax-free, we have to prove the following lemma.

Lemma 3: Let J be an infinitesimal message along any $\lambda \in \mathscr{L}$ and define $U \in \mathscr{R}_\lambda$ by

$$U(s) = \exp\left(-\int_0^s \frac{g(\lambda'(x),\nabla_{\lambda'(x)} V)}{g(\lambda'(x),V_{\lambda(x)})}\,dx\right) V_{\lambda(s)}$$

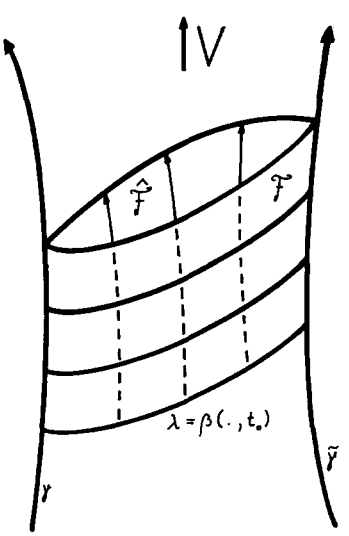


FIG. 3. The "parallactic bag" formed by the two-surfaces $\mathscr{F}$ and $\hat{\mathscr{F}}$.

for all $s \in [0,1]$. Then the following properties are satisfied:

(i) $g(\lambda',U') = 0$,

(ii) $g(\lambda'(s),U(s)) = g(\lambda'(0),V_{\lambda(0)})$ for all $s \in [0,1]$,

(iii) $(J-U)(0) = (J-U)(1) = 0$,

(iv) $J - U \in \mathscr{P}_\lambda^\perp$.

Proof: (i) The proof follows by the usual derivation rules from the definition of U.

(ii) As a consequence of (i), $g(\lambda',U):[0,1]\to\mathbb{R}$ has to be a constant function, i.e., $g(\lambda'(s),U(s)) = g(\lambda'(0),U(0))$ for all $s \in [0,1]$.

(iii) Here $U(0) = V_{\lambda(0)}$ and hence, $U(0) = J(0)$ is obvious; $U(1) = e^{r(\lambda)}V_{\lambda(1)}$ and hence, $U(1) = J(1)$ can be read from the proof of Theorem 3.

(iv) From $g(\lambda',U') = 0$ and $g(\lambda',J') = 0$, we know that $g(\lambda',J-U)$ has to be a constant function: As a result of (iii) it has to vanish. ■

Lemma 3 will be used to show the following thereom.

Theorem 4: Let J be an infinitesimal message along any light signal $\lambda \in \mathscr{L}$ and let $U \in \mathscr{P}_\lambda$ be the vector field defined in Lemma 3. Then the following two conditions are equivalent:

(i) J is parallax-free,

(ii) $J - U$ is a multiple of λ'.

Proof: Since "(ii) $\Rightarrow$ (i)" of Theorem 4 is implied by Definition 8 in a trivial manner, we only have to prove "(i) $\Rightarrow$ (ii)." By assumption, there is a map $v:[0,1]\to\mathbb{R}$ such that for all $s \in [0,1]$

$$J(s) - v(s)V_{\lambda(s)} \text{ is parallel to } \lambda'. \tag{5.1}$$

The boundary condition $J(0) = V_{\lambda(0)}$ implies

$$v(0) = 1. \tag{5.2}$$

The condition $g(\lambda',J') = 0$ implies

$$v'(s)g(\lambda'(s),V_{\lambda(s)}) = -v(s)g(\lambda'(s),\nabla_{\lambda'(s)}V) \tag{5.3}$$

for all $s \in [0,1]$. By integrating (5.3) with the initial condition (5.2), we end up with

$$v(s)V_{\lambda(s)} = U(s) \text{ for all } s \in [0,1],$$

which completes the proof as a result of (5.1). ■

Using the concept of Jacobi classes (see Definition 2), Theorem 4 can be reformulated in the following way.

Theorem 5: Let $\lambda \in \mathscr{L}$ be any light signal and define $U \in \mathscr{P}_\lambda$ as in Lemma 3. Then the following conditions are equivalent:

(i) λ is parallax-free,

(ii) $[U]$ is a Jacobi class.

Proof: Since "(i) $\Rightarrow$ (ii)" of Theorem 5 is an immediate consequence of Theorem 4, we only have to prove the implication "(ii) $\Rightarrow$ (i)." Without loss of generality, we assume λ to be regular. Our assumption guarantees the existence of a Jacobi field J such that

$$J - U \text{ is parallel to } \lambda'. \tag{5.4}$$

For the present, J is unique only up to the transformations $J(s) \mapsto J(s) + (as+b)\lambda'(s)$ with arbitrary real constants a and b. We remove this ambiguity by the requirements $(J-U)(0) = 0$ and $(J-U)(1) = 0$. Moreover, this unique J satisfies $g(\lambda',J') = g(\lambda',U') = 0$, where the first equality follows from (5.4) and the second was proven in Lemma 3. Thus we have shown that J is the (unique) infinitesimal message along J. This, together with (5.4), completes the proof. ■

Theorem 5 gives a terse and easily remembered criterion, but we want to obtain another which is more appropriate to practical calculations. To that end, let us bring the index form $\mathscr{J}_\lambda$ into play. First, we have to prove a lemma, which relates $\mathscr{J}_\lambda$ to kinematical quantities.

Lemma 4: Let J be an infinitesimal message along any light signal $\lambda \in \mathscr{L}$ and define $U \in \mathscr{P}_\lambda$ as in Lemma 3. Let X be a lightlike vector field having λ as an integral curve and define a two-form ζ by

$$\zeta := d\left(\rho - 2\frac{\sigma(X,.)}{g(V,X)} + \frac{\sigma(X,X)}{g(V,X)^2}g(V,.)\right),$$

where ρ denotes the one-form defined in (4.4), σ denotes the shear of V, and d stands for the exterior derivative. Then the index form $\mathscr{J}_\lambda$ satisfies, for all $A \in \mathscr{P}_\lambda^\perp$,

$$\begin{aligned}\mathscr{J}_\lambda(A,J-U) &= g(A(1),J'(1)-U'(1)) \\ &\quad - g(A(0),J'(1)-U'(1)) \\ &\quad + 2g(\lambda'(0),V_{\lambda(0)})\int_0^1 \zeta(\lambda'(s),A(s))ds.\end{aligned}$$

Proof: From representation (3.1) of the index form and Lemma 3 (ii) we find

$$\mathscr{J}_\lambda(A,J-U) = g(A(1),J'(1)-U'(1)) - g(A(0),J'(0)-U'(0)) + g(\lambda'(0),V_{\lambda(0)})\int_0^1 \frac{g(A(s),R(\lambda'(s),U(s),\lambda'(s)) - U''(s)}{g(\lambda'(s),U(s))}ds.$$

Let us concentrate upon the integrand in the last term of the above equation. By calculating explicitly the second derivative U'' of the vector field U from its definition (given in Lemma 3), this integrand becomes

$$\frac{g(R(\lambda'(s),U(s),\lambda'(s)) - U''(s),.)}{g(\lambda'(s),U(s))} = \left[\frac{g(R(X,V,X),.)}{g(X,V)}\right.$$

$$
\left.
-\frac{g(\nabla_X(\nabla_X V-[g(X,\nabla_X V)/g(X,V)]V)-[g(X,\nabla_X V)/g(X,V)](\nabla_X V-[g(X,\nabla_X V)/g(X,V)]V),.)}{g(X,V)}
\right]_{\lambda(s)}
$$

$$
=\left[\frac{g(\nabla_X\nabla_V X-\nabla_{[X,V]}X,.)}{g(X,V)}-\nabla_X\frac{g(\nabla_X V-[g(X,\nabla_X V)/g(X,V)]V,.)}{g(X,V)}\right]_{\lambda(s)}
$$

$$
=\left[\nabla_X\frac{g(L_V X+[g(X,\nabla_X V)/g(X,V)]V,.)}{g(X,V)}+\frac{g(X,\nabla_X V)g(\nabla_V X,.)}{g(X,V)^2}\right.
$$

$$
\left.-\nabla_{[V,X]}\frac{g(X,.)}{g(X,V)}+\text{multiples of } g(X,.)\right]_{\lambda(s)}
$$

[using the identity $\nabla_X\alpha=L_X\alpha-\alpha(\nabla.X)$, which is valid for every one-form α]

$$
=\left[L_X\frac{g(L_V X+[g(X,\nabla_X V)/g(X,V)]V,.)}{g(X,V)}-\frac{g(L_V X+[g(X,\nabla_X V)/g(X,V)]V,\nabla.X)}{g(X,V)}+\frac{g(X,\nabla_X V)}{g(X,V)}\nabla_V\frac{g(X,.)}{g(X,V)}\right.
$$

$$
\left.+L_{[V,X]}\frac{g(X,.)}{g(X,V)}-\frac{g(X,\nabla.L_V X)}{g(X,V)}+\text{multiples of } g(X,.)\right]_{\lambda(s)}
$$

(using the identity $L_{[V,X]}=i_{[V,X]}d+di_{[V,X]}$, where i denotes the contraction operator)

$$
=\left[L_X\frac{g(L_V X+[g(X,\nabla_X V)/g(X,V)]V,.)}{g(X,V)}-\frac{g(X,\nabla_X V)g(V,\nabla.X)}{g(X,V)^2}+\frac{g(X,\nabla_X V)}{g(X,V)}\left(L_V\frac{g(X,.)}{g(X,V)}-\frac{g(X,\nabla.V)}{g(X,V)}\right)\right.
$$

$$
\left.+i_{[V,X]}d\frac{g(X,.)}{g(V,X)}+d\frac{g(X,[V,X])}{g(X,V)}-\frac{g([V,X],\nabla.X)}{g(X,V)}-\frac{g(X,\nabla.L_V X)}{g(X,V)}+\text{multiples of } g(X,.)\right]_{\lambda(s)}
$$

[the terms marked by ... cancel since X is lightlike and hence, $g(X,\nabla_V X)=0$; furthermore, we use the identities $L_X=i_X d+di_X$ and $i_{[V,X]}=L_V i_X-i_X L_V$]

$$
=\left[i_X d\frac{g(L_V X+[g(X,\nabla_X V)/g(X,V)]V,.)}{g(X,V)}+0+\frac{g(X,\nabla_X V)}{g(X,V)}L_V\frac{g(X,.)}{g(X,V)}\right.
$$

$$
\left.+L_V i_X d\frac{g(X,.)}{g(X,V)}-i_X L_V d\frac{g(X,.)}{g(X,V)}-\text{multiples of } g(X,.)\right]_{\lambda(s)}
$$

[using $g(X,\nabla.X)=0$ in the ... term and the identity $L_V d=dL_V$ in the - - - term]

$$
=\left[i_X d\left(\frac{g(L_V X,.)}{g(X,V)}-L_V\frac{g(X,.)}{g(X,V)}+\frac{g(X,\nabla_X V)g(V,.)}{g(X,V)^2}\right)\right.
$$

$$
\left.+\frac{g(X,\nabla_X V)}{g(X,V)}L_V\frac{g(X,.)}{g(X,V)}+L_V\nabla_X\frac{g(X,.)}{g(X,V)}+\text{multiples of } g(X,.)\right]_{\lambda(s)}
$$

[the term marked by ... is a multiple of $g(X,.)$; using the representation (4.3) of $L_V g$ results after an easy calculation in the following]

$$
=\left[i_X d\left(g(\nabla_V V,.)-\frac{\theta}{3}g(V,.)-2\frac{\sigma(X,.)}{g(V,X)}+\frac{\sigma(X,X)g(V,.)}{g(V,X)^2}+\text{multiples of } g(X,.)\right)\right]_{\lambda(s)}
$$

$$
=2\xi(\lambda'(s),.)+\text{multiples of } g(\lambda'(s),.).
$$

■

Having finished this laborious proof, we are ready to give a criterion for a light signal to be parallax-free by kinematical quantities.

Theorem 6: Assume $\lambda \in \mathscr{L}$ to be any light signal and define the two-form ζ as in Lemma 4. Then the following two conditions are equivalent:

(i) λ is parallax-free,

(ii) $\zeta(\lambda',.)$ is a multiple of $g(\lambda',.)$.

Proof: Without loss of generality, we assume λ to be regular. We denote the unique infinitesimal message along λ by J and introduce the vector field $U \in \mathscr{P}_\lambda$ as in Lemma 3.

For "(i) $\Rightarrow$ (ii)" of Theorem 6: Our assumption guarantees $J - U$ to be a multiple of λ' as a result of Theorem 4. Thus from the definition of the index form, we find

$$\mathscr{J}_\lambda(A,J-U)=0 \tag{5.5}$$

for all $A \in \mathscr{P}^\perp_\lambda$. On the other hand, as a consequence of Lemma 4,

$$\mathscr{J}_\lambda(A,J-U)=2g(\lambda'(0),V_{\lambda(0)})\int_0^1 \zeta(\lambda'(s),A(s))ds \tag{5.6}$$

has to hold for all $A \in \mathscr{P}^\perp_\lambda$. (The boundary terms vanish as a result of $J - U$ being parallel to λ', which implies $J' - U'$ being parallel to λ'.) Since this consideration remains valid for arbitrarily short parts of the light signal, (5.5) and (5.6) require $\zeta(\lambda',.)$ to be a multiple of $g(\lambda',.)$.

For "(ii) $\Rightarrow$ (i)" of Theorem 6: As a result of Lemma 4, the assumption implies $\mathscr{J}_\lambda(J-U,J-U)=0$. (This time the boundary terms vanish since $J - U$ vanishes in the boundary points, cf. Lemma 3.) From Lemma 1 we know that $J - U$ has to be a multiple of λ' in this case, which completes the proof resulting from Theorem 4. ∎

Theorem 6 gives a criterion for a light signal to be parallax-free in terms of an eigenvalue equation along λ. Please note that the two-form ζ has to be constructed for each $\lambda \in \mathscr{L}$ individually (except in the case of vanishing shear), but that the one-form $\zeta(\lambda',.)$ along λ does not depend on the extension X chosen for the tangent field. This is not obvious from the definition of ζ, but it becomes manifest if, by a rather tedious calculation, all derivatives are expressed by covariant derivatives.

In concluding this section, let us reformulate the main result of our earlier paper[1] in the terminology of the present paper.

Theorem 7: For our kinematical world model (M,g,V), the following two conditions are equivalent:

(i) every $\lambda \in \mathscr{L}$ is parallax-free,

(ii) on each simply connected open submanifold of M there is a scalar function f such that $e^f V$ is a conformal Killing field (this implies that the shear has to vanish and that the one-form ρ defined in (4.4) is given by $\rho = df$).

For a proof, the reader is referred to Ref. 1, with the additional remark that this proof can be markedly shortened if Theorem 6 is available.

VI. MODELS WITH TEMPORALLY CONSTANT REDSHIFT

In this section and for Sec. V we shall apply the foregoing results to models (M,g,V) with special redshift properties. The present section is devoted to models in which the redshift is temporally constant for each pair of observers in V. In order to find a characterization of such models, we proceed in two steps: First, we characterize the situation that an observer sees every infinitesimally close neighbor under some redshift remaining constant in time. In the second step, we prove that in the case of temporally constant redshift no parallaxes can occur. The combination of these two results will give the desired characterization.

Theorem 8: For our kinematical world model (M,g,V), the following two conditions are equivalent:

(i) every observer γ in V sees every infinitesimally neighboring observer under temporally constant redshift,

(ii) the shear σ vanishes and the one-form ρ defined in (4.4) satisfies $L_V\rho = -\rho(V)\rho$.

Proof: For "(i) $\Rightarrow$ (ii)" of Theorem 8: Consider a lightlike vector field X with $g(X,V)>0$ and $L_V X$ parallel to V. In this case, the vector field $t \mapsto X_{\gamma(t)}$ along any integral curve γ of V describes a one-parameter family of light signals connecting γ with an infinitesimally neighboring integral curve of V. From the representation (4.5) of the red shift function we can read that our assumption requires, for every X with the aforementioned properties,

$$\begin{aligned}0 &= V\left(\rho(X)-\frac{\sigma(X,X)}{g(V,X)}\right)\\ &= (L_V\rho)(X)+\rho(L_V X)\\ &\quad -\frac{(L_V\sigma)(X,X)}{g(V,X)}-2\frac{\sigma(X,L_V X)}{g(V,X)}\\ &\quad +\frac{\sigma(X,X)}{g(V,X)^2}((L_V g)(V,X)+g(V,L_V X)).\end{aligned} \tag{6.1}$$

We know that $L_V X = wV$ holds with some scalar function w. From $0 = Vg(X,X) = (L_V g)(X,X) + 2g(X,L_V X)$ we find $w = \rho(X) - \sigma(X,X)/g(V,X)$ and (6.1) becomes

$$(L_V\rho)(X)+\rho(X)\rho(V)-\frac{(L_V\sigma)(X,X)}{g(V,X)}+\frac{\sigma(X,X)^2}{g(V,X)^3}=0. \tag{6.2}$$

Since (6.2) is purely algebraic in X, it has to hold for all lightlike X. Let (E_1,E_2,E_3) denote an orthonormal basis in the orthocomplement of V. Then $X_\pm = E_i \pm V$ may be inserted into (6.2) for $i = 1,2,3$:

$$\begin{aligned}&(L_V\rho)(E_i)\pm(L_V\rho)(V)+\rho(E_i)\rho(V)\pm\rho(V)^2\\ &\quad\pm(L_V\sigma)(E_i,E_i)\pm\sigma(E_i,E_i)^2=0.\end{aligned}$$

The sum and the difference, respectively, of the + and − versions of the above equation reads as

$$(L_V\rho)(E_i)+\rho(E_i)\rho(V)=0, \tag{6.3}$$

$$(L_V\rho)+\rho(V)^2+(L_V\sigma)(E_i,E_i)+\sigma(E_i,E_i)^2=0. \tag{6.4}$$

From (6.4) we can deduce that the term

$$\sum_{i=1}^{3} \sigma(E_i,E_i)^2$$

has to take the same value in every orthonormal basis (E_1,E_2,E_3). Since σ is trace-free, this implies that the shear vanishes. The condition $L_V \rho = -\rho(V)\rho$ can then be read from (6.3) and (6.4). The implication "(ii) $\Rightarrow$ (i)" of Theorem 8 is shown to hold true by reading the above proof in the reverse direction. ■

If we require the vector field V to be complete, then the condition $L_V \rho = -\rho(V)\rho$ can only be satisfied by $\rho(V) = 0$ and hence, $L_V \rho = 0$ (since the only solution to the differential equation $y' = -y^2$, which is defined on all of $\mathbf{R}$, is $y \equiv 0$).

It is our aim to characterize those models (M,g,V) in which every observer is seen under temporally constant redshift (not only the infinitesimally neighboring ones). The reader might think that this aim has already been achieved by Theorem 8, having been misled by the following superficial argument: If every observer in V sees his infinitesimal neighbors under temporally constant redshifts, then this has to also hold true for neighbors in finite distances as a result of the additivity of the red shift function r along every light ray. This argument is erroneous since the possible appearance of parallaxes has not been taken into account. Indeed, not only is the argument erroneous, but so is its result. This is shown by the following example:

$$M = \mathbb{R}^4,$$

$$g = dx^2 + dy^2 + dz^2 - dt^2 - F(x,y,z)^2 t^2\, dz^2 + 2F(x,y,z)t\, dz \vee dt,$$

$$V = \frac{\partial}{\partial t}.$$

Here, $\vee$ denotes the symmetrized tensor product and F stands for an arbitrary smooth function of x, y, and z. In this case, the shear vanishes and the one-form ρ is given by $\rho = F\,dz$. Hence, condition (ii) of Theorem 8 is satisfied. On the other hand, $d\rho = dF \wedge dz$ does not vanish if F is assumed to be nonconstant with respect to the arguments x and y. As a consequence of Theorem 7, this model cannot be free of parallaxes. As a result of the following theorem, it turns out to be impossible that the redshift in this model is temporally constant for every pair of observers.

Theorem 9: Assume the redshift in (M,g,V) to be temporally constant for every pair of observers. Then every $\lambda \in \mathscr{L}$ is parallax-free.

Proof: Our assumption guarantees the redshift to be temporally constant, in particular for every pair of infinitesimally neighboring observers. Hence, Theorem 8 implies

$$\sigma = 0 \tag{6.5}$$

and the redshift function resulting from (4.5) reduces to

$$r(\lambda) = \int_\lambda \rho. \tag{6.6}$$

Now let us fix two integral curves $\gamma: I \to M$ and $\tilde{\gamma}: \tilde{I} \to M$ of V, along with a message β from γ to $\tilde{\gamma}$ with the message function h. For convenience, we assume the map β to be an embedding. By assumption,

$$\frac{d}{dt}\int_{\beta(.,t)} \rho = 0 \text{ for all } t \in I \tag{6.7}$$

as a result of (6.6). Fix any $t_0 \in I$ and apply the Stokes theorem to the two-surface $\mathscr{F}(\Delta t)$ bounded by γ, $\tilde{\gamma}$, $\beta(.,t_0)$, and $\beta(.,t_0 + \Delta t)$ and lying under the image of β:

$$\int_{\mathscr{F}(\Delta t)} d\rho = \int_{t_0}^{t_0+\Delta t} \rho(\gamma'(t))dt + \int_{\beta(.,t_0+\Delta t)} \rho - \int_{t_0}^{t_0+\Delta t} \rho((\tilde{\gamma}\circ h)'(t))dt - \int_{\beta(.,t_0)} \rho. \tag{6.8}$$

As a result of (6.7), the second term on the rhs of (6.8) cancels with the last term. Dividing (6.8) by Δt and taking the limit $\Delta t \to 0$ results in

$$\int_0^1 d\rho(J(s),\lambda'(s))ds = \rho(J(0)) - \rho(J(1)), \tag{6.9}$$

with $\lambda := \beta(.,t_0)$ and J denoting the variational vector field along λ, which generates the variation β. Recall that β is an arbitrary message up to the requirement that it be an embedding. Hence, (6.9) has to hold for every $\lambda \in \mathscr{L}$ which is free of conjugate points, with J denoting the unique infinitesimal message along λ. Keeping this result in mind, we are now going to carry out another construction with the same $\lambda = \beta(.,t_0)$. Define a map

$$\delta: [0,1] \times I \to M, \quad (s,t) \mapsto \delta(s,t)$$

by the following requirements:

(i) $\delta(s,.)$ is some reparametrization of an integral curve of V for all $s \in [0,1]$,

(ii) $\delta(.,t)$ is lightlike for all $t \in I$,

(iii) $\delta(.,t_0) = \beta(.,t_0)$.

Let us denote by $\hat{\mathscr{F}}(\Delta t)$ that part of the image of δ that is bounded by γ, $\tilde{\gamma}$, $\delta(.,t_0)$, and $\delta(.,t_0 + \Delta t)$. Here $\hat{\mathscr{F}}(\Delta t)$ coincides with $\mathscr{F}(\Delta t)$ if β is parallax-free, which is unknown to us at this time. (The symbols $\mathscr{F}$ and $\hat{\mathscr{F}}$ are here used in a way analogous to Fig. 3.) Note that the integral

$$\int_{\delta(.,t)} \rho \tag{6.10}$$

is not allowed to be interpreted as redshift since $\delta(.,t)$ is not known to be geodesic. However, any sufficiently short part of $\delta(.,t)$ differs from a light signal arbitrarily little. Thus by writing the integral (6.10) as a limit of summations over step functions,

$$\frac{d}{dt}\int_{\delta(.,t)} \rho = 0 \tag{6.11}$$

turns out to hold true as a consequence of (6.7). As in the first part of the proof with $\mathscr{F}(\Delta t)$, we carry out the Stokes theorem argument with $\hat{\mathscr{F}}(\Delta t)$ analogously. We arrive at an equation similar to (6.9), with J replaced by the vector field generating the variation δ. It can easily be checked that this generating vector field is given by the vector field U defined in Lemma 3. Thus

$$\int_0^1 d\rho(U(s),\lambda'(s))ds = \rho(U(0)) - \rho(U(1)) \tag{6.12}$$

has been established. From (6.9) and (6.12) we find

$$\int_0^1 d\rho(J(s) - U(s), \lambda'(s))ds = 0 \tag{6.13}$$

since $J - U$ vanishes in the boundary points; see Lemma 3. Using Lemma 4, (6.5), and (6.13) we arrive at

$$\mathscr{J}_\lambda(J - U, J - U) = 0.$$

The above equation has to hold for every $\lambda \in \mathscr{L}$ without conjugate points and with J denoting the infinitesimal message along λ: As a result of Lemma 1, this equation requires $J - U$ to be parallel to λ', which proves J to be parallax-free according to Theorem 4. Since every light signal can be cut into pieces which are free of conjugate points, this completes the proof. ■

The combination of Theorems 8 and 9 results in the following characterization of models with temporally constant redshift.

Theorem 10: For our kinematical world model (M,g,V), the following two conditions are equivalent:

(i) the redshift is temporally constant for every pair of observers in V,

(ii) on every simply connected open submanifold of M there is a scalar function f such that $e^f V$ is a conformal Killing field and f satisfies $L_V\, df = -df(V)df$.

Proof: See Theorems 7–9. ■

If V is complete, $L_V\, df = -df(V)df$ implies $df(V) = 0$, which means that $e^f V$ is a Killing field rather than a conformal Killing field. Consequently, a complete model with temporally constant redshift has to be stationary.

VII. MODELS WITH ISOTROPIC REDSHIFT

This section is devoted to models (M,g,V) in which the redshift under which an observer sees his infinitesimal neighbors is isotropic. From the representation (4.5) of the red shift function we can read that this is the case iff V is freely falling and shear-free. This result has been known since 1961; see Ehlers.[8] In this special situation the redshift function reduces to

$$r(\lambda) = -\int_\lambda \frac{\theta}{3} g(V,.). \tag{7.1}$$

For convenience we introduce the following terminology.

Definition 10: Here (M,g,V) is called a "Hubble model" iff $\sigma = 0$ and $\nabla_V V = 0$. A Hubble model is called "proper" iff its volume expansion θ (and in consequence its redshift function r) is strictly positive.

Note that (M,g,V) is a Hubble model with $\theta = 0$ iff V is a Killing field. It should be emphasized that by Definition 10 the redshift in a Hubble model is ensured to be isotropic only in the infinitesimal regime. It is not even clear what should be meant by isotropy in the large regime since there is no natural measure for the distance between emitter and receiver.

The most evident example for a Hubble model is given by a Robertson–Walker space-time with its standard observer field. However, this is a very special example since it is irrotational and free of parallaxes. The rotation (= vorticity) of an arbitrary Hubble model need not vanish (we shall illustrate this by an example below), but the rotation is somewhat restricted: From the generally valid representation

$$\omega = -dg(V,.) - g(V,.) \wedge g(\nabla_V V,.) \tag{7.2}$$

of the rotation two-form ω [cf. Ehlers,[8] Eq. (1.1.15)] we can read that in every Hubble model

$$d\omega = 0 \tag{7.3}$$

has to be satisfied, which may be viewed as a conservation law for the rotation. Furthermore, the following theorem holds true.

Theorem 11: Let (M,g,V) be a proper Hubble model. If every $\lambda \in \mathscr{L}$ is parallax-free, then the rotation has to vanish.

Proof: In the case of a Hubble model, the one-form ρ defined in (4.4) reduces to $\rho = -(\theta/3)g(V,.)$, which implies

$$d\rho = -\frac{d\theta}{3} \wedge g(V,.) - \frac{\theta}{3} dg(V,.). \tag{7.4}$$

Since $d\rho$ is assumed to vanish as a result of Theorem 7, wedging of (7.4) by $g(V,.)$ results in

$$0 = (\theta/3)g(V,.) \wedge dg(V,.). \tag{7.5}$$

As a result of the generally valid equation (7.2), on the other hand we find

$$g(V,.) \wedge dg(V,.) = -g(V,.) \wedge \omega. \tag{7.6}$$

Inserting (7.6) into (7.5) and contracting with V results in

$$0 = (\theta/3)\omega. \tag{7.7}$$

Since θ is assumed to be strictly positive, (7.7) implies that the rotation vanishes. ■

Note that the converse of Theorem 11 does not hold true. An irrotational proper Hubble model might admit parallaxes.

As a result of Theorem 11, a nonvanishing rotation implies either an isotropy in redshift or the occurrence of parallaxes (or both). This does not contradict our earlier finding[1] that observations of parallaxes alone (!) do not yield any information about the rotation at all.

Whereas there are obviously many examples of rotating Hubble models with vanishing θ (e.g., Gödel's cosmos with its standard observer field), the reader might doubt whether proper (!) rotating Hubble models do exist. For this reason, an explicit example is given:

$$M = \mathbb{R}^4,$$

$$g = e^t(dx^2 + dy^2 + dz^2) + F(x,y,z)^2(e^t - 1)dz^2 - dt^2 + 2F(x,y,z)dz \vee dt,$$

$$V = \frac{\partial}{\partial t}.$$

Here, $\vee$ denotes the symmetrized tensor product and F stands for an arbitrary smooth function of x, y, and z. This (M,g,V) turns out to be a Hubble model with $\theta = \frac{3}{2}$ and $\omega = dz \wedge dF$. Hence, the rotation does not vanish if we assume F to be nonconstant with respect to the arguments x and y. As a result of Theorem 11, this model cannot be parallax-free.

We end up with a theorem showing the impossibility of constructing an example with analogous properties on an underlying space-time with constant curvature.

Theorem 12: Assume (M,g,V) to be a Hubble model (not necessarily a proper one) and (M,g) to be a space of constant curvature. Then V has to be irrotational and every $\lambda \in \mathscr{L}$ has to be parallax-free (i.e., the model has to be conformally static).

Proof: We start from the well-known propagation equation for the deformation tensor along any integral curve $t \mapsto \gamma(t)$ of V. In index notation, this equation reads [cf. Eq. (4.25) in Ref. 9] as

$$\frac{d}{dt}\theta_{ij} = -R_{i4j4} - \omega_{ik}\omega_{kj} - \theta_{ik}\theta_{kj} + \dot{V}_{(i;j)} + \dot{V}_i\dot{V}_j, \tag{7.8}$$

where an orthonormal Fermi basis has been introduced (i,j, and k running from 1–3 and the summation over k are understood). Since (M,g) is assumed to be of constant curvature, the curvature tensor has to satisfy

$$R(X,Y,Z) = K(g(X,Z)Y - g(Y,Z)X) \tag{7.9}$$

with some constant K, which implies that R_{i4j4} is a multiple of δ_{ij}. Since (M,g,V) is assumed to be a Hubble model, $\dot{V}_i = (\nabla_V V)_i = 0$ and $\theta_{ij} = (\theta/3)\delta_{ij}$. Hence, (7.8) forces $\omega_{ik}\omega_{kj}$ to be a multiple of δ_{ij}. Since the antisymmetrical 3×3 matrix (ω_{kj}) must have a nontrivial kernel, this requires $\omega_{ik}\omega_{kj} = 0$ and, again as a consequence of the antisymmetry, $\omega_{ij} = 0$. In order to prove that every $\lambda \in \mathscr{L}$ is parallax-free, consider the vector field U defined in Lemma 3 for an arbitrary $\lambda \in \mathscr{L}$. Since the shear, acceleration, and rotation of V are known to vanish, U reduces to

$$U(s) = \exp\left(-\int_0^s \frac{1}{3}\theta_{\lambda(x)}g(V_{\lambda(x)},\lambda'(x))dx\right)V_{\lambda(s)}.$$

Explicit calculation of the second derivative of U shows $U''(s)$ to be a multiple of $\lambda'(s)$. On the other hand, $R(\lambda',U,\lambda')$ is a multiple of λ' as a result of (7.9). Hence, $[U]$ is a Jacobi class, which proves our arbitrary $\lambda \in \mathscr{L}$ to be parallax-free as a result of Theorem 5. ■

ACKNOWLEDGMENTS

I would like to thank Professor K.-E. Hellwig (TU Berlin) and his co-workers, Wolfgang Hasse in particular, for helpful discussions.

[1]W. Hasse and V. Perlick, J. Math. Phys. **29**, 2064 (1988).
[2]J. Beem and P. Ehrlich, *Global Lorentzian Geometry* (Dekker, New York, 1981).
[3]H. Weyl, *Space, Time, Matter* (Dover, New York, 1961); translation from *Raum, Zeit, Materie* (Springer, Berlin, 1923).
[4]W. Kermack, W. McCrea, and E. Whittacker, Proc. R. Soc. Edinburgh **53**, 31 (1932).
[5]E. Schrödinger, *Expanding Universes* (Cambridge U. P., Cambridge, 1956).
[6]D. Brill, Lect. Notes Phys. **14**, 45 (1972).
[7]R. Herrman, *Topics in General Relativity* (Interdisciplinary Mathematics V, private press, 1973).
[8]J. Ehlers, Akad. Wiss. Lit. (Mainz), Abh. Math. Nat. Kl. **1961**(11), 791 (1961).
[9]S. Hawking and G. Ellis, *The Large Scale Structure of Space Time* (Cambridge U. P., Cambridge, 1973).

Internal and external metrics for a perfect fluid cylinder in general relativity

W. Davidson[a)]
Mathematics Department, University of Otago, Dunedin, New Zealand

(Received 4 January 1990; accepted for publication 4 April 1990)

A solution to Einstein's equations describing a perfect fluid cylinder of finite radius is presented. The proper density μ and pressure p of the fluid are physically well behaved in the radial coordinate range $0 \leqslant r \leqslant r_1$. On the axis ($r = 0$) the solution is regular and μ and p are finite and positive. As r increases μ and p decrease steadily through positive values, p vanishing at $r = r_1$. The ratio p/μ (< 1) is also monotonically decreasing, as is also the velocity of sound a (< 1) in the fluid. The equation of state is $p = \frac{1}{7}\mu - N\mu^{3/10}$, where N is a positive constant. The matching metric for the vacuum exterior to the cylinder is given, so that the space-time is complete and nonsingular.

I. INTRODUCTION

Stationary cylindrically symmetric nonvacuum solutions of Einstein's equations for nonrotating fluid have been quite small in number,[1] and few of these have related to a finite radial distribution. Evans[2] succeeded in finding three such solutions [his models (ii)–(iv)]. Model (ii) had the equation of state $p = \frac{1}{3}\mu + \text{const}$. Models (iii) and (iv) were difficult to analyze with equations of state that were far from simple. Kramer[3] has solved the problem of perfect fluid having equation of state $p = (\gamma - 1)\mu$ ($\gamma = \text{const}$), but the fluid cylinders are radially infinite (see also Ref. 4).

Here we give an example of solution for a perfect fluid cylinder that has finite radius and reasonable physical characteristics.

II. FIELD EQUATIONS

The metric for stationary, cylindrically symmetric space-time will be taken in the general static form

$$ds^2 = A\,dr^2 + B\,dz^2 + C\,d\phi^2 - D\,dt^2, \tag{1}$$

where A, B, C, and D are functions of the radial coordinate r only. The metric admits an Abelian group G_3 acting on timelike orbits T_3, the three Killing vectors being ∂_t, ∂_ϕ, and ∂_z. For a perfect fluid the Einstein equations are (primes indicating differentiation with respect to r)

$$K_0 p = \frac{1}{4}A^{-1}\left[\frac{B'C'}{BC} + \frac{C'D'}{CD} + \frac{D'B'}{DB}\right], \tag{2}$$

$$= \frac{1}{4}A^{-1}\left[2\left(\frac{C''}{C} + \frac{D''}{D}\right) - \left(\frac{C'^2}{C^2} + \frac{D'^2}{D^2} + \frac{A'C'}{AC} + \frac{A'D'}{AD} - \frac{C'D'}{CD}\right)\right], \tag{3}$$

$$= \frac{1}{4}A^{-1}\left[2\left(\frac{B''}{B} + \frac{D''}{D}\right) - \left(\frac{B'^2}{B^2} + \frac{D'^2}{D^2} + \frac{A'B'}{AB} + \frac{A'D'}{AD} - \frac{B'D'}{BD}\right)\right], \tag{4}$$

$$-K_0\mu = \frac{1}{4}A^{-1}\left[2\left(\left(\frac{B'}{B}\right)' + \left(\frac{C'}{C}\right)'\right) + \left(\frac{B'}{B} + \frac{C'}{C}\right)\times\left(\frac{B'}{B} + \frac{C'}{C} - \frac{A'}{A}\right) - \frac{B'C'}{BC}\right]. \tag{5}$$

III. A SOLUTION

A particular solution to these equations gives for the metric coefficients

$$A = f^{6/7}g^{-1}, \quad B = f^{-3/7}g, \quad C = f^{-3/7}r^2, \quad D = f^{12/7}, \tag{6}$$

where

$$f = \tfrac{64}{9} + \beta^2 r^2, \quad f_1 = \tfrac{64}{9} + \beta^2 r_1^2, \quad g = f_1^2 - 4f, \tag{7}$$

r_1 and β being positive constants. The pressure and density are given by

$$K_0 p = \tfrac{64}{7}\beta^2 f^{-20/7}(f_1^2 - f^2), \tag{8}$$

$$K_0\mu = \tfrac{64}{3}\beta^2 f^{-20/7} f_1^2. \tag{9}$$

It is clear that the metric is regular at the axis $r = 0$, and since $g > 0$ in the range $0 \leqslant r \leqslant r_1$ the space-time is free of singularity.

IV. PHYSICAL PROPERTIES

On the axis $\mu = \mu_0 > 0$, $p = p_0 > 0$. Moreover, $dp/dr < 0$, $d\mu/dr < 0$ in the range $0 \leqslant r \leqslant r_1$. Thus p and μ monotonically decrease as r increases in this range, while p vanishes at $r = r_1$ that is therefore the boundary of the fluid cylinder. At the boundary the density $\mu = \mu_1 > 0$.

For the ratio p/μ we derive

$$\frac{p}{\mu} = \frac{3}{7}\left(1 - \frac{f^2}{f_1^2}\right) > 0. \tag{10}$$

Thus

$$\tfrac{3}{7} > p/\mu \geqslant 0 \tag{11}$$

and p/μ steadily decreases as r increases, vanishing at $r = r_1$. At the axis the maximum of the ratio p/μ never exceeds $\frac{3}{7}$ regardless of r_1.

The velocity of sound in the fluid, given by $a = (dp/$

[a)] Permanent address: 25 Paddock Close, Edwinstowe, Notts, NG21 9LP, England.

 0022-2488/90/081972-02$03.00

$dr)^{1/2}/(d\mu/dr)^{1/2}$ has the expression

$$a=\left(\frac{3}{7}\right)^{1/2}\left(1-\frac{3}{10}\frac{(\frac{64}{9}+\beta^2r^2)^2}{(\frac{64}{9}+\beta^2r_1^2)^2}\right)^{1/2}, \tag{12}$$

and so is also monotonically decreasing as r increases, lying in the range

$$(\tfrac{3}{7})^{1/2}>a\geqslant(\tfrac{3}{10})^{1/2} \tag{13}$$

(velocity of light $c=1$). At the axis a never exceeds the value $(\frac{3}{7})^{1/2}$.

The equation of state of the fluid will be seen to be

$$p=\tfrac{3}{7}\mu-N\mu^{3/10}, \tag{14}$$

where $N>0$ is a constant.

The proper radius R of the fluid cylinder is

$$R=\int_0^{r_1}\frac{f^{3/7}}{g^{1/2}}\,dr. \tag{15}$$

Since the function g remains positive so that the integrand in (15) is finite in the range $0\leqslant r\leqslant r_1$, the integral is well behaved and finite so that R is finite.

V. THE EXTERIOR METRIC

The interior metric to be matched at the boundary $r=r_1$ is

$$\begin{aligned}ds^2=&(\tfrac{64}{9}+\beta^2r^2)^{6/7}((\tfrac{64}{9}+\beta^2r_1^2)^2-4(\tfrac{64}{9}+\beta^2r^2))^{-1}dr^2\\&+(\tfrac{64}{9}+\beta^2r^2)^{-3/7}((\tfrac{64}{9}+\beta^2r_1^2)^2\\&-4(\tfrac{64}{9}+\beta^2r^2))dz^2+(\tfrac{64}{9}+\beta^2r^2)^{-3/7}r^2\,d\phi^2\\&-(\tfrac{64}{9}+\beta^2r^2)^{12/7}\,dt^2.\end{aligned} \tag{16}$$

We seek a cylindrically symmetric vacuum exterior metric of the form (1). At the boundary $r=r_1$ the metric coefficients A, B, C, and D must be continuous. Since the pressure p of the fluid vanishes at $r=r_1$, we see from (2) that the first derivatives B', C', and D' must also be continuous, but it is not required that A' be continuous. This is the same situation as in spherical symmetry; in particular, the Schwarzschild interior and exterior metrics have a discontinuous first derivative in their radial metric coefficients.

The required external metric having A, B, C, D, B', C', and D' continuous at $r=r_1$ is found to be

$$\begin{aligned}ds^2=&(\tfrac{28}{9}+\beta^2r_1^2)^{-1}(\tfrac{64}{9}+\beta^2r_1^2)^{(-6b^2+25b-24)/7b}(\tfrac{64}{9}+\beta^2r^2)^{(6b^2-26b+24)/7b}r_1^{-2}r^2\,dr^2\\&+(\tfrac{28}{9}+\beta^2r_1^2)(\tfrac{64}{9}+\beta^2r_1^2)^{(2/7)(8-3b)}(\tfrac{64}{9}+\beta^2r^2)^{-(6/7)(2-b)}\,dz^2+(\tfrac{64}{9}+\beta^2r_1^2)^{-3(8-3b)/7b}\\&\times(\tfrac{64}{9}+\beta^2r^2)^{12(2-b)/7b}r_1^2\,d\phi^2-(\tfrac{64}{9}+\beta^2r^2)^{12/7}\,dt^2,\end{aligned} \tag{17}$$

where b is a constant defined by

$$b=\tfrac{3}{2}\beta^2r_1^2(\tfrac{28}{9}+\beta^2r_1^2)^{-1}. \tag{18}$$

VI. CONCLUSION

The interior solution presented here is physically well behaved, without singularity, and appropriately matched at the fluid boundary by an exterior vacuum space-time.

[1]D. Kramer, H. Stephani, E. Herlt, and M. MacCallum, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980), Chap. 20.
[2]A. B. Evans, J. Phys. A **10**, 1303 (1977).
[3]D. Kramer, Class. Quantum Grav. **5**, 393 (1988).
[4]A. F. da F. Teixeira and I. Wolk, Nuovo Cimento B **41**, 387 (1977).

A spherical collapse solution with neutrino outflow

E. N. Glass
Physics Department, University of Windsor, Windsor, Ontario N9B 3P4, Canada

(Received 19 September 1989; accepted for publication 21 March 1990)

A three-parameter family of solutions of Einstein's field equations is given that represents a collapsing perfect fluid with outgoing neutrino flux. Solutions with "naked" singularities are exhibited. They can be forbidden by requiring pressure less than or equal to the density as a condition of cosmic censorship.

I. INTRODUCTION

In this work, Wesson's[1] spherical, time-dependent, perfect-fluid solution is extended to include outgoing neutrino flux. The neutrino intrinsic angular momentum is neglected, and it is assumed that there is no microscopic interaction between the fluid matter and neutrinos. The neutrino flux interacts gravitationally with the fluid and affects the character of the fluid collapse.

The energy–momentum tensor for the neutrino flux is chosen to be $fl^\mu l^\nu$, where l^μ is an outgoing principal null vector of the type D Weyl tensor belonging to the space-times considered here. This form is exact[2] for a neutrino field adapted to l^μ, when l^μ is geodesic, shear-free, and twist-free (the case here). More generally, one can regard $fl^\mu l^\nu$ as a short-wavelength or "geometrical optics" limit.

In Sec. II the metric and relevant field equations are presented, and in Sec. III a three-parameter family of solutions is given. A set of first-order equations (similar to those given by Misner[3]) for the pressure, density, and mass is given in Sec. IV. In Sec. V, restrictions are placed on the three parameters of the solution such that it represents a physically valid system undergoing gravitational collapse. Proof that pressure less than or equal to the density implies the presence of a horizon is given in Sec. VI. In the discussion that follows it is noted that, when the condition that the pressure be smaller than the density is relaxed, the solution admits "naked" singularities. We conclude that a necessary condition for cosmic censorship is a restriction on the equation of state such that the pressure be less than or equal to the density. The nonzero components of the Einstein tensor are given in Appendix A and, for completeness, the condition for matching the space-time here to the Vaidya space-time is developed in Appendix B.

II. METRIC AND FIELD EQUATIONS

We consider a space-time with spherically symmetric metric

$$g_{\mu\nu}\,dx^\mu\,dx^\nu = A^2\,dt^2 - B^2\,dr^2 - R^2\,d\Omega^2, \tag{1}$$

where $A = A(r,t)$, $B = B(r,t)$, $R = R(r,t)$, and where $d\Omega^2$ is the metric of the unit sphere. The energy–momentum tensor for a perfect fluid with neutrino flux is given by ($G = c = 1$)

$$T^{\mu\nu} = wu^\mu u^\nu - p(g^{\mu\nu} - u^\mu u^\nu) + fl^\mu l^\nu, \tag{2}$$

where p is the isotropic pressure and w is the fluid mass–energy density. Time is comoving with $u^\mu\partial_\mu = A^{-1}\partial_t$. Here, l^μ is an outgoing null vector $l^\mu\partial_\mu = A^{-1}\partial_t + B^{-1}\partial_r$, and f is the neutrino energy density in the fluid rest frame. In the following, primes denote $\partial/\partial r$ and overdots denote $\partial/\partial t$.

The field equation (see Appendix A) arising from pressure isotropy is

$$\begin{aligned} &A^{-2}\left[\frac{\ddot R}{R} - \frac{\ddot B}{B} + \frac{\dot A\dot B}{AB} + \frac{\dot R}{R}\left(\frac{\dot R}{R} - \frac{\dot A}{A} - \frac{\dot B}{B}\right)\right] \\ &+ B^{-2}\left[\frac{R''}{R} + \frac{A''}{A} - \frac{A'}{A}\frac{B'}{B}\right. \\ &\left. - \frac{R'}{R}\left(\frac{R'}{R} + \frac{A'}{A} + \frac{B'}{B}\right)\right] + R^{-2} \\ &+ 2A^{-1}B^{-1}\left[\frac{\dot R'}{R} - \frac{A'}{A}\frac{\dot R}{R} - \frac{\dot B R'}{BR}\right] = 0. \end{aligned} \tag{3}$$

The neutrino energy density is given by

$$8\pi f = 2A^{-1}B^{-1}\left[\frac{\dot R'}{R} - \frac{A'}{A}\frac{\dot R}{R} - \frac{\dot B R'}{BR}\right]. \tag{4}$$

The fluid density and pressure are given, respectively, by

$$\begin{aligned} 8\pi w = {} & A^{-2}\left[\left(\frac{\dot R}{R}\right)^2 + 2\frac{\dot B\dot R}{BR}\right] \\ & - B^{-2}\left[2\frac{R''}{R} + \left(\frac{R'}{R}\right)^2 - 2\frac{B'}{B}\frac{R'}{R}\right] \\ & + R^{-2} - 8\pi f, \end{aligned} \tag{5}$$

$$\begin{aligned} 8\pi p = {} & B^{-2}\left[\left(\frac{R'}{R}\right)^2 + 2\frac{A'}{A}\frac{R'}{R}\right] \\ & - A^{-2}\left[2\frac{\ddot R}{R} + \left(\frac{\dot R}{R}\right)^2 - 2\frac{\dot A\dot R}{AR}\right] \\ & - R^{-2} - 8\pi f. \end{aligned} \tag{6}$$

III. THE SOLUTION

A three-parameter family (α, k_0, k_1) of solutions of the pressure isotropy equation (3) is given by

$$A = r/t_0, \quad B = h(t), \quad R = rS(t), \tag{7}$$

where t_0 scales the time coordinate, and

$$h = 1 + \alpha t/t_0, \tag{8a}$$

$$S^2 = k_0 h^{2+2/\alpha} + k_1 h^{-2/\alpha} + \tfrac{1}{2}(1+\alpha)^{-1}h^2, \quad \alpha \neq -1, \tag{8b}$$

 0022-2488/90/081974-04$03.00

$$S^2 = k_0 + k_1 h^2 - h^2 \ln(h), \quad \alpha = -1. \tag{8c}$$

Neutrino flux is present when the parameter α is nonzero. In the limit $\alpha \to 0$, $B \to B_w$, and $S \to S_w$ where

$$B_w = 1, \quad S_w^2 = k_0 e^{2t/t_0} + k_1 e^{-2t/t_0} + \tfrac{1}{2}.$$

The $\alpha \to 0$ limit is Wesson's perfect-fluid solution[1] with the equation of state $p = w$.

The neutrino energy density is

$$4\pi f = -\alpha R^{-2} S^2 h^{-2}. \tag{9}$$

The mass [see Eq. (18) below] is given by

$$\begin{aligned} 2m = {} & RS^{-2}(2+\alpha)\left[\alpha k_0^2 h^{2+4/\alpha}\right. \\ & + \alpha(1+\alpha)^{-1} k_0 h^{2+2/\alpha} \\ & \left. + \tfrac{1}{4}(2+\alpha)^{-1} h^2 - 2k_0 k_1\right], \quad \alpha \neq -1, \end{aligned} \tag{10}$$

$$\begin{aligned} 2m = {} & RS^{-2}\left[k_0 - k_0^2 h^{-2}\right. \\ & \left. - 2k_0 k_1 + 2k_0 \ln(h) + h^2/4\right], \quad \alpha = -1. \end{aligned}$$

For this solution, the mass, pressure, and fluid density are related by

$$2m = 4\pi(p+w)R^3, \tag{11}$$

where the pressure and density are given, respectively, for $\alpha \neq -1$ by

$$\begin{aligned} 8\pi p = {} & -R^{-2} S^{-2}\left[\alpha(\alpha+2) k_0^2 h^{2+4/\alpha} + \alpha k_0 h^{2+2/\alpha}\right. \\ & + \alpha k_1 h^{-2/\alpha} + \tfrac{1}{4}(\alpha-1)(1+\alpha)^{-1} h^2 \\ & \left. + 2(\alpha+1)(\alpha+2) k_0 k_1\right], \end{aligned} \tag{12}$$

$$\begin{aligned} 8\pi w = {} & R^{-2} S^{-2}\left[3\alpha(\alpha+2) k_0^2 h^{2+4/\alpha} + \alpha(5+3\alpha)\right. \\ & \times (1+\alpha)^{-1} k_0 h^{2+2/\alpha} + \alpha k_1 h^{-2/\alpha} \\ & + \tfrac{1}{4}(3\alpha+1)(1+\alpha)^{-1} h^2 \\ & \left. + 2(\alpha-1)(\alpha+2) k_0 k_1\right], \end{aligned} \tag{13}$$

and for $\alpha = -1$ by

$$\begin{aligned} 8\pi p = {} & R^{-2} S^{-2}\left[k_0^2 h^{-2} + k_1 h^2\right. \\ & \left. + 2k_0 - h^2 \ln(h) + h^2/4\right], \end{aligned} \tag{12'}$$

$$\begin{aligned} 8\pi w = {} & R^{-2} S^{-2}\left[4k_0 \ln(h) - 3k_0^2 h^{-2} - k_1 h^2 - 4k_0 k_1\right. \\ & \left. + h^2 \ln(h) + h^2/4\right]. \end{aligned} \tag{13'}$$

The expressions for p and w make clear the presence of a curvature singularity at $R = 0$.

In the limit $\alpha \to 0$ the pressure and density equalize to

$$8\pi p = 8\pi w = R_w^{-2} S_w^{-2}\left(\tfrac{1}{4} - 4k_0 k_1\right),$$

with $R_w = rS_w$.

IV. DYNAMICAL EQUATIONS

It is interesting to observe the effect of the neutrino flux on the changes in pressure and density. To that end, we write the dynamical equations

$$\nabla_\nu T^{\mu\nu} = 0, \tag{14}$$

and the mass change equations. These equations, with an equation of state, form a first-order differential set equivalent to Eqs. (3)–(6). We will write u_μ and n_μ components of $T^{\mu\nu}{}_{;\nu} = 0$, where $n_\mu\, dx^\mu = B\, dr$ is a spacelike unit vector normal to $r = \text{const}$ surfaces. It is useful to first establish the components of

$$e^\mu = \nabla_\nu (f l^\mu l^\nu),$$

where $l^\mu = u^\mu - n^\mu$. Direct calculation yields $e^\mu u_\mu = e^\mu n_\mu = E$, where

$$E = \frac{\dot f}{A} + \frac{f'}{B} + 2f\left[A^{-1}\left(\frac{\dot B}{B} + \frac{\dot R}{R}\right) + B^{-1}\left(\frac{A'}{A} + \frac{R'}{R}\right)\right]. \tag{15}$$

Equation (15) can alternatively be written in terms of the neutrino flux $N = 4\pi R^2 f$ as

$$4\pi R^2 E = l^\mu \nabla_\mu N - 2N l^\mu l^\nu \nabla_\mu u_\nu,$$

where E gives the rate at which neutrino outflow removes energy from the system. The u_μ component of Eq. (14) is

$$\dot w + (p+w)(\dot B/B + 2\dot R/R) + AE = 0, \tag{16}$$

and the n_μ component is

$$p' + (p+w)(A'/A) + BE = 0. \tag{17}$$

The mass is given by

$$2m = R\left[1 + A^{-2}(\dot R)^2 - B^{-2}(R')^2\right], \tag{18}$$

where

$$m' = 4\pi R^2 [R' w + Bf(\dot R/A + R'/B)], \tag{19}$$

$$\dot m = -4\pi R^2 [\dot R p + Af(\dot R/A + R'/B)]. \tag{20}$$

The last terms in both Eqs. (19) and (20) represent outgoing neutrino energy with $l^\mu \nabla_\mu R = \dot R/A + R'/B$.

V. PHYSICAL PARAMETERS

The values of α, k_0, and k_1 must be restricted to yield physically valid solutions. For the neutrino energy density to be positive, Eq. (9) implies that α be negative. The fluid density and pressure must both be positive [with Eq. (11) then assuring positive mass], and the pressure less than or equal to the density (causal sound speed). To describe collapse rather than expansion, we require the fluid rate of expansion θ to be negative, where

$$\theta = \nabla_\mu u^\mu = A^{-1}(\dot B/B + 2\dot R/R). \tag{21}$$

Finally, R must be real and so S^2 must be positive. Thus we require the solution to have parameters such that

$$f > 0, \quad p > 0, \quad w > 0, \quad p \leqslant w, \quad \theta < 0, \quad S^2 > 0. \tag{22}$$

We will call a parameter set "valid" when restrictions (22) are satisfied at $t = 0$. A computer search through the values of the parameter set (α, k_0, k_1), using Eqs. (8)–(13) at $t = 0$, reveals many valid sets (there are no valid sets for $\alpha = -2$). Some examples are given in Table I, where solution (7) implies that all the physical quantities listed are independent of the r coordinate, and so hold for the entire spacelike surface at $t = 0$. Rows 1 and 2 of Table I do not contain valid sets.

VI. HORIZONS

The fluid rate of expansion θ can be written as

$$R\theta = \alpha S/h + 2t_0 \dot S. \tag{23}$$

Since α must be negative, $\dot S$ need not be negative for $\theta < 0$,

TABLE I. Parameter sets (α,k_0,k_1) and physical values at $t=0$. All physical quantities are dimensionless ($c=G=1$) and independent of coordinate r.

α	k_0	k_1	$4\pi R^2 p$	$4\pi R^2 w$	$2m/R$	$R\theta$	$4\pi R^2 f$
−0.1	0.00	0.1	0.24	0.14	0.38	−0.5	0.07
−0.8	−0.41	0.0	0.50	0.49	0.98	−4.0	1.67
−0.8	−0.42	0.0	0.50	0.51	1.01	−4.0	1.66
−1.0	−1.00	1.5	0.75	1.75	2.50	−3.5	0.50
−1.0	−1.00	100.0	0.50	1.50	2.00	−30.0	99.00
−1.0	−100.00	1.0E10	0.50	199.50	200.00	−3.0E05	1.0E10
−3.0	0.50	0.0	0.50	0.50	1.00	−2.5	0.75
−10.0	3.00	−2.0	81.37	551.37	632.74	−60.0	9.44
−10.0	1000.00	−800.0	8.8E04	2.5E05	3.3E05	−1301.1	1999.44
−100.0	3.00	−2.0	1.4E04	7.3E04	8.7E04	−690.2	99.49
−100.0	100.00	−52.0	3.0E04	2.0E06	2.0E06	−3535.6	4799.50

but when $\dot{S}<0$ then θ is always negative. Thus $\dot{S}<0$ is a sufficient condition for $\theta<0$.

To find an expression for the difference between p and w, Eq. (5) is subtracted from Eq. (6) and the isotropy equation (3) is substituted along with solution (7), yielding

$$2\pi(p-w)R^2=(-\alpha S/h)(t_0\dot{S}+S/h). \tag{24}$$

Since S/h is positive, the condition for $p\leqslant w$ is

$$\dot{S}<0, \quad |t_0\dot{S}|\geqslant S/h. \tag{25}$$

The mass equation (18) can be written as

$$2m/R=1+(t_0\dot{S})^2-(S/h)^2. \tag{26}$$

When $p\leqslant w$, condition (25) implies that $2m/R\geqslant 1$. Thus solutions with an equation of state restricted to $p\leqslant w$ are always collapsing and always within a trapped surface. Since $2m/R$ is independent of r, the trapped surface is a horizon.

VII. DISCUSSION

All the valid parameter sets maintain $2m\geqslant R$. Thus each valid set evolves within a horizon which shields the singularity at $R=0$ (cosmic censorship).

In rows 1 and 2 of Table I, the condition $p\leqslant w$ has been relaxed. Ultrabaric pressure allows $R>2m$ at $t=0$, and consequently those two space-times exhibit a "naked" singularity. (There is an unresolved dispute[4,5] over the physical necessity of the condition $p\leqslant w$ for equilibrium configurations.) The parameter set on row 1 determines a spacetime which maintains $p>w$ and $R>2m$ for the entire time evolution leading up to $h=0$, where the description of the evolution breaks down. At that time, the r coordinate loses its meaning since $g^{\mu\nu}r_{,\mu}r_{,\nu}=-h^{-2}$. (A coordinate labelling null surfaces can be used instead of t to continue the evolution.[8])

The space-time associated with the parameter set on row 2 has a different evolution. At $t/t_0=0.05$, $2m/R=1.09$, $4\pi R^2 p=0.51$, and $4\pi R^2 w=0.58$. Therefore, early in the evolution a horizon appears and the pressure falls to less than the density. The remainder of the evolution maintains $p<w$ and takes place within the horizon.

We conclude, because of the examples of "naked" singularities given above, that a necessary condition for cosmic censorship is a restriction on the equation of state such that $p\leqslant w$.

APPENDIX A: EINSTEIN TENSOR

Sign conventions are fixed by Ricci's identity $2u_{\nu;[\alpha\beta]}\equiv u_\mu R^\mu{}_{\nu\alpha\beta}$, and $R_{\mu\nu}=R^\alpha{}_{\mu\nu\alpha}$. The field equations are $G^\mu{}_\nu=-8\pi T^\mu{}_\nu$. The nonzero components of the Einstein tensor for metric (1) are given by

$$G^t{}_t=-A^{-2}\left[\left(\frac{\dot{R}}{R}\right)^2+2\frac{\dot{B}\dot{R}}{BR}\right]+B^{-2}\left[2\frac{R''}{R}+\left(\frac{R'}{R}\right)^2-2\frac{B'}{B}\frac{R'}{R}\right]-R^{-2}, \tag{A1}$$

$$G^t{}_r=2A^{-2}\left[\frac{\dot{R}'}{R}-\frac{A'}{A}\frac{\dot{R}}{R}-\frac{\dot{B}R'}{BR}\right], \tag{A2}$$

$$G^r{}_t=-A^2B^{-2}G^t{}_r, \tag{A3}$$

$$G^r{}_r=-A^{-2}\left[2\frac{\ddot{R}}{R}+\left(\frac{\dot{R}}{R}\right)^2-2\frac{\dot{A}\dot{R}}{AR}\right]+B^{-2}\left[\left(\frac{R'}{R}\right)^2+2\frac{A'}{A}\frac{R'}{R}\right]-R^{-2}, \tag{A4}$$

$$G^\vartheta{}_\vartheta=G^\varphi{}_\varphi=-A^{-2}\left[\frac{\ddot{R}}{R}+\frac{\ddot{B}}{B}-\frac{\dot{A}\dot{B}}{AB}-\frac{\dot{R}}{R}\left(\frac{\dot{A}}{A}-\frac{\dot{B}}{B}\right)\right]+B^{-2}\left[\frac{R''}{R}+\frac{A''}{A}-\frac{A'}{A}\frac{B'}{B}+\frac{R'}{R}\left(\frac{A'}{A}-\frac{B'}{B}\right)\right]. \tag{A5}$$

APPENDIX B: MATCH TO VAIDYA

The interior metric (1) can be matched[6,7] to an exterior Vaidya metric given by

$$ds^2=[1-2M(\tau)/y]d\tau^2+2\,d\tau\,dy-y^2\,d\Omega^2. \tag{B1}$$

The boundary three-surface b is expressed in interior coordinates by

$$r=r_b, \quad r_b=\text{const}$$

with unit spacelike normal

$$n_\mu^-\,dx^\mu=B_b\,dr. \tag{B2}$$

[Subscript b indicates evaluation on boundary b, i.e., $B_b=B(r_b,t)$. Interior objects will be indicated by a minus and exterior objects by a plus.] In the exterior region the boundary equation is

$$y = y_b(\tau),$$

with unit normal

$$n_\mu^+ \, dx^\mu = \left[2\left(\frac{dy_b}{d\tau}\right) + 1 - \frac{2M_b}{y_b}\right]^{-1/2}\left[dy - \left(\frac{dy_b}{d\tau}\right)d\tau\right]. \tag{B3}$$

The intrinsic metric on b is

$$ds_b^2 = dT^2 - R_b^2 \, d\Omega^2. \tag{B4}$$

The junction conditions on b are equality of the interior and exterior first and second fundamental forms. For the interior

$$ds_-^2 = A_b^2 \, dt^2 - R_b^2 \, d\Omega^2, \tag{B5}$$

and for the exterior

$$ds_+^2 = \left[2\left(\frac{dy_b}{d\tau}\right) + 1 - \frac{2M_b}{y_b}\right]d\tau^2 - y_b^2 \, d\Omega^2. \tag{B6}$$

Demanding $ds_-^2 = ds_b^2 = ds_+^2$ yields

$$A_b \, dt = dT,$$
$$R_b(t) = y_b(\tau), \tag{B7}$$
$$\frac{d\tau}{dT} = \left[2\left(\frac{dy_b}{d\tau}\right) + 1 - \frac{2M_b}{y_b}\right]^{-1/2}.$$

The unit normal (B3) can now be written as

$$n_\mu^+ \, dx^\mu = \left(\frac{d\tau}{dT}\right)dy - \left(\frac{dy_b}{dT}\right)d\tau. \tag{B8}$$

The mass $m(r,t)$ within $r = \text{const}$, $t = \text{const}$ two-surfaces is given by

$$2m = R\,[1 + g^{\mu\nu}(\partial_\mu R)(\partial_\nu R)],$$

and $M(\tau)$ within $\tau = \text{const}$, $y = \text{const}$ two-surfaces is given by

$$2M = y[1 + g_+^{\mu\nu}(\partial_\mu y)(\partial_\nu y)].$$

It follows from (B7) and continuity of the second fundamental form that

$$M_b(\tau) = m(r_b, t). \tag{B9}$$

Junction conditions equivalent to continuity of the first and second fundamental forms are

$$(G_{\mu\nu} n^\mu n^\nu)_b^+ = (G_{\mu\nu} n^\mu n^\nu)_b^-, \tag{B10}$$
$$(G_{\mu\nu} n^\mu e_j^\nu)_b^+ = (G_{\mu\nu} n^\mu e_j^\nu)_b^-, \tag{B11}$$

where three linearly independent e_j^ν span the boundary surface. For the interior

$$G_{\mu\nu}^- = -8\pi[(w+p)u_\mu u_\nu - p g_{\mu\nu} + f l_\mu l_\nu], \tag{B12}$$

where $l_\mu \, dx^\mu = A \, dt - B \, dr$, and for the exterior

$$G_{\mu\nu}^+ = F L_\mu L_\nu, \tag{B13}$$

where $L_\mu \, dx^\mu = d\tau$. (The luminosity observed at infinity is $-\frac{1}{2}y^2 F = -dM/d\tau$.) Equations (B8) and (B13) yield

$$(G^{\mu\nu} n_\mu n_\nu)_b^+ = F_b\left(\frac{d\tau}{dT}\right)^2,$$

and Eqs. (B2) and (B12) yield

$$(G^{\mu\nu} n_\mu n_\nu)_b^- = -8\pi(p_b + f_b).$$

It follows from condition (B10) that

$$8\pi(p_b + f_b) = -F_b\left(\frac{d\tau}{dT}\right)^2. \tag{B14}$$

This relation expresses conservation of momentum across the boundary.

In the exterior region, a timelike unit vector v^μ orthogonal to n_μ^+ is given by

$$v_+^\mu \, \partial_\mu = \left(\frac{d\tau}{dT}\right)\partial_\tau + \left(\frac{dy_b}{dT}\right)\partial_y. \tag{B15}$$

The exterior and interior vectors v^μ and u^μ coincide on b:

$$(u_\mu \, dx^\mu)_b^- = (v_\mu \, dx^\mu)_b^+ = dT.$$

Equations (B2) and (B12) yield

$$(G^\mu{}_\nu n_\mu u^\nu)_b^- = -8\pi f_b. \tag{B16}$$

Equations (B8) and (B13) yield

$$(G^\mu{}_\nu n_\mu v^\nu)_b^+ = F_b\left(\frac{d\tau}{dT}\right)^2. \tag{B17}$$

It follows from Eq. (B14) and the equality of (B16) and (B17) that

$$p_b = 0. \tag{B18}$$

[1] P. S. Wesson, J. Math. Phys. **19**, 2283 (1978).
[2] J. Wainwright, J. Math. Phys. **12**, 828 (1971).
[3] C. W. Misner, Phys. Rev. **137**, B 1360 (1965).
[4] E. N. Glass, Phys. Rev. D **28**, 2693 (1983).
[5] G. Caparaso and K. Brecher, Phys. Rev. D **28**, 2694 (1983).
[6] P. C. Vaidya, Nature **171**, 260 (1953).
[7] The first match of an interior to the Vaidya metric is given in J. L. Synge, Proc. R. Ir. Acad. Sec. A **59**, 1 (1957). In that work an incoherent shell of radiation is joined to Vaidya. The first correct derivation of the matching condition for a collapsing fluid with radial heat flow to Vaidya is given in N. O. Santos, Mon. Not. R. Astron. Soc. **216**, 403 (1985).
[8] The t coordinate can be replaced by the null coordinate $u = -\alpha^{-1} r^{-\alpha}(1 + \alpha t/t_0)$. Evalualtion of $2m/R$ at constant u reveals the naked singularity.

Numerical procedures for sample structures on stochastic differential equations

Hiroshi Nakazawa
Department of Physics, Kyoto University, Kyoto 606, Japan

(Received 19 September 1989; accepted for publication 28 March 1990)

Procedures are discussed for the numerical construction of sample processes on stochastic differential equations. Structures of a relevant class of Gaussian white-noise functionals are analyzed to some closed forms, and new numerical algorithms for single noise problems are presented up to the accuracy of $O(h^3)$, or within errors of $O(h^{7/2})$, in the sense of the mean-square norm.

I. INTRODUCTION

Stochastic differential equations (SDEs) generate continuous Markov processes,[1–6] and give versatile representations in a variety of physical problems[4,5,7–11] that involve additive as well as multiplicative noise terms. Computational accesses to these problems have also been established by several numerical algorithms,[12–21] that have so far been constructed to give the highest efficiency in reproducing expectation values associated with the solution processes. Recent physical trends show, however, the enhanced interest in high-precision sample structures, raising problems that seem to call for some new approaches. We discuss aspects of these problems to show possibilities as well as difficulties in the numerical integration procedures on SDEs for sample structures and high precision.

The SDEs of our interest relate realizations of the noise to those of the solution processes. In a time step h, the increment of the solution process is composed of a set of noise functionals that are characteristic to the numerical algorithm chosen. The strategies of the existing algorithms are almost unanimous in constructing the set of approximating random variables that reproduce relevant expectation values, but branch off to two separate groups in the way of their implementation. In one line, originated by Rao *et al.*[13] for single noise problems, one constructs approximate noise functionals so as for all expectation values associated with the increment to be reproduced accurately to $O(h^2)$ or within errors of $O(h^3)$. These approximate noise functionals are then substituted into the precise expressions constituting the increment to a preassigned order in h. The algorithm gives

$$\lim_{h\to 0} \text{Prob}\{|\text{one time step error}| > \epsilon\}/h^2 = 0, \quad \forall \epsilon > 0,$$

by the Tchebychev inequality

$$\text{Prob}\{|A| > c\} \leqslant \langle A^2 \rangle / c^2, \quad \forall c > 0, \tag{1}$$

with $\langle \cdots \rangle$ for expectation values.

In the other line, originated notably by Mil'shtein,[14] one constructs the process increment as a whole[14–21] to a prescribed degree of accuracy in h. This strategy is often realized by Runge–Kutta type formulations, and culminates in the algorithms of Helfand[18] and of Greenside and Helfand[19] that give expectation values correctly to $O(h^3)$ or $O(h^4)$ for additive noise cases. Even the difficult problems with multiplicative, multiple noises are now under the control of the $O(h^2)$-accurate algorithms of Klauder and Petersen[20] and of Mil'shtein[21] in this line for expectation values.

The computational efficiency and the powerful development of the latter strategy are conspicuous on comparison. As (1) indicates typically, the accuracy in sample structures is related to the accuracy of expectation values (though with intrications and distances). There are also a few established statements in this regard.[22,20] It is therefore possible in principle to use the algorithms in the latter line to obtain sample structures (cf. Sec. IV further). However, these methods in their original use suffer from one serious drawback in practice in that they intrinsically link the noise-solution interrelation with the h chosen. Hence their solution processes for different h cannot refer to the same realization of the noise in a simple way. An accuracy test of the adopted scheme for the choice of the optimal h or of the very scheme itself, would require long time or ensemble averages. This is not only circuitous and inefficient but also can form a true hindrance when sample structures are of primary importance or the system of SDEs is large.[23]

The points remind us anew of a merit in the method of Rao *et al.* that was in fact set forth to give sample structures. By its nature the method preserves the exact noise-solution interrelation. With the noise functionals obtained samplewise on some fine time points, the method constructs sample solutions, for varied choices of h or the algorithm, that correspond to the same realization of this noise process. We shall then be able to infer firmly on the optimal scheme or h for long time runs, with only a few sample runs just as in the cases of ordinary differential equations (ODEs). The informations so obtained will also have value for the use in Runge–Kutta type procedures. The problems in this plot are, as stated by many authors, whether the necessary noise functionals can be obtained with sufficient accuracy at a reasonable numerical cost, and whether the intrications arising in high precision schemes could be reduced to a manageable size.

This paper shows that they are surmountable in principle, with the analysis of a pertinent class of Gaussian white-noise functionals. We shall find useful structures in these functionals, in the general point of view established by Wiener[24] and Itô.[25] The knowledge will give ways of approximations for SDEs, ending in the procedures for single noise problems up to $O(h^3)$ accuracy in the root-mean-square

 0022-2488/90/081978-13$03.00

sense and revealing another merit of the Rao–Borwankar–Ramkrishna formulation, the implementability of the main procedure in forms independent of the specific SDE chosen.

The paper is constructed as follows. Section II recapitulates the perturbation expansion procedure on SDEs, taking a way[26] to manage intrications and identifying necessary noise functionals to be analyzed. The structures of these functionals will be clarified in the central Sec. III by the case of scalar noise SDEs. The results will be summarized in Sec. IV to construct algorithms that extend the Rao–Borwankar–Ramkrishna scheme. An application will give insights into the possibilities and the limitations of the methods by an explicit comparison of accuracy levels for choices of h. Extensions to more general SDEs will also be discussed, including some observations on multiplicative multiple noise problems.

Throughout this work Itô's definition will be adopted for stochastic integrals by its fundamental facilities. We would say that a stochastic process $x(t)$ has an increment $x(t+h)-x(t)$ of magnitude $O(h^p)$, implying

$$\lim_{h\to 0} \text{Prob}\{|x(t+h)-x(t)| > h^{p-\delta}\} = 0, \quad \forall \delta > 0. \tag{2}$$

If $y(t)$ is an approximation to this $x(t)$ and if the error $e(t)=x(t)-y(t)$ is a process with an increment of $O(h^{p+1/2})$ or smaller, then we would state that $y(t)$ is the pth degree accurate or has a $(p+\frac{1}{2})$th or higher error. By the Tchebychev inequality (1) this is ensured by $\langle e^2(h)\rangle = O(h^{2p})$. We denote M_2 for the space of random variables with finite second moments generated from a Wiener process $B(t)$ (or such processes) for $t\in T\equiv[0,h]$. The M_2 norm is $\|A\| = \langle A^2\rangle^{1/2}$. Our aim is the procedures accurate up to $O(h^3)$ in this norm. We shall always assume that stochastic processes to arise are adapted to the Wiener process(es), with all necessary moments bounded uniformly in t.

II. PERTURBATION EXPANSION

Let $B(t)$ denote exclusively a standard Wiener process characterized by

$$B(0)=0, \quad \langle B(t)\rangle = 0, \quad \langle B(s)B(t)\rangle = \min(s,t). \tag{3}$$

Consider an Itô SDE,[1-3,6]

$$dx(t) = a[x(t),t]dt + b[x(t),t]dB(t), \tag{4}$$

where functions $a(x,t)$ and $b(x,t)$ are assumed to have continuous partial derivatives to any necessary degrees and give solutions of (4) without explosion to times of our concern. The Itô stochastic integral equation (SIE)

$$x(t) = x_0 + \int_{t_0}^{t} a[x(s),s]ds + \int_{t_0}^{t} b[x(s),s]dB(s), \tag{5}$$

for $t_0 \leqslant t \leqslant t_0 + h$ with x_0 independent of $B(t)-B(t_0)$, poses the problem of the integration of (4) in the time step $t_0 \to t_0 + h$. Since $B(t+t_0)-B(t_0)$ is another standard Wiener process, we take hereafter $t_0 = 0$ in (5) by redefinitions of $a(x,t)$, $b(x,t)$, and $B(t)$. The integral $\int_0^t b[x(s),s]dB(s)$ is $O(b[x(t),t])\times O(h^{1/2})$ by the Tchebychev inequality (1) and a fundamental property of Itô stochastic integrals

$$\left\langle \int \phi(s)dB(s) \int \psi(t)dB(t) \right\rangle = \int \langle \phi(s)\psi(s)\rangle ds. \tag{6}$$

In contrast, the drift term $\int_0^t a[x(s),s]ds$ is $O(a[x(t),t])\times O(h)$. Itô SIE (5) admits perturbation expansions in powers of $h^{1/2}$, as is well known.

An $O(h^n)$ procedure on SDE's thus corresponds to a far more difficult $O(h^{2n})$ problem on ODE's. A systematic process of analysis is important.[27] We adopt the explicit ϵ-expansion method. Introduce a parameter $\epsilon > 0$ to be put eventually $\epsilon = 1$ and take another SIE for $0 \leqslant t \leqslant h$,

$$x(t) = x_0 + \epsilon \int_0^t b[x(s),s]dB(s) + \epsilon^2 \int_0^t a[x(s),s]ds. \tag{5'}$$

Assuming the expansion

$$x(t) = x_0(t) + \epsilon x_1(t) + \cdots + \epsilon^n x_n(t) + O(\epsilon^{n+1}), \quad x_0(t) \equiv x_0, \quad n \geqslant 2, \tag{7}$$

substituting (7) into (5'), taking formal Taylor expansions of $a[x(s),s]$ and $b[x(s),s]$ about (x_0,s) and comparing both sides, we obtain a system of equations:

$$x_0(t) \equiv x_0, \quad x_1(t) = \int_0^t b(x_0,s)dB(s),$$

$$x_2(t) = \int_0^t b_x(x_0,s)x_1(s)dB(s) + \int_0^t a(x_0,s)ds, \tag{8a}$$

$$x_n(t) = \int_0^t b^{(n-1)}(s)dB(s) + \int_0^t a^{(n-2)}(s)ds, \quad n \geqslant 2. \tag{8b}$$

Here, $a^{(p)}(s)$ and $b^{(p)}(s)$, respectively, denote coefficients of ϵ^p in $a[x(s),s]$ and $b[x(s),s]$ for $x(s)$ of (7), and all stochastic integrals in (8a) and (8b) are in Itô's sense.

Equations (8a) and (8b) may be solved recursively to obtain explicit forms of $\{x_k(t)\}$. To elucidate the process we take for the time being the time homogeneous SDE (4) or SIE (5) with $a(x,s)$ and $b(x,s)$ independent of s. Introduce

$$a_p \equiv \frac{d^p a(x_0)}{dx_0^p}, \quad b_p \equiv \frac{d^p b(x_0)}{dx_0^p},$$

$$I_{ij}(t) \equiv \int_0^t s^i B^j(s)dB(s), \quad i,j \geqslant 0. \tag{9}$$

The corresponding solutions of (8a) and (8b) are summarized below:

$$x_0(t) = x_0,$$
$$x_1(t) = b_0 B(t),$$
$$x_2(t) = (\tfrac{1}{2}b_0 b_1)B^2(t) + (a_0 - \tfrac{1}{2}b_0 b_1)t,$$
$$x_3(t) = b_0(\tfrac{1}{6}b_0 b_2 + \tfrac{1}{6}b_1^2)B^3(t) + b_0(a_1 - \tfrac{1}{2}b_1^2 - \tfrac{1}{2}b_0 b_2)B(t)t + (a_0 b_1 - a_1 b_0 - \tfrac{1}{2}b_0^2 b_2)I_{10}(t),$$
$$x_4(t) = b_0(\tfrac{1}{24}b_0^2 b_3 + \tfrac{1}{24}b_1^3 + \tfrac{1}{6}b_0 b_1 b_2)B^4(t) + b_0(\tfrac{1}{2}a_1 b_1 - b_0 b_1 b_2 - \tfrac{1}{4}b_1^3 - \tfrac{1}{4}b_0^2 b_3 + \tfrac{1}{2}a_2 b_0)B^2(t)t$$
$$+ (a_1 b_0 b_1 + a_0 b_0 b_2 + \tfrac{1}{2}b_0^2 b_1 b_2 - a_0 b_1^2 + \tfrac{1}{2}b_0^3 b_3 - a_2 b_0^2)I_{11}(t) - b_1(a_1 b_0 - a_0 b_1 - \tfrac{1}{2}b_0^2 b_2)I_{10}(t)B(t)$$
$$+ (\tfrac{1}{2}a_0 a_1 + \tfrac{1}{4}b_0^2 b_1 b_2 - \tfrac{1}{2}a_0 b_1^2 + \tfrac{1}{8}b_0 b_1^3 + \tfrac{1}{8}b_0^3 b_3 - \tfrac{1}{4}a_2 b_0^2)t^2,$$
$$x_5(t) = c_5^{(5)}B^5(t) + c_3^{(5)}B^3(t)t + c_1^{(5)}B(t)t^2 + d_{20}^{(5)}I_{20}(t) + d_{12}^{(5)}I_{12}(t) + e_{11}^{(5)}I_{11}(t)B(t) + e_{10}^{(5)}I_{10}(t)B^2(t) + g_{10}^{(5)}I_{10}(t)t,$$
$$x_6(t) = c_6^{(6)}B^6(t) + c_4^{(6)}B^4(t)t + c_2^{(6)}B^2(t)t^2 + d_{13}^{(6)}I_{13}(t) + d_{21}^{(6)}I_{21}(t) + e_{12}^{(6)}I_{12}(t)B(t)$$
$$+ e_{20}^{(6)}I_{20}(t)B(t) + e_{11}^{(6)}I_{11}(t)B^2(t) + e_{10}^{(6)}I_{10}(t)B^3(t) + f_3^{(6)}t^3 + g_{11}^{(6)}I_{11}(t)t + h_{10}^{(6)}[I_{10}(t)]^2 + i_{10}^{(6)}I_{10}(t)B(t)t. \tag{10}$$

The derivation will be described further in the Appendix, with a list of the coefficients in (10) for the important case of linear $b(x)$.

If (10) is substituted into (7) and ϵ is put to 1, approximate one-time-step solutions are obtained in the form

$$x^{(n)}(h) = x_0 + x_1(h) + \cdots + x_n(h), \tag{11}$$

for $2 \leqslant n \leqslant 6$, which are expressed in terms of

$$B(h) = I_{00}(h), \quad I_{10}(h), \quad I_{20}(h),$$
$$I_{11}(h), \quad I_{12}(h), \quad I_{13}(h), \quad I_{21}(h). \tag{12}$$

If $a(x,s)$ and $b(x,s)$ depend on s, their Taylor expansions in s should be taken in (8a) and (8b) and (10) is modified accordingly. The expression of $x^{(4)}(h)$ for such a general case was given by Rao *et al.*[13] as their Eq. (3.2), and (10) reproduces its specialization. The ϵ expansion to $n \leqslant 4$ for time homogeneous (4) is thus justified.

The procedure in fact has a wider domain. Let $\{a(x,s), b(x,s)\}$ satisfy for any relevant p and $(x,s) \in \mathbb{R} \times [0,h]$,

$$\left|\frac{\partial^p a(x,s)}{\partial x^p}\right| \leqslant C[1 + |a(x,s)|]^q, \quad \exists q, C > 0,$$

just as polynomials. Take a C^∞ function $-1 \leqslant \mu(X) \leqslant 1$ with $\mu(X) \equiv X$ for $|X| \leqslant \frac{1}{2}$ and $\equiv \operatorname{sign}(X)$ for $|X| \geqslant 1$. Define $\binom{a_N}{b_N}(x,s) \equiv N\mu[\binom{a}{b}(x,s)/N]$, and

$$x_N(t) = x_0 + \int_0^t a_N[x_N(s),s]ds + \int_0^t b_N[x_N(s),s]dB(s).$$

Finally, assume on (4) the uniform convergence $\|x(t) - x_N(t)\| \to 0$ for $N \to \infty$, $\forall t \in [0,h]$, as one of the computability condition.

Proposition 1: Let SDE (4) generally be time inhomogeneous. (i) For any n and $0 \leqslant t \leqslant h$ $x^{(n)}(t)$ of (11) approximates $x(t)$ of (5) within an error of $O(h^{(n+1)/2})$ in the M_2 norm, provided $a(x_0,s)$, $b(x_0,s)$ and their derivatives give $O(1)$ coefficients of noise functionals in $x^{(n)}(t)$. (ii) The noise functionals of (12) exhaust those that specify the increment $x(h)$ to $O(h^3)$ by (8a) and (8b).

The proof will be presented at the end of the Appendix, showing the applicability of (a) to many component SDEs.

In order to see the basic structure of (10), to measure to some extent the performance levels of algorithms based on it and to motivate the efforts for higher-precision algorithms, a numerical result will now be presented on an example.[15] Consider an SDE for $x(0) = x_0$,

$$dx(t) = (\tfrac{1}{2}b^2)x(t)dt + bx(t)dB(t), \tag{13}$$

with a real constant b. Without loss of generality, $b = x_0 = 1$ may be assumed by substitutions $b^2 t \to t$, $x_0 x \to x$. The exact solution of (13) then takes the form

$$x(t) = \exp[B(t)], \tag{14}$$

by Itô's formula. An induction with (8a) and (8b) or a comparison of the list in the Appendix shows that (10) implies the following algorithm for $n \geqslant 2$, $t = kh$, $k = 1,2,\ldots$ with $x^{(n)}(0) = 1$:

$$x^{(n)}(t+h) = x^{(n)}(t)\left\{\sum_{i=0}^{n}[B(t+h) - B(t)]^i/i!\right\}. \tag{15}$$

Generating standard, normal, and mutually independent random variables (SNIRV's, for short) as increments of $B(t)$, we may compare the exact (14) with (15) for $2 \leqslant n \leqslant 6$ on one and the same realization $\{B(t); t = k\delta, k = 0,1,\ldots\}$ of a Wiener process. Table I shows the result of a run in double precision for $0 \leqslant t \leqslant 32$ and $\delta = 2^{-13} = \frac{1}{8192}$.

III. WHITE-NOISE FUNCTIONALS

The structure of M_2, known by the name of Wiener–Itô decomposition,[24,25,28,4,5] is the basis of our analysis. The relevant facts are summarized in the statements below for later convenience.

Proposition 2: Let $\{\phi_i(t); \quad i = 0,1,\ldots\}$ be a complete orthonormal (CON) set of functions in the real function space $L_2(T)$ with $T \equiv [0,h]$, and define SNIRV's $\{\xi_i\}$ by Wiener integrals

$$\xi_i \equiv \int_0^h \phi_i(s)dB(s), \quad i = 0,1,\ldots. \tag{16}$$

TABLE I. Precision of algorithms for various h on the problem $dx = (x/2)dt + x\,dB$ with $x(0) = 1$.

Procedure	h	$t = 10$	$t = 20$	$t = 30$
exact		0.321 036 946 08	0.185 696 188 23	0.096 766 328 22
$x^{(2)}$	2^{-4}	0.435 029 307 59	0.300 741 105 31	0.201 555 500 18
	2^{-5}	0.387 512 482 43	0.242 154 037 68	0.144 667 882 14
	2^{-6}	0.332 172 726 35	0.202 075 261 34	0.108 618 964 38
	2^{-13}	0.321 343 528 88	0.185 943 927 65	0.096 954 189 48
$x^{(3)}$	2^{-4}	0.297 376 511 30	0.161 565 034 95	0.076 638 050 24
	2^{-5}	0.305 662 475 76	0.172 217 033 27	0.086 127 013 67
	2^{-6}	0.312 870 697 32	0.177 241 892 61	0.090 768 928 73
$x^{(4)}$	2^{-4}	0.322 700 102 50	0.187 007 669 73	0.097 785 453 76
	2^{-5}	0.321 516 065 98	0.186 046 203 30	0.097 047 026 55
	2^{-6}	0.321 016 002 83	0.185 721 392 43	0.096 787 291 74
	2^{-7}	0.321 064 246 58	0.185 706 601 70	0.096 779 919 47
	2^{-8}	0.321 043 763 43	0.185 697 357 52	0.096 768 158 22
	2^{-9}	0.321 037 999 94	0.185 696 472 47	0.096 766 403 62
$x^{(5)}$	2^{-4}	0.320 800 347 53	0.185 485 043 10	0.096 556 649 94
	2^{-5}	0.320 962 973 98	0.185 635 727 37	0.096 717 332 33
	2^{-6}	0.321 014 878 37	0.185 673 436 95	0.096 750 759 66
$x^{(6)}$	2^{-4}	0.321 052 792 68	0.185 707 164 51	0.096 774 606 61
	2^{-5}	0.321 038 854 86	0.185 697 492 23	0.096 767 419 38
	2^{-6}	0.321 036 706 65	0.185 696 156 20	0.096 766 326 98
exact		0.321 036 946 08	0.185 696 188 23	0.096 766 328 22

(i) Denote $H_n(\xi)$ for the nth degree Hermite polynomial,

$$H_n(\xi) \equiv (-1)^n \exp(\tfrac{1}{2}\xi^2) d^n \exp(-\tfrac{1}{2}\xi^2)/d\xi^n, \quad n \geqslant 0,$$
$$H_0(\xi) = 1, \quad H_1(\xi) = \xi, \quad H_2(\xi) = \xi^2 - 1, \tag{17}$$
$$H_3(\xi) = \xi^3 - 3\xi, \quad H_4(\xi) = \xi^4 - 6\xi^2 + 3.$$

Let $\{i(1), i(2), \ldots, i(m)\}$ be non-negative indices with $n = \Sigma_{k=1}^{m} i(k)$. The random variables

$$\gamma^{(n)}[i(1), i(2), \ldots, i(m)] \equiv \prod_{k=1}^{m} \frac{H_{i(k)}(\xi_k)}{i(k)!^{1/2}}, \tag{18}$$

for all such possible indices form a CON basis of a subspace $M_2^{(n)}$ of M_2, which in turn gives the orthogonal decomposition (Wiener–Itô decomposition) of M_2,

$$M_2 = \oplus_{n=0}^{\infty} M_2^{(n)}. \tag{19}$$

(ii) The orthonormal basis $\gamma^{(n)}[i(1), i(2), \ldots, i(m)]$ of $M_2^{(n)}$ in (18) has the representation,

$$\gamma^{(n)} = [i(1)! \cdots i(m)!]^{-1/2} \int_0^h dB(s_1) \int_0^{s_1} dB(s_2) \cdots \times \int_0^{s_{n-1}} dB(s_n) \sum_{\pi \in S_n} \phi_1(s_{\pi(1)}) \phi_1(s_{\pi(2)}) \cdots \times \phi_1(s_{\pi[i(1)]}) \phi_2(s_{\pi[i(1)+1]}) \phi_2(s_{\pi[i(1)+2]}) \cdots \times \phi_m(s_{\pi[n-i(m)+1]}) \cdots \phi_m(s_{\pi(n)}), \tag{20}$$

where $\pi \in S_n$ moves on all permutations in the nth degree symmetric group S_n.

(iii) Let $\chi(s_1, \ldots, s_n) \in L_2(T^n)$ be an arbitrary nonrandom function. An iterated stochastic integral

$$\eta = \int_0^h dB(s_1) \int_0^{s_1} dB(s_2) \cdots \int_0^{s_{n-1}} dB(s_n) \chi(s_1, \ldots, s_n), \tag{21}$$

belongs to $M_2^{(n)}$, and exhausts the form of its constituents.

Notes: Since the subjects are known widely, only some notes will be in order. (i) refers first to the fact that the space $M_2^{(1)}$ of Wiener integrals is spanned by SNIRV's $\{\xi_0, \xi_1, \ldots\}$, which is obvious by the correspondence between $M_2^{(1)}$ and $L_2(T)$ given by (6) for Wiener integrals. For $n \geqslant 2$ (i) follows[24,25,28] from the structure of $M_2^{(1)}$ and the complete orthonormality of Hermite polynomials $\{H_n(\xi)/n!^{1/2}\}$ with respect to the weight $\exp(-\tfrac{1}{2}\xi^2)/(2\pi)^{1/2}$. (ii) is Theorems 3.1 and 5.1 of Itô;[25,29] the representation (20) is apparently intricate but very useful as we shall find. (iii) is the restatement of this Theorem 5.1 itself; η may be identified with the orthogonal polynomial functional $G_n(K)$ of Wiener[24] by regarding $\chi(s_1, \ldots, s_n)$ as the restriction of a permutation invariant $n!K(s_1, \ldots, s_n)$ to $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_n$. ■

Hereafter we fix the CON set $\{\phi_i(t); i = 0, 1, \ldots\}$ to be the one based on Legendre polynomials:

$$\phi_i(s) \equiv [(2i+1)/h]^{1/2} P_i[2(s-h/2)/h],$$
$$P_i(t) \equiv (2^i i!)^{-1} \frac{d^i (t^2-1)^i}{dt^i}. \tag{22}$$

The Gaussian functionals $I_{00}(h)$, $I_{10}(h)$, and $I_{20}(h)$ are now obtained:

Corollary 3: Let

$$I_{00}(h) \equiv B(h) = h^{1/2}\xi_0,$$
$$I_{10}(h) = h^{3/2}(\xi_0/2 + \xi_1/12^{1/2}),$$
$$I_{20}(h) = h^{5/2}\left(\frac{\xi_0}{3} + \frac{\xi_1}{12^{1/2}} + \frac{\xi_2}{180^{1/2}}\right). \tag{23}$$

Proof: These are manifest by $P_0(t) = 1$, $P_1(t) = t$, $P_2(t) = \frac{3}{2}t^2 - \frac{1}{2}$ and (22). ■

The approximations $x^{(2)}(h)$ and $x^{(3)}(h)$ for $x(h)$ are thus realized precisely by generating SNIRV's $\{\xi_0\}$ and $\{\xi_0, \xi_1\}$, respectively, as is well known.

The next approximation $x^{(4)}(h)$ requires of the nonlinear functional $I_{11}(h) = O(h^2)$. This is by itself a twice iterated stochastic integral,

$$I_{11}(h) = \int_0^h s\, dB(s) \int_0^s dB(t) \in M_2^{(2)}. \tag{24}$$

By Proposition 2, it has an expansion in terms of the second-degree Hermite polynomials of $\{\xi_0, \xi_1, \ldots\}$. The series notably enjoys analytical expressions for coefficients:

Theorem 4: Define

$$\{ij\cdots m\} \equiv \int_0^h \phi_i(s) dB(s) \int_0^s \phi_j(t) dB(t) \times \cdots \int_0^u \phi_m(v) dB(v). \tag{25}$$

The functional $I_{11}(h)$ has the construction,

$$I_{11}(h) = \lim_{k\to\infty} I_{11}^{(k)}(h),$$
$$I_{11}^{(0)}(h) = h^2[\tfrac{1}{3}\gamma_{00}^{(2)} 2^{1/2} + \gamma_{01}^{(2)}/(3^{1/2}4)],$$
$$I_{11}^{(k)}(h) = I_{11}^{(0)}(h) + h^2\left\{ -\sum_{i=1}^{k} \frac{\gamma_{ii}^{(2)}}{8^{1/2}(2i+3)(2i-1)} + \sum_{i=1}^{k} \frac{\gamma_{i-1,i+1}^{(2)}}{4(2i+1)(2i+3)^{1/2}(2i-1)^{1/2}} \right\},$$
$$k \geqslant 1, \tag{26}$$

where $\gamma_{ii}^{(2)}$ and $\gamma_{ij}^{(2)} \equiv \gamma_{ij}^{(2)}$ $(i \neq j)$ are second-degree normalized Hermite polynomials in a modified notation of (18),

$$\gamma_{ii}^{(2)} \equiv 2!^{-1/2} H_2(\xi_i) = 2!^{-1/2}(\xi_i^2 - 1) = (2!)^{1/2}\{ii\},$$
$$\gamma_{ij}^{(2)} = \gamma_{ji}^{(2)} \equiv H_1(\xi_i) H_1(\xi_j) = \xi_i \xi_j = \{ij\} + \{ji\}. \tag{27}$$

Proof: The random variables $\{\gamma_{jk}^{(2)}; j \leqslant k\}$ of (27) exhaust the case $n = 2$ of (18) and their expressions as iterated stochastic integrals follow from (20). They form a CON system in $M_2^{(2)}$, so that $I_{11}(h)$ has the expansion $I_{11}(h) = h^2 \Sigma_{j<k} D_{jk} \gamma_{jk}^{(2)}$ with $D_{jk} = \langle h^{-2} I_{11}(h) \gamma_{jk}^{(2)} \rangle$. Evaluation of D_{ii} with (6), (25), and (27) gives

$$D_{ii} = h^{-2}(2!)^{1/2} \int_0^h s\phi_i(s) ds \int_0^s \phi_i(t) dt$$
$$= 2^{-5/2}(2i+1) \int_{-1}^{1} (s+1) P_i(s) U_i(s) ds$$
$$= 2^{-5/2}(2i+1)\left[4\delta_{i0} - 2^{-1} \int_{-1}^{1} U_i^2(s) ds\right],$$

where we introduced

$$U_i(s) \equiv \int_{-1}^{s} P_i(t) dt. \tag{28}$$

Rodrigues' formula (22) proves

$$U_0 = P_0 + P_1, \quad U_i = (-P_{i-1} + P_{i+1})/(2i+1), \quad i \geqslant 1, \tag{29}$$

resulting in the $\gamma_{ii}^{(2)}$ parts of (26) by the orthogonality of Legendre polynomials. Similarly, $0 \leqslant i < j$ yield

$$D_{ij} = D_{ji} = 8^{-1} h^2 [(2i+1)(2j+1)]^{1/2} \times \int_{-1}^{1} (s+1)[P_i(s) U_j(s) + U_i(s) P_j(s)] ds$$
$$= -8^{-1}[(2i+1)(2j+1)]^{1/2} \int_{-1}^{1} U_i(s) U_j(s) ds.$$

In view of (29), the proof of (26) is completed. ■

The next functional $I_{12}(h)$ is $\frac{5}{2}$th degree in h, and has the decomposition into iterated stochastic integrals:

$$I_{12}(h) = I_{20}(h) + 2J(h), \quad I_{20}(h) \in M_2^{(1)},$$
$$J(h) \equiv \int_0^h s\, dB(s) \int_0^s dB(t) \int_0^t dB(u) \in M_2^{(3)}. \tag{30}$$

By (20) the subspace $M_2^{(3)}$ is spanned by the following CON basis for distinct i, j, and k in all possible combinations:

$$\gamma_{iii}^{(3)} = 3!^{-1/2} H_3(\xi_i) = 3!^{-1/2}(\xi_i^3 - 3\xi_i) = 6^{1/2}\{iii\},$$
$$\gamma_{iij}^{(3)} = 2!^{-1/2} H_2(\xi_i) H_1(\xi_j) = 2!^{-1/2}(\xi_i^2 - 1)\xi_j$$
$$= 2^{1/2}(\{iij\} + \{iji\} + \{jii\}),$$
$$\gamma_{ijk}^{(3)} = H_1(\xi_i) H_1(\xi_j) H_1(\xi_k) = \xi_i \xi_j \xi_k$$
$$= \{ijk\} + \{jki\} + \{kij\} + \{ikj\} + \{kji\} + \{jik\}. \tag{31}$$

Sorry to say, systematic expressions could be found for only a portion of expansion coefficients of $J(h)$. Several significant ones among them have thus to be evaluated, respectively. In order to make this and the subsequent procedures reproducible, we prepare two auxiliary statements.

Auxiliary Lemma A:

(i) $P_0 P_i = P_i, \quad i \geqslant 0,$

(ii) $P_1 P_i = [iP_{i-1} + (i+1)P_{i+1}]/(2i+1), \quad i \geqslant 1,$

(iii) $P_2 P_i = 3i(i-1)P_{i-2}/[2(2i+1)(2i-1)] + (i+1)iP_i/[(2i+3)(2i-1)] + 3(i+2)(i+1)P_{i+2}/[2(2i+3)(2i+1)], \quad i \geqslant 2.$

Proof: These are Neumann's formula proved in pp. 83–87 of Hobson.[30] The recurrence formula of Legendre polynomials will also prove (ii) and (iii) readily with simple algebra. ■

Auxiliary Lemma B: There hold

$$U_0 U_i = -(i-1)P_{i-2}/[(2i+1)(2i-1)] - P_{i-1}/(2i+1) - P_i/[(2i+3)(2i-1)] + P_{i+1}/(2i+1) + (i+2)P_{i+2}/[(2i+3)(2i+1)], \quad i \geqslant 0,$$

$$U_1^2 = \tfrac{2}{15}P_0 - \tfrac{4}{21}P_2 + \tfrac{2}{35}P_4,$$

$$U_1 U_2 = \tfrac{2}{35}P_1 - \tfrac{4}{35}P_3 + \tfrac{2}{63}P_5,$$

$$U_1 U_3 = -\tfrac{1}{105}P_0 + \tfrac{1}{21}P_2 - \tfrac{23}{385}P_4 + \tfrac{5}{231}P_6,$$

$$U_2^2 = \tfrac{2}{105}P_0 - \tfrac{2}{55}P_4 + \tfrac{4}{231}P_6.$$

Proof: Straightforward applications of the preceding lemma and (29). ■

We are now at a position to prove the following theorem.

Theorem 5: The functional $J(h)$ in $I_{12}(h) = I_{20}(h) + 2J(h)$ has the following orthogonal expansion in terms of the basis (31), arranged in the order of decreasing magnitude for coefficients:

$$J(h) = h^{5/2}\left[\frac{\gamma_{000}^{(3)}3^{1/2}}{2^{1/2}4} + \frac{\gamma_{001}^{(3)}3^{1/2}}{2^{1/2}20} - \frac{\gamma_{011}^{(3)}}{2^{1/2}20} + \frac{\gamma_{002}^{(3)}}{10^{1/2}12} + \frac{\gamma_{111}^{(3)}3}{2^{1/2}140} - \frac{\gamma_{012}^{(3)}}{15^{1/2}28} - \frac{\gamma_{022}^{(3)}}{2^{1/2}84} + \frac{\gamma_{013}^{(3)}}{21^{1/2}40}\right.$$
$$+ \frac{\gamma_{122}^{(3)}}{6^{1/2}84} - \frac{\gamma_{113}^{(3)}}{14^{1/2}60} - \frac{\gamma_{033}^{(3)}}{2^{1/2}180} + \frac{\gamma_{003}^{(3)}}{10^{1/2}84} + \frac{\gamma_{024}^{(3)}}{5^{1/2}168} - \frac{\gamma_{124}^{(3)}3^{1/2}}{5^{1/2}308} + \frac{\gamma_{133}^{(3)}}{6^{1/2}165} - \frac{\gamma_{023}^{(3)}}{35^{1/2}72}$$
$$\left. - \frac{\gamma_{044}^{(3)}}{2^{1/2}308} + \frac{\gamma_{014}^{(3)}}{3^{1/2}280} + \epsilon_{12}\zeta_{12}\right],$$

$$\epsilon_{12} = (12553/384199200)^{1/2} \doteqdot 5.7160 \times 10^{-3}. \tag{32}$$

Here, $\zeta_{12} \in M_2^{(3)}$ is a random variable orthogonal to the rest of (32) with $\langle \zeta_{12} \rangle = 0$ and $\langle \zeta_{12}^2 \rangle = 1$.

Proof: Equations (6) and (31) imply

$$\langle J\gamma_{iii}^{(3)} \rangle = h^{5/2}(2i+1)^{3/2}6^{-1/2} \times \left[\delta_{i0} - 16^{-1}\int_{-1}^{1} U_i^3(s)ds\right],$$

$$\langle J\gamma_{iij}^{(3)} \rangle = -h^{5/2}(2i+1)(2j+1)^{1/2}2^{-9/2} \times \int_{-1}^{1} U_i^2(s)U_j(s)ds,$$

$$\langle J\gamma_{ijk}^{(3)} \rangle = -h^{5/2}[(2i+1)(2j+1)(2k+1)]^{1/2}2^{-4} \times \int_{-1}^{1} U_i(s)U_j(s)U_k(s)ds.$$

It is straightforward to evaluate coefficients in (32) consecutively with (29) and Auxiliary Lemma B. The expression (32) was constructed by choosing terms that do not involve $\{\xi_i; i \geqslant 6\}$. Subtraction of the squares of obtained coefficients from $\langle J^2(h) \rangle = h^5/10$ gives ϵ_{12}^2. Any residual term does not exceed ϵ_{12} in magnitude and is omitted. Among these are

$$h^{-5/2}\langle J\gamma_{055}^{(3)} \rangle = -1/(2^{1/2}468),$$
$$h^{-5/2}\langle J\gamma_{025}^{(3)} \rangle = 1/(55^{1/2}126)$$

and so forth, derived by extending Auxiliary Lemma B or by direct integrations of products of Legendre polynomials. Some others, such as $\langle J\gamma_{112}^{(3)} \rangle$ or $\langle J\gamma_{123}^{(3)} \rangle$, vanish identically by symmetry. ■

The next functional $I_{21}(h) = O(h^3)$ is a twice-iterated stochastic integral,

$$I_{21}(h) = \int_0^h s^2\, dB(s) \int_0^s dB(t) \in M_2^{(2)}. \tag{33}$$

In effect, its expansion need be considered only in terms of $\{\xi_0, \xi_1, \xi_2\}$. The result is given below.

Theorem 6: The functional $I_{21}(h)$ admits an orthogonal decomposition,

$$I_{21}(h) = h^3\left[\sum_{0<i<j<2} E_{ij}\gamma_{ij}^{(2)} + \epsilon_{21}\zeta_{21}\right],$$

$$E_{00} = \frac{1}{8^{1/2}}, \quad E_{01} = \frac{3^{1/2}}{10}, \quad E_{02} = \frac{1}{5^{1/2}6},$$

$$E_{11} = -1/(2^{1/2}10), \quad E_{12} = -1/(15^{1/2}14),$$

$$E_{22} = -1/(2^{1/2}84),$$

$$\epsilon_{21} = (247/352800)^{1/2} \doteqdot 2.646 \times 10^{-2}, \tag{34}$$

where ζ_{21} is a random variable in $M_2^{(2)}$ with $\langle \zeta_{21} \rangle = 0$ and $\langle \zeta_{21}^2 \rangle = 1$.

Proof: The orthonormal basis $\{\gamma_{ij}^{(2)}\}$ of (27) for $M_2^{(2)}$ gives

$$E_{ii} = \langle h^{-3} I_{21} \gamma_{ii}^{(2)} \rangle$$
$$= 2!^{1/2} h^{-3} \int_0^h s^2 \phi_i(s) ds \int_0^s \phi_i(t) dt$$
$$= \frac{(2i+1)}{2^{7/2}} \left[8\delta_{i0} - \int_{-1}^{1} (s+1) U_i^2(s) ds \right]$$
$$= \frac{(2i+1)}{2^{7/2}} \left[8\delta_{i0} - \int_{-1}^{1} U_0(s) U_i(s) U_i(s) ds \right],$$
$$E_{ij} = \langle h^{-3} I_{21} \gamma_{ij}^{(2)} \rangle$$
$$= -\frac{(2i+1)^{1/2}(2j+1)^{1/2}}{8} \int_{-1}^{1} U_0(s) U_i(s) U_j(s) ds.$$

Auxiliary Lemma B at once proves (34). ■

The final functional $I_{13}(h) = O(h^3)$ is decomposed as

$$I_{13}(h) = 3I_{21}(h) + 6K(h),$$
$$K(h) = \int_0^h s\, dB(s) \int_0^s dB(t) \int_0^t dB(u) \int_0^u dB(v) \in M_2^{(4)}. \tag{35}$$

Again, those Hermite polynomials composed of $\{\xi_0, \xi_1, \xi_2\}$ only are needed practically for the representation of $K(h)$. The following forms, for distinct i, j, and k obtained from (20), exhaust such relevant orthonormal basis of $M_2^{(4)}$:

$$\gamma_{iiii}^{(4)} = 4!^{-1/2}(\xi_i^4 - 6\xi_i^2 + 3) = 4!^{1/2}\{iiii\},$$
$$\gamma_{iiij}^{(4)} = 3!^{-1/2}(\xi_i^3 - 3\xi_i)\xi_j$$
$$= 3!^{1/2}(\{iiij\} + \{iiji\} + \{ijii\} + \{jiii\}),$$
$$\gamma_{iijj}^{(4)} = 2!^{-1}(\xi_i^2 - 1)(\xi_j^2 - 1)$$
$$= 2!(\{iijj\} + \{ijij\} + \{jiij\}$$
$$+ \{jjii\} + \{jiji\} + \{ijji\}), \tag{36}$$
$$\gamma_{iijk}^{(4)} = 2!^{-1/2}(\xi_i^2 - 1)\xi_j\xi_k$$
$$= 2!^{1/2}(\{iijk\} + \{iikj\} + \{ijik\} + \{ikij\}$$
$$+ \{ijki\} + \{ikji\} + \{jiik\} + \{kiij\}$$
$$+ \{jiki\} + \{kiji\} + \{jkii\} + \{kjii\}).$$

Theorem 7: The functional $K(h)$ in $I_{13}(h) = 3I_{21}(h) + 6K(h)$ admits the orthogonal decomposition

$$K(h) = h^3 \left[\sum_{0<i<j<k<m<2} F_{ijkm} \gamma_{ijkm}^{(4)} + \epsilon_{13} \zeta_{13} \right] \in M_2^{(4)},$$

where the random variable ζ_{13} has $\langle \zeta_{13} \rangle = 0$ and $\langle \zeta_{13}^2 \rangle = 1$. In decreasing order of magnitude, the coefficients are as follows:

$$F_{0000} = 2/(6^{1/2}5), \quad F_{0001} = 1/(2^{1/2}30), \quad F_{0011} = -\tfrac{1}{70},$$
$$F_{0002} = 1/(30^{1/2}14), \quad F_{0111} = 3/(2^{1/2}280),$$
$$F_{0012} = -1/(30^{1/2}28), \quad F_{0022} = -1/252,$$
$$F_{1111} = -1/(6^{1/2}140), \quad F_{0122} = 1/(6^{1/2}168),$$
$$F_{0112} = 1/(10^{1/2}168), \quad F_{1122} = -1/924,$$
$$F_{2222} = -5/(6^{1/2}12012), \quad F_{0222} = 5^{1/2}/(6^{1/2}1848),$$
$$F_{1112} = F_{1222} = 0,$$
$$\epsilon_{13} = (3\,009\,367/64\,929\,664\,800)^{1/2} \doteq 6.808 \times 10^{-3}. \tag{37}$$

Proof: The procedure is now a routine. By (6) we have

$$F_{iiii} = 4!^{-1/2} 2^{-5} (2i+1)^2 \left[2^5 \delta_{i0} - \int_{-1}^{1} U_i^4(s) ds \right],$$
$$F_{iiij} = -3!^{-1/2} 2^{-5} (2i+1)^{3/2} (2j+1)^{1/2}$$
$$\times \int_{-1}^{1} U_i^3(s) U_j(s) ds,$$
$$F_{iijj} = -2^{-6}(2i+1)(2j+1) \int_{-1}^{1} U_i^2(s) U_j^2(s) ds,$$
$$F_{iijk} = -2^{-11/2}(2i+1)(2j+1)^{1/2}(2k+1)^{1/2}$$
$$\times \int_{-1}^{1} U_i^2(s) U_j(s) U_k(s) ds.$$

Auxiliary Lemma B consecutively proves (37). ■

IV. NUMERICAL PROCEDURES AND GENERALIZATIONS

Define $\epsilon_{11}^{(k)} = h^{-2} \| I_{11}(h) - I_{11}^{(k)}(h) \|$ with $I_{11}^{(k)}(h)$ of (26). Consider the $O(h^2)$ approximation $x^{(4)}(h)$ to $x(h)$.[31] An algorithm for the integration of SDEs is obtained by replacing $I_{11}(h)$ in $x_4(h)$ with $I_{11}^{(k)}(h)$. The error is then $O(h^{5/2}) + O(h^2 \epsilon_{11}^{(k)})$ by Proposition 1 (i) and (10), with $I_{00}(h)$ and $I_{10}(h)$ tenable free of error by Corollary 3. This algorithm is thus valid as an $O(h^2)$ scheme if $h^2 \epsilon_{11}^{(k)} \leqslant h^{5/2}$ or $h \geqslant \epsilon_{11}^{(k)2}$ holds true:

Proposition 8: The $O(h^2)$ algorithm $x^{(4)}(h)$ of (11) is obtained by generating $k+2$ SNIRV's $\{\xi_0, \xi_1, \ldots, \xi_{k+1}; k \geqslant 0\}$, constructing $I_{00}(h)$ and $I_{10}(h)$ by (23) and approximating $I_{11}(h)$ with $I_{11}^{(k)}(h)$ of (26). For a fixed integer $k \geqslant 0$ the available number $1/h$ of division of the unit time interval should be in the range

$$1/h \leqslant 1/h_{\min}^{(4)}(k) \equiv 1/\epsilon_{11}^{(k)2}. \tag{38}$$

Conversely, the scheme for a given h, i.e., for the assigned nominal precision of $O(h^2)$ or within the nominal error of $O(h^{5/2})$, is realized by taking $I_{11}^{(k)}(h)$ of (26) for $I_{11}(h)$ with k fulfilling (38).

Table II shows $1/h_{\min}^{(4)}(k)$ for some relevant values of k.

Take the algorithm based on $x^{(5)}(h)$ with the nominal error of $O(h^3)$. The error in the approximation of $I_{11}(h)$ in $x_4(h)$ by $I_{11}^{(k)}(h)$ should now satisfy $h^2 \epsilon_{11}^{(k)} \leqslant h^3$, or

$$1/h \leqslant 1/h_{\min}^{(5)}(k) \equiv 1/\epsilon_{11}^{(k)}. \tag{39}$$

The rhs of (39) is also shown in Table II. For simplicity we discuss only the case $k \geqslant 3$ for $I_{11}^{(k)}$, assuming SNIRV's $\{\xi_0, \xi_1, \ldots, \xi_4\}$ to be generated always. Functionals needed anew are $I_{20}(h)$ and $I_{12}(h)$ in $x_5(h)$. The former is given by (23) precisely. The approximation (32) for $J(h)$ in $I_{12}(h)$ involves an error of $O(2h^{5/2}\epsilon_{12})$, requiring $1/h \leqslant 1/(2\epsilon_{12})^2 \doteq 7652$. This is practically not a restriction. Also, the error arising from $I_{11}^{(k)}(h) B(h)$ in $x_5(h)$ may be ignored by $B(h) = O(h^{1/2})$. Altogether we have the following proposition.

TABLE II. The maximum number of division of a unit time interval admitted in the approximation $I_{11}^{(k)}$ of (26) for I_{11}.

k	$1/h_{\min}^{(4)}(k)$	$1/h_{\min}^{(5)}(k)$	$1/h_{\min}^{(6)}(k)$
0	144	12	5.24
1	1 800	42.43	12.16
2	6 533	80.83	18.69
3	15 876	126	25.13
4	31 363	177.10	31.54
5	54 531	233.52	37.92
6	86 914	294.81	44.30
7	130 050	360.62	50.66
8	185 474	430.67	57.03
9	254 722	504.70	63.39
10	339 329	582.52	69.75
11	440 833	663.95	76.11
12	560 769	748.85	82.46
13	700 673	837.06	88.82
14	862 081	928.48	95.17
15	1 046 529	1023	101.53
16	1 255 553	1120.51	107.88
17	1 490 689	1220.94	114.23
18	1 753 473	1324.19	120.59
19	2 045 441	1430.19	126.94
20	2 368 129	1538.87	133.29
21	2 723 073	1650.17	139.64
22	3 111 809	1764.03	145.99
23	3 535 873	1880.39	152.35
30	7 622 529	2760.89	196.80
40	17 638 528	4199.82	260.31
50	33 952 128	5826.85	323.81

Proposition 9: The $O(h^{5/2})$ algorithm $x^{(5)}(h)$ of (11) is implemented by generating SNIRV's $\{\xi_0,\xi_1,...,\xi_{k+1};k\geqslant 3\}$, constructing $I_{00}(h)$, $I_{10}(h)$, and $I_{20}(h)$ by (23), replacing $I_{11}(h)$ with $I_{11}^{(k)}(h)$ of (26) and constructing $J(h)$ in $I_{12}(h)=I_{20}(h)+2J(h)$ by (32). The restriction (39) puts the lower bound, of available h for a given k or of k for a given h.

Consider $x^{(6)}(h)$ as the final of the algorithms. The error from $I_{11}^{(k)}(h)$ in $x_4(h)$ is now required stringently to be less than $O(h^{7/2})$, resulting in

$$1/h\leqslant 1/h_{\min}^{(6)}(k)\equiv 1/\epsilon_{11}^{(k)2/3}, \tag{40}$$

whose rhs is given in Table II. Similarly, the error from the approximation (32) for $J(h)$ in $I_{12}(h)$ should fulfill $1/h\leqslant 1/(2\epsilon_{12})=87.47$. For $I_{21}(h)$ and $I_{13}(h)$ in $x_6(h)$ the errors in the approximations (34) and (37) give rise to restrictions $1/h\leqslant 1/\epsilon_{21}^2=1428$ or $1/(6\epsilon_{13})^2=599$, which are in effect not restrictive. We conclude:

Proposition 10: The $O(h^3)$ algorithm $x^{(6)}(h)$ of (11) is implemented by generating SNIRV's $\{\xi_0,\xi_1,...,\xi_{k+1};k\geqslant 3\}$, constructing $I_{00}(h)$, $I_{10}(h)$, and $I_{20}(h)$ by (23), approximating $I_{11}(h)$ with $I_{11}^{(k)}(h)$ of (26), $J(h)$ in $I_{12}(h)$ with (32), and $I_{21}(h)$ and $I_{13}(h)$ with (34) and (37), respectively. The restriction (40) together with $1/h\leqslant 1/(2\epsilon_{12})=87.47$ limits the available smallest h or the tenable highest total precision $O(h^3)\geqslant 1.494\times 10^{-6}$ for a given k.

Proposition 10 calls for a suitable prescaling of time $t\to t_0 t$ for the implementation of the $x^{(6)}(h)$ scheme, in order to make the transformed (4), $dx(t)=t_0 a(x,t_0 t)dt+t_0^{1/2}b(x,t_0 t)dB(t)$, fulfill the proviso in Proposition 1 (i). The number of terms in $x^{(6)}(h)$ depends strongly on SDEs, usually increasing very rapidly from that of $x^{(5)}(h)$ or $x^{(4)}(h)$ as shown in the Appendix. Therefore, a rather small t_0 may be necessitated, spoiling the virtue of $x^{(6)}(h)$.[20] With this reservation, we may summarize that our $O(h^3)$ procedure in general allows for the largest h if a nominal precision 1.494×10^{-6} suffices. The $O(h^{5/2})$ or $O(h^2)$ procedures do not give the largest h for this or lower level of accuracy, but their nominal precision may be made very high by a small enough h but for the accumulation of errors due to many time steps.[20]

We now observe an example, in order to see these and other problems in rescaling SDEs, to present a systematic comparison of the actual precision associated with procedures $x^{(4)}(h)$, $x^{(5)}(h)$, and $x^{(6)}(h)$, and also to show a method to construct $x^{(6)}(h)$ for very small h beyond the limit $h_{\min}^{(6)}$ in (40). We consider the Stratonovich model,[8,32]

$$dx(t)=[(c+b^2/2)x-x^3]dt+bx\,dB(t)\,,\quad x(0)=x_0\,, \tag{41}$$

for the case $b=c=x_0=1$. This corresponds to the Stratonovich SDE $dx=(cx-x^3)dt+bx\,dB(t)$. The parameter values stipulate $x(t)>0$ for all t, with the stationary probability density peaked at $x=1/2^{1/2}$.

In order to compare the results for a wide range of h and also to realize the procedure $x^{(2)}(h)$ for comparison on a very fine time step $\delta=2^{-13}=1/8192$ that divides the aimed h into $N=h/\delta$ subintervals, noise functional increments were generated at every δ. The precise construction with $C(s)\equiv B(t+s)-B(t)$,

$$I_{ij}(t+\delta)=I_{ij}(t)+\int_0^\delta (t+s)^i[B(t)+C(s)]^j dC(s), \tag{42}$$

of noise functionals was used to obtain them at $t=k\delta$, $k=1,2,...,N$. The increments $\Delta I_{ij}(\delta)\equiv\int_0^\delta s^i C(s)^j\,dC(s)$ were approximated by $I_{11}^{(22)}(\delta)$ of (26), (32), (34), or (37). These were then substituted into (42) to obtain $I_{ij}(h)$ for $h=N\delta$. The error in $I_{11}(h)$ is now $\delta^2\epsilon_{11}^{(22)}N=\delta h\epsilon_{11}^{(22)}$. This is smaller than $h^{7/2}$ if $1/h\leqslant(\delta\epsilon_{11}^{(22)})^{-2/5}=731$. Similarly, the error in $I_{12}(h)$ is $2\delta^{5/2}\epsilon_{12}N=2\delta^{3/2}\epsilon_{12}h$ and should be smaller than $h^{7/2}$. This gives a nonrestrictive condition $1/h\leqslant(2\delta^{3/2}\epsilon_{12})^{-2/5}=1333$. Errors in I_{21} and I_{13} are estimated likewise, and seen to be negligible comparatively. Thus the accuracy of $x^{(6)}(h)$, not to mention $x^{(4)}(h)$ and $x^{(5)}(h)$, for $h\geqslant 1/256=2^{-8}$ could be examined without being marred by the errors in the noise functionals.

Table III shows the results obtained in double precision. The third-degree scheme $x^{(6)}(h)$ shows convergence at the time step $h=1/256$ with excellent stability. Other procedures showing traces of oscillations are still on their way to convergence in this range of h. For a long run, therefore, this problem should be rescaled by $t\to t/4$, and the approximation $x^{(6)}(h)$ with $I_{11}^{(10)}(h)$ (say) and $h=1/64$ will be the best, though its excellence is not so overwhelming as the example (13).

TABLE III. Results for the stochastic model of Stratonovich $dx = (3x/2 - x^3)dt + x\,dB$ with $x(0) = 1$.

Procedure	h	$t = 10$	$t = 20$	$t = 30$
$x^{(2)}(t)$	2^{-13}	0.256 382 469 24	0.601 457 256 65	0.432 718 411 24
	2^{-4}	0.255 981 508 81	0.595 832 987 09	0.426 646 571 69
	2^{-5}	0.255 650 626 58	0.600 449 150 39	0.432 146 870 08
$x^{(4)}(t)$	2^{-6}	0.256 147 132 34	0.601 212 588 07	0.432 525 031 63
	2^{-7}	0.256 294 341 55	0.601 272 189 86	0.432 709 008 71
	2^{-8}	0.256 320 961 89	0.601 395 083 17	0.432 680 583 51
	2^{-4}	0.255 406 890 22	0.597 028 656 14	0.433 724 088 46
	2^{-5}	0.256 324 185 42	0.601 006 203 14	0.431 800 304 17
$x^{(5)}(t)$	2^{-6}	0.256 289 102 97	0.601 449 342 43	0.432 511 385 58
	2^{-7}	0.256 302 139 97	0.601 390 592 25	0.432 597 845 83
	2^{-8}	0.256 314 616 06	0.601 417 216 45	0.432 666 576 76
	2^{-4}	0.255 653 716 94	0.598 601 874 09	0.433 605 266 42
	2^{-5}	0.255 767 091 39	0.600 798 600 11	0.431 658 621 38
$x^{(6)}(t)$	2^{-6}	0.256 278 889 20	0.601 331 401 20	0.432 596 268 50
	2^{-7}	0.256 294 239 80	0.601 383 434 85	0.432 658 312 90
	2^{-8}	0.256 313 506 09	0.601 416 303 73	0.432 673 848 02

In so far as the SDEs contain only a single noise, the procedures of this section apply even if many degrees of freedom enter. As the list in the Appendix shows, all functionals of (12) arise in a general additive noise problem. However, a class of physical Brownian motion processes obeying

$$dx(t) = u(t)dt, \quad du(t) = -[\beta u - f(x)]dt + \sigma\, dB(t),$$

for possibly nonlinear $f(x)$ with constants $\beta > 0$ and σ, enjoy a drastic simplification; only Gaussian functionals $B(h)$, $I_{10}(h)$, and $I_{20}(h)$ are needed for $x^{(6)}(h)$. This scheme is undoubtedly the best to be adopted for them.

Multiplicative multiple noise problems in general still remain difficult,[21] however, in the present formulation by the appearance of stochastic integrals

$$L(h) \equiv \int_0^h dB(s) \int_0^s dC(t)$$

and

$$M(h) \equiv \int_0^h dC(s) \int_0^s dB(t)$$

in the $O(h)$ terms of the perturbation series, where $C(t)$ is now a standard Wiener process independent of $B(t)$. Define $\eta_i \equiv \int_0^h \phi_i(s)dC(s)$, besides $\{\xi_i\}$ of (16).

Proposition 11: L(h) has the orthogonal decomposition,

$$L(h) = L^{(k)}(h) + h\epsilon^{(k)}\xi_k,$$

$$L^{(k)}(h) \equiv h\{2^{-1}\xi_0\eta_0 + \sum_{i=1}^{k} [4(2i+1)(2i-1)]^{-1/2} \times (\xi_i\eta_{i-1} - \xi_{i-1}\eta_i)\}, \tag{43}$$

$$\epsilon^{(k)} = [2(2k+1)]^{-1/2}, \quad \langle \xi_k \rangle = 0, \quad \langle \xi_k^2 \rangle = 1.$$

The integral $M(h)$ is obtained by interchanging $\{\xi_i\}$ and $\{\eta_i\}$.

Proof: The functional $L(h)$ is bilinear in $\{\xi_i\}$ and $\{\eta_j\}$, and the set $\{\xi_i\eta_j;\ i,j \geqslant 0\}$ forms a CON basis for the representation $L(h) = \Sigma_{ij} G_{ij}\xi_i\eta_j$ with $G_{ij} = \langle L(h)\xi_i\eta_j \rangle$. Itô's formula and (6) give

$$\xi_i\eta_j = \int_0^h \phi_i(s)dB(s) \int_0^s \phi_j(t)dC(t) + \int_0^h \phi_j(s)dC(s) \int_0^s \phi_i(t)dB(t),$$

$$G_{ij} = \{4^{-1}h\,[(2i+1)(2j+1)]^{1/2}\} \int_{-1}^{1} P_i(s)U_j(s)ds.$$

The coefficients of $L^{(k)}(h)$ in (43) are obtained by (29), together with

$$(\epsilon^{(k)})^2 = \langle [L(h) - L^{(k)}(h)]^2 \rangle / h^2 = 2^{-1}\{1 - \tfrac{1}{2} - \sum_{i=1}^{k} [(2i+1)(2i-1)]^{-1}\} = [2(2k+1)]^{-1}. \quad \blacksquare$$

Sorry to say, the series (43) converges extremely slowly. Even for the use in $x^{(2)}(h)$, a given h for the nominal error of $O(h^{3/2})$ requires k to fulfill $h \geqslant 1/[2(2k+1)]$, or a k chosen inevitably introduces an error greater than $[2(2k+1)]^{-3/2}$. The approximation based on (43) is thus not useful. Though similar closed forms for $L(h)$ may be obtained for different choices of CON system $\{\phi_i(t)\}$,[33] the circumstance is not improved.

Klauder and Petersen[20] gave an algorithm accurate to $O(h^{3/2})$ in the M_2 norm, which includes the SDEs

$$dL(s) = C(s)dB(s), \quad dM(s) = B(s)dC(s),$$

for $L(s)$ and $M(s)$. The algorithm gives an approximation,

$$L^{KP}(t+h) = L^{KP}(t) + h^{1/2}[B(t) + (h/2)^{1/2}\chi]\omega,$$
$$M^{KP}(t+h) = M^{KP}(t) + h^{1/2}[C(t) + (h/2)^{1/2}\omega]\chi, \tag{44}$$

with SNIRV's $\{\chi,\omega\}$ for the increments of $B(t)$ and $C(t)$. As a numerical procedure the improvement from (43) is drastic, as with other possible formulations of Mil'shtein.[21]

If we need to compare, samplewise and at any cost, the accuracy of schemes for SDEs with multiple, multiplicative noises, it might be advisable to use (44) in the context of the preceding example to obtain noise functionals, and apply them to $O(h^2)$ Rao–Borwankar–Ramkrishna type formulation. The existence of Klauder–Petersen and of Mil'shtein algorithms is invaluable. The procedures for noise functionals are independent of the structure of the SDE to be treated. Their direct installation in computing libraries is possible and will greatly facilitate the integration of SDEs as discussed in this work, evoking wider applications.

ACKNOWLEDGMENT

Numerical computations in this work were conducted on the FACOM M-780/30 System of the Data Processing Center, Kyoto University.

APPENDIX: PERTURBATION SERIES

Since the system is time homogeneous, the first two equations of (8a) are solved at once by $x_0(t)$ and $x_1(t)$ of (10). The last equation of (8a) then gives

$$x_2(t) = c_2^{(2)}B^2(t) + f_1^{(2)}t,$$
$$c_2^{(2)} = \tfrac{1}{2}b_0b_1, \quad f_1^{(2)} = a_0 - c_2^{(2)}. \tag{A1}$$

The explicit form of the equation for $x_3(t)$ is

$$x_3(t) = \int_0^t \left[b_1x_2(s) + \frac{b_2x_1^2(s)}{2!}\right]dB(s) + \int_0^t a_1x_1(s)ds.$$

Using Itô's formula,

$$B^2(s)dB(s) = \tfrac{1}{3}dB^3(s) - d[B(s)s] + sB(s)dB(s),$$

we obtain

$$x_3(t) = c_3^{(3)}B^3(t) + c_1^{(3)}B(t)t + d_{10}^{(3)}I_{10}(t),$$

with

$$c_3^{(3)} = (2b_1c_2^{(2)} + b_0^2b_2)/3!,$$
$$c_1^{(3)} = -b_1c_2^{(2)} - \tfrac{1}{2}b_0^2b_2 + a_1b_0, \tag{A2}$$
$$d_{10}^{(3)} = -c_1^{(3)} + b_1f_1^{(2)}.$$

The next equation is

$$x_4(t) = \int_0^t \left[b_1x_3(s) + b_2x_1(s)x_2(s) + \frac{b_3x_1^3(s)}{3!}\right]dB(s) + \int_0^t \left[a_1x_2(s) + \frac{a_2x_1^2(s)}{2!}\right]ds.$$

Substituting the preceding results and reducing all (stochastic) integrals to combinations of the functionals of (12) by Itô's formula, we arrive at the following:

$$x_4(t) = c_4^{(4)}B^4(t) + c_2^{(4)}B^2(t)t + d_{11}^{(4)}I_{11}(t) + e_{10}^{(4)}I_{10}(t)B(t) + f_2^{(4)}t^2,$$

$$c_4^{(4)} = (b_0^3b_3 + 6b_1c_3^{(3)} + 6b_0b_2c_2^{(2)})/4!,$$
$$c_2^{(4)} = -6c_4^{(4)} + a_1c_2^{(2)} + \tfrac{1}{2}a_2b_0^2, \quad e_{10}^{(4)} = b_1d_{10}^{(3)},$$
$$f_2^{(4)} = \tfrac{1}{2}(-c_2^{(4)} - e_{10}^{(4)} + a_1f_1^{(2)}),$$
$$d_{11}^{(4)} = -2c_2^{(4)} + b_1c_1^{(3)} + b_0b_2f_1^{(2)} - e_{10}^{(4)}. \tag{A3}$$

Equations (A1)–(A3) are solved to the first half of (10), the result of Rao *et al.*[13]

The procedure continues straight, but terms increase very rapidly. We note only the generating formulas for coefficients. The functionals for $x_5(t)$ are as noted in (10). Their coefficients satisfy:

$$c_5^{(5)} = \tfrac{1}{5}\left[b_1c_4^{(4)} + b_0b_2c_3^{(3)} + \tfrac{1}{2}b_2(c_2^{(2)})^2 + \tfrac{1}{2}b_0^2b_3c_2^{(2)} + b_0^4b_4/4!\right],$$
$$u_1^{(5)} = a_1c_3^{(3)} + a_2b_0c_2^{(2)} + \tfrac{1}{6}a_3b_0^3,$$
$$c_3^{(5)} = -10c_5^{(5)} + u_1^{(5)}, \quad e_{10}^{(5)} = \tfrac{1}{2}(b_1e_{10}^{(4)} + b_0b_2d_{10}^{(3)}),$$
$$e_{11}^{(5)} = b_1d_{11}^{(4)}, \quad u_2^{(5)} = \tfrac{1}{2}(a_1c_1^{(3)} + a_2b_0f_1^{(2)}),$$
$$c_1^{(5)} = 15c_5^{(5)} - \tfrac{1}{2}e_{11}^{(5)} - e_{10}^{(5)} - \tfrac{3}{2}u_1^{(5)} + u_2^{(5)},$$
$$g_{10}^{(5)} = -e_{10}^{(5)} + a_1d_{10}^{(3)},$$
$$d_{20}^{(5)} = -15c_5^{(5)} + b_1f_2^{(4)} + \tfrac{1}{2}b_2(f_1^{(2)})^2 + \tfrac{1}{2}e_{11}^{(5)} + 2e_{10}^{(5)} + \tfrac{3}{2}u_1^{(5)} - u_2^{(5)} - a_1d_{10}^{(3)},$$
$$d_{12}^{(5)} = 30c_5^{(5)} + b_1c_2^{(4)} + b_0b_2c_1^{(3)} + b_2c_2^{(2)}f_1^{(2)} + \tfrac{1}{2}b_0^2b_3f_1^{(2)} - \tfrac{1}{2}e_{11}^{(5)} - e_{10}^{(5)} - 3u_1^{(5)}. \tag{A4}$$

Finally, using one intricate result of Itô's formula

$$I_{10}(s)B^2(s)dB(s) = d\{-\tfrac{1}{3}I_{13}(s) + 2I_{21}(s) - \tfrac{1}{2}B^2(s)s^2 + \tfrac{1}{3}s^3 + \tfrac{1}{3}I_{10}(s)B^3(s) - I_{10}(s)B(s)s + \tfrac{1}{2}[I_{10}(s)]^2\},$$

which is confirmed by taking the differential of the rhs, we arrive at the following:

$$c_6^{(6)} = \tfrac{1}{6}\left[b_1c_5^{(5)} + b_0b_2c_4^{(4)} + b_2c_2^{(2)}c_3^{(3)} + \tfrac{1}{2}b_0^2b_3c_3^{(3)} + \tfrac{1}{2}b_0b_3(c_2^{(2)})^2 + \tfrac{1}{6}b_0^3b_4c_2^{(2)} + b_0^5b_5/5!\right],$$
$$u_1^{(6)} = a_1c_4^{(4)} + a_2b_0c_3^{(3)} + \tfrac{1}{2}a_2(c_2^{(2)})^2 + \tfrac{1}{2}a_3b_0^2c_2^{(2)} + a_4b_0^4/4!,$$
$$c_4^{(6)} = -15c_6^{(6)} + u_1^{(6)}, \quad e_{11}^{(6)} = \tfrac{1}{2}(b_1e_{11}^{(5)} + b_0b_2d_{11}^{(4)}),$$
$$e_{10}^{(6)} = \tfrac{1}{3}(b_1e_{10}^{(5)} + b_0b_2e_{10}^{(4)} + b_2c_2^{(2)}d_{10}^{(3)} + \tfrac{1}{2}b_0^2b_3d_{10}^{(3)}),$$
$$u_2^{(6)} = a_1c_2^{(4)} + a_2b_0c_1^{(3)} + a_2c_2^{(2)}f_1^{(2)} + \tfrac{1}{2}a_3b_0^2f_1^{(2)},$$
$$e_{12}^{(6)} = b_1d_{12}^{(5)}, \quad e_{20}^{(6)} = b_1d_{20}^{(5)},$$
$$c_2^{(6)} = 45c_6^{(6)} - \tfrac{1}{2}e_{12}^{(6)} - e_{11}^{(6)} - \tfrac{3}{2}e_{10}^{(6)} - 3u_1^{(6)} + \tfrac{1}{2}u_2^{(6)},$$
$$g_{11}^{(6)} = -e_{11}^{(6)} + a_1d_{11}^{(4)},$$
$$u_3^{(6)} = \tfrac{1}{2}(b_1g_{10}^{(5)} + b_2f_1^{(2)}d_{10}^{(3)}),$$

$$u_4^{(6)} = a_1 e_{10}^{(4)} + a_2 b_0 d_{10}^{(3)}, \quad i_{10}^{(6)} = -e_{10}^{(6)} + u_4^{(6)},$$
$$h_{10}^{(6)} = u_3^{(6)} - \tfrac{1}{2} i_{10}^{(6)},$$
$$d_{13}^{(6)} = 60c_6^{(6)} + b_1 c_3^{(5)} + b_0 b_2 c_2^{(4)} + b_2 c_2^{(2)} c_1^{(3)} + b_2 c_3^{(3)} f_1^{(2)} + \tfrac{1}{2} b_0^2 b_3 c_1^{(3)} + b_0 b_3 c_2^{(2)} f_1^{(2)} + \tfrac{1}{6} b_0^3 b_4 f_1^{(2)} - e_{12}^{(6)} - e_{11}^{(6)} - e_{10}^{(6)} - 4u_1^{(6)},$$
$$d_{21}^{(6)} = -90c_6^{(6)} + b_1 c_1^{(5)} + b_0 b_2 f_2^{(4)} + b_2 f_1^{(2)} c_1^{(3)} + \tfrac{1}{2} b_0 b_3 (f_1^{(2)})^2 - e_{20}^{(6)} + e_{12}^{(6)} + 3e_{11}^{(6)} + 6e_{10}^{(6)} + 6u_1^{(6)} - u_2^{(6)} - a_1 d_{11}^{(4)} - u_4^{(6)},$$
$$f_3^{(6)} = -15c_6^{(6)} - \tfrac{1}{3} e_{20}^{(6)} + \tfrac{1}{6} e_{12}^{(6)} + \tfrac{1}{3} e_{11}^{(6)} + e_{10}^{(6)} - \tfrac{1}{3} u_3^{(6)} + u_1^{(6)} - \tfrac{1}{6} u_2^{(6)} + \tfrac{1}{3} a_1 f_2^{(4)} + \tfrac{1}{6} a_2 (f_1^{(2)})^2 - \tfrac{1}{6} u_4^{(6)}. \tag{A5}$$

Generating Eqs. (A1)–(A5) were solved with REDUCE[26] on a FACOM M-780/30 System at Kyoto University. Since the expressions are massive, we note below only the significant case of linear $b(x)$, with $b_k = 0$ for $k \geqslant 2$:

$$c_2^{(2)} = b_1 b_0/2!, \quad f_1^{(2)} = a_0 - \tfrac{1}{2} b_1 b_0,$$
$$c_3^{(3)} = b_1^2 b_0/3!, \quad c_1^{(3)} = b_0 (a_1 - \tfrac{1}{2} b_1^2),$$
$$d_{10}^{(3)} = -a_1 b_0 + a_0 b_1,$$
$$c_4^{(4)} = b_1^3 b_0/4!, \quad c_2^{(4)} = b_0 (\tfrac{1}{2} a_1 b_1 - \tfrac{1}{4} b_1^3 + \tfrac{1}{2} b_0 a_2),$$
$$d_{11}^{(4)} = a_1 b_1 b_0 - a_0 b_1^2 - b_0^2 a_2,$$
$$e_{10}^{(4)} = -b_1 (a_1 b_0 - a_0 b_1),$$
$$f_2^{(4)} = \tfrac{1}{2} a_1 a_0 - \tfrac{1}{2} a_0 b_1^2 + \tfrac{1}{8} b_1^3 b_0 - \tfrac{1}{4} b_0^2 a_2,$$
$$c_5^{(5)} = b_1^4 b_0/5!,$$
$$c_3^{(5)} = b_0 (\tfrac{1}{6} a_1 b_1^2 - \tfrac{1}{12} b_1^4 + \tfrac{1}{2} b_1 b_0 a_2 + \tfrac{1}{6} b_0^2 a_3),$$
$$c_1^{(5)} = b_0 (\tfrac{1}{2} a_1^2 - \tfrac{1}{2} a_1 b_1^2 + \tfrac{1}{2} a_0 a_2 + \tfrac{1}{8} b_1^4 - \tfrac{1}{2} b_1 b_0 a_2 - \tfrac{1}{4} b_0^2 a_3),$$
$$d_{20}^{(5)} = \tfrac{1}{2} a_1^2 b_0 - \tfrac{1}{2} a_1 a_0 b_1 - \tfrac{1}{2} a_0 b_0 a_2 + \tfrac{1}{4} b_1 b_0^2 a_2 + \tfrac{1}{4} b_0^3 a_3,$$
$$d_{12}^{(5)} = -\tfrac{1}{2} b_0^2 (b_1 a_2 + b_0 a_3),$$
$$e_{11}^{(5)} = b_1 (a_1 b_1 b_0 - a_0 b_1^2 - b_0^2 a_2),$$
$$e_{10}^{(5)} = -\tfrac{1}{2} b_1^2 (a_1 b_0 - a_0 b_1),$$
$$g_{10}^{(5)} = -a_1^2 b_0 + a_1 a_0 b_1 + \tfrac{1}{2} a_1 b_1^2 b_0 - \tfrac{1}{2} a_0 b_1^3,$$
$$c_6^{(6)} = b_1^5 b_0/6!,$$
$$c_4^{(6)} = \tfrac{1}{48} b_0 (2a_1 b_1^3 - b_1^5 + 14 b_1^2 b_0 a_2 + 12 b_1 b_0^2 a_3 + 2 b_0^3 a_4),$$
$$c_2^{(6)} = \tfrac{1}{4} a_1^2 b_1 b_0 - \tfrac{1}{2} a_1 b_1^3 b_0 + \tfrac{3}{4} a_1 b_0^2 a_2 + \tfrac{1}{4} a_0 b_1^4 + \tfrac{1}{4} a_0 b_1 b_0 a_2 + \tfrac{1}{4} a_0 b_0^2 a_3 + \tfrac{1}{16} b_1^5 b_0 - \tfrac{1}{2} b_1^2 b_0^2 a_2 - \tfrac{5}{8} b_1 b_0^3 a_3 - \tfrac{1}{8} b_0^4 a_4,$$
$$d_{21}^{(6)} = -\tfrac{1}{2} a_1^2 b_1 b_0 + \tfrac{1}{2} a_1 a_0 b_1^2 + \tfrac{1}{2} a_1 b_1^3 b_0 + \tfrac{1}{2} a_1 b_0^2 a_2 - \tfrac{1}{2} a_0 b_1^4 - \tfrac{1}{2} a_0 b_1 b_0 a_2 - \tfrac{1}{2} a_0 b_0^2 a_3 - \tfrac{1}{4} b_1^2 b_0^2 a_2 + \tfrac{3}{4} b_1 b_0^3 a_3 + \tfrac{1}{4} b_0^4 a_4,$$
$$d_{13}^{(6)} = -\tfrac{1}{3} a_1 b_1^3 b_0 + \tfrac{1}{3} a_0 b_1^4 + \tfrac{1}{3} b_1^2 b_0^2 a_2 - \tfrac{1}{3} b_1 b_0^3 a_3 - \tfrac{1}{6} b_0^4 a_4,$$
$$e_{20}^{(6)} = b_1 (\tfrac{1}{2} a_1^2 b_0 - \tfrac{1}{2} a_1 a_0 b_1 - \tfrac{1}{2} a_0 b_0 a_2 + \tfrac{1}{4} b_1 b_0^2 a_2 + \tfrac{1}{4} b_0^3 a_3),$$
$$e_{12}^{(6)} = -\tfrac{1}{2} b_1 b_0^2 (b_1 a_2 + b_0 a_3),$$
$$e_{11}^{(6)} = \tfrac{1}{2} b_1^2 (a_1 b_1 b_0 - a_0 b_1^2 - b_0^2 a_2),$$
$$e_{10}^{(6)} = -\tfrac{1}{6} b_1^3 (a_1 b_0 - a_0 b_1),$$
$$f_3^{(6)} = \tfrac{1}{6} a_1^2 a_0 + \tfrac{1}{12} a_1^2 b_1 b_0 - \tfrac{1}{3} a_1 a_0 b_1^2 + \tfrac{1}{24} a_1 b_1^3 b_0 - \tfrac{1}{6} a_1 b_0^2 a_2 + \tfrac{1}{6} a_0^2 a_2 + \tfrac{1}{12} a_0 b_1^4 - \tfrac{1}{4} a_0 b_1 b_0 a_2 - \tfrac{1}{12} a_0 b_0^2 a_3 - \tfrac{1}{48} b_1^5 b_0 + \tfrac{1}{8} b_1^2 b_0^2 a_2 + \tfrac{1}{8} b_1 b_0^3 a_3 + \tfrac{1}{24} b_0^4 a_4,$$
$$g_{11}^{(6)} = a_1^2 b_1 b_0 - a_1 a_0 b_1^2 - \tfrac{1}{2} a_1 b_1^3 b_0 - a_1 b_0^2 a_2 + \tfrac{1}{2} a_0 b_1^4 + \tfrac{1}{2} b_1^2 b_0^2 a_2,$$
$$h_{10}^{(6)} = \tfrac{1}{2} b_0 a_2 (a_1 b_0 - a_0 b_1)$$
$$i_{10}^{(6)} = -a_1^2 b_1 b_0 + a_1 a_0 b_1^2 + \tfrac{1}{2} a_1 b_1^3 b_0 - a_1 b_0^2 a_2 - \tfrac{1}{2} a_0 b_1^4 + a_0 b_1 b_0 a_2.$$

The work is closed by the proof of Proposition 1.

(i) The formal Taylor expansion used was

$$a[x_0 + \epsilon x_1 + \epsilon^2 x_2 + \cdots + \epsilon^n x_n + \cdots, s] = a(x_0, s) + \sum_{j \geqslant 1} a^{(j)}(s)(\epsilon x_1 + \epsilon^2 x_2 + \cdots + \epsilon^n x_n + \cdots)^j \tag{A6}$$
$$\equiv \sum_{k \geqslant 0} \epsilon^k a_k(s),$$
$$a_0(s) = a(x_0, s), \quad a_1(s) = a^{(1)}(s) x_1,$$
$$a_k(s) = \sum_{p+q+\cdots+r=k} c_{pq\ldots r}(s) x_p x_q \cdots x_r,$$

with $a^{(j)}(s) \equiv \partial^j a(x_0, s)/\partial x_0^j$, where $c_{pq\ldots r}(s)$ is a linear combination of a $a^{(j)}(s)$'s. Similarly we have

$$b[x_0 + \epsilon x_1 + \epsilon^2 x_2 + \cdots + \epsilon^n x_n + \cdots, s] = \sum_{k \geqslant 0} \epsilon^k b_k(s),$$
$$b_k(s) = \sum_{p+q+\cdots+r=k} d_{pdq\ldots r}(s) x_p x_q \cdots x_r. \tag{A7}$$

These give (8b) an explicit form:

$$x_n(s) = \sum_{p+\cdots+r=n-2} \int_0^s c_{p\ldots r}(t)\, x_p(t) \cdots x_r(t) dt + \sum_{p+\cdots+r=n-1} \int_0^s d_{p\ldots r}(t)\, x_p(t) \cdots x_r(t) dB(t). \tag{A8}$$

Denote $\delta \equiv h^{(n+1)/2}$. By assumption there is N_0 that gives $\|x(t) - x_N(t)\| = O(\delta)$ for $0 \leqslant t \leqslant h$, $\forall N \geqslant N_0$. We fix any such N, and show that the nth iteration $x_N^{(n)}(t)$ based on (8a,b) for $\{a_N(x,s), b_N(x,s)\}$ approximates $x_N(t)$ within an error of $O(\delta)$. This suffices, at least for a fixed realization of x_0, because the coefficients $\{c_{p\ldots r}(s), d_{p\ldots r}(s)\}$ constructed with $\{a_N(x_0,s), b_N(x_0,s)\}$ by (A6) and (A7) for $s \in [0,h]$

become identical for a large enough N with those of $\{a(x_0,s),b(x_0,s)\}$ implying $x_N^{(n)}(s)\equiv x^{(n)}(s)$.

By the assumptions on $\{a(y,s),b(y,s),\mu(X)\}$ the functions $\{a_N(y,s),b_N(y,s)\}$ and their y-derivatives are uniformly bounded in $(y,s)\in\mathbb{R}\times[0,h]$. From now on the subscript N will be deleted from $a_N(y,s)$, $b_N(y,s)$ and $x_N^{(n)}(t)$. Induction shows that the corresponding $x_p(t)$ determined consecutively by (A8) is $O(h^{p/2})$ for $0\leqslant t\leqslant h$, proving in turn that $a_p(s)$ and $b_p(s)$ are $O(h^{p/2})$ by (A6) and (A7). Call the rhs of (5) as the Picard mapping of $\{x(s);0\leqslant s\leqslant t\leqslant h\}$ and denote (5) as $x(t)=P[x(s)](t)$. Just as the existence proofs[1-3,6] for SDEs we transform $x^{(n)}(t)$ by P:

$$x'(t)\equiv P[x^{(n)}(s)](t)$$
$$=x_0+\int_0^t a[x^{(n)}(s),s]ds+\int_0^t b[x^{(n)}(s),s]dB(s).$$

Define $R(t)\equiv x'(t)-x^{(n)}(t)$, the residual term,

$$R(t)\equiv\int_0^t\{a[x^{(n)}(s),s]-\sum_{p=0}^{n-2}a^{(p)}(s)\}ds$$
$$+\int_0^t\{b[x^{(n)}(s),s]-\sum_{p=0}^{n-1}b^{(p)}(s)\}dB(s).$$

Taylor's formula applied on $a[x^{(n)}(s),s]$ and $b[x^{(n)}(s),s]$ about (x_0,s) proves $|R(t)|\leqslant C_N\delta$, with N dependent but finite C_N. Thus $P[x^{(n)}(s)](t)=x^{(n)}(t)+O(\delta)$ holds true. Further P-iterations on $x'(t)$ change the form of $R(t)$, but not its order of magnitude. Since $P^k[x^{(n)}(t)]$ is known to converge in M_2 t uniformly to the solution $x(t)\equiv x_N(t)$ of (5) as $k\to\infty$, a sufficiently large k gives for $0\leqslant t\leqslant h$,

$$P^k[x^{(n)}(t)]=x(t)+O(\delta)=x^{(n)}(t)+O(\delta).$$

The above was for infinitesimal h. A small but finite h manifestly necessitates the proviso.

(ii) The preceding analysis and Itô's formula clarify that the general form of a term in the iteration (8a) and (8b) or (A8) with Taylor expanded $a_p(s)$ and $b_p(s)$ is as follows:[27]

$$\int_0^h s^{k(1)}\,dB_1(s)\int_0^s s^{k(2)}\,dB_2(t)\int_0^t\cdots\int_0^u s^{k(m)}\,dB_m(v).$$

Here, $dB_j(u)$ is either du or $dB(u)$ with $k(j)\geqslant 0$. Time integrals may be deleted by partial integration or by

$$\int_0^s t^{k-1}\,dt\int_0^t dB(u)f(u)$$
$$=\frac{1}{k}\left[s^k\int_0^s dB(t)f(t)-\int_0^s t^k\,dB(t)f(t)\right],$$

for $k\geqslant 1$. Thus any term in the expansion (11) up to $O(h^3)$ is expressed exclusively by combinations of

$$\int_0^h s^j\,dB(s),\quad j=1,2,$$
$$\int_0^h s^j\,dB(s)\int_0^s t^k\,dB(t),\quad j+k=1,2,$$
$$\int_0^h s^j\,dB(s)\int_0^s t^k\,dB(t)\int_0^t u^m\,dB(u),\quad j+k+m=1,$$
$$\int_0^h s^j\,dB(s)\int_0^s t^k\,dB(t)\int_0^t u^m\,dB(u)\int_0^u v^n\,dB(v),$$
$$j+k+m+n=1,$$

and iterated integrals up to the sixth degree of $dB(s)$ only. The latter are Hermite polynomials of $B(h)$ by Proposition 2 (ii). All of the former are transformed by Itô's formula to the combinations of the type

$$\int_0^h s^j\,dB(s)\int_0^s dB(t)\cdots\int_0^u dB(v)\int_0^v dB(w),$$
$$j\geqslant 0,$$

which are exhausted to $O(h^3)$ by $B(h)$, $I_{10}(h)$, $I_{20}(h)$, $I_{11}(h)$, $I_{21}(h)$, $J(h)$, and $K(h)$ of Sec. III. Only the most lengthy one of these results will suffice for the proof:

$$\int_0^h dB(s)\int_0^s dB(t)\int_0^t dB(u)\int_0^u v\,dB(v)$$
$$=\tfrac{1}{6}B^3(h)I_{10}(h)+\tfrac{1}{2}B(h)I_{12}(h)+\tfrac{1}{2}[h-B^2(h)]$$
$$\times I_{11}(h)-\tfrac{1}{6}I_{13}(h)-\tfrac{1}{2}hB(h)I_{10}(h)+\tfrac{1}{6}h^3.\quad\blacksquare$$

[1] K. Itô, *On Stochastic Differential Equations*, Memoires of Amer. Math. Soc. No. 4 (1951); *Lectures on Stochastic Processes* (Tata Institute, Bombay, 1961).
[2] A. H. Jazwinsky, *Stochastic Processes and Filtering Theory* (Academic, New York, 1970).
[3] A. Friedman, *Stochastic Differential Equations* (Academic, New York, 1975).
[4] T. Hida, *Brown Undô* (Iwanami Shoten, Tokyo, 1975); English translation *Brownian Motion* (Academic, New York, 1980).
[5] B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
[6] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes* (North-Holland, Amsterdam, 1981).
[7] R. Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962).
[8] A. Schenzle and H. Brand, Phys. Rev. A **20**, 1628 (1979).
[9] J. M. Sancho, M. San Miguel, and J. D. Gunton, Phys. Rev. A **26**, 1589 (1982).
[10] H. Gausterer and J. R. Klauder, Phys. Rev. D **33**, 3678 (1986).
[11] M. McClendon and H. Rabitz, Phys. Rev. A **37**, 3479 (1988).
[12] G. Maruyama, Rend. Circ. Mat. Palermo, Ser. 2 **4**, 48 (1955).
[13] N. J. Rao, J. D. Borwankar, and D. Ramkrishna, SIAM J. Control **12**, 124 (1974).
[14] G. N. Mil'shtein, SIAM Theory Prob. Its Appl. **19**, 557 (1974).
[15] D. J. Wright, The digital simulation of stochastic differential equations. IEEE Trans. Autom. Control **AC19**, 75 (1974).
[16] P. E. Kloeden and R. A. Pearson, J. Aust. Math. Soc. Ser. B **20**, 8 (1977).
[17] G. N. Mil'shtein, SIAM Theory Prob. Its Appl. **23**, 396 (1978).
[18] E. Helfand, Bell Syst. Tech. J. **58**, 2289 (1979).
[19] H. S. Greenside and E. Helfand, Bell Syst. Tech. J. **60**, 1927 (1981).
[20] J. R. Klauder and W. P. Petersen, SIAM J. Numer. Anal. **22**, 1153 (1985).
[21] G. N. Mil'shtein, Theory Prob. Its Appl. **30**, 750 (1986).
[22] W. Rümelin, SIAM J. Numer. Anal. **19**, 604 (1982).

[23] Such a large system of SDEs can typically arise as stochastic partial differential equations representing the motion of a continuum undergoing random injection of energy in one of its Fourier modes. The problem is to

obtain transient structures which necessarily vanish on the static, averaged field profiles. Spatial discretization results in a huge system of coupled nonlinear SDEs with an additive noise. It is absolutely necessary to choose an algorithm with the highest accuracy, in order to reduce the total number of integration steps by the largest h. The numerical cost for the accurate evaluation of noise functionals is negligible in this sort of problems.

[24] N. Wiener, The homogeneous chaos, Am. J. Math. **55**, 897 (1938); *Nonlinear Problems in Random Theory* (MIT Press and Wiley, New York, 1958).

[25] K. Itô, J. Math. Soc. Jpn. **3–1**, 156 (1951).

[26] In the Appendix we derive recursive equations for expansion coefficients and solve them on computer with REDUCE. The method seems to have been pioneered by J. Leblond and D. Talay, in *Proceedings of the French–Belgian Statisticians' Congress*, Rouen, November 1986, edited by J. P. Raoult; cf. Ref. 34.

[27] An abstract form of expansion including that for the time variable has been given with convergence proofs for systems of SDEs by Wagner and Platen and by Platen. See E. Platen, in *Lecture Notes in Control and Information Sciences No. 25* (Springer, New York, 1980); An approximation method for a class of Itô processes, Litovsk. Matem. Sb. **21**, 121 (1981).

[28] R. H. Cameron and W. T. Martin, Ann. Math. **48**, 385 (1947).

[29] The Hermite polynomial [denote it $H_n^*(x)$] used in Ref. 25 is related to the present one by $H_n(x) = H_n^*(x/2^{1/2})/2^{n/2}$.

[30] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge U.P., London, 1931).

[31] Rao, Borwankar, and Ramkrishna[13] gave (cf. pp. 136–137 of Ref. 13) the following approximation for their functional $Z_{1n}(h)$:

$$Z_{1n}(h) \equiv \int_0^h dB(s) \int_0^s dB(t)(s-t)$$
$$= 2I_{11}(h) - B(h)I_{10}(h) + \tfrac{1}{2}h^2$$
$$= h^2\zeta/12^{1/2},$$

where ζ forms a SNIRV with $\{\xi_0,\xi_1\}$. This reproduces $\langle Z_{1n}^2\rangle = h^4/12$ precisely, but $\langle B^2(h)Z_{1n}\rangle = h^3/3$ is then ignored entirely as an irrelevant quantity smaller than $O(h^2)$. Thus, an implication of the above approximation

$$I_{11}(h) \doteq h^2[(\xi_0^2 - 1)/4 + (\xi_0\xi_1 + \zeta)/48^{1/2}],$$

gives $\langle I_{11}(h)^2\rangle \doteq h^4/6$ for the precise $h^4/4$.

[32] A. Greiner, W. Strittmatter, and J. Honerkamp, J. Stat. Phys. **51**, 95–108 (1988).

[33] Various Fourier bases may be chosen for $\{\phi_i(s)\}$, which can be different for $\{\xi_i\}$ and $\{\eta_i\}$. For example, Eq. (51) of Ref. 34 gives an expansion of $L(h)$ with $\{\phi_i(s) \propto \cos(\pi is/h)\}$, convergent in the mean-square like the series $\zeta(2)$, the ζ function of Riemann. This is essentially the same as (45). Fourier type bases may likewise be chosen for $\{\phi_i(s)\}$ in the analysis of Sec. III. However, $I_{10}(h) = O(h^{3/2})$ then takes series expressions that converge in the mean square as $\zeta(4)$, requiring unpractically $O(10^2)$ ξ_i's for the use in $O(h^3)$ schemes. Here is the reason to choose the Legendre basis.

[34] P. E. Kloeden and E. Platen, Stochastic Hydrol. Hydraul. **3**, 155 (1989).

Bispinor geometry for even-dimensional space-time

J. P. Crawford
Penn State/Fayette, Uniontown, Pennsylvania 15401

(Received 19 September 1989; accepted for publication 11 April 1990)

The geometric properties of Dirac spinor fields defined over even-dimensional space-time are explored with the aim of formulating the associated nonlinear sigma models. A spinor field Ψ may be uniquely reconstructed from the real bispinor densities $\rho_i = \bar{\Psi}\Gamma_i\Psi$, apart from an overall phase, so that the ρ_i constitute an alternate representation of the physical information contained in Ψ. For space-time of dimension $N = 2n$, the corresponding Dirac spinor has $D = 2^n$ complex components, and the bispinor densities satisfy a system of $(D-1)^2$ homogeneous quadratic algebraic equations. The basis elements of the Clifford algebra $\{\Gamma_i\}$ span a $D^2 = 2^N$-dimensional space whose Cartan metric is flat pseudo-Riemannian; the bispinor densities reside in the $(2D-1)$-dimensional curved subspace induced as an embedding by the algebraic constraints. The explicit geometric structure of the bispinor spaces are examined and found to be generalizations of Robertson–Walker space. In particular, the line element may be written as $dS^2 = D\,d\sigma^2 + \sigma^2\,d\Omega^2$, where $\sigma = \bar{\Psi}\Psi$ is the scalar density and $d\Omega$ is the line element for the homogeneous space: $\mathrm{SU}(D/2,D/2)/S(U(1)\otimes U(D/2-1,D/2))$.

I. INTRODUCTION

Nonlinear sigma models (or, equivalently, harmonic mappings[1]) have been a popular topic in the physics literature since their introduction by Gell-Mann and Levy in 1960,[2] most recently, but certainly not exclusively, in the context of string theory.[3] Aside from their practical and technical attractiveness, nonlinear sigma models also exhibit great aesthetic appeal as the natural generalization of the notion of geodesic, and thereby have some similarity with general relativity. By a nonlinear sigma model, we mean a minimal mapping of space-time into a "target space" whose geometric structure is given. Usually (but not necessarily) the target space is a homogeneous (or coset) space: G/H, G being the isometry group and H being the isotropy group, and one natural extension of these models is to gauge the isometry group.

One of the interesting features of many of these models is the existence of topological solitons. For example, the Skyrme model[4] is essentially a nonlinear sigma model for the meson fields in which the baryons arise as topological solitons (though, admittedly, one must add an additional term to the Lagrangian to stabilize the solitons). The possibility of generating additional particle states (solitons) without the introduction of additional fields is particularly attractive in light of the continuing proliferation of "elementary" particles.

Supersymmetry notwithstanding, our current understanding of the physical realm is based on the complementary notions of matter (e.g., electrons and quarks) and interaction (e.g., electromagnetic forces), the former represented by Dirac (spinor) fields and the latter by Yang–Mills (gauge) fields. Therefore, it seems worthwhile to pursue the construction of potential physical theories based of the idea of a harmonic mapping of the matter fields and then possibly adding the gauge fields of the isometry group, the hope being that the spectrum of states (including especially the solitons) will be rich enough to account for the observed particles without the introduction of an inordinate number of "fundamental" matter fields.

If we are to take this point of view, an immediate and important question arises: What specific geometric form should we choose for the space associated with the matter fields? Fortunately, there is a natural and promising possibility: bispinor geometry. If we assume that the matter fields are fundamentally represented as Dirac spinors Ψ, then the bispinor densities $\rho_i = \bar{\Psi}\Gamma_i\Psi$ reside in a curved subspace of the space spanned by the Clifford algebra. This "bispinor space" may also be viewed as the space of physical observables, for the bispinor densities, by their definition, are quadratic functionals of the quantum wave function.

The intent of this paper is to find and explore the geometric structure of the bispinor space. Not only is this a crucial step in the construction of the associated nonlinear sigma models, but the results are intrinsically interesting in that they elucidate the geometric structure of the space of physical observables of spin-$\frac{1}{2}$ particles. The construction and analysis of the nonlinear sigma models are left for a future investigation.

We organize the material as follows: Section II contains a discussion of Clifford, Fierz, and bispinor algebras, including a discourse on the Dirac spinor metric and automorphism group. This material sets the stage for the general discussion of bispinor geometry in Sec. III where the case of two-dimensional space-time is included as a simple and instructive example. Section IV contains a summary of the main results and a discussion of some outstanding issues.

II. BISPINOR ALGEBRA

We begin with some definitions and properties of Clifford, Fierz, and bispinor algebras and spinors. Though some of this material has appeared elsewhere,[5] this discussion will serve to establish notation and make this exposition reasona-

 0022-2488/90/081991-07$03.00

bly self-contained. A point of note: the physical condition of reality for the bispinor densities and the corresponding Dirac spinor metric and Dirac normalization of the Clifford algebra basis elements are crucial factors for obtaining the Clifford algebra metric and structure factors, the bispinor algebra, and Dirac normalization preserving automorphism group.

A. Real Clifford algebra

The basic defining relation for the Clifford algebra[6] reads:

$$\{\gamma_\mu,\gamma_\nu\} = 2g_{\mu\nu}I, \quad \mu,\nu = 0,1,2,...,N-1, \tag{2.1}$$

where we shall take $g_{\mu\nu}$ to be the metric of an underlying $N = 2n$ even-dimensional space-time, which in Cartesian coordinates has the form diag$(1,-1,...,-1)$. The elements γ_μ (vector) and I (identity and scalar) are the generators of the full Clifford algebra and may be represented faithfully by $D\times D$ matrices where $D = 2^n$. The basis for the *real* Clifford algebra $\mathscr{R}_{1,N-1}$ may be completed by adding elements of the following form to the set:

$$\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-M}} \equiv (1/M!)\epsilon_{\mu_1\mu_2\cdots\mu_N}\gamma^{\mu_{N-M+1}}\gamma^{\mu_{N-M+2}}\cdots\gamma^{\mu_N}, \tag{2.2}$$

where $M = 2,3,...,N$. Then a set of elements forming a basis for the real Clifford algebra may be chosen to be

$$\{\Gamma_i\} \equiv \{I,\gamma_\mu,\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-2}},...,\tilde{\gamma}_\mu,\tilde{\gamma}\}, \tag{2.3}$$

where we have defined $\tilde{\gamma} \equiv \tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-N}}$. However, since we will be concerned exclusively with Dirac spinors in the following, a different normalization for the $\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-M}}$ shall prove to be more convenient [see Eq. (2.8)]. We return to this point after considering Dirac spinors and the Dirac spinor metric.

B. Dirac spinor metric

Spinors may be defined in several different fashions. In the context of Clifford algebra, the spinors are defined to be elements of a left minimal ideal, whereas in the context of group theory we say that the spinors are carriers of the fundamental representation of the group. The connection between the two may be found in the observation that the spinors are the carriers of the fundamental representation of the automorphism group of the Dirac normalized Clifford algebra (see Sec. II F). Here we simply make the pedestrian observation that the spinors are $D\times 1$ complex matrices upon which the $D\times D$ matrices of the Clifford algebra act.

Physically, a spinor represents the quantum wave function of a spin-$\frac{1}{2}$ particle, and as such the physical observables are obtained from real quadratic functionals of these fields. It is the physical requirement of reality that forces us to introduce the spinor metric because the matrices γ_μ are anti-Hermitian for $\mu = 1,2,...,N-1$. We define the spinor metric γ as follows:

Invariant length: $$\sigma \equiv \Psi^+\gamma\Psi \equiv \overline{\Psi}\Psi = \sigma^*, \tag{2.4a}$$

Current density: $$J_\mu \equiv \overline{\Psi}\gamma_\mu\Psi = J_\mu^*. \tag{2.4b}$$

The indicated requirement of reality then gives

$$\gamma^+ = \gamma, \tag{2.5a}$$

$$\gamma^{-1}\gamma_\mu^+\gamma = \gamma_\mu. \tag{2.5b}$$

The unique solution of these equations is found to be $\gamma = \gamma_0$, and we note for future reference that we can always choose a representation in which γ_0 takes the diagonal form:

$$\gamma_0 = \gamma = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} = \gamma^{-1} = \gamma^+, \tag{2.6}$$

where here I is the $(D/2)\times(D/2)$ unit matrix. The correct interpretation of this equation requires comment. In particular, if $a,b,c,...$ represent spinor indices so that Ψ^a are the spinor components, then γ_0 has matrix components labeled as ${}^a[\gamma_0]_b$ whereas the spinor metric γ has matrix components labeled as ${}_a[\gamma]_b$ and the Dirac conjugate spinor has components $\overline{\Psi}_a = \Psi^{*b}{}_b[\gamma]_a$.

Now consider the remaining bispinor densities. We now easily find [see Eq. (2.2)]

$$(\overline{\Psi}\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-M}}\Psi)^* = (-1)^{M(M-1)/2}\overline{\Psi}\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-M}}\Psi, \tag{2.7}$$

and therefore the bispinor densities defined in this way are not all real. Thus, it is appropriate to redefine the basis elements of the algebra as follows:

$$\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-M}} \equiv (i^{M(M-1)/2}/M!)\epsilon_{\mu_1\mu_2\cdots\mu_N}\gamma^{\mu_{N-M+1}}\cdots\gamma^{\mu_{N-1}}\gamma^{\mu_N}, \tag{2.8}$$

and with this normalization we have

$$\overline{\Gamma}_i \equiv \gamma^{-1}\Gamma_i^+\gamma = \Gamma_i, \tag{2.9a}$$

$$\rho_i \equiv \overline{\Psi}\Gamma_i\Psi = \rho_i^*. \tag{2.9b}$$

Hence, defined in this way, all of the basis elements $\{\Gamma_i\}$ are Dirac self-adjoint and therefore all of the bispinor densities $\{\rho_i\}$ are real. Notice that as defined in Eq. (2.8) the $\{\Gamma_i\}$ form a basis for the *complex* Clifford algebra, and we shall refer to this particular choice of basis as Dirac normalization.

As a specific example, consider the familiar case of $N = 4$, where we may compare the Dirac normalized basis elements with those of Bjorken and Drell.[7] We have

$$\begin{aligned} M=4 \quad & \tilde{\gamma} = -i\gamma_5, \quad \text{where } \gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3, \\ M=3 \quad & \tilde{\gamma}_\mu = -\gamma_5\gamma_\mu, \\ M=2 \quad & \tilde{\gamma}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\lambda\kappa}\sigma^{\lambda\kappa}, \quad \text{where } \sigma^{\lambda\kappa} = (i/2)[\gamma^\lambda,\gamma^\kappa]. \end{aligned} \tag{2.10}$$

This example suggests the following complementary definition to Eq. (2.8):

$$\tilde{\gamma}_{\mu_1\mu_2\cdots\mu_{N-M}} \equiv (1/M!)\epsilon_{\mu_1\mu_2\cdots\mu_N}\gamma^{\mu_{N-M+1}\cdots\mu_{N-1}\mu_N}, \tag{2.11a}$$

$$\gamma_{\mu_1\mu_2\cdots\mu_M} = [(-1)^{M(N-M)+1}/(N-M)!] \times \epsilon_{\mu_1\mu_2\cdots\mu_N}\tilde{\gamma}^{\mu_{M+1}\mu_{M+2}\cdots\mu_N}. \tag{2.11b}$$

The choice between Eq. (2.8) and Eq. (2.11b) is strictly a matter of convenience, and often a subset of each is most useful.

C. Clifford algebra metric and structure factors

Henceforth we shall always use Dirac normalization [Eq. (2.8)] as our defining relation for the basis elements of the Clifford algebra. The dimension of the Clifford algebra (total number of elements in the basis) is easily found to be $d = 2^N = D^2$, and the natural metric on this space (defined à la Cartan[8]) is given by

$$G_{ij} \equiv (1/D)\mathrm{tr}(\Gamma_i \Gamma_j), \tag{2.12}$$

which is found to have the signature of SO($d/2$,$d/2$). Notice that the choice of Dirac normalization for the basis elements is crucial in determining the signature of this metric. The metric may be used to raise and lower the indices of the Γ_i in the usual manner.

We will define the structure factors for the Clifford algebra as follows:

$$\Gamma_i \Gamma_j \equiv C_{ijk}\Gamma^k \Leftrightarrow C_{ijk} = (1/D)\mathrm{tr}(\Gamma_i \Gamma_j \Gamma_k), \tag{2.13a}$$

$$\Gamma_i \Gamma_j \Gamma_k \equiv C_{ijkl}\Gamma^l \Leftrightarrow C_{ijkl} = (1/D)\mathrm{tr}(\Gamma_i \Gamma_j \Gamma_k \Gamma_l). \tag{2.13b}$$

Clearly, the four index structure factor is not independent of the three index structure factor and has been defined as a matter of convenience (see Sec. II D). Also note that both of these structure factors are invariant under cyclic permutations of their indices.

D. Bispinor algebra

We are now in position to construct the algebraic system that the bispinor densities satisfy. The basic idea may be stated very simply. The number of bispinor densities that can be constructed from a given spinor [Eq. (2.9b)] is the dimension of the Clifford algebra $d = 2^N = D^2$. However, as the spinors are composed of $D = 2^n$ complex functions, and furthermore, as the overall phase of the spinor does not affect the values of the bispinor densities, only $(2D - 1)$ of the bispinor densities may be considered independent. Therefore the bispinor densities must satisfy a system of $(D-1)^2$ equations, which we will call the bispinor algebra. Note that, strictly speaking, the use of the terminology "bispinor algebra" is incorrect, since this "bispinor algebra" does not form a vector space in which the vector product is defined. However, for the sake of expedient discussion, we will retain this imprecise notation.

The derivation of the bispinor algebra has been presented elsewhere,[5] but we include it here for the sake of completeness. Consider the Fierz rearrangement[9] of the outer product of two elements of the Clifford algebra basis:

$$^a[\Gamma^i]_b \otimes {}^c[\Gamma^j]_d = F^{ij}{}_{kl}\, {}^a[\Gamma^k]_d \otimes {}^c[\Gamma^l]_b, \tag{2.14}$$

where a, b, c, and d are spinor indices. If we now contract Eq. (2.14) with $^b[\Gamma^m]_a \otimes {}^d[\Gamma^n]_c$ and use Eq. (2.12) and (2.13), we obtain

$$F_{ijkl}C_{m\ n}^{\ k\ l} = DG_{im}G_{jn}, \tag{2.15}$$

and similarly,

$$F_{ijkl} = (1/D)C_{kilj}. \tag{2.16}$$

Finally, if we contract Eq. (2.14) with $\bar{\Psi}_a \Psi^b \bar{\Psi}_c \Psi^d$ and use Eq. (2.16), we obtain

$$\rho_i \rho_j = (1/D)C_{kilj}\rho^k \rho^l. \tag{2.17}$$

This system of quadratic homogeneous equations constitutes the bispinor algebra. Note that we are treating the spinors as classical fields so that, in particular, we have not assumed that they anticommute.

As a summary and example, consider again the familiar case of $N = 4$-dimensional space-time. In this case, the Clifford algebra has $d = 16$ basis elements [Eqs. (2.10)] and the spinors are composed of $D = 4$ complex functions. Let us define the bispinor densities $\{\rho_i\}$ as follows:

$$\sigma \equiv \bar{\Psi}\Psi, \quad \Sigma_{\mu\nu} \equiv \bar{\Psi}\sigma_{\mu\nu}\Psi, \quad \pi \equiv \bar{\Psi}\tilde{\gamma}\Psi,$$
$$J_\mu \equiv \bar{\Psi}\gamma_\mu \Psi, \quad K_\mu \equiv \bar{\Psi}\tilde{\gamma}_\mu \Psi. \tag{2.18}$$

Then, a complete set of $(D-1)^2 = 9$ equations may be written as

$$J^\mu J_\mu = \sigma^2 + \pi^2 = -K^\mu K_\mu, \quad J^\mu K_\mu = 0,$$
$$\sigma\Sigma_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\lambda\tau}\{\pi\Sigma^{\lambda\tau} + (J^\lambda K^\tau - J^\tau K^\lambda)\}. \tag{2.19}$$

These equations have been found and discussed previously by several authors.[10,11]

E. Reconstruction theorem

We now prove a theorem concerning the reconstruction of a spinor from the associated bispinor densities.[5,11,12] With the definition

$$\rho \equiv \rho^i \Gamma_i, \tag{2.20}$$

where $\{\rho_i\}$ is any set of functions satisfying the bispinor algebra, we have the identity

$$\rho\Gamma_i\rho = \rho^k\rho^l\Gamma_k\Gamma_i\Gamma_l = \rho^k\rho^l C_{kilj}\Gamma^j = D\rho_i\rho_j\Gamma^j = D\rho_i\rho, \tag{2.21}$$

and here we have made use of Eq. (2.17). Furthermore, any spinor may be written in the following form:

$$\Psi = e^{-i\phi}R\eta, \tag{2.22}$$

with R being an element of the bispinor algebra [as in Eq. (2.20)] and η being an arbitrary constant spinor. If we now construct the bispinor densities,

$$\rho_i = \bar{\Psi}\Gamma_i\Psi = \bar{\eta}R\Gamma_i R\eta = DR_i\bar{\eta}R\eta, \tag{2.23}$$

we see that

$$R_i = \Omega\rho_i, \quad \Omega^2 = \{D\bar{\eta}\rho\eta\}^{-1}, \tag{2.24}$$

and therefore the R_i are uniquely determined from the ρ_i and the arbitrary constant spinor η. In other words, the spinor Ψ may be uniquely reconstructed, apart from an overall phase, from the associated bispinor densities, and therefore the $\{\rho_i\}$ constitute an equivalent representation for the physical content of the spinor. Of course, this must be the case, since the bispinor densities are precisely the densities of the physical observables of the spinor field.

F. Dirac automorphism group

Finally, we close this section with a discussion of the Lie algebra and Lie group associated with the Clifford algebra, the basis elements of the Clifford algebra having Dirac normalization [Eq. (2.8)]. The full automorphism group of the complex Clifford algebra is the group of transformations

leaving Eq. (2.1) form invariant, and therefore is easily seen to be $Gl(D,c)$. However, an element of this group will not in general leave Eq. (2.9a) invariant. The construction of the automorphism group preserving the Dirac normalization follows.

The structure factors for the Lie algebra associated with the Clifford algebra are defined in the usual manner as

$$[\Gamma_i,\Gamma_j] \equiv \iota f_{ijk}\Gamma^k. \tag{2.25}$$

Now observe that with the normalization of Eq. (2.8), the structure factors of the Clifford algebra [Eq. (2.13a)] are not all real, and we may write

$$C_{ijk} \equiv R_{ijk} + \iota I_{ijk}, \tag{2.26}$$

and furthermore,

$$(\overline{\Gamma_i\Gamma_j}) = \Gamma_j\Gamma_i = C^*_{ijk}\Gamma^k = C_{jik}\Gamma^k, \tag{2.27}$$

from which we obtain

$$f_{ijk} = 2I_{ijk}. \tag{2.28}$$

A general element of the Lie group is given by

$$A = e^{\iota\lambda_k\Gamma^k}, \tag{2.29}$$

where the λ_i are real parameters, and therefore

$$\overline{A} = e^{-\iota\lambda_k\Gamma^k} = A^{-1}. \tag{2.30}$$

Now consider the transformation

$$\Gamma_i' = \overline{A}\Gamma_i A. \tag{2.31}$$

These transformations generate automorphisms of the Clifford algebra because the forms of Eqs. (2.1) and (2.8) are invariant. Furthermore, Eq. (2.9a) is also invariant, and therefore these transformations preserve the Dirac self-adjoint property that guarantees that the bispinor densities (physical observables) remain real. Hence these transformations constitute the automorphism group of the Dirac normalized Clifford algebra.

Equivalently, we may consider the transformation of the spinors,

$$\Psi' = A\Psi, \quad \overline{\Psi}' = \overline{\Psi A}, \tag{2.32}$$

so that, as expected,

$$\rho_i' = \overline{\Psi}'\Gamma_i\Psi' = \overline{\Psi A}\Gamma_i A\Psi = \overline{\Psi}\Gamma_i'\Psi. \tag{2.33}$$

As a special case of this equation, we have

$$\sigma' = \overline{\Psi}'\Psi' = \overline{\Psi A}A\Psi = \overline{\Psi}\Psi = \sigma. \tag{2.34}$$

This result justifies the terminology "invariant length" used in Eq. (2.4a), and along with inspection of Eq. (2.6), allows the identification of the group as $U(D/2,D/2)$ (Ref. 13). Finally, observe that the $U(1)$ factor does not effect a transformation of the $\{\Gamma_i\}$, though it does impose a phase shift on Ψ, and therefore the full effective automorphism group maintaining Dirac normalization is identified as $SU(D/2,D/2)$. This will be important for our discussion of the bispinor geometry. Also note that for the case of four-dimensional space-time, this group is the conformal group SU(2,2) (Ref. 14).

III. BISPINOR GEOMETRY

The basic notion of the bispinor space is rather straightforward. As we have seen, the dimension of the Clifford space is $d = 2^N = D^2$, and the physical quantity $\rho = \rho_i\Gamma^i$ is a vector in this space. However, as the bispinor densities ρ_i satisfy a system of $(D-1)^2$ equations (the bispinor algebra), the bispinor densities actually reside in the $(2D-1)$-dimensional subspace induced as an embedding. It is the structure of this space that we now wish to explore.

A. Two-dimensional space-time

We will first consider the case of two-dimensional space-time as a simple example.[15] It has the advantage of being amenable to the intuitive approach to non-Euclidean geometry first outlined by Gauss.[16] However, before constructing the metric on this space, we need to examine the bispinor algebra.

We begin with an explicit representation for the matrices γ_μ satisfying Eq. (2.1):

$$\gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{3.1}$$

and making use of Eq. (2.8), we have

$$\tilde{\gamma} = \frac{i}{2}\epsilon_{\mu\nu}\gamma^\mu\gamma^\nu = i\gamma_0\gamma_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \tag{3.2}$$

so the basis elements for the Clifford algebra and the corresponding bispinor densities are

$$\{\Gamma_i\} = \{I,\gamma_\mu,\tilde{\gamma}\}, \tag{3.3a}$$

$$\{\rho_i\} = \{\sigma,J_\mu,\pi\}. \tag{3.3b}$$

Physically, σ is a scalar density, J_μ is a vector density, and π is a pseudoscalar density.

The bispinor algebra [Eq. (2.17)] for this case consists of only one equation and may be explicitly constructed from Eq. (2.13b) and the representation given in Eq. (3.1) and (3.2). We find

$$J_\mu J^\mu = \sigma^2 + \pi^2. \tag{3.4}$$

Of course, the explicit form of the bispinor algebra may be constructed without recourse to a specific representation of the Clifford algebra.

We see then that with four bispinor densities satisfying one equation, the bispinor space (the space of physical observables) is three dimensional. To obtain the geometric structure of this space, we follow the procedure first set forth by Gauss.[16]

Consider a p-dimensional space P with coordinates $x^i (i = 1,2,...,p)$ and metric G_{ij} so that the distance between two neighboring points dS is obtained from

$$dS^2 = G_{ij}\,dx^i\,dx^j. \tag{3.5}$$

Now consider a q-dimensional space Q (where $q < p$) that is embedded in P and with coordinates y^a $(a = 1,2,...,q)$. In space Q we can write the coordinates of space P (at least locally) as

$$x^i = x^i(y^a), \quad dx^i = \frac{\partial x^i}{\partial y^a}\,dy^a. \tag{3.6}$$

The distance between two neighboring points in Q is obtained by substitution of Eq. (3.6) into Eq. (3.5):

$$dS^2 = G_{ij}\frac{\partial x^i}{\partial y^a}\frac{\partial x^i}{\partial y^b}dy^a\,dy^b$$
$$= g_{ab}\,dy^a\,dy^b, \tag{3.7}$$

and this generates the metric of the embedded space g_{ab} as indicated. The choice of coordinates y^a is dictated by the algebraic constraints (embedding functions):

$$F_A(x^i) = 0, \quad A = q+1, q+2, \ldots, p. \tag{3.8}$$

Specifically, the parameterization in Eq. (3.6) is chosen such that Eqs. (3.8) are satisfied identically.

The metric of the Dirac normalized Clifford algebra is obtained from Eq. (2.12) and we find

$$dS^2 = d\sigma^2 + dJ_\mu\,dJ^\mu - d\pi^2$$
$$= d\sigma^2 + dJ_0^2 - dJ_1^2 - d\pi^2. \tag{3.9}$$

One can now easily verify that the bispinor algebra [Eq. (3.4), the embedding function] is satisfied identically with the following parameterization (choice of coordinates) of the bispinor densities:

$$J_0 = \pm\sigma\cosh(\lambda), \quad J_1 = \pm\sigma\sinh(\lambda)\cos(\phi),$$
$$\pi = \pm\sigma\sinh(\lambda)\sin(\phi). \tag{3.10}$$

Finally, substitution of Eq. (3.10) into Eq. (3.9) generates the metric of the bispinor space:

$$dS^2 = 2d\sigma^2 - \sigma^2[d\lambda^2 + \sinh^2(\lambda)d\phi^2]. \tag{3.11}$$

This is the main result for the case of two-dimensional space-time. Observe that this space is a three-dimensional hyperbolic Robertson–Walker space.[16] More specifically, we may write

$$dS^2 = 2\,d\sigma^2 + \sigma^2\,d\Omega^2, \tag{3.12}$$

where the quantity $d\Omega$ is the line element for the homogeneous space $SO(1,2)/SO(2) \cong SU(1,1)/U(1)$ (Ref. 17).

This example contains the two principle elements of the general result to follow. In particular, the "factorization" of σ from the other bispinor densities as in Eq. (3.10), and the appearance of the Robertson–Walker metric and internal homogeneous space as in Eq. (3.12), are common elements of the general case.

Finally, we comment on the choice of sign in Eq. (3.10). In particular, if $\sigma < 0$, then the minus sign must be chosen to insure that $J_0 > 0$. This may be easily verified as inspection of Eqs. (2.4b) and (2.6) yields $J_0 = \bar{\Psi}\gamma_0\Psi = \Psi^+\Psi$.

B. Even-dimensional space-time

To explore the geometric structure for the general case, we need to examine

$$dS^2 = G_{ij}\,d\rho^i\,d\rho^j, \tag{3.13}$$

where the Dirac normalized Clifford algebra metric G_{ij} is given in Eq. (2.12). If we now attempt to proceed in the same fashion as in the example above, we encounter the rather difficult problem of the general parameterization of the ρ_i such that not only the bispinor algebra [Eq. (2.17)] is satisfied identically, but also the geometric content of the resulting metric is reasonably transparent. Therefore we shall pursue a different method, taking clues from the structure of the metric found in the two-dimensional case.

First, consider the potential "factorization" of σ from the remaining bispinor densities as in Eq. (3.10). To this end, we make the following definitions:

$$i,j \equiv s,1,2,\ldots,D^2-1, \quad I,J \equiv 1,2,\ldots,D^2-1,$$
$$\Gamma_s \equiv I, \quad \rho_s \equiv \bar{\Psi}\Psi \equiv \sigma. \tag{3.14}$$

To find the length of the vector ρ_i, consider

$$\rho_s\rho_s = (1/D)C_{ksls}\rho^k\rho^l = (1/D)G_{kl}\rho^k\rho^l = \sigma^2, \tag{3.15}$$

where we have used Eqs. (2.12), (2.13b), (2.17), and (3.14b). It is now straightforward to obtain the length of ρ_I as

$$G_{IJ}\rho^I\rho^J = G_{ij}\rho^i\rho^j - G_{ss}\rho^s\rho^s = (D-1)\sigma^2 \tag{3.16}$$

so we make the definition (factorization)

$$\rho_I \equiv \pm\sigma\Phi_I \;\Rightarrow\; G_{IJ}\Phi^I\Phi^J = (D-1). \tag{3.17}$$

Next we consider

$$d\rho_I = \pm(\Phi_I\,d\sigma + \sigma\,d\Phi_I), \quad G_{IJ}\Phi^I\,d\Phi^J = 0, \tag{3.18}$$

and substitution of Eq. (3.18) into Eq. (3.13) yields

$$dS^2 = D\,d\sigma^2 + \sigma^2 G_{IJ}\,d\Phi^I\,d\Phi^J$$
$$\equiv D\,d\sigma^2 + \sigma^2\,d\Omega^2. \tag{3.19}$$

So the factorization of σ has already generated an expression similar to Eq. (3.12).

It remains to show that $d\Omega$ is the line element of a homogeneous space $S = G/H$, where G is the isometry group and H is the isotropy group of this subspace.[17] As to the isometry group, it is clear from the discussion in Sec. II F that this is $SU(D/2, D/2)$, the full Dirac automorphism group, since this group acts transitively on the subspace.

Now the isotropy group is the largest group of transformations leaving any one point of a homogeneous space invariant. Though we may work with the scaled bispinor densities $\{\Phi^I\}$ directly, the identification of this group is more easily obtained from an examination of the spinor. First note that the factorization of σ from the remaining bispinor densities may be implemented at the spinor level as follows:

$$\Psi = \sigma^{1/2}\psi, \quad \Phi^I = \pm\bar{\psi}_\pm\Gamma^I\psi_\pm, \tag{3.20}$$

and we may choose the invariant point (the "origin") to be

$$\psi_{0+} = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} \quad \text{or} \quad \psi_{0-} = \begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}. \tag{3.21}$$

Now it is clear that the largest subgroup of the isometry group leaving either point invariant is $S(U(1)\otimes U(D/2-1, D/2))$. As a check, we calculate the dimension of this space as

$$(D^2-1) - (D-1)^2 = 2(D-1), \tag{3.22}$$

which is correct.

A comment is in order concerning the two possible choices for the "origin" in Eq. (3.21). The bispinor space actually splits into two subspaces whose only intersection is the point $\sigma = 0$. However, the geometric structures of these two subspaces are identical, and the isometry group acts transitively in each separate slice (σ = constant) regardless of the sign of σ.

We summarize this result by stating a theorem whose proof is given by the above construction.

Theorem: The geometric structure of the bispinor space is described by the metric $dS^2 = D\,d\sigma^2 + \sigma^2\,d\Omega^2$, where $d\Omega$ is the line element of the homogeneous space $SU(D/2,D/2)/S(U(1)\otimes U(D/2-1,D/2))$.

IV. CONCLUDING REMARKS

We begin this section with a summary of the basic ideas and results. For space-time of dimension $N = 2n$, the Dirac normalized basis elements of the Clifford algebra $\{\Gamma_i\}$ span a $d = 2^N$-dimensional space whose Cartan metric has the signature of $SO(d/2,d/2)$. The corresponding Dirac spinor Ψ has $D = 2^n$ complex components and may be uniquely reconstructed, apart from an overall phase, from the bispinor densities (physical observables) $\rho_i = \bar{\Psi}\Gamma_i\Psi$. Now only $(2D-1)$ of the $d = D^2$ bispinor densities may be considered independent and in fact satisfy a system $(D-1)^2$ quadratic homogeneous equations (bispinor algebra). These equations may be viewed as embedding functions from the space spanned by the Dirac normalized basis elements of the Clifford algebra into the $(2D-1)$ curved subspace in which the bispinor densities reside (bispinor space).

Finally, we have found that the bispinor space is a Robertson–Walker-type space in that it has metric given by Eq. (3.19) where the line element $d\Omega$ is that of the homogeneous space $SU(D/2,D/2)/S(U(1)\otimes U(D/2-1,D/2))$.

As stated in the Introduction, the construction and analysis of the associated nonlinear sigma models has yet to be completed, but several general comments are in order. In particular, the addition of gauge fields of the isometry group now seems to be entirely natural, as this group is simply the automorphism group. Demanding that this group be the gauge group is equivalent to the condition that the physical theory should be independent of the specific representation of the Clifford algebra and that this representation may be chosen locally. Such a gauge theory has already been considered in a different context by several authors.[18]

This also suggests that we consider the inclusion of internal symmetries. In this case, the construction of the bispinor algebra will proceed generally along the same lines as developed here, with the interesting additional feature that the geometric structure may depend upon the particular representation of the internal symmetry group that the Dirac spinor carries. As an example, if we consider the simplest case of two-dimensional space-time with the spinor carrying the fundamental representation of $SU(2)$, then the bispinor algebra is isomorphic to the bispinor algebra for the case of four-dimensional space-time with no internal symmetry, and therefore the bispinor spaces are identical.

A note concerning the case of odd dimensional space-time is also in order. The additional difficulty here is that the basis for the center of the Clifford algebra contains both I and $\tilde{\gamma}$, and this makes the description of the representations slightly more complicated. This problem is currently being addressed, but at this point it appears that the bispinor algebra for space-time of dimension $N = 2n+1$ splits into two disjoint algebras, each one isomorphic to the bispinor algebra for space-time of dimension $N = 2n$, and that the bispinor space for $N = 2n+1$ is identical to two separate $N = 2n$ bispinor spaces.

Finally, we note here with interest, but in passing, that in the case of four-dimensional space-time, the isometry group is the conformal group $SU(2,2)$, and the homogeneous subspace of the bispinor space is geometrically equivalent to positive projective twistor space.[19] This intriguing connection shall be discussed in a future paper.

The currently prevalent folklore dictates that each distinct elementary particle of matter is represented by an individual spinor field. It is hoped that spectrum of states of the sigma model based on bispinor geometry will be rich enough to be considered as an alternative.

ACKNOWLEDGMENTS

The author is indebted to Professor Murat Gunaydin and Professor Don Page for several valuable discussions concerning homogeneous spaces, and grateful to Professor Nicholas Ford, Professor Peter Ostrander, and Professor John Sokol for critically appraising the manuscript while in preparation.

This work was supported in part by a grant from the Penn State/Fayette Organized Research–Scholarly Activities Fund.

[1] The notion of harmonic mapping was first introduced by F. B. Fuller, Proc. Natl. Acad. Sci. USA **40**, 987 (1954). See also J. Eells and J. H. Sampson, Am. J. Math. **86**, 109 (1964); J. Eells and L. Lemaire, Bull. London Math. Soc. **10**, 1 (1978); C. W. Misner; Phys. Rev. D **18**, 4510 (1978).

[2] M. Gell-Mann and M. Levy, Nuovo Cimento **16**, 705 (1960).

[3] The literature on sigma models is vast. For a review of the sigma model approach to string theory, see A. A. Tseytlin, Int. J. Mod. Phys. A **4**, 1257 (1989).

[4] T. H. Skyrme, Proc. R. Soc. London Ser. A **260**, 127 (1961); Nucl. Phys. **31**, 556 (1962). See also G. S. Adkins, C. R. Nappi, and E. Witten, Nucl. Phys. B **228**, 552 (1983); G. S. Adkins and C. R. Nappi, *ibid.* **233**, 109 (1984).

[5] Most of the material contained in Secs. II C–II E has appeared in J. P. Crawford, "Dirac equation for bispinor densities," in *Clifford Algebras and Their Applications in Mathematical Physics*, edited by J. S. R. Chisholm and A. K. Common, NATO ASI Series C (Reidel, Dordrecht, 1986), Vol. 183.

[6] W. K. Clifford, Am. J. Math. **1**, 350 (1878). See also C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U.P., New York, 1954); M. Riesz, *Clifford Numbers and Spinors* (The Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, 1958); I. R. Porteous, *Topological Geometry* (Cambridge U.P., Cambridge, 1981), pp. 240–276; D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics* (Kluwer, Dordrecht, 1987).

[7] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[8] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific, Warsaw, 1977).

[9] W. Pauli, Ann. Inst. H. Poincaré **6**, 109 (1936); M. Fierz, Z. Phys. **104**, 553 (1937). See also V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory* (Addison–Wesley, Reading, MA, 1971), p. 88.

[10] F. A. Kaempffer, Phys. Rev. D **23**, 918 (1981).

[11] J. P. Crawford, J. Math. Phys. **26**, 1439 (1985).

[12] Y. Takahashi, Phys. Rev. D **26**, 2169 (1982); Prog. Theor. Phys. **69**, 369 (1983). See also V. A. Zhelnorovich, Sov. Phys. Dokl. **24**, 899 (1979); A. A. Campolataro, Int. J. Theor. Phys. **19**, 99 and 127 (1980). These articles and that of Ref. 11 consider only the case of four-dimensional space-time. Here and in Ref. 5 the general case of even dimensional space-time is considered.

[13] P. Lounesto, Found. Phys. **11**, 721 (1981).

[14] *Conformal Groups and Related Symmetries: Physical Results and Mathematical Background*, Proceedings of a Symposium Held at the Arnold Sommerfeld Institute for Mathematical Physics, edited by A. O. Barut and H.-D. Doebner (Springer-Verlag, Berlin, 1986).
[15] J. P. Crawford, Proc. Penn. Acad. Sci. **61**, 202 (1987).
[16] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).
[17] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974). See also S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, Orlando, 1978); S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969), Vol. 2.
[18] J. S. R. Chisholm and R. S. Farwell, J. Phys. A: Math. Gen. **22**, 1059 (1989) and references therein; A. O. Barut and J. McEwan, Lett. Math. Phys. **11**, 67 (1986) and references therein; J. McEwan, Nuovo Cimento A **98**, 491 (1987).
[19] R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U.P., Cambridge, 1986), Vol. 2. See also R. Ablamowicz, Z. Oziewicz, and J. Rzewuski, J. Math. Phys. **23**, 231 (1982); M. Nakashima, Lett. Math. Phys. **16**, 237 (1988).

Extension of two-body Dirac equations to general covariant interactions through a hyperbolic transformation

H. W. Crater
University of Tennessee Space Institute, Tullahoma, Tennessee 37388

P. Van Alstine
Pacific Sierra Research Corporation, 12340 Santa Monica Boulevard, Los Angeles, California 90025

(Received 30 October 1989; accepted for publication 14 March 1990)

In previous work, Dirac's constraint mechanics and supersymmetry were used to construct two-body Dirac equations for particles interacting through world scalar and vector potentials. The resulting compatible 16-component wave equations, $\mathscr{S}_1\psi=\gamma_{51}(\gamma_1\cdot(p_1-\tilde{A}_1)+m_1+\tilde{S}_1)\psi=0$, $\mathscr{S}_2\psi=\gamma_{52}(\gamma_2\cdot(p_2-\tilde{A}_2)+m_2+\tilde{S}_2)\psi=0$, yield Schrödinger-like equations in the center-of-momentum (c.m.) system with simpler structure than that possessed by the Bethe–Salpeter equation or any of its standard three-dimensional truncations. For strong interactions, these equations have yielded a relativistic quark model for meson spectroscopy, while for electromagnetic interactions they have yielded a family of exact singlet positronium solutions. This paper uncovers a hyperbolic interaction structure in these equations. This structure is used to generalize these equations to include pseudoscalar, pseudovector, and tensor interactions.

I. INTRODUCTION

In a series of papers,[1-3] we have used Dirac's constraint mechanics and supersymmetry to derive a pair of coupled but compatible relativistic wave equations that generalize Dirac's equation for a single spin one-half particle in an external field to a system of two spinning particles interacting through world scalar and vector potentials. We subsequently applied these equations to the quark–antiquark bound state[4,5]—system thought to be governed primarily by vector and scalar interactions. The potentials that appeared in that application were suggested by phenomenological considerations. However, our equations possess a connection to relativistic quantum field theory in which their relativistic potentials are determined by field theory.[4] A given field theory through its characteristic dependences on the spins of matter and exchanged particles generates a family of relativistic potentials that transform in different ways under the Lorentz group. For example, electrodynamics (in the annihilation channel) generates, through the Fierz transformation, pseudoscalar and psuedovector pieces of the interaction in addition to vector and scalar interactions. Thus, if two-body Dirac equations really do exist for a given field theory, it must be possible to construct such equations for interactions and systems of interactions beyond the scalar and vector. In a recent paper, Sadzjian[6] showed how to achieve this formally by proposing two-body Dirac equations of the form

$$(\gamma_1\cdot p_1+m_1-(\gamma_2\cdot p_2-m_2)\mathscr{U})\Psi=0, \tag{1a}$$

$$(\gamma_2\cdot p_2+m_2-(\gamma_1\cdot p_1-m_1)\mathscr{U})\Psi=0. \tag{1b}$$

These equations are weakly compatible for arbitrary $\mathscr{U}$. Sadzjian then parametrized his choices for $\mathscr{U}$ as $\mathscr{U}\equiv\gamma_{51}\gamma_{52}\Delta=S$, $\gamma_{51}\gamma_{52}\mathscr{P}$, $\gamma_{\mu 1}\gamma_2^{\mu}\mathscr{V}$, $\gamma_{51}\gamma_{52}\gamma_{\mu 1}\gamma_2^{\mu}A$, $\sigma_{1\mu\nu}\sigma_2^{\mu\nu}\mathscr{T}$, for scalar, pseudoscalar, vector, pseudovector, and tensor interactions, structures suggested directly by the corresponding interactions of the Bethe–Salpeter equation. On the other hand, the relativistic potential structures of our equations are two-body extensions of ordinary external potentials of the one-body Dirac equation (to which they automatically degenerate in the infinite-mass limit). In this paper, we develop a new form of our equations for scalar interactions closely related to the form we generated with the use of supersymmetry. This new form of the equations depends on an invariant matrix function Δ that can be extended to interactions beyond the scalar (by modifying the matrix structure of Δ). We then investigate the connection between Sazdjian's coupled Dirac equations and ours. We find that the two sets of equations are connected by a transformation that depends nonlinearly on the potential (in fact, a hyperbolic transformation). Because of this nonlinear dependence, Sadzjian's equations are equivalent to ours for arbitrary interactions only in the weak potential limit. However, due to the matrix structure of the interactions we find that for interactions that do not possess spacelike parts (like scalar, pseudoscalar, timelike vector, and timelike pseudovector) the equations are identical after a redefinition of the invariant potential function [i.e., $\Delta\rightarrow\tanh(\Delta)$ since, for those interactions, Δ^{2n+1} and Δ have the same matrix coefficients, those matrices being square roots of unity]. For interactions that contain spacelike pieces (and whose matrix coefficients are not roots of unity), the higher-order terms in our equation produce additional pieces of the interaction with a different Lorentz character than those appearing in Sazdjian's equation. For example, our vector equation, in the next to lowest order in Δ, produces what turns out to be a pseudovector interaction in addition to the vector interaction of Sazdjian's approach. In the same fashion, our new form of the equations explains (through its hyperbolic structure) Sadzjian's need to include unusual cubic structures in his treatment of spacelike interactions in order to produce

 0022-2488/90/081998-17$03.00

equations devoid of complicated momentum-dependent tensor terms in the corresponding reduced Schrödinger-like forms. Furthermore, because of their hyperbolic structure, the conserved norm defined by our equations does not contain nonlinear dependences on the interaction such as appear in the norm defined by Sazdjian's equations.[7,8]

In Sadzjian's form, the equations are almost trivially compatible (weakly) but difficult to reduce to the usual heavy-particle limits without additional direction from field theory. In our equations, the interactions take the form of external potentials (hence, we will say that our equations are in "external potential form") but render the verification of compatibility more difficult. One would like to be able to realize Sazdjian's compatibility for arbitrary forms of interaction—but in the equations written in our external potential form (so that they easily generate the usual the heavy particle limits). We find that we can achieve this result by generalizing the spin-dependent potential structure Δ that serves as the argument of the nonlinear transformation that connects our equations with Sazdjian's. As a result, we are able to extend the interaction structures of our equations to include pseudoscalar, pseudovector, and tensor interactions.

In order to make the particular forms of our equations corresponding to each of these interactions intelligible to the reader, we first review the basic variables and structures that appear in the analogous wave equations for spinless particles. We then review an algebraic formalism for dealing with manipulations of combinations of Dirac matrices that one encounters in the construction of two-body Dirac equations. We then use the method to construct the known equations that govern scalar and vector interactions. After noting the presence of hyperbolic structure in these equations, we use its generalization to construct two-body Dirac equations for a collection of relativistic interactions of physical interest.

II. REVIEW OF CONSTRAINT MECHANICS: SPINLESS PARTICLES

Following Todorov,[3,9] we use the following dynamical and kinematical variables in the constraint description of the relativistic two-body problem:

(i) Relative position, $x_1 - x_2$.

(ii) Relative momentum, $p = (1/w)(\epsilon_2 p_1 - \epsilon_1 p_2)$.

(iii) Total c.m. energy, $w = \sqrt{-P^2}$.

(iv) Total momentum, $P = p_1 + p_2$.

(v) (Conserved) constituent c.m. energies,

$$\epsilon_1 = (w^2 + m_1^2 - m_2^2)/2w,$$

$$\epsilon_2 = (w^2 + m_2^2 - m_1^2)/2w.$$

(In terms of these, $p_1 = \epsilon_1 \hat{P} + p$, $p_2 = \epsilon_2 \hat{P} - p$, where $\hat{P} = P/w$.)

(vi) Relativistic reduced mass and energy of a fictitious particle of relative motion,

$$m_w = m_1 m_2/w, \quad \epsilon_w = (w^2 - m_1^2 - m_2^2)/2w.$$

(vii) On-shell value of the relative momentum squared,

$$b^2(w) = \epsilon_w^2 - m_w^2 = \epsilon_1^2 - m_1^2 = \epsilon_2^2 - m_2^2$$
$$= (1/4w^2)(w^4 - 2w^2(m_1^2 + m_2^2) + (m_1^2 - m_2^2)^2).$$

To see how one introduces relativistic dynamics into these kinematic forms through the constraint approach[3,10-14] consider first two-body Klein–Gordon equations. In quantum constraint mechanics these equations are treated as simultaneous generalized mass shell conditions on the wave function,

$$\mathscr{H}_1\psi = (p_1^2 + m_1^2 + \Phi_1(x,p_1,p_2))\psi = 0, \tag{2a}$$

$$\mathscr{H}_2\psi = (p_2^2 + m_2^2 + \Phi_2(x,p_1,p_2))\psi = 0. \tag{2b}$$

Classically, $\mathscr{H}_1$ and $\mathscr{H}_2$ are covariant constraints on the dynamical variables $p_1, p_2, x = x_1 - x_2$. The system Hamiltonian $\mathscr{H} = \lambda_1\mathscr{H}_1 + \lambda_2\mathscr{H}_2$ generates motion in an evolution parameter τ. The classical requirement that $\mathscr{H}_1$ and $\mathscr{H}_2$ be conserved in τ implies that the Poisson bracket, $\{\mathscr{H}_1, \mathscr{H}_2\}$, vanish on the surface of solution (weakly). The corresponding quantum condition—that $[\mathscr{H}_1, \mathscr{H}_2]$ vanish on ψ—guarantees that the original Klein–Gordon equations (operator mass-shell conditions) form a compatible set without additional operator conditions on the wave function. This compatibility condition restricts the way in which two-particle interactions can appear in such equations. In particular, one finds that the vanishing of this commutator as an operator (strong compatibility) requires that the interaction function depend only on the component of x perpendicular to the total four momentum:[3,10-12]

$$x_\perp^\mu = (g^{\mu\nu} + \hat{P}^\mu\hat{P}^\nu)x_\nu, \tag{3}$$

where $\hat{P} = P/w$, $P = p_1 + p_2$, $w^2 = -P^2$. One further finds that the interaction functions (referred to here as quasipotentials) must be equal,

$$\Phi_1 = \Phi_2 = \Phi_w(x_\perp, p_1, p_2), \tag{4}$$

a relativistic analog of Newton's third law. The invariant

$$r \equiv \sqrt{x_\perp^2} \tag{5}$$

appearing in these potentials is the spatial interparticle separation only in the center-of-momentum system. The fact that x may only appear as $x_\perp$ means that constraint mechanics controls the relative time in a covariant way. Although the quasipotential Φ_w may depend on other invariant combinations of $x_\perp, p_1, p_2$, in this paper, we limit our attention to invariant functions of r and $l \equiv \sqrt{(x_\perp \times p)^2}$, where $(a\times b)^\mu = \epsilon^{\nu\kappa\lambda\mu}\hat{P}_\nu a_\kappa b_\lambda$.

The two quantum constraint equations can be recombined in two independent ways. Differencing leads to a wave equation that controls the relative energy:

$$(\mathscr{H}_1 - \mathscr{H}_2)\psi = 2P\cdot p\psi = 0. \tag{6}$$

The other independent combination, which we define[3,10] as $\mathscr{H} \equiv (\epsilon_2\mathscr{H}_1 + \epsilon_1\mathscr{H}_2)/w$, leads to the Schrödinger-like form

$$\mathscr{H}\psi = 0 \rightarrow (p^2 + \Phi_w)\psi = b^2(w)\psi = (\epsilon_w^2 - m_w^2)\psi. \tag{7}$$

In the c.m. system, Eq. (6) implies the relation

$$p^2\psi = p_\perp^2\psi = p^{-2}\psi,$$

so that, in this frame, Eq. (7) has a three-dimensional Schrödinger-like form. However, Eq. (7) is a fully covariant equation (with a c.m. energy-dependent potential).

In our applications, we have taken the point of view that

each particle's Φ_i is constructed from constituent scalar and vector potentials produced by the presence of the other particle. We introduce vector and scalar interactions through the minimal momentum and mass substitutions

$$p_1^\mu = p^\mu + \epsilon_1 \hat{P}^\mu \to p_1^\mu - A_1^\mu$$
$$= G_1(r,l)p^\mu + E_1(r,l)\hat{P}^\mu \equiv \pi_1^\mu, \tag{8a}$$
$$p_2^\mu = -p^\mu + \epsilon_2 \hat{P}^\mu \to p_2^\mu - A_2^\mu$$
$$= -G_2(r,l)p^\mu + E_2(r,l)\hat{P}^\mu \equiv \pi_2^\mu, \tag{8b}$$
$$m_1 \to m_1 + S_1 \equiv M_1(r,l), \tag{9a}$$
$$m_2 \to m_2 + S_2 \equiv M_2(r,l), \tag{9b}$$

performed on $\mathcal{H}_i^0 = p_i^2 + m_i^2$. These minimal substitutions are straightforward two-body extensions of those generated by standard covariant coupling of a single particle to vector or scalar external fields. Thus the original free-mass shell forms

$$\mathcal{H}_i^0 = p_i^2 + m_i^2, \quad i = 1,2 \tag{10}$$

become

$$\mathcal{H}_i = \pi_i^2 + M_i^2, \quad i = 1,2. \tag{11}$$

This procedure determines the Φ_i in (2a) and (2b) in terms of constituent vector and scalar potentials.

The decomposition of Eqs. (8a)–(9b) associates the invariant functions G_1, G_2 with spacelike vector potentials, E_1, E_2 with timelike vector potentials, and M_1, M_2 with scalar potentials. These six scalar functions are not independent. In fact, the assumption of separate "third law" conditions on the scalar (S), timelike (TL) and spacelike (SL) vector parts,

$$\Phi_{1S} = \Phi_{2S}, \tag{12}$$
$$\Phi_{1A(TL)} = \Phi_{2A(TL)}, \tag{13}$$
$$\Phi_{1A(SL)} = \Phi_{2A(SL)}, \tag{14}$$

implies that $M_1^2 - M_2^2 = m_1^2 - m_2^2$, $E_1^2 - E_2^2 = \epsilon_1^2 - \epsilon_2^2$, and $G_1^2 = G_2^2 \equiv G^2$ so that there are only two invariant functions for the vector interaction, one for the scalar. When vector interactions are generated through the coupling to a field—as in QED—there will be further relations among the potentials. In that case, both E_i and G_i become functions of an underlying (generalized) Coulombic potential $\mathscr{A}$. In our applications of this technique (for particles with spin) to the phenomenological treatment of the meson system, we allowed for the presence of both a short-distance (electromagneticlike) or gauge-vector (containing both timelike and spacelike parts) and a long distance timelike vector by parametrizing E_i in terms of two invariant functions $\mathscr{A}$ and $\mathscr{V}$, and G in terms of $\mathscr{A}$ alone. Thus we took

$$A_i^\mu = A_i^\mu(\mathscr{A}(r,l), \mathscr{V}(r,l)). \tag{15}$$

For scalar interactions,

$$S_i = S_i(S(r,l), \mathscr{A}(r,l)). \tag{16}$$

The dependence of S_i on $\mathscr{A}$ is such that each S_i vanishes if S, the underlying scalar interaction vanishes.

III. CONSTRAINT MECHANICS FOR TWO SPIN ONE-HALF PARTICLES: A REVIEW OF THE RESULTS OF SUPERSYMMETRY

In a recently published work,[3] we showed how to use supersymmetry to find compatible Dirac operators for two spinning particles interacting through a system of relativistic scalar and vector interactions. (We had studied the scalar interaction alone in earlier work.[1-2]) In order to examine the connection of the resulting forms to those of Sadzjian,[6] we review these results. For two spin one-half particles, we start from two (compatible) free Dirac equations in the forms

$$\mathscr{S}_{10}\psi = (\theta_1 \cdot p_1 + m_1 \theta_{51})\psi = 0, \tag{17a}$$
$$\mathscr{S}_{20}\psi = (\theta_2 \cdot p_2 + m_2 \theta_{52})\psi = 0, \tag{17b}$$

which become

$$\mathscr{S}_{10}\psi = (\theta_1 \cdot p + \epsilon_1 \theta_1 \cdot \hat{P} + m_1 \theta_{51})\psi = 0, \tag{17a'}$$
$$\mathscr{S}_{20}\psi = (-\theta_2 \cdot p + \epsilon_2 \theta_2 \cdot \hat{P} + m_2 \theta_{52})\psi = 0, \tag{17b'}$$

respectively, in the effective one-body variables of Todorov. [In Eqs. (17'), we have divided the momenta p_i into timelike and spacelike parts $p_1 = \epsilon_1 \hat{P} + p$ and $p_2 = \epsilon_2 \hat{P} - p$.] We have written the matrix coefficients of these ordinary Dirac equations not in terms of gamma matrices but in terms of products of gamma matrices whose algebraic properties permit more efficient calculation of the commutation relations appropriate to two spinning bodies. These "theta" matrices

$$\theta_i^\mu \equiv i\sqrt{\tfrac{1}{2}}\gamma_{5i}\gamma_i^\mu, \quad \mu = 0,1,2,3, \quad i = 1,2, \tag{18a}$$
$$\theta_{5i} \equiv i\sqrt{\tfrac{1}{2}}\gamma_{5i} \tag{18b}$$

satisfy the fundamental anticommutation relations

$$[\theta_i^\mu, \theta_i^\nu]_+ = -g^{\mu\nu}, \tag{19a}$$
$$[\theta_{5i}, \theta_i^\mu]_+ = 0, \tag{19b}$$
$$[\theta_{5i}, \theta_{5i}]_+ = -1. \tag{19c}$$

[Projected theta matrices then satisfy

$$[\theta_i \cdot \hat{P}, \theta_i \cdot \hat{P}]_+ = 1, \tag{20a}$$
$$[\theta_i \cdot \hat{P}, \theta_{i\perp}^\mu]_+ = 0, \tag{20b}$$

where $\theta_{i\perp}^\mu = \theta_{i\nu}(g^{\mu\nu} + \hat{P}^\mu \hat{P}^\nu)$.] The algebraic significance of the theta matrices in the dynamical description provided by Eqs. (17a) and (17b) is that they ensure that the Dirac operators $\mathscr{S}_{10}$ and $\mathscr{S}_{20}$ are exact operator square roots of the corresponding mass-shell forms $-\frac{1}{2}(p_1^2 + m_1^2)$ and $-\frac{1}{2}(p_2^2 + m_2^2)$. [On occasion we will use covariant (c.m. projected) versions of the Dirac α and β matrices here defined by

$$\beta_i = -\gamma_i \cdot \hat{P} = 2\theta_{5i}\theta_i \cdot \hat{P}, \tag{21}$$
$$\alpha_i^\mu = 2\theta_{i\perp}^\mu \theta_i \cdot \hat{P}, \tag{22}$$

and

$$\sigma_i^{\mu\nu} = \gamma_{5i}\alpha_i^\mu = i2\sqrt{2}\theta_{5i}\theta_i \cdot \hat{P}\theta_{\perp i}. \tag{23}$$

We have introduced the important but unfamiliar theta matrices[15] in order to take advantage of their remarkable algebraic properties to simplify the otherwise complicated consequences of compatibility ($[\mathscr{S}_1, \mathscr{S}_2]_- \psi = 0$) when interactions are present. The fundamental anticommutation relations Eqs. (19a)–(19c) of the theta matrices lead directly through a "pseudoclassical" correspondence limit[2] to a

graded symplectic structure in which the thetas become two commuting sets of Grassmann variables. This space possesses a graded Poisson bracket that takes the differential form of Berezin and Marinov.[15] When quantized, this bracket becomes a generalized quantum bracket that is sometimes a commutator, sometimes an anticommutator depending on the nature of its operator arguments. In terms of this bracket, all necessary commutation relations involving the quantum thetas can be carried out through operations that are isomorphic to those involving the classical brackets. The quantum bracket apes the classical graded structure by sorting quantum operators into even and odd classes. For dynamical variables A_α and A_β that have well-defined character (odd or even) with respect to each spin or Grassmann space, the generalized quantum bracket takes the form

$$[A_\alpha,A_\beta]_{-\eta_{\alpha\beta}} = A_\alpha A_\beta - \eta_{\alpha\beta}A_\beta A_\alpha, \tag{24}$$

where[2] $\eta_{\alpha\beta} = (-)^{(\epsilon_{\alpha1}\epsilon_{\beta1}+\epsilon_{\alpha2}\epsilon_{\beta2})}$. The variable $\epsilon_{\alpha1}$ is 0 if A_α is even in space one (like $p,x,\theta_{51}\theta_1\cdot\hat{P}$) and is 1 if A_α is odd in space one (like $\theta_1\cdot x,\theta_{51}\theta_2\cdot\hat{P}$). (Similarly, $\epsilon_{\alpha2}$ keeps track of parity in space two.) Note that the last variable is then odd in both spaces—doubly odd. This sorts the variables into those that are even in both spaces, odd in both, even in space one while odd in space two, and odd in space one while even in space two. In addition, there are additive combinations that do not have well-defined character (e.g., $\theta_1\cdot x + x\cdot p,\theta_1\cdot x + \theta_2\cdot p$). For expressions that contain only one set of spin variables, when inserted as pairs of arguments of the quantum bracket, for two even variables, or one odd and one even, $-\eta_{\alpha\beta} = -$ and the bracket is a commutator. For two odd variables, $-\eta_{\alpha\beta} = +$ and the bracket is an anticommutator. We define the product quantum bracket such that the bracket of $A_\alpha A_\beta$ with A_γ is

$$[A_\alpha A_\beta,A_\gamma]_{-\eta_{\alpha\gamma}\eta_{\beta\gamma}}.$$

This implies that within the Grassmann space of a single particle, the product of an odd with an odd is an even, the product of an even with an odd is an odd, and that the product of an even with an even is an even. Using the definition in (24), one finds that

$$[A_\alpha A_\beta,A_\gamma]_{-\eta_{\alpha\gamma}\eta_{\beta\gamma}} = A_\alpha[A_\beta,A_\gamma]_{-\eta_{\beta\gamma}} + \eta_{\beta\gamma}[A_\alpha,A_\gamma]_{-\eta_{\alpha\gamma}}A_\beta \tag{25}$$

We now use this bracket to construct pairs of compatible Dirac equations for interacting particles. Consider what happens when we attempt to introduce scalar interactions into the two free-particle equations (17′). If we make the minimal substitutions (8) of the spinless case, we do not obtain compatible two-body Dirac equations. That is, in the brackets (25),

$$\mathscr{S}_1\psi = (\theta_1\cdot p + \epsilon_1\theta_1\cdot\hat{P} + M_1\theta_{51})\psi = 0, \tag{26a}$$

$$\mathscr{S}_2\psi = (-\theta_2\cdot p + \epsilon_2\theta_2\cdot\hat{P} + M_2\theta_{52})\psi = 0 \tag{26b}$$

produce an operator that does not vanish on ψ:

$$[\mathscr{S}_1,\mathscr{S}_2]_-\psi = [\theta_1\cdot p,M_2\theta_{52}] + [M_1\theta_{51},-\theta_2\cdot p]_- = -i(\partial M_1\cdot\theta_1\theta_{52} + \partial M_2\cdot\theta_2\theta_{51})\psi \neq 0.$$

In our earlier work, we used supersymmetry arguments to extend the naive $\mathscr{S}_1$ and $\mathscr{S}_2$ to forms that are compatible. The procedure we used had four steps. (a) We found supersymmetries of the pseudoclassical limit of an ordinary free one-body Dirac equation. (b) We introduced interactions of a single Dirac particle with external potentials that preserved these supersymmetries. For scalar interactions this required the coordinate replacement

$$x^\mu \to \tilde{x}^\mu \equiv x^\mu + i\{\theta^\mu\theta_5/[m + S(\tilde{x})]\}.$$

(Since the Grassmann variables satisfy $\theta^2 = 0$ this self-referent relation has a terminating Taylor expansion.) (c) We found that if we maintained the one-body supersymmetries for each spinning particle through the replacement

$$x_1^\mu = (x_1 - x_2)_1^\mu \to (\tilde{x}_1 - \tilde{x}_2)_1^\mu$$

in the relativistic potentials S_i, we obtained compatible classical constraints. (d) Finally, we canonically quantized these constraints to obtain compatible two-body Dirac equations of the form

$$\mathscr{S}_1\psi = (\theta_1\cdot p + \epsilon_1\theta_1\cdot\hat{P} + M_1\theta_{51} - i\partial L_1\cdot\theta_2\theta_{52}\theta_{51})\psi = 0, \tag{27a}$$

$$\mathscr{S}_2\psi = (-\theta_2\cdot p + \epsilon_2\theta_2\cdot\hat{P} + M_2\theta_{52} + i\partial L_2\cdot\theta_1\theta_{52}\theta_{51})\psi = 0. \tag{27b}$$

Note that the requirement of compatibility leads to terms of relatively simple structure [in fact, recoil terms which are functions of $x_\perp$ that add to the naive forms of (26a) and (26b)]. Here we will assume the existence of such terms and use the requirement of compatibility to determine relations among the coefficients (potentials) that multiply them. The only additional nontrivial commutators needed to check compatibility ($[\mathscr{S}_1,\mathscr{S}_2]_-\psi = 0$) are

$$[-i\partial L_1\cdot\theta_2\theta_{51}\theta_{52},-\theta_2\cdot p]_- = (i\partial L_1\cdot p + \tfrac{1}{2}\partial^2 L_1)\theta_{51}\theta_{52},$$

$$[\theta_1\cdot p,\partial L_2\cdot\theta_1\theta_{51}\theta_{52}]_- = (-i\partial L_2\cdot p - \tfrac{1}{2}\partial^2 L_2)\theta_{51}\theta_{52},$$

$$[-i\partial L_1\cdot\theta_2\theta_{51}\theta_{52},M_2\theta_{52}]_- = iM_2\theta_{51}\partial L_1\cdot\theta_2,$$

$$[M_1\theta_{51},+i\partial L_2\cdot\theta_1\theta_{51}\theta_{52}]_- = iM_1\theta_{52}\partial L_2\cdot\theta_1,$$

$$[\epsilon_1\theta_1\cdot\hat{P},+i\partial L_2\cdot\theta_1\theta_{51}\theta_{52}]_- = [\epsilon_1\theta_1\cdot\hat{P},+i\partial L_2\cdot\theta_1]_+\theta_{51}\theta_{52} = -i\epsilon_1\hat{P}\cdot\partial L_2\theta_{51}\theta_{52} = 0,$$

$$[-i\partial L_1\cdot\theta_2\theta_{51}\theta_{52},\epsilon_2\theta_2\cdot\hat{P}]_- = -[-i\partial L_2\cdot\theta_2,\epsilon_2\theta_2\cdot\hat{P}]_+\theta_{51}\theta_{52} = -i\epsilon_2\hat{P}\cdot\partial L_1\theta_{51}\theta_{52} = 0.$$

[The last two commutators vanish since $L_i = L_i(x_\perp)$ and $\hat{P}\cdot x_\perp = 0$.] Note that we have used the product rule (25) to determine whether to compute commutators or anticommutators. Collecting coefficients of independent matrices, we find the simple differential equations

$$\partial M_1 = M_2\partial L_1, \tag{28a}$$

$$\partial M_2 = M_1\partial L_2, \tag{28b}$$

$$\partial L_1 = \partial L_2. \tag{29}$$

(Note that in the static-limit $\partial L \to 0$ so that each of our equations reduces to the standard one-body equation for interaction with an external scalar potential.) If we solve these while identifying the free-particle rest-masses $M_i(L=0) = m_i$, we obtain

$$M_1 = m_1 \operatorname{ch} L + m_2 \operatorname{sh} L \tag{30a}$$

$$M_2 = m_2 \operatorname{ch} L + m_1 \operatorname{sh} L. \tag{30b}$$

In our earlier supersymmetry treatment, the recoil terms at the end of Eqs. (27a) and (27b) appeared as the quantum remnants of the classical Grassmann–Taylor expansion of the mass potential generated by its argument—the supersymmetric position variable $\tilde{x}_\perp$. Note that this solution of the compatibility condition implies

$$M_1^2 - M_2^2 = m_1^2 - m_2^2, \tag{31}$$

the third law condition already discussed for spinless particles. As we found in our treatment of the scalar interaction using supersymmetries, such methods reduce the problem of compatibility for spinning particles to those conditions that are already needed for compatibility for spinless ones.

Elsewhere,[3] we have extended our supersymmetric treatment to the case of timelike vector interactions. Just as in the scalar case, the naive replacement

$$\epsilon_i \to E_i(r,l)$$

does not lead to compatible two-body Dirac equations. That is, when

$$\mathscr{S}_1\psi = (\theta_1 \cdot p + E_1\theta_1 \cdot \hat{P} + m_1\theta_{51})\psi = 0,$$

$$\mathscr{S}_2\psi = (-\theta_2 \cdot p + E_2\theta_2 \cdot \hat{P} + m_2\theta_{52})\psi = 0,$$

we find

$$[\mathscr{S}_1, \mathscr{S}_2]_-\psi$$
$$= [\theta_1\partial_p, E_2\theta_2 \cdot \hat{P}]_- + [E_1\theta_1 \cdot \hat{P}, -\theta_2\partial_p]_-$$
$$= -i(\partial E_1 \cdot \theta_1\theta_2 \cdot \hat{P} + \partial E_2 \cdot \theta_1\theta_1 \cdot \hat{P})\psi \neq 0.$$

This time enforcement of supersymmetries for each spinning particle leads to the recoil-corrected forms

$$\mathscr{S}_1\psi = (\theta_1 \cdot p + E_1\theta_1 \cdot \hat{P} + m_1\theta_{51} + i\partial J_1 \cdot \theta_2\theta_2 \cdot \hat{P}\theta_1 \cdot \hat{P})\psi = 0, \tag{32a}$$

$$\mathscr{S}_2\psi = (-\theta_2 \cdot p + E_2\theta_2 \cdot \hat{P} + m_2\theta_{52} - i\partial J_2 \cdot \theta_1\theta_1 \cdot \hat{P}\theta_2 \cdot \hat{P})\psi = 0. \tag{32b}$$

The requirement of compatibility ($[\mathscr{S}_1, \mathscr{S}_2]_-\psi = 0$), then yields the simple differential equations:

$$\partial E_1 = E_2\partial J_1, \tag{33a}$$

$$\partial E_2 = E_1\partial J_2, \tag{33b}$$

$$\partial J_1 = \partial J_2. \tag{34}$$

(Note again that in the static limit $\partial J \to 0$, so that each of our equations reduces to the standard one-body equation for a spinning particle in an external timelike vector potential.) Solution of these equations with identification of the usual free-particle energies $E_i(J=0) = \epsilon_i$ then gives

$$E_1 = \epsilon_1 \operatorname{ch} J + \epsilon_2 \operatorname{sh} J, \tag{35a}$$

$$E_2 = \epsilon_2 \operatorname{ch} J + \epsilon_1 \operatorname{sh} J. \tag{35b}$$

Just as for the scalar interaction, the recoil terms at the end of Eqs. (32a) and (32b) appear in the supersymmetric treatment as the quantum remnants of the Grassmann–Taylor expansion of the energy potential generated by its argument—the supersymmetric position variable $\tilde{x}_\perp$. Note that this solution of the compatibility condition leads to

$$E_1^2 - E_2^2 = \epsilon_1^2 - \epsilon_2^2, \tag{36}$$

the third law condition for spinless particles interacting through a timelike vector potential.

Finally, for spacelike interaction, the naive choices

$$\mathscr{S}_1\psi = (G_1\theta_1 \cdot p + \epsilon_1\theta_1 \cdot \hat{P} + m_1\theta_{51})\psi = 0,$$

$$\mathscr{S}_2\psi = (-G_2\theta_2 \cdot p + \epsilon_2\theta_2 \cdot \hat{P} + m_2\theta_{52})\psi = 0,$$

prove incompatible:

$$[\mathscr{S}_1, \mathscr{S}_2]\psi = [G_1\theta_1 \cdot p, -G_2\theta_2 \cdot p]_-$$
$$= (i(G_2\partial G_1 \cdot \theta_1\theta_2 \cdot p - G\partial G_2 \cdot \theta_2\theta_1 \cdot p))\psi \neq 0.$$

However,

$$\mathscr{S}_1\psi = (G_1\theta_1 \cdot p + \epsilon_1\theta_1 \cdot \hat{P} + m_1\theta_{51} + i\theta_2 \cdot \partial G_1\theta_{1\perp} \cdot \theta_{2\perp})\psi = 0, \tag{37a}$$

$$\mathscr{S}_2\psi = (-G_2\theta_2 \cdot p + \epsilon_2\theta_1 \cdot \hat{P} + m_2\theta_{52} - i\theta_1\partial G_2\theta_{1\perp} \cdot \theta_{2\perp})\psi = 0 \tag{37b}$$

are compatible provided that

$$G_1\partial G_2 = G_2\partial G_1. \tag{38}$$

Thus G_1 and G_2 differ by at most a multiplicative constant. If the corresponding Dirac equations are to become the usual free-particle Dirac equations when the interaction vanishes, G_1 and G_2 must each become unity (in this limit). Hence, the constant of proportionality must be one. Thus

$$G_1 = G_2 \equiv G. \tag{39}$$

(Note that an attempt to introduce additional spacelike vector interactions by including the additional piece $i\partial Q \cdot \theta_1$ in $\mathscr{S}_1$ and $-i\partial Q \cdot \theta_2$ in $\mathscr{S}_2$ would not alter the vector interactions already introduced through G because such terms can be eliminated by the scale change $\psi \to e^Q\psi$.)

When both scalar and timelike four-vector interactions are present,[3] the compatible two-body Dirac equations turn out to be

$$\mathscr{S}_1\psi = (\theta_1 \cdot p + E_1\theta_1 \cdot \hat{P} + M_1\theta_{51} + i\partial J \cdot \theta_2\theta_1 \cdot \hat{P}\theta_2 \cdot \hat{P} - i\partial L \cdot \theta_2\theta_{51}\theta_{52})\psi = 0, \tag{40a}$$

$$\mathscr{S}_2\psi = (-\theta_2 \cdot p + E_2\theta_2 \cdot \hat{P} + M_2\theta_{52} - i\partial J \cdot \theta_1\theta_1 \cdot \hat{P}\theta_2 \cdot \hat{P} + i\partial L \cdot \theta_1\theta_{52}\theta_{51})\psi = 0, \tag{40b}$$

in which M_1, M_2, E_1, E_2, L, and J are related by Eqs. (30) and (35).

When all three interactions are turned on at once, the solutions (28)–(30), (33)–(35), and (38)–(39) yield the compatible two-body Dirac equations[3]

$$\mathscr{S}_1\psi = (G\theta_1 \cdot p + E_1\theta_1 \cdot \hat{P} + M_1\theta_{51} + iG(\theta_2 \cdot \partial \ln G\theta_{1\perp} \cdot \theta_{2\perp} + \theta_2 \cdot \partial J\theta_1 \cdot \hat{P}\theta_2 \cdot \hat{P} - \theta_2 \cdot \partial L\theta_{51}\theta_{52}))\psi = 0, \tag{41a}$$

$$\mathscr{S}_2\psi=(-G\theta_2\cdot p+E_2\theta_2\cdot\hat{P}+M_2\theta_{52}$$
$$-iG(\theta_1\cdot\partial\ln G\theta_{1\perp}\cdot\theta_{2\perp}$$
$$+\theta_1\cdot\partial J\theta_1\cdot\hat{P}\theta_2\hat{P}-\theta_1\cdot\partial L\theta_{51}\theta_{52})\psi=0. \tag{41b}$$

Note that the requirement of compatibility generates three spin-dependent recoil terms at the end of each Dirac equation, which can be written compactly as

$$\theta_2\cdot\partial\begin{pmatrix}-L\theta_{51}\theta_{52}\\ J\theta_1\cdot\hat{P}\theta_2\cdot\hat{P}\\ \ln G\theta_{1\perp}\cdot\theta_{2\perp}\end{pmatrix}, \tag{42a}$$

$$\theta_1\cdot\partial\begin{pmatrix}L\theta_{51}\theta_{52}\\ -J\theta_1\cdot\hat{P}\theta_2\cdot\hat{P}\\ -\ln G\theta_{1\perp}\cdot\theta_{2\perp}\end{pmatrix}. \tag{42b}$$

The physically important case of electromagneticlike interactions (related timelike and spacelike interactions) deserves special mention. In that case, our compatible two-body Dirac equations reduce to

$$\mathscr{S}_1\psi=(G\theta_1\cdot p+E_1\theta_1\cdot\hat{P}+m_1\theta_{51}$$
$$+i\theta_2\cdot\partial G\theta_1\cdot\theta_2)\psi=0, \tag{43a}$$
$$\mathscr{S}_2\psi=(-G\theta_2\cdot p+E_2\theta_2\cdot\hat{P}+m_2\theta_{52}$$
$$-i\theta_1\cdot\partial G\theta_1\cdot\theta_2)\psi=0, \tag{43b}$$

in which the compatibility restrictions of (35) and (39) lead to

$$E_1=G(\epsilon_1-\epsilon_2)/2+w/2G, \tag{44a}$$
$$E_2=G(\epsilon_2-\epsilon_1)/2+w/2G. \tag{44b}$$

Note that in (43), the recoil terms have combined to yield the characteristic factor $\theta_1\cdot\theta_2=\theta_{1\perp}\cdot\theta_{2\perp}-\theta_1\cdot\hat{P}\theta_2\cdot\hat{P}$.

IV. GENERALIZATION OF THE SUPERSYMMETRIC FORMS TO OTHER COVARIANT INTERACTIONS

So far, we have been able to determine appropriate modifications of free Dirac equations that lead to compatible two-body Dirac equations in the presence of scalar, timelike, and spacelike vector interactions. The resulting dynamical forms of the two-body Dirac equations are identical to their one-body counterparts in corresponding external fields except for the presence of recoil terms [see Eqs. (42a) and (42b)] that vanish when either of the particles becomes very heavy. But, how can we determine the corresponding corrections needed for the construction of compatible Dirac equations containing pseudoscalar, timelike pseudovector, spacelike pseudovector, and tensor interactions? Unfortunately, our earlier treatments of the scalar and vector interactions employed alterations of classical relativistic properties—minimal mass and four-momentum substitutions—not available for the pseudovector and pseudoscalar interactions. However, we see that regardless of the details of origin of the interaction terms for each of the cases treated so far (supersymmetry, minimal substitution) they share a common algebraic—in fact, hyperbolic structure. In each case, the interactions are generated by hyperbolic functions of the potential whose gradient determines the magnitude of the corresponding recoil term. [As we saw in our derivation of Eqs. (28)–(31) and (33)–(36), these structures arise from the solution of the compatibility problem and enforce generalized third law conditions on the interactions.] As we shall see, if we use the hyperbolic structure to rewrite our solutions for $\mathscr{S}_i$'s for the three interactions introduced so far in a compact form, we find that such hyperbolic structures can be readily generalized to incorporate their axial counterparts as well as the tensor interactions.

We first note that using Eqs. (28)–(30), the scalar equations (27a) and (27b) can be written in the form

$$\mathscr{S}_1\psi=(\theta_1\cdot p+\epsilon_1\theta_1\cdot\hat{P}+(m_1\,\mathrm{ch}\,L+m_2\,\mathrm{sh}\,L)\theta_{51}$$
$$-i\theta_2\cdot\partial L\theta_{51}\theta_{52})\psi, \tag{45a}$$
$$\mathscr{S}_2\psi=(-\theta_2\cdot p+\epsilon_2\theta_2\cdot\hat{P}+(m_2\,\mathrm{ch}\,L+m_1\,\mathrm{sh}\,L)\theta_{52}$$
$$+i\theta_1\cdot\partial L\theta_{51}\theta_{52})\psi. \tag{45b}$$

These two Dirac equations can be brought to a more general form through the introduction of the matrix

$$\mathscr{O}_1=2\theta_{51}\theta_{52}, \tag{46}$$

which is a root of unity, $\mathscr{O}_1^2=1$, that is odd in each theta space. We then rewrite (45a) and (45b) as

$$\mathscr{S}_1\psi=(\mathscr{S}_{10}+m_1(\mathrm{ch}(2\Delta)-1)\theta_{51}+m_2\,\mathrm{sh}(2\Delta)\theta_{52}$$
$$+i\theta_2\cdot\partial\Delta)\psi, \tag{47a}$$
$$\mathscr{S}_2\psi=(\mathscr{S}_{20}+m_2(\mathrm{ch}(2\Delta)-1)\theta_{52}+m_1\,\mathrm{sh}(2\Delta)\theta_{51}$$
$$-i\theta_1\cdot\partial\Delta)\psi, \tag{47b}$$

where

$$\Delta=-\mathscr{O}_1L/2. \tag{48}$$

If we rearrange these equations, we find that the combinations

$$\mathbf{S}_1\psi=(\mathrm{ch}(\Delta)\mathscr{S}_1-\mathrm{sh}(\Delta)\mathscr{S}_2)\psi=0, \tag{49a}$$
$$\mathbf{S}_2\psi=(\mathrm{ch}(\Delta)\mathscr{S}_2-\mathrm{sh}(\Delta)\mathscr{S}_1)\psi=0 \tag{49b}$$

take the general forms

$$\mathbf{S}_1\psi=(\mathscr{S}_{10}\,\mathrm{ch}(\Delta)+\mathscr{S}_{20}\,\mathrm{sh}(\Delta))\psi=0, \tag{50a}$$
$$\mathbf{S}_2\psi=(\mathscr{S}_{20}\,\mathrm{ch}(\Delta)+\mathscr{S}_{10}\,\mathrm{sh}(\Delta))\psi=0, \tag{50b}$$

after we have used simple hyperbolic identities and brought the matrices on the left of each $\mathscr{S}_i$ to the right.

Since the new constraints (49a) and (49b) are nothing but algebraic rearrangements of linear combinations of the old compatible constraints $\mathscr{S}_i$, they must themselves be compatible. However, we shall verify the compatibility explicitly. We already know that the constraints $\mathscr{S}_1$ and $\mathscr{S}_2$ are compatible $[\mathscr{S}_1,\mathscr{S}_2]_-\psi=0$. The commutator $[\mathbf{S}_1,\mathbf{S}_2]_-$ is a sum of four commutators. The first,

$$[\mathrm{ch}(\Delta)\mathscr{S}_1,\mathrm{ch}(\Delta)\mathscr{S}_2]_-$$
$$=\mathrm{ch}(\Delta)(\mathrm{ch}(\Delta)[\mathscr{S}_1,\mathscr{S}_2]_-+[\mathscr{S}_1,\mathrm{ch}(\Delta)]_-\mathscr{S}_2$$
$$+[\mathrm{ch}(\Delta),\mathscr{S}_2]_-\mathscr{S}_1)\approx 0, \tag{51a}$$

vanishes weakly [we need to use the constraints $\mathscr{S}_i\approx 0$ (i.e., $\mathscr{S}_1\psi=0$)]. Likewise,

$$[\mathrm{sh}(\Delta)\mathscr{S}_2,\mathrm{sh}(\Delta)\mathscr{S}_i]_-\approx 0. \tag{51b}$$

We are left with

$$
\begin{aligned}
&-[\mathrm{ch}(\Delta)\mathscr{S}_1,\mathrm{sh}(\Delta)\mathscr{S}_1]_- - [\mathrm{sh}(\Delta)\mathscr{S}_2,\mathrm{ch}(\Delta)\mathscr{S}_2]_- \\
&\quad = -(\mathrm{ch}(\Delta)(\mathrm{sh}(\Delta)[\mathscr{S}_1,\mathscr{S}_1]_- \\
&\qquad + [\mathscr{S}_1,\mathrm{sh}(\Delta)]_-\mathscr{S}_1) \\
&\qquad - \mathrm{sh}(\Delta)[\mathrm{ch}(\Delta),\mathscr{S}_1]_-\mathscr{S}_1) + (1\to 2) \approx 0. \quad (51c)
\end{aligned}
$$

Thus

$$[\mathbf{S}_1,\mathbf{S}_2]_- \approx 0.$$

We now conjecture that the constraints in the general forms (50a) and (50b) are the proper forms for the introduction of relativistic interactions in the sense that all interactions known to us appear simply as choices for the invariant form Δ. It is our aim to find compatible constraints of the "external potential" form $\mathscr{S}_1$ and $\mathscr{S}_2$ for more general interactions from the $\mathbf{S}_1$ and $\mathbf{S}_2$ constraints. To do this, we must first show that the new forms ($\mathbf{S}_1$ and $\mathbf{S}_2$) in (50a) and (50b) are compatible for arbitrary $\Delta(x_\perp)$. Then, the proof that the constraints $\mathscr{S}_1$ and $\mathscr{S}_2$ which we uncover from the $\mathbf{S}_1$ and $\mathbf{S}_2$ constraints are compatible is isomorphic to the proof given in (51) but with the roles of $\mathscr{S}_i$ and $\mathbf{S}_i$ interchanged.

The new forms (50a) and (50b) of the two-body Dirac equations can be related to ones recently proposed by Sazdjian[6] of the form (here written in terms of the theta matrices)

$$S_1\Psi = (\mathscr{S}_{10} + \mathscr{S}_{20}\mathscr{W})\Psi, \quad (52a)$$

$$S_2\Psi = (\mathscr{S}_{10} + \mathscr{S}_{10}\mathscr{W})\Psi. \quad (52b)$$

If we identify

$$\Psi = \mathrm{ch}(\Delta)\psi \quad (53a)$$

and

$$\mathscr{W} = \mathrm{th}(\Delta), \quad (53b)$$

then they are in equivalent form. Note, however, that in Sazdjian's equations for a given interaction (say vector) the potential $\mathscr{W}$ has a simple matrix structure (i.e., $\theta_1\cdot\theta_2$). On the other hand, when our equations are written in his form with a Δ that has the same matrix structure, our $\mathscr{W}$ may contain additional matrix structure since the hyperbolic tangent is a nonlinear function of Δ. These additional terms will not appear for interactions whose matrix structures are roots of unity since, for such interactions (e.g., scalar, pseudoscalar, timelike vector and pseudovector), the matrix structures of Δ and $\mathrm{th}(\Delta)$ are the same. But, for those interactions whose Δ's are not multiples of roots of unity (i.e., those for spacelike interactions), our equations are not equivalent to Sadzjian's. Hence, in general, Sazdjian's form of the two-body Dirac equations is a weak-potential version (small Δ) of ours (50a) and (50b). Now, Sazdjian's forms of the two-body Dirac equations [and our generalized version (52a) and (52b) are compatible for arbitrary $\mathscr{W}$ provided that $\mathscr{W} = \mathscr{W}(x_\perp)$. We use this fact to show that our general hyperbolic forms (50a) and (50b) are compatible for arbitrary $\Delta(x_\perp)$. [We have slightly altered Sadzjian's proof of compatibility of (52a) and (52b).] First, note that the relative energy constraint $P\cdot p\Psi = 0$ follows from

$$
\begin{aligned}
(\mathscr{S}_{10}S_1 - \mathscr{S}_{20}S_2)\Psi &= (\mathscr{S}_{10}^2 - \mathscr{S}_{20}^2)\Psi \\
&= -\tfrac{1}{2}(p_1^2 + m_1^2 - p_2^2 - m_2^2)\Psi \\
&= -P\cdot p\Psi = 0. \quad (54)
\end{aligned}
$$

In order to demonstrate the (weak) compatibility of the two constraints, one must calculate

$$
\begin{aligned}
[S_1,S_2]\Psi &= [\mathscr{S}_{10},\mathscr{S}_{20}]_-\Psi \\
&\quad + [\mathscr{S}_{10},\mathscr{S}_{10}\mathscr{W}]_-\Psi + [\mathscr{S}_{20}\mathscr{W},\mathscr{S}_{20}]_-\Psi \\
&\quad + [\mathscr{S}_{20}\mathscr{W},\mathscr{S}_{10}\mathscr{W}]_-\Psi \\
&= (\mathscr{S}_{10}^2 - \mathscr{S}_{20}^2)\mathscr{W}\Psi - \mathscr{S}_{10}\mathscr{W}\mathscr{S}_{10}\Psi \\
&\quad - \mathscr{S}_{10}\mathscr{W}\mathscr{S}_{20}\mathscr{W}\Psi \\
&\quad + \mathscr{S}_{20}\mathscr{W}\mathscr{S}_{20}\Psi + \mathscr{S}_{20}\mathscr{W}\mathscr{S}_{10}\mathscr{W}\Psi \\
&= -P\cdot p\mathscr{W}(x_\perp)\Psi - \mathscr{S}_{10}\mathscr{W}\mathscr{S}_1\Psi \\
&\quad + \mathscr{S}_{20}\mathscr{W}\mathscr{S}_2\Psi.
\end{aligned}
$$

Using (52a) and (52b), (54), and $[P\cdot p,\mathscr{W}(x_\perp)]_- = 0$, one then finds that each of the terms vanishes. Thus S_1 and S_2 are weakly compatible.

Next we show that compatibility of Sadzjian's constraints ($[S_1,S_2]_-\Psi = 0$) plus the constraints themselves ($S_i\Psi = 0$) imply the compatibility of our forms (50a) and (50b): $[\mathbf{S}_1,\mathbf{S}_1]\psi = 0$. First, we observe that

$$\mathbf{S}_i\psi = S_i\,\mathrm{ch}(\Delta)\psi = S_i\Psi.$$

Therefore,

$$
\begin{aligned}
[\mathbf{S}_1,\mathbf{S}_2]\psi &= (S_1\,\mathrm{ch}(\Delta)S_2 - S_2\,\mathrm{ch}(\Delta)S_1)\Psi \\
&= [S_1,\mathrm{ch}(\Delta)]S_2\Psi - [S_2,\mathrm{ch}(\Delta)]S_1\Psi \\
&\quad + \mathrm{ch}(\Delta)[S_1,S_2]\Psi = 0. \quad (55)
\end{aligned}
$$

Hence, our forms (50a) and (50b) of the two-body Dirac equations are compatible for arbitrary $\Delta(x_\perp)$.

Now, from (49a) and (49b), we see that the new constraints $\mathbf{S}_i$ are related to our original "external potential" ones by

$$\mathscr{S}_1 = \mathrm{ch}(\Delta)\mathbf{S}_1 + \mathrm{sh}(\Delta)\mathbf{S}_2, \quad (56a)$$

$$\mathscr{S}_2 = \mathrm{ch}(\Delta)\mathbf{S}_2 + \mathrm{sh}(\Delta)\mathbf{S}_1. \quad (56b)$$

Even though we used the scalar interaction to carry out the compatibility check in (51a)–(51c), the proof that the "external potential" $\mathscr{S}_i$ constraints are compatible *for arbitrary* $\Delta(x_\perp)$ given the compatibility of the $\mathbf{S}_i$ *for arbitrary* Δ is virtually identical to (51a)–(51c).

As we shall show for eight invariant forms for $\Delta(x_\perp)$, the corresponding "external potential" form $\mathscr{S}_i$ constraints can actually be written in a form that looks like that of a one-body Dirac equation; that is,

$$
\begin{aligned}
\mathscr{S}_1\psi &= (\mathscr{S}_{10} + Z_1(x_\perp,p))\psi \\
&= (\theta_1\cdot p + \epsilon_1\theta_1\cdot\hat{P} + m_1\theta_{51} + Z_1(x_\perp,p)) \\
&\equiv (\mathscr{Z}(x_\perp)\theta_1\cdot p + \mathscr{Z}_1(x_\perp))\psi = 0, \quad (57a) \\
\mathscr{S}_2\psi &= (\mathscr{S}_{20} + Z_2(x_\perp,p))\psi \\
&= (-\theta_2\cdot p + \epsilon_1\theta_2\cdot\hat{P} + m_2\theta_{52} + Z_2(x_\perp,p)) \\
&= (-\mathscr{Z}(x_\perp)\theta_2\cdot p + \mathscr{Z}_2(x_\perp))\psi = 0. \quad (57b)
\end{aligned}
$$

(Unlike the S_i or $\mathbf{S}_i$ forms of the two-body Dirac equation, these external potential forms have no cross kinetic terms depending on $\theta_j \cdot p$ $i \neq j$. This property simplifies the reduction to Schrödinger-like form.) Even at this stage, we see the importance of the hyperbolic structure of the equations (which the reader will recall emerged automatically for the scalar in our supersymmetry approach) in bringing Eqs. (56a) and (56b) to the "external potential" form through the identity $\mathrm{ch}^2(\Delta) - \mathrm{sh}^2(\Delta) = 1$. Recall also the importance of this structure in guaranteeing a physical principle—the third law.

Such properties of the hyperbolic structure are classical in that they are necessary to guarantee consistency even at the (relativistic) classical level. But, the hyperbolic structure has a relativistic quantum-mechanical consequence as well. In two recent papers,[7-8] Sazdjian has shown how to construct scalar products that accompany his form of the two-body Dirac equations given in Eqs. (52a) and (52b). The result he obtains is (rewritten here in the notation of our paper)

$$\begin{aligned}\langle \Psi_{P',a}, \Psi_{P,b} \rangle &= (2\pi)^3 \delta^3(\vec{P}' - \vec{P}) \\ &\quad \times \int d^3 \left[\Psi_a^\dagger(x) \left(1 - \mathscr{W}^2 - 4w^2 \gamma_{10}\gamma_{20} \frac{\partial \mathscr{W}}{\partial P^2} \right) \Psi_b(x) \right] \\ &= (2\pi)^3 w \delta^3(\vec{P}' - \vec{P}) \delta_{ab} f_a(w).\end{aligned}$$

Note that (as pointed out by Sazdjian) this scalar product is potential-dependent even if $\mathscr{W}$ is energy independent. However, using the transformation (53a) and (53b) and a simple hyperbolic identity we find that the scalar product that accompanies our form (50a) and (50b) of the two-body Dirac equations is given by

$$\begin{aligned}\langle \psi_{P',a}, \psi_{P,b} \rangle &= (2\pi)^3 \delta^3(\vec{P}' - \vec{P}) \\ &\quad \times \int d^3 \left[\psi_a^\dagger(x) \left(1 - 4w^2 \gamma_{10}\gamma_{20} \frac{\partial \Delta}{\partial P^2} \right) \psi_b(x) \right] \\ &= (2\pi)^3 w \delta^3(\vec{P}' - \vec{P}) \delta_{ab} f_a(w).\end{aligned}$$

Note that for energy-independent potentials, this scalar product is of the same potential-independent form as that for the one-body Dirac equation with energy-independent potentials. Perhaps the hyperbolic structure of the two-body Dirac equations will turn out to be a consequence of the requirement that the scalar product take the simple $\psi\dagger\psi$ form for energy-independent potentials.

We now investigate the constraints of external potential form S_i generated by eight choices for $\Delta(x_\perp)$. (The first three of these will merely reproduce our results for scalar, timelike vector, and spacelike vector interactions.) In each case, we first construct the new general hyperbolic constraints (50) and then pass to the corresponding "external potential" constraints through (56). For scalar interactions, we shall verify that the choice

$$\Delta = -\mathscr{O}_1 L(x_\perp)/2 = -[1_1 1_2 L(x_\perp)/2]\mathscr{O}_1 \qquad (48')$$

(where $\mathscr{O}_1 = 2\theta_{51}\theta_{52}$) leads to the result given in (27)–(30). For timelike vector interactions, the choice

$$\Delta = \mathscr{O}_2 J(x_\perp)/2 = [\gamma_1 \cdot \hat{P} \gamma_1 \cdot \hat{P} J(x_\perp)/2]\mathscr{O}_1, \qquad (58)$$

where $\mathscr{O}_2 = 2\theta_1 \cdot \hat{P}\theta_2 \cdot \hat{P}$, will lead to the result (32)–(35). For spacelike vector interaction, the choice

$$\Delta = \mathscr{O}_3 \mathscr{G}(x_\perp)/2 = [(\gamma_{1\perp} \cdot \gamma_{2\perp} \mathscr{G}(x_\perp))/2]\mathscr{O}_1, \qquad (59)$$

where $\mathscr{O}_3 = 2\theta_{1\perp} \cdot \theta_{2\perp}$, will lead to (37)–(39). The matrices $\mathscr{O}_1$, $\mathscr{O}_2$, and $\mathscr{O}_3$ are "doubly odd" (odd in each spin space) and symmetric in the labels of the two spinning particles. Here, $\mathscr{O}_1$ and $\mathscr{O}_2$ are roots of unity, $\mathscr{O}_1^2 = \mathscr{O}_2^2 = 1$. However, since $\mathscr{O}_3 = -\beta_1\beta_1 \sigma_1 \cdot \sigma_2$ and $(\sigma_1 \cdot \sigma_2)^2 = 3 - 2\sigma_1 \cdot \sigma_2$, one finds that

$$\mathscr{O}_3^2 = 3\mathscr{C}_1 + 2\mathscr{C}_2 \mathscr{O}_3,$$

where $\mathscr{C}_1 = 1$, and $\mathscr{C}_2 = 4\theta_{51}\theta_{52}\theta_1 \cdot \hat{P}\theta_2 \cdot \hat{P} = \beta_1\beta_2$. There exists a fourth doubly odd matrix combination

$$\mathscr{O}_4 = \mathscr{C}_2 \mathscr{O}_3,$$

which, like $\mathscr{O}_3$, is not a root of unity ($\mathscr{O}_4^2 = \mathscr{O}_3^2 \neq 1$). So, we uncover a fourth odd interaction, the covariant "polar" part of the full tensor interaction:

$$\Delta = \mathscr{F}(x_\perp)\mathscr{O}_4/2 = [\alpha_1 \cdot \alpha_2 \mathscr{F}(x_\perp)/2]\mathscr{O}_1. \qquad (60)$$

We now construct the two external potential constraints $\mathscr{S}_1$ and $\mathscr{S}_2$ corresponding to each of these four interactions. In this construction, the theta combinations of the Dirac gamma matrices again prove useful. Their characters (even or odd) in the general brackets (24) and (25) dictate whether one should employ commutators or anticommutators to obtain the external potential forms from (56) and (50). First note that for each of these four interactions, $\mathrm{sh}(\Delta)$ is a doubly odd function of the doubly odd variable Δ. Using (24) to guide us to the proper bracket, we obtain

$$\begin{aligned}\mathscr{S}_1 &= \mathrm{ch}(\Delta)\mathbf{S}_1 + \mathrm{sh}(\Delta)\mathbf{S}_2 \\ &= \mathrm{ch}(\Delta)\mathscr{S}_{10}\,\mathrm{ch}(\Delta) + \mathrm{ch}(\Delta)\mathscr{S}_{20}\,\mathrm{sh}(\Delta) \\ &\quad + \mathrm{sh}(\Delta)\mathscr{S}_{20}\,\mathrm{ch}(\Delta) + \mathrm{sh}(\Delta)\mathscr{S}_{10}\,\mathrm{sh}(\Delta) \\ &= \mathrm{ch}^2(\Delta)\mathscr{S}_{10} + \mathrm{ch}(\Delta)[\mathscr{S}_{10},\mathrm{ch}(\Delta)]_- \\ &\quad + \mathrm{ch}(\Delta)[\mathscr{S}_{20},\mathrm{sh}(\Delta)]_+ \\ &\quad + \mathrm{sh}(\Delta)[\mathscr{S}_{20},\mathrm{ch}(\Delta)]_- \\ &\quad + \mathrm{sh}(\Delta)[\mathscr{S}_{10},\mathrm{sh}(\Delta)]_+ - \mathrm{sh}^2(\Delta)\mathscr{S}_{10},\end{aligned} \qquad (61)$$

with a similar expression for $\mathscr{S}_2$. Note how the plus sign in conjunction with the odd–odd nature of $\mathrm{sh}(\Delta)$ combine to give a negative coefficient for $\mathrm{sh}^2(\Delta)$, which in turn allows one to use the simple hyperbolic identity $\mathrm{ch}^2(\Delta) - \mathrm{sh}^2(\Delta) = 1$ in the construction of the external potential form (57a) and (57b) [see (63) below]. We need to compute the four quantum brackets $[\mathscr{S}_{i0},\mathrm{ch}(\Delta)]_-$ and $[\mathscr{S}_{i0},\mathrm{sh}(\Delta)]_+$ for $i = 1,2$. First, we isolate the derivative parts of the constraints by using the product rule (25) to decompose the following parts of these four quantum brackets:

$$\begin{aligned}[\theta_1 \cdot p, \mathrm{ch}(\Delta)]_- &= -i\theta_1 \cdot \partial(\Delta)\mathrm{sh}(\Delta) \\ &\quad + [\theta_1^\mu, \mathrm{ch}(\Delta)]_- p_\mu,\end{aligned} \qquad (62a)$$

$$[-\theta_2\cdot p,\mathrm{ch}(\Delta)]_- = i\theta_2\cdot\partial(\Delta)\mathrm{sh}(\Delta) - [\theta_2^\mu,\mathrm{ch}(\Delta)]_- p_\mu, \quad (62b)$$

$$[\theta_1\cdot p,\mathrm{sh}(\Delta)]_+ = -i\theta_1\cdot\partial(\Delta)\mathrm{ch}(\Delta) + [\theta_1^\mu,\mathrm{sh}(\Delta)]_+ p_\mu. \quad (62c)$$

$$[-\theta_2\cdot p,\mathrm{sh}(\Delta)]_+ = i\theta_2\cdot\partial(\Delta)\mathrm{ch}(\Delta) - [\theta_2^\mu,\mathrm{sh}(\Delta)]_+ p_\mu. \quad (62d)$$

Thus the derivative parts of (61) are

$$\begin{aligned}&\mathrm{ch}(\Delta)(-i\theta_1\cdot\partial(\Delta)\mathrm{sh}(\Delta)+i\theta_2\cdot\partial(\Delta)\mathrm{ch}(\Delta))\\&+\mathrm{sh}(\Delta)(i\theta_2\cdot\partial(\Delta)\mathrm{sh}(\Delta)-i\theta_1\cdot\partial(\Delta)\mathrm{ch}(\Delta))\\&=i\theta_2\cdot\partial(\Delta)-i([\mathrm{ch}(\Delta),\theta_1^\mu]_-\mathrm{sh}(\Delta)\\&-[\mathrm{ch}(\Delta),\theta_2^\mu]_-\mathrm{ch}(\Delta)-[\mathrm{sh}(\Delta),\theta_2^\mu]_+\mathrm{sh}(\Delta)\\&+[\mathrm{sh}(\Delta),\theta_1^\mu]_+\mathrm{ch}(\Delta))\partial_\mu(\Delta).\end{aligned}$$

Note that the choice of commutators versus anticommutators is dictated by the facts that Δ is odd in both particles' theta matrices and that the hyperbolic sine is an odd function (while the hyperbolic cosine is an even function). (Note also that $[\partial(\Delta),\Delta]_- = 0$.) As a result,

$$\begin{aligned}\mathscr{S}_1 = \mathscr{S}_{10} &+ i\theta_2\cdot\partial(\Delta) - i([\mathrm{ch}(\Delta),\theta_1^\mu]_-\mathrm{sh}(\Delta)\\&-[\mathrm{ch}(\Delta),\theta_2^\mu]_-\mathrm{ch}(\Delta)-[\mathrm{sh}(\Delta),\theta_2^\mu]_+\mathrm{sh}(\Delta)+[\mathrm{sh}(\Delta),\theta_1^\mu]_+\mathrm{ch}(\Delta))\partial_\mu(\Delta)\\&+\mathrm{ch}(\Delta)([\theta_1^\mu,\mathrm{ch}(\Delta)]_- p_\mu+[\epsilon_1\theta_1\cdot\hat{P},\mathrm{ch}(\Delta)]_-+[m_1\theta_{51},\mathrm{ch}(\Delta)]_--[\theta_2^\mu,\mathrm{sh}(\Delta)]_+ p_\mu+[\epsilon_2\theta_2\cdot\hat{P},\mathrm{sh}(\Delta)]_+\\&+[m_2\theta_{52},\mathrm{sh}(\Delta)]_+)+\mathrm{sh}(\Delta)(-[\theta_2^\mu,\mathrm{ch}(\Delta)]_- p_\mu+[\epsilon_2\theta_2\cdot\hat{P},\mathrm{ch}(\Delta)]_-+[m_2\theta_{52},\mathrm{ch}(\Delta)]_-\\&+[\theta_1^\mu,\mathrm{sh}(\Delta)]_+ p_\mu+[\epsilon_1\theta_1\cdot\hat{P},\mathrm{sh}(\Delta)]_++[m_1\theta_{51},\mathrm{sh}(\Delta)]_+), \quad (63)\end{aligned}$$

along with a similar expression for $\mathscr{S}_2$.

In each of the brackets of (63) which contain θ_i^μ, that matrix may be replaced by $\theta_{i\perp}$ since it is contracted with either $\partial_\mu f(x_\perp)$ or p_μ (which satisfies $P\cdot p\approx 0$).

Case (i) scalar: $\Delta = -\frac{1}{2}\mathscr{O}_1 L(x_\perp)$. Consequently,

$$\mathrm{ch}(\Delta) = \mathscr{C}_1\,\mathrm{ch}(L/2),\quad \mathrm{sh}(\Delta) = -\mathscr{O}_1\,\mathrm{sh}(L/2). \quad (64)$$

To construct the $\mathscr{S}_i$, one needs to know the elementary brackets:

$$[\mathscr{C}_1,\theta_{i\perp}^\mu]_- = [\mathscr{C}_1,\theta_i\cdot\hat{P}]_- = [\mathscr{C}_1,\theta_{51}]_- = 0, \quad (65)$$

$$[\mathscr{O}_1,\theta_{i\perp}^\mu]_+ = [\mathscr{O}_1,\theta_i\cdot\hat{P}]_+ = 0, \quad (66)$$

$$[\mathscr{O}_1,\theta_{5i}]_+ = -2\theta_{5j},\quad i\neq j. \quad (67)$$

These imply that

$$\begin{aligned}[\mathrm{ch}(\Delta),\theta_{i\perp}^\mu]_- &= [\mathrm{ch}(\Delta),\theta_1\cdot\hat{P}]_-\\&= [\mathrm{ch}(\Delta),\theta_{51}]_- = 0, \quad (68)\\[\mathrm{sh}(\Delta),\theta_{i\perp}^\mu]_+ &= [\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_+\\&= O[\mathrm{sh}(\Delta),\theta_{5i}]_+ = 2\,\mathrm{sh}(\Delta)\theta_{5i}. \quad (69)\end{aligned}$$

To perform the remaining multiplications, we use the facts that

$$\mathscr{O}_1\theta_{51} = -\theta_{52},\quad \mathscr{O}_1\theta_{52} = -\theta_{51},$$

along with hyperbolic identities, to obtain

$$\mathscr{S}_1 = \theta_1\cdot p+\epsilon_1\theta_1\cdot\hat{P}+m_1\,\mathrm{ch}(L\mathscr{O}_1)\theta_{51} - m_2\,\mathrm{sh}(L\mathscr{O}_1)\theta_{52} - i\theta_2\cdot\frac{\partial L}{2}\mathscr{O}_1. \quad (70a)$$

Similarly, we find that

$$\mathscr{S}_2 = -\theta_1\cdot p+\epsilon_2\,\mathrm{th}_1\cdot\hat{P}+m_2\,\mathrm{ch}(L\mathscr{O}_1)\theta_{52} - m_1\,\mathrm{sh}(L\mathscr{O}_1)\theta_{51}+i\theta_1\cdot\frac{\partial L}{2}\mathscr{O}_1. \quad (70b)$$

Since $\mathscr{O}_1$ is a root of unity, $\mathrm{ch}(\mathscr{O}_1 L) = \mathrm{ch}(L)$, $\mathrm{sh}(\mathscr{O}_1 L) = \mathscr{O}_1\,\mathrm{sh}(L)$. Thus these equations are just the scalar equations (27)–(30) that we originally derived through supersymmetric techniques.

Case (ii) timelike four vector: $\Delta = \frac{1}{2}\mathscr{O}_2 J$. Consequently,

$$\mathrm{ch}(\Delta) = \mathscr{C}_1\,\mathrm{ch}(J/2),\quad \mathrm{sh}(\Delta) = \mathscr{O}_2\,\mathrm{sh}(J/2). \quad (71)$$

Carrying out steps similar to those given above for the scalar interaction, we obtain

$$\mathscr{S}_1 = \theta_1\cdot p+\epsilon_1\,\mathrm{ch}(J\mathscr{O}_2)\theta_2\cdot\hat{P}+\epsilon_2\,\mathrm{sh}(J\mathscr{O}_2)\theta_2\cdot\hat{P} + m_1\theta_{51}+i\theta_2\cdot\frac{\partial J}{2}\mathscr{O}_2. \quad (72a)$$

Similarly,

$$\mathscr{S}_2 = -\theta_2\cdot p+\epsilon_2\,\mathrm{ch}(J\mathscr{O}_2)\theta_2\cdot\hat{P}+\epsilon_1\,\mathrm{sh}(J\mathscr{O}_2)\theta_1\cdot\hat{P} + m_2\theta_{52}-i\theta_2\cdot\frac{\partial J}{2}\mathscr{O}_2. \quad (72b)$$

Since $\mathscr{O}_2$ is a root of unity, these equations are just those generated by supersymmetric techniques: (32)–(35).

Case (iii) Spacelike vector: $\Delta = \frac{1}{2}\mathscr{O}_3\mathscr{G}(x_\perp)$. This case is more complex algebraically since $\mathscr{O}_3^2\neq 1$. However, we can write

$$\mathscr{O}_3 = \mathscr{C}_2(\mathscr{C}_1 - 2\mathscr{R}), \quad (73)$$

where

$$\mathscr{R} = \tfrac{1}{2}(\mathscr{C}_1 - \mathscr{O}_4)\,(=\tfrac{1}{2}(1+\sigma_2\cdot\sigma_2)) \quad (74)$$

is a root of unity ($\mathscr{R}^2=1$). Thus

$$\begin{aligned}\mathrm{ch}(\Delta)&=\mathrm{ch}\left(\frac{\mathscr{E}_1\mathscr{G}}{2}-\mathscr{R}\mathscr{G}\right)\\&=\mathrm{ch}\left(\frac{\mathscr{G}}{2}\right)\mathrm{ch}(\mathscr{G})-\mathscr{R}\,\mathrm{sh}\left(\frac{\mathscr{G}}{2}\right)\mathrm{sh}(\mathscr{G})\\&=\mathrm{ch}^3\left(\frac{\mathscr{G}}{2}\right)+\frac{1}{2}\mathscr{O}_4\,\mathrm{sh}\left(\frac{\mathscr{G}}{2}\right)\mathrm{sh}(\mathscr{G}),\end{aligned}\tag{75}$$

$$\begin{aligned}\mathrm{sh}(\Delta)&=\mathscr{E}_2\,\mathrm{sh}(\mathscr{E}_1\mathscr{G}/2-\mathscr{R}\mathscr{G})\\&=\mathscr{E}_2\left(\mathrm{sh}\left(\frac{\mathscr{G}}{2}\right)\mathrm{ch}(\mathscr{G})-\mathscr{R}\,\mathrm{ch}\left(\frac{\mathscr{G}}{2}\right)\mathrm{sh}(\mathscr{G})\right)\\&=\mathscr{E}_2\,\mathrm{sh}^3\left(\frac{\mathscr{G}}{2}\right)+\frac{1}{2}\mathscr{O}_3\,\mathrm{ch}\left(\frac{\mathscr{G}}{2}\right)\mathrm{sh}(\mathscr{G}).\end{aligned}\tag{76}$$

One needs to know the elementary brackets

$$[\mathscr{E}_2,\theta^\mu_{21}]_+=2\theta_{21}\mathscr{E}_2=2\mathscr{E}_2\theta^\mu_{21},\tag{77}$$

$$[\mathscr{E}_2,\theta_i\cdot\hat{P}]_+=[\mathscr{E}_2,\theta_{5i}]_+=0.\tag{78}$$

Since $\mathscr{E}_2\mathscr{O}_4=\mathscr{O}_3$, one also needs to know

$$[\mathscr{O}_3,\theta^\mu_{i1}]_+=-2\theta^\mu_{j1},\quad i\neq j,\tag{79}$$

$$[\mathscr{O}_3,\theta_1\cdot\hat{P}]_+=[\mathscr{O}_3,\theta_{5i}]_+=0.\tag{80}$$

In addition, one must use

$$\begin{aligned}[\mathscr{O}_4,\theta^\mu_{i1}]_-&=-2\theta^\mu_{i1}\mathscr{O}_4+[\mathscr{O}_4,\theta^\mu_{i1}]_+\\&=-2\theta^\mu_{i1}\mathscr{O}_4-2\theta^\mu_{j1}\mathscr{E}_2\\&=2\mathscr{O}_4\theta^\mu_{i1}+2\mathscr{E}_2\theta^\mu_{j1},\quad i\neq j\end{aligned}\tag{81}$$

$$[\theta_i\cdot\hat{P},\mathscr{O}_4]_-=[\theta_{5i},\mathscr{O}_4]_-=0.\tag{82}$$

Consequently,

$$\begin{aligned}&[\mathrm{ch}(\Delta),\theta^\mu_{i1}]_-\\&\quad=-\mathrm{sh}(\mathscr{G}/2)\mathrm{sh}(\mathscr{G})(\theta^\mu_{i1}\mathscr{O}_4+\theta^\mu_{j1}\mathscr{E}_2)\\&\quad=\mathrm{sh}(\mathscr{G}/2)\mathrm{sh}(\mathscr{G})(\mathscr{O}_4\theta^\mu_{i1}+\mathscr{E}_2\theta^\mu_{j1}),\end{aligned}\tag{83}$$

$$[\mathrm{ch}(\Delta),\theta_i\cdot\hat{P}]_-=[\mathrm{ch}(\Delta),\theta_{51}]_-=0,\tag{84}$$

while

$$\begin{aligned}&[\mathrm{sh}(\Delta),\theta^\mu_{i1}]_+\\&\quad=2\,\mathrm{sh}^3(\mathscr{G}/2)\theta^\mu_{i1}\mathscr{E}_2-\mathrm{ch}(\mathscr{G}/2)\mathrm{sh}(\mathscr{G})\theta^\mu_{j1}\\&\quad=2\,\mathrm{sh}^3(\mathscr{G}/2)\mathscr{E}_2\theta^\mu_{i1}-\mathrm{ch}(\mathscr{G}/2)\mathrm{sh}(\mathscr{G})\theta^\mu_{j1},\end{aligned}\tag{85}$$

$$[\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_+=0=[\mathrm{sh}(\Delta),\theta_{5i}]_+.\tag{86}$$

One then uses

$$\mathscr{O}_4\mathscr{E}_2=\mathscr{O}_3,\quad\mathscr{O}_4\mathscr{O}_3=\mathscr{E}_2\mathscr{O}_3^2=3\mathscr{E}_2+2\mathscr{O}_3,$$

$$\mathscr{O}_4^2=3+2\mathscr{O}_4,\tag{87}$$

along with the identities

$$\theta^\mu_{1\perp}\mathscr{O}_3+\theta^\mu_{1\perp}\mathscr{E}_2+\theta^\mu_{21}\mathscr{O}_4+\theta^\mu_{21}=0,\tag{88}$$

$$\mathscr{O}_4\theta^\mu_{1\perp}+\mathscr{E}_2\theta^\mu_{21}+\mathscr{O}_3\theta^\mu_{21}+\theta^\mu_{1\perp}=0,\tag{89}$$

to perform the remaining multiplications. After using numerous hyperbolic identities one finds

$$\begin{aligned}&-[\mathrm{ch}(\Delta),\theta^\mu_{1\perp}]_-\,\mathrm{sh}(\Delta)-[\mathrm{sh}(\Delta),\theta^\mu_{1\perp}]_+\,\mathrm{ch}(\Delta)\\&\quad=2\,\mathrm{sh}(\mathscr{G}/2)\mathrm{ch}(\mathscr{G}/2)\theta^\mu_{21},\end{aligned}\tag{90}$$

$$\begin{aligned}&[\mathrm{ch}(\Delta),\theta^\mu_{21}]_-\,\mathrm{ch}(\Delta)+[\mathrm{sh}(\Delta),\theta^\mu_{21}]_+\,\mathrm{sh}(\Delta)\\&\quad=2\,\mathrm{sh}^2(\mathscr{G}/2)\theta^\mu_{21},\end{aligned}\tag{91}$$

$$\begin{aligned}&\mathrm{ch}(\Delta)[\theta^\mu_{1\perp},\mathrm{ch}(\Delta)]_-+\mathrm{sh}(\Delta)[\theta^\mu_{21},\mathrm{sh}(\Delta)]_+\\&\quad=2\,\mathrm{sh}^2(\mathscr{G}/2)\theta^\mu_{1\perp},\end{aligned}\tag{92}$$

$$\begin{aligned}&-\mathrm{ch}(\Delta)[\theta^\mu_{21},\mathrm{sh}(\Delta)]_+-\mathrm{sh}(\Delta)[\theta^\mu_{21},\mathrm{ch}(\ )]_-\\&\quad=2\,\mathrm{sh}(\mathscr{G}/2)\mathrm{ch}(\mathscr{G}/2)\theta^\mu_{1\perp},\end{aligned}\tag{93}$$

so that

$$\begin{aligned}\mathscr{S}_1&=\exp(\mathscr{G})\theta_1\cdot p+\epsilon_1\theta_1\cdot\hat{P}+m_1\theta_{51}\\&\quad+i\exp(\mathscr{G})\theta_2\cdot\frac{\partial\mathscr{G}}{2}\mathscr{O}_3.\end{aligned}\tag{94a}$$

Similarly,

$$\begin{aligned}\mathscr{S}_2&=-\exp(\mathscr{G})\theta_2\cdot p+\epsilon_2\theta_2\cdot\hat{P}+m_2\theta_{52}\\&\quad-i\exp(\mathscr{G})\theta_1\cdot\frac{\partial\mathscr{G}}{2}\mathscr{O}_3.\end{aligned}\tag{94b}$$

The external potential forms (94a) and (94b) [with $\mathscr{G}=\ln(\mathscr{G})$], are just the constraints (37)–(39) that we had derived earlier.

Case (iv) tensor (Polar): $\Delta=\frac{1}{2}\mathscr{O}_4\mathscr{F}$. In Appendix A, we show our method applied to this interaction yields

$$\begin{aligned}\mathscr{S}_1&=\exp(\mathscr{F}\mathscr{E}_2)\theta_1\cdot p+\epsilon_1\,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_1\cdot\hat{P}\\&\quad+\epsilon_2\,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_2\cdot\hat{P}\\&\quad+m_1\,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_{51}+m_2\,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_{52}\\&\quad+i\exp(\mathscr{F}\mathscr{E}_2)\theta_2\cdot\frac{\partial\mathscr{F}}{2}\mathscr{O}_4,\end{aligned}\tag{95a}$$

$$\begin{aligned}\mathscr{S}_2&=-\exp(\mathscr{F}\mathscr{E}_2)\theta_2\cdot p+\epsilon_2\,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_2\cdot\hat{P}\\&\quad+\epsilon_1\,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_1\cdot\hat{P}\\&\quad+m_2\,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_{52}+m_2\,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_{51}\\&\quad-i\exp(\mathscr{F}\mathscr{E}_2)\theta_1\cdot\frac{\partial\mathscr{F}}{2}\mathscr{O}_4,\end{aligned}\tag{95b}$$

for the two compatible constraints. The pair of Dirac equations (95a) and (95b) for the polar tensor interactions and the four pairs of Dirac equations that we shall derive below for the axial interactions are new forms which accompany the three pairs of Dirac equations that we had found previously through quantization of supersymmetric pseudoclassical forms.

The axial counterparts to the constraints (56a) and (56b) in the case of polar interactions are

$$\mathscr{S}_1=\mathrm{ch}(\Delta)\mathbf{S}_1-\mathrm{sh}(\Delta)\mathbf{S}_2,\tag{96a}$$

$$\mathscr{S}_2=\mathrm{ch}(\Delta)\mathbf{S}_2-\mathrm{sh}(\Delta)\mathbf{S}_1,\tag{96b}$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are still given by (50a) and (50b). Just as for the polar (56a) and (56b), the compatibility of these two constraints follows from that of the $\mathbf{S}_i$. Note that the minus sign [as opposed to the plus sign in (56a) and (56b)] combines with the fact that Δ is even–even for the axial interactions to give a minus sign coefficient for $\mathrm{sh}^2(\Delta)$ [see (101) below], which, in turn, will allow one to use the simple hyperbolic identity $\mathrm{ch}^2(\Delta)-\mathrm{sh}^2(\Delta)=1$ in the construction of the external potential form (57a) and (57b) [see (102) below]. These axial Δ's are

$$\Delta = \frac{\mathscr{C}_1 C}{2} = \frac{C}{2} = \frac{\mathscr{O}_1 C}{2}\mathscr{O}_1 = -\frac{\gamma_{51}\gamma_{52}C}{2}\mathscr{O}_1, \tag{97}$$

for the pseudoscalar interaction,

$$\Delta = -\frac{\mathscr{C}_2 H}{2} = -\frac{\mathscr{O}_2 H}{2}\mathscr{O}_1 = \frac{\gamma_{51}\gamma_1\cdot\hat{P}\gamma_{52}\gamma_2\cdot\hat{P}H}{2}\mathscr{O}_1, \tag{98}$$

for timelike pseudovector interaction,

$$\Delta = -\frac{\mathscr{C}_3 I}{2} = -\frac{\mathscr{O}_3 I}{2}\mathscr{O}_1 = \frac{\gamma_{51}\gamma_{1\perp}\cdot\gamma_{52}\gamma_{2\perp}I}{2}\mathscr{O}_1, \tag{99}$$

for the spacelike pseudovector interaction, and

$$\Delta = -\frac{\mathscr{C}_4 Y}{2} = \frac{\mathscr{O}_4 Y}{2}\mathscr{O}_1 = -\frac{\sigma_1\cdot\sigma_2 Y}{2}\mathscr{O}_1, \tag{100}$$

for the axial part of the tensor interaction. Recall that as in the polar cases C, H, I, and Y are functions of $x_\perp$. We have chosen a minus sign in (96a) and (96b) because $\mathrm{sh}\Delta$ is even in the number of theta matrices. Since this quantity will appear in commutators instead of anticommutators, we find [in contrast to Eq. (61) for polar interactions] that

$$\begin{aligned}\mathscr{S}_1 = {}& \mathrm{ch}^2(\Delta)\mathscr{S}_{10} + \mathrm{ch}(\Delta)[\mathscr{S}_{10},\mathrm{ch}(\Delta)]_- \\ & + \mathrm{ch}(\Delta)[\mathscr{S}_{20},\mathrm{sh}(\Delta)]_- - \mathrm{sh}(\Delta)[\mathscr{S}_{20},\mathrm{ch}(\Delta)]_- \\ & - \mathrm{sh}(\Delta)[\mathscr{S}_{10},\mathrm{sh}(\Delta)]_- - \mathrm{sh}^2(\Delta)\mathscr{S}_{10}.\end{aligned} \tag{101}$$

Steps analogous to those below (61) (with commutators appearing instead of anticommutators at appropriate places) show that the general form of the Dirac operator for the axial interaction analogous to (63) for the polar is

$$\begin{aligned}\mathscr{S}_1 = {}& \mathscr{S}_{10} + i\theta_2\cdot\partial(\Delta) - i([\mathrm{ch}(\Delta),\theta_1^\mu]_-\mathrm{sh}(\Delta) \\ & - [\mathrm{ch}(\Delta),\theta_2^\mu]_-\mathrm{ch}(\Delta) + [\mathrm{sh}(\Delta),\theta_2^\mu]_-\mathrm{sh}(\Delta) - [\mathrm{sh}(\Delta),\theta_1^\mu]_-\mathrm{ch}(\Delta))\partial_\mu(\Delta) \\ & + \mathrm{ch}(\Delta)([\theta_1^\mu,\mathrm{ch}(\Delta)]_- p_\mu + [\epsilon_1\theta_1\cdot\hat{P},\mathrm{ch}(\Delta)]_- + [m_1\theta_{51},\mathrm{ch}(\Delta)]_- - [\theta_2^\mu,\mathrm{sh}(\Delta)]_- p_\mu + [\epsilon_2\theta_2\cdot\hat{P},\mathrm{sh}(\Delta)]_- \\ & + [m_2\theta_{52},\mathrm{sh}(\Delta)]_-) + \mathrm{sh}(\Delta)(+[\theta_2^\mu,\mathrm{ch}(\Delta)]_- p_\mu - [\epsilon_2\theta_2\cdot\hat{P},\mathrm{ch}(\Delta)]_- \\ & - [m_2\theta_{52},\mathrm{ch}(\Delta)]_- - [\theta_1^\mu,\mathrm{sh}(\Delta)]_- p_\mu - [\epsilon_1\theta_1\cdot\hat{P},\mathrm{sh}(\Delta)]_- - [m_1\theta_{51},\mathrm{sh}(\Delta)]_-),\end{aligned} \tag{102}$$

along with a similar expression for $\mathscr{S}_2$.

Case (v) pseudoscalar: $\Delta = \frac{1}{2}C$. Consequently,

$$\mathrm{ch}(\Delta) = \mathrm{ch}(C/2), \quad \mathrm{sh}(\Delta) = \mathrm{sh}(C/2). \tag{103}$$

As a result,

$$\begin{aligned}[\mathrm{ch}(\Delta),\theta_{i\perp}^\mu]_- &= [\mathrm{ch}(\Delta),\theta_i\cdot\hat{P}]_- \\ &= [\mathrm{ch}(\Delta),\theta_{5i}]_- = 0,\end{aligned} \tag{104}$$

$$\begin{aligned}[\mathrm{sh}(\Delta),\theta_{i\perp}^\mu]_- &= [\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_- \\ &= [\mathrm{sh}(\Delta),\theta_{51}]_- = 0,\end{aligned} \tag{105}$$

so that

$$\mathscr{S}_1 = \theta_1\cdot p + \epsilon_1\theta_1\cdot\hat{P} + m_1\theta_{51} + i\theta_2\cdot\partial(C/2). \tag{106a}$$

Similarly,

$$\mathscr{S}_2 = -\theta_2\cdot p + \epsilon_2\theta_2\cdot\hat{P} + m_2\theta_{52} - i\theta_1\cdot\partial(C/2). \tag{106b}$$

Case (vi) Timelike pseudovector: $\Delta = \frac{1}{2}\mathscr{C}_2 H$. Then

$$\mathrm{ch}(\Delta) = \mathrm{ch}(H), \quad \mathrm{sh}(\Delta) = \mathscr{C}_2\,\mathrm{sh}\,H. \tag{107}$$

Thus

$$\begin{aligned}[\mathrm{ch}(\Delta),\theta_{i\perp}^\mu]_- &= [\mathrm{ch}(\Delta),\theta_i\cdot\hat{P}]_- \\ &= [\mathrm{ch}(\Delta),\theta_{5i}]_- = [\mathrm{sh}(\Delta),\theta_{i\perp}^\mu]_- = 0,\end{aligned} \tag{108}$$

$$\begin{aligned}&[\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_- = 2\,\mathrm{sh}(\Delta)\theta_i\cdot\hat{P}, \\ &[\mathrm{sh}(\Delta),\theta_{51}]_- = 2\,\mathrm{sh}(\Delta)\theta_{5i}.\end{aligned} \tag{109}$$

In addition,

$$\begin{aligned}&[\mathscr{C}_2,\theta_{i\perp}^\mu]_- = 0, \quad [\mathscr{C}_2,\theta_i\cdot\hat{P}]_- = 2\mathscr{C}_2\theta_i\cdot\hat{P}, \\ &[\mathscr{C}_2,\theta_{5i}]_- = 2\mathscr{C}_2\theta_{51}\end{aligned} \tag{110}$$

implies that

$$[\mathrm{sh}(\Delta),\theta_{i\perp}^\mu]_- = 0, \tag{111}$$

and that

$$\begin{aligned}&[\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_- = 2\,\mathrm{sh}(\Delta)\theta_i\cdot\hat{P}, \\ &[\mathrm{sh}(\Delta),\theta_{5i}]_- = 2\,\mathrm{sh}(\Delta)\theta_{5i}.\end{aligned} \tag{112}$$

When we substitute these brackets into (102), we find

$$\begin{aligned}\mathscr{S}_1 = {}& \theta_1\cdot p + \epsilon_1\,\mathrm{ch}(H\mathscr{C}_2)\theta_1\cdot\hat{P} + \epsilon_2\,\mathrm{sh}(H\mathscr{C}_2)\theta_2\cdot\hat{P} \\ & + m_1\,\mathrm{ch}(H\mathscr{C}_2)\theta_{51} + m_2\,\mathrm{sh}(H\mathscr{C}_2)\theta_{52} \\ & - i\theta_2\cdot\frac{\partial H}{2}\mathscr{C}_2.\end{aligned} \tag{113a}$$

Similarly,

$$\begin{aligned}\mathscr{S}_2 = {}& -\theta_2\cdot p + \epsilon_2\,\mathrm{ch}(H\mathscr{C}_2)\theta_2\cdot\hat{P} + \epsilon_1\,\mathrm{sh}(H\mathscr{C}_2)\theta_1\cdot\hat{P} \\ & + m_2\,\mathrm{ch}(H\mathscr{C}_2)\theta_{52} + m_1\,\mathrm{sh}(H\mathscr{C}_2)\theta_{51} \\ & + i\theta_1\cdot\frac{\partial H}{2}\mathscr{C}_2.\end{aligned} \tag{113b}$$

Case (vii) spacelike pseudovector: $\Delta = -\frac{1}{2}I(x_\perp)\mathscr{C}_3$. Using the identity $\mathscr{C}_3 = \mathscr{O}_1\mathscr{O}_3 = \mathscr{O}_2(\mathscr{C}_1 - 2\mathscr{R})$, we find that steps similar to those given for the spacelike vector interaction and the polar part of the tensor interaction given in Appendix A yield

$$\begin{aligned}\mathscr{S}_1 = {}& \exp(\mathscr{O}_1 I)\theta_1\cdot p + \epsilon_1\theta_1\cdot\hat{P} + m_1\,\mathrm{ch}(I\mathscr{C}_3)\theta_{51} \\ & + m_2\,\mathrm{sh}(I\mathscr{C}_3)\theta_{52} - i\exp(\mathscr{O}_1 I)\theta_2\cdot\frac{\partial I}{2}\mathscr{C}_3.\end{aligned} \tag{114a}$$

Similarly,

$$\mathscr{S}_2 = -\exp(\mathscr{O}_1 I)\theta_2\cdot p + \epsilon_2\theta_2\cdot\hat{P} + m_2\,\mathrm{ch}(I\mathscr{C}_3)\theta_{52}$$

$$+ m_1 \,\mathrm{sh}(I\mathscr{E}_3)\theta_{51} + i\exp(\mathscr{O}_1 I)\theta_1 \cdot \frac{\partial I}{2}\mathscr{E}_3. \tag{114b}$$

Case (viii) tensor–axial: $\Delta = \frac{1}{2}Y(x_\perp)\mathscr{E}_4$.
Using the identity $\mathscr{E}_4 = \mathscr{O}_2\mathscr{O}_3 = \mathscr{O}_1(\mathscr{E}_1 - 2\mathscr{R})$, we find that steps similar to those given for the spacelike vector interaction and the polar part of the tensor interaction given in Appendix A yield

$$\begin{aligned}\mathscr{S}_1 = {} & \exp(\mathscr{O}_2 Y)\theta_1 \cdot p + \epsilon_1 \,\mathrm{ch}(Y\mathscr{E}_4)\theta_1 \cdot \hat{P} \\ & + \epsilon_2 \,\mathrm{sh}(Y\mathscr{E}_4)\theta_2 \cdot \hat{P} + m_1\theta_{51} \\ & - i\exp(\mathscr{O}_2 Y)\theta_2 \cdot \frac{\partial Y}{2}\mathscr{E}_4.\end{aligned} \tag{115a}$$

In a similar fashion, we obtain

$$\begin{aligned}\mathscr{S}_2 = {} & -\exp(\mathscr{O}_2 Y)\theta_2 \cdot p + \epsilon_2 \,\mathrm{ch}(Y\mathscr{E}_4)\theta_2 \cdot \hat{P} \\ & + \epsilon_1 \,\mathrm{sh}(Y\mathscr{E}_4)\theta_1 \cdot \hat{P} + m_2\theta_{52} \\ & + i\exp(\mathscr{O}_2 Y)\theta_1 \cdot \frac{\partial Y}{2}\mathscr{E}_4.\end{aligned} \tag{115b}$$

V. CHARACTERISTIC FORMS OF TWO-BODY DIRAC EQUATIONS FOR SINGLE COVARIANT INTERACTIONS

For each of the eight relativistic interactions that we have considered above, our methods lead to a pair of coupled but compatible Dirac equations. Each of these "external potential" forms contains characteristic minimal extensions of structures of the free Dirac equation accompanied by an appropriate spin-dependent recoil term that depends on the gradient of the potential. For scalar interactions $[\Delta = -\frac{1}{2}L(x_\perp)\mathscr{O}_1]$, the mass terms become $x_\perp$ dependent:

$$\begin{aligned}\mathscr{S}_1\psi = {} & \Big(\theta_1 \cdot p + \epsilon_1\theta_1 \cdot \hat{P} + m_1 \,\mathrm{ch}(L\mathscr{O}_1)\theta_{51} \\ & - m_2 \,\mathrm{sh}(L\mathscr{O}_1)\theta_{52} - i\theta_2 \cdot \frac{\partial L}{2}\mathscr{O}_1\Big)\psi,\end{aligned} \tag{70a'}$$

$$\begin{aligned}\mathscr{S}_2\psi = {} & \Big(-\theta_1 \cdot p + \epsilon_2\theta_1 \cdot \hat{P} + m_2 \,\mathrm{ch}(L\mathscr{O}_1)\theta_{52} \\ & - m_1 \,\mathrm{sh}(L\mathscr{O}_1)\theta_{51} + i\theta_1 \cdot \frac{\partial L}{2}\mathscr{O}_1\Big)\psi.\end{aligned} \tag{70b'}$$

For timelike vector interactions $[\Delta = \frac{1}{2}J(x_\perp)\mathscr{O}_2]$, the timelike momentum terms become $x_\perp$ dependent:

$$\begin{aligned}\mathscr{S}_1\psi = {} & \Big(\theta_1 \cdot p + \epsilon_1 \,\mathrm{ch}(J\mathscr{O}_2)\theta_1 \cdot \hat{P} + \epsilon_2 \,\mathrm{sh}(J\mathscr{O}_2)\theta_2 \cdot \hat{P} \\ & + m_1\theta_{51} + i\theta_2 \cdot \frac{\partial J}{2}\mathscr{O}_2\Big)\psi,\end{aligned} \tag{72a'}$$

$$\begin{aligned}\mathscr{S}_2\psi = {} & \Big(-\theta_2 \cdot p + \epsilon_2 \,\mathrm{ch}(J\mathscr{O}_2)\theta_2 \cdot \hat{P} + \epsilon_1 \,\mathrm{sh}(J\mathscr{O}_2)\theta_1 \cdot \hat{P} \\ & + m_2\theta_{52} - i\theta_2 \cdot \frac{\partial J}{2}\mathscr{O}_2\Big).\end{aligned} \tag{72b'}$$

For spacelike vector interactions $[\Delta = \frac{1}{2}\mathscr{G}(x_\perp)\mathscr{O}_3]$, the spacelike momentum terms become $x_\perp$ dependent:

$$\begin{aligned}\mathscr{S}_1\psi = {} & \Big(\exp(\mathscr{G})\theta_1 \cdot p + \epsilon_1\theta_1\partial\hat{P} + m_1\theta_{51} \\ & + i\exp(\mathscr{G})\theta_2 \cdot \frac{\partial\mathscr{G}}{2}\mathscr{O}_3\Big),\end{aligned} \tag{94a'}$$

$$\begin{aligned}\mathscr{S}_2\psi = {} & \Big(-\exp(\mathscr{G})\theta_2 \cdot p + \epsilon_2\theta_2\partial\hat{P} + m_2\theta_{52} \\ & - i\exp(\mathscr{G})\theta_1 \cdot \frac{\partial\mathscr{G}}{2}\mathscr{O}_3\Big)\psi.\end{aligned} \tag{94b'}$$

For "polar" tensor interactions $[\Delta = \frac{1}{2}\mathscr{F}(x_\perp)\mathscr{O}_4]$ the mass, timelike, and spacelike momentum terms become not only $x_\perp$ dependent but also spin dependent (through the appearance of $\theta_{1\perp}\cdot\theta_{2\perp}$ in $\mathscr{O}_4$):

$$\begin{aligned}\mathscr{S}_1\psi = {} & \Big(\exp(\mathscr{F}\mathscr{E}_2)\theta_1 \cdot p + \epsilon_1 \,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_1 \cdot \hat{P} \\ & + \epsilon_2 \,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_2 \cdot \hat{P} + m_1 \,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_{51} \\ & + m_2 \,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_{52} + i\exp(\mathscr{F}\mathscr{E}_2)\theta_2 \cdot \frac{\partial\mathscr{F}}{2}\mathscr{O}_4\Big)\psi\end{aligned} \tag{95a'}$$

$$\begin{aligned}\mathscr{S}_2\psi = {} & \Big(-\exp(\mathscr{F}\mathscr{E}_2)\theta_2 \cdot p + \epsilon_2 \,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_2 \cdot \hat{P} \\ & + \epsilon_1 \,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_1 \cdot \hat{P} + m_2 \,\mathrm{ch}(\mathscr{F}\mathscr{O}_4)\theta_{52} \\ & + m_2 \,\mathrm{sh}(\mathscr{F}\mathscr{O}_4)\theta_{51} - i\exp(\mathscr{F}\mathscr{E}_2)\theta_1 \cdot \frac{\partial\mathscr{F}}{2}\mathscr{O}_4\Big)\psi.\end{aligned} \tag{95b'}$$

For pseudoscalar interactions $[\Delta = \frac{1}{2}C(x_\perp)\mathscr{E}_1]$ mass, timelike, and spacelike momentum terms are independent of $x_\perp$—the potential is entirely contained in the gradient-dependent recoil term:

$$\mathscr{S}_1\psi = \Big(\theta_1 \cdot p + \epsilon_1\theta_1 \cdot \hat{P} + m_1\theta_{51} + i\theta_2 \cdot \partial\frac{C}{2}\Big)\psi \tag{106a'}$$

$$\mathscr{S}_2\psi = \Big(-\theta_2 \cdot p + \epsilon_2\theta_2 \cdot \hat{P} + m_2\theta_{52} - i\theta_1 \cdot \partial\frac{C}{2}\Big)\psi. \tag{106b'}$$

For timelike pseudovector interactions $[\Delta = -\frac{1}{2}H(x_\perp)\mathscr{E}_2]$ mass and timelike momentum terms become $x_\perp$ dependent but not spin dependent ($\theta_{1\perp}\cdot\theta_{2\perp}$ does not appear in $\mathscr{E}_2$):

$$\begin{aligned}\mathscr{S}_1\psi = {} & \Big(\theta_1 \cdot p + \epsilon_1 \,\mathrm{ch}(H\mathscr{E}_2)\theta_1 \cdot \hat{P} + \epsilon_2 \,\mathrm{sh}(H\mathscr{E}_2)\theta_2 \cdot \hat{P} \\ & + m_1 \,\mathrm{ch}(H\mathscr{E}_2)\theta_{51} + m_2 \,\mathrm{sh}(H\mathscr{E}_2)\theta_{52} \\ & - i\theta_2 \cdot \frac{\partial H}{2}\mathscr{E}_2\Big)\psi,\end{aligned} \tag{113a'}$$

$$\begin{aligned}\mathscr{S}_2\psi = {} & \Big(-\theta_2 \cdot p + \epsilon_2 \,\mathrm{ch}(H\mathscr{E}_2)\theta_2 \cdot \hat{P} + \epsilon_1 \,\mathrm{sh}(H\mathscr{E}_2)\theta_1 \cdot \hat{P} \\ & + m_2 \,\mathrm{ch}(H\mathscr{E}_2)\theta_{52} + m_1 \,\mathrm{sh}(H\mathscr{E}_2)\theta_{51} \\ & + i\theta_1 \cdot \frac{\partial H}{2}\mathscr{E}_2\Big)\psi.\end{aligned} \tag{113b'}$$

For spacelike pseudovector interactions $[\Delta = -\frac{1}{2}I(x_\perp)\mathscr{E}_3]$, the spacelike momentum terms become $x_\perp$ dependent while the mass terms become $x_\perp$ dependent and spin dependent (through the appearance of $\theta_{1\perp}\cdot\theta_{2\perp}$ in $\mathscr{E}_3$):

$$\mathscr{S}_1\psi=\Big(\exp(\mathscr{O}_1 I)\theta_1\cdot p+\epsilon_1\theta_1\cdot\hat{P}+m_1\ \mathrm{ch}(I\mathscr{C}_3)\theta_{51}$$
$$+m_2\ \mathrm{sh}(I\mathscr{C}_3)\theta_{52}-i\exp(\mathscr{O}_1 I)\theta_2\cdot\frac{\partial I}{2}\mathscr{C}_3\Big)\psi, \tag{114a'}$$

$$\mathscr{S}_2\psi=\Big(-\exp(\mathscr{O}_1 I)\theta_2\cdot p+\epsilon_2\theta_2\cdot\hat{P}+m_2\mathrm{ch}(I\mathscr{C}_3)\theta_{52}$$
$$+m_1\ \mathrm{sh}(I\mathscr{C}_3)\theta_{51}+i\exp(\mathscr{O}_1 I)\theta_1\cdot\frac{\partial I}{2}\mathscr{C}_3\Big)\psi. \tag{114b'}$$

For "axial" tensor interactions $[\Delta=-\frac{1}{2}Y(x_\perp)\mathscr{C}_4]$ the spacelike momentum terms become $x_\perp$ dependent while the timelike momentum terms become $x_\perp$ dependent and spin dependent (through the appearance of $\theta_{1\perp}\cdot\theta_{2\perp}$ in $\mathscr{C}_4$):

$$\mathscr{S}_1\psi=\Big(\exp(\mathscr{O}_2 Y)\theta_1\cdot p+\epsilon_1\ \mathrm{ch}(Y\mathscr{C}_4)\theta_1\cdot\hat{P}$$
$$+\epsilon_2\ \mathrm{sh}(Y\mathscr{C}_4)\theta_2\cdot\hat{P}+m_1\theta_{51}$$
$$-i\exp(\mathscr{O}_2 Y)\theta_2\cdot\frac{\partial Y}{2}\mathscr{C}_4\Big)\psi, \tag{115a'}$$

$$\mathscr{S}_2\psi=\Big(-\exp(\mathscr{O}_2 Y)\theta_2\cdot p+\epsilon_2\ \mathrm{ch}(Y\mathscr{C}_4)\theta_2\cdot\hat{P}$$
$$+\epsilon_1\ \mathrm{sh}(Y\mathscr{C}_4)\theta_1\cdot\hat{P}+m_2\theta_{52}$$
$$+i\exp(\mathscr{O}_2 Y)\theta_1\cdot\frac{\partial Y}{2}\mathscr{C}_4\Big)\psi. \tag{115b'}$$

Each of these equations has the form of (57a) and (57b)—the generic "external potential" form. Note that while the minimal extensions appearing in the scalar and vector equations are simple alterations of classical relativistic properties, those appearing in the tensor and axial equations could not have been anticipated from simple classical considerations.

VI. TWO-BODY DIRAC EQUATIONS FOR COMBINATIONS OF COVARIANT INTERACTIONS

In physical applications in which the relativistic potentials appearing in two-body Dirac equations result from quantum field theory, two or more of these eight interaction occur in combination. Accordingly, we now generalize these equations to include certain important pairs of interactions. First, as a pedagogical aid to the reader, we review the treatment of the simplest combination, scalar plus timelike four-vector that plays an important role in relativistic phenomenological treatments[4-5] of the long-range confining part of the chromodynamic interactions of heavy and light quarks. [We have treated this pair elsewhere[3] using supersymmetric methods to obtain Eqs. (40a) and (40b)]. Then, we examine the case of additive timelike and spacelike vector interactions—the electromagnetic case—which we have solved exactly[16] for parapositronium (with field theoretic spectrum correct to order α^4). Finally, we treat the case of additive scalar and pseudoscalar interactions that appears as part of the Fierz transformed annihilation channel of electrodynamics—our principle motivation for the generalized treatment of interactions that appear in this paper.

We return to the general hyperbolic constraint forms (50a) and (50b) which we know to be compatible for arbitrary Δ:

$$\mathbf{S}_1=\mathscr{S}_{10}\ \mathrm{ch}(\Delta)+\mathscr{S}_{20}\ \mathrm{sh}(\Delta), \tag{50a'}$$
$$\mathbf{S}_2=\mathscr{S}_{20}\ \mathrm{ch}(\Delta)+\mathscr{S}_{10}\ \mathrm{sh}(\Delta). \tag{50b'}$$

We examine the case of additive scalar and timelike vector interactions:

$$\Delta\equiv\Delta_J+\Delta_L=\tfrac{1}{2}(\mathscr{O}_2 J(x_\perp)-\mathscr{O}_1 L(x_\perp)). \tag{116}$$

Since both $\mathscr{O}_1$ and $\mathscr{O}_2$ are doubly odd matrices, the general form given in (61) for the $\mathscr{S}_1$ constraint (similarly for the $\mathscr{S}_2$ constraint) is still valid. We use (64), (71), and $\mathscr{O}_1\mathscr{O}_2=\mathscr{C}_2$ to obtain

$$\mathrm{ch}(\Delta)=\mathrm{ch}(\Delta_J)\mathrm{ch}(\Delta_L)+\mathrm{sh}(\Delta_J)\mathrm{sh}(\Delta_L)$$
$$=\mathrm{ch}\left(\frac{J}{2}\right)\mathrm{ch}\left(\frac{L}{2}\right)-\mathscr{C}_2\ \mathrm{sh}\left(\frac{J}{2}\right)\mathrm{sh}\left(\frac{L}{2}\right), \tag{117}$$

$$\mathrm{sh}(\Delta)=\mathrm{sh}(\Delta_J)\mathrm{ch}(\Delta_L)+\mathrm{ch}(\Delta_J)\mathrm{sh}(\Delta_L)$$
$$=\mathscr{O}_2\ \mathrm{sh}\left(\frac{\Delta}{J}\right)\mathrm{ch}\left(\frac{L}{2}\right)-\mathscr{O}_1\ \mathrm{ch}\left(\frac{J}{2}\right)\mathrm{sh}\left(\frac{L}{2}\right). \tag{118}$$

We then make use of (68)–(69) and similar relations for the timelike vector interactions to obtain

$$[\mathrm{ch}(\Delta),\theta_{i1}^{\mu}]_-=0=[\mathrm{sh}(\Delta),\theta_{i1}^{\mu}]_+, \tag{119}$$
$$[\mathrm{ch}(\Delta),\theta_i\cdot\hat{P}]_-=2\ \mathrm{sh}(\Delta_J)\mathrm{sh}(\Delta_L)\theta_i\cdot\hat{P}, \tag{120}$$
$$[\mathrm{ch}(\Delta),\theta_{5i}]_-=2\ \mathrm{sh}(\Delta_J)\mathrm{sh}(\Delta_L)\theta_{5i}, \tag{121}$$
$$[\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_+=2\ \mathrm{sh}(\Delta_J)\mathrm{ch}(\Delta_L)\theta_i\cdot\hat{P}, \tag{122}$$
$$[\mathrm{sh}(\Delta),\theta_{5i}]_+=2\ \mathrm{ch}(\Delta_J)\mathrm{sh}(\Delta_L)\theta_{5i}. \tag{123}$$

Substitution of these results into (63) then yields after simplification

$$\mathscr{S}_1=\theta_1\cdot p+\epsilon_1\ \mathrm{ch}(\mathscr{O}_2 J)\theta_1\cdot\hat{P}+\epsilon_2\ \mathrm{sh}(\mathscr{O}_2 J)\theta_2\cdot\hat{P}$$
$$+m_1\ \mathrm{ch}(\mathscr{O}_1 L)\theta_{51}-m_2\ \mathrm{sh}(\mathscr{O}_1 L)\theta_{52}$$
$$+i\theta_2\cdot\frac{\partial}{2}(J\mathscr{O}_2-L\mathscr{O}_1). \tag{124a}$$

Similarly, we find that

$$\mathscr{S}_2=-\theta_2\cdot p+\epsilon_2\ \mathrm{ch}(\mathscr{O}_2 J)\theta_2\cdot\hat{P}+\epsilon_1\ \mathrm{sh}(\mathscr{O}_2 J)\theta_1\cdot\hat{P}$$
$$+m_2\ \mathrm{ch}(\mathscr{O}_1 L)\theta_{52}-m_1\ \mathrm{sh}(\mathscr{O}_1 L)\theta_{51}$$
$$-i\theta_1\cdot\frac{\partial}{2}(J\mathscr{O}_2-L\mathscr{O}_1). \tag{124b}$$

When we use the fact that $\mathscr{O}_i$'s are roots of unity in each hyperbolic function, we see that these are just our earlier results Eqs. (40a) and (40b).

Next, we treat the more complicated case of additive timelike and spacelike interaction (important for the electromagnetic case) for which

$$\Delta=\Delta_J+\Delta_{\mathscr{G}}$$
$$=\tfrac{1}{2}(\mathscr{O}_2 J(x_\perp)+\mathscr{O}_3\mathscr{G}(x_\perp)). \tag{125}$$

When we use

$$\mathrm{ch}(\Delta)=\mathrm{ch}(\Delta_{\mathscr{G}})\mathrm{ch}(\Delta_J)+\mathrm{sh}(\Delta_{\mathscr{G}})\mathrm{sh}(\Delta_J),$$

$$\mathrm{sh}(\Delta)=\mathrm{sh}(\Delta_{\mathscr{G}})\mathrm{ch}(\Delta_J)+\mathrm{ch}(\Delta_{\mathscr{G}})\mathrm{sh}(\Delta_J),$$

(68), (84), and (86), we find that

$$[\mathrm{ch}(\Delta),\theta^{\mu}_{i1}]_- = [\mathrm{ch}(\Delta_{\mathscr{G}}),\theta^{\mu}_{i1}]_-\,\mathrm{ch}(\Delta_J) - [\mathrm{sh}(\Delta_{\mathscr{G}}),\theta^{\mu}_{i1}]_+\,\mathrm{sh}(\Delta_J), \tag{126}$$

$$[\mathrm{ch}(\Delta),\theta_i\cdot\hat{P}]_- = \mathrm{sh}(\Delta_{\mathscr{G}})[\mathrm{sh}(\Delta_J),\theta_i\cdot\hat{P}]_+, \tag{127}$$

$$[\mathrm{ch}(\Delta),\theta_{5i}]_- = 0, \tag{128}$$

$$[\mathrm{sh}(\Delta),\theta^{\mu}_{i1}]_+ = [\mathrm{sh}(\Delta_{\mathscr{G}}),\theta^{\mu}_{i1}]_+\,\mathrm{ch}(\Delta_J) - [\mathrm{ch}(\Delta_{\mathscr{G}}),\theta^{\mu}_{i1}]_-\,\mathrm{sh}(\Delta_J), \tag{129}$$

$$[\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_- = \mathrm{sh}(\Delta_{\mathscr{G}})[\mathrm{sh}(\Delta_J),\theta_i\cdot\hat{P}]_+, \tag{130}$$

$$[\mathrm{sh}(\Delta),\theta_{5i}]_+ = 0. \tag{131}$$

Substitution of (126)–(131) into (63) with additional simplification yields

$$\mathscr{S}_1=\exp(\mathscr{G})\theta_1\cdot p+\mathrm{ch}(\mathscr{O}_2 J)\epsilon_1\theta_1\cdot\hat{P}+\mathrm{sh}(\mathscr{O}_2 J)\epsilon_2\theta_2\cdot\hat{P} + m_1\theta_{51}+\exp(\mathscr{G})i\theta_2\cdot(\partial/2)\,(\mathscr{G}\mathscr{O}_3+J\mathscr{O}_2). \tag{132a}$$

Similarly,

$$\mathscr{S}_2=-\exp(\mathscr{G})\theta_2\cdot p+\mathrm{ch}(\mathscr{O}_2 J)\epsilon_2\theta_2\cdot\hat{P} + \mathrm{sh}(\mathscr{O}_2 J)\epsilon_1\theta_1\cdot\hat{P} + m_2\theta_{52}-\exp(\mathscr{G})i\theta_1\cdot(\partial/2)\,(\mathscr{G}\mathscr{O}_3+J\mathscr{O}_2). \tag{132b}$$

For electromagneticlike interactions the potentials are related through $J=-\mathscr{G}$ so that

$$\mathscr{S}_1=\exp(\mathscr{G})\theta_1\cdot p+\mathrm{ch}(\mathscr{O}_2\mathscr{G})\epsilon_1\theta_1\cdot\hat{P} - \mathrm{sh}(\mathscr{O}_2\mathscr{G})\epsilon_2\theta_2\cdot\hat{P} + m_1\theta_{51}+\exp(\mathscr{G})i\theta_2\cdot(\partial/2)\,(\mathscr{G}\theta_1\cdot\theta_2) \tag{133a}$$

and

$$\mathscr{S}_2=\exp(\mathscr{G})\theta_2\cdot p+\mathrm{ch}(\mathscr{O}_2\mathscr{G})\epsilon_2\theta_2\cdot\hat{P} - \mathrm{sh}(\mathscr{O}_2\mathscr{G})\epsilon_1\theta_1\cdot\hat{P} + m_2\theta_{52}-\exp(\mathscr{G})i\theta_1\cdot(\partial/2)\,(\mathscr{G}\theta_1\cdot\theta_2) \tag{133b}$$

[since the combination $\Delta=(\mathscr{G}\mathscr{O}_3+J\mathscr{O}_2)/2=(\mathscr{G}\theta_1\cdot\theta_2)/2$]. If we identify $G=\exp(\mathscr{G})$, we reproduce (43)–(44) a result that we had derived in an earlier paper using supersymmetry methods.

A weak potential form of Eqs. (133a) and (133b) has been derived by Sazdjian from a form equivalent to Eqs. (52a) and (52b). Starting with $\mathscr{W}(x_1)=(\frac{1}{2}\mathscr{O}_3\mathscr{G})+\frac{1}{3}(\frac{1}{2}\mathscr{O}_3\mathscr{G})^3$ (in our notation), he works his way to a form like Eqs. (133a) and (133b). Viewed from the standpoint of our hyperbolic structure, he constructed the Dirac constraint that contains just the first two terms of the hyperbolic tangent. As he pointed out, if the first term is regarded as arising from single-vector exchange, the second term (an axial vector term) can be regarded as arising (in its covariant structure) from a triple-vector exchange. In his interpretation, our combination of hyperbolic functions would correspond to a sum over vector exchanges of all orders. In fact, viewed from the Bethe–Salpeter equation, however, our structure is an extrapolated extension of a single-vector exchange.

Next, we examine the still more complex structure generated by addition of polar and axial interactions (such as produced by electrodynamics when the Fierz transformated annihilation channel is included). We wish to construct $\mathscr{S}_i$ constraints from combinations of $\mathbf{S}_i$ that yield the simple external potential forms. For the polar interactions, Δ is an odd–odd matrix and $\mathscr{S}_i$ is given by (63), whereas for the axial interactions, Δ is an even–even interaction and $\mathscr{S}_i$ is given by (102). We still start from the general constraints

$$\mathbf{S}_1=\mathscr{S}_{10}\,\mathrm{ch}(\Delta)+\mathscr{S}_{20}\,\mathrm{sh}(\Delta), \tag{50a'}$$

$$\mathbf{S}_2=\mathscr{S}_{20}\,\mathrm{ch}(\Delta)+\mathscr{S}_{10}\,\mathrm{sh}(\Delta), \tag{50b'}$$

but with $\Delta=\Delta_{\mathscr{O}}+\Delta_{\mathscr{G}}\equiv\Delta_+$. Since Δ_+ has a mixed "parity," we define $\Delta_-=\Delta_{\mathscr{O}}-\Delta_{\mathscr{G}}$ and are forced to take

$$\mathscr{S}_1=\mathrm{ch}(\Delta_-)\mathbf{S}_1+\mathrm{sh}(\Delta_-)\mathbf{S}_2, \tag{134a}$$

$$\mathscr{S}_2=\mathrm{ch}(\Delta_-)\mathbf{S}_2+\mathrm{sh}(\Delta_-)\mathbf{S}_1, \tag{134b}$$

in order to use the simple hyperbolic identity $\mathrm{ch}^2(\Delta_-)-\mathrm{sh}^2(\Delta_-)=1$ to bring (192a) and (192b) to external potential form. That is, the plus sign coefficient of $\Delta_{\mathscr{O}}$ and the minus sign coefficient of $\Delta_{\mathscr{G}}$ in conjunction with the odd–odd nature of $\Delta_{\mathscr{O}}$ and $\mathrm{sh}(\Delta_{\mathscr{O}})$ for the polar interactions and the even–even nature of $\Delta_{\mathscr{G}}$ and $\mathrm{sh}(\Delta_{\mathscr{G}})$ for the axial interactions combine to give a minus sign coefficient for $\mathrm{sh}^2(\Delta_-)$, which, in turn, allows us to use the simple hyperbolic identity $\mathrm{ch}^2(\Delta_-)-\mathrm{sh}^2(\Delta_-)=1$ in the construction of the external potential form (57a) and (57b) [see (135) below]. Note that (134a) and (134b) generalize our two earlier forms (56a) and (56b) and (96a) and (96b) reducing to them when either $\Delta_{\mathscr{O}}=0$ or $\Delta_{\mathscr{G}}=0$. Note that the compatibility of these two constraints follows from those of the $\mathbf{S}_i$ just as did that of (56a) and (56b). Next, we consider how to generalize (63) and (102), the equations for the external potential forms of the constraints. In Appendix B, we show that this generalization is given by

$$\begin{aligned}\mathscr{S}_1=\mathscr{S}_{10}+i\theta_2\cdot\partial(\Delta_+)-i([&\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}}),\theta^{\mu}_{11}]_-\,\mathrm{sh}(\Delta_+)-[\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}}),\theta^{\mu}_{11}]_+\mathrm{sh}(\Delta_+)\\ &-[\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}}),\theta^{\mu}_{21}]_-\mathrm{ch}(\Delta_+)+[\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}}),\theta^{\mu}_{21}]_+\,\mathrm{ch}(\Delta_+)\\ &-[\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}}),\theta^{\mu}_{21}]_-\,\mathrm{sh}(\Delta_+)\\ &+[\mathrm{sh}(\Delta_{\mathscr{G}})\mathrm{ch}(\Delta_{\mathscr{O}}),\theta^{\mu}_{21}]_-\,\mathrm{sh}(\Delta_+)+[\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}}),\theta^{\mu}_{11}]_+\,\mathrm{ch}(\Delta_+)\end{aligned}$$

$$
\begin{aligned}
&- [\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}}),\theta_{1\perp}^{\mu}]_{-}\mathrm{ch}(\Delta_{+}))\partial_{\mu}(\Delta_{+})\\
&+ \mathrm{ch}(\Delta_{-})([\theta_{1\perp}^{\mu},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{-}p_{\mu} + [\theta_{1\perp}^{\mu},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{+}p_{\mu} + [\epsilon_1\theta_1\cdot\hat{P},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{-}\\
&+ [\epsilon_1\theta_1\cdot\hat{P},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{+} + [m_1\theta_{51},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{-} + [m_1\theta_{51},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{+}\\
&- [\theta_{2\perp}^{\mu},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+}p_{\mu}\\
&- [\theta_{2\perp}^{\mu},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{-}p_{\mu} + [\epsilon_2\theta_2\cdot\hat{P},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+} + [\epsilon_2\theta_2\cdot\hat{P},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{-}\\
&+ [m_2\theta_{52},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+} + [m_2,\theta_{52},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{-}) + \mathrm{sh}(\Delta_{-})(-[\theta_{2\perp}^{\mu},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{-}p_{\mu}\\
&- [\theta_{2\perp}^{\mu},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+}p_{\mu} + [\epsilon_2\theta_2\cdot\hat{P},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{-} + [\epsilon_2\theta_2\cdot\hat{P},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+}\\
&+ [m_2\theta_{52},\mathrm{ch}(\Delta_{\mathscr{O}})\mathrm{sh}(\Delta_{\mathscr{G}})]_{-} + [m_2\theta_{52},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+}\\
&+ [\theta_{1\perp}^{\mu},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+}p_{\mu} + [\theta_{1\perp}^{\mu},\mathrm{sh}(\Delta_{\mathscr{G}})\mathrm{ch}(\Delta_{\mathscr{O}})]_{-}p_{\mu} + [\epsilon_1\cdot\theta_1\cdot\hat{P},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+}\\
&+ [\epsilon_1\theta_1\cdot\hat{P},\mathrm{sh}(\Delta_{\mathscr{G}})\mathrm{ch}(\Delta_{\mathscr{O}})]_{-} + [m_1\theta_{51},\mathrm{sh}(\Delta_{\mathscr{O}})\mathrm{ch}(\Delta_{\mathscr{G}})]_{+} + [m_1\theta_{51},\mathrm{sh}(\Delta_{\mathscr{G}})\mathrm{ch}(\Delta_{\mathscr{O}})]_{-}), \qquad (135)
\end{aligned}
$$

with a similar expression for $\mathscr{S}_2$. Note that this complicated expression simplifies to either (63) or (102) if either $\Delta_{\mathscr{G}}$ or $\Delta_{\mathscr{O}}$ vanishes.

We specialize this result to the additive scalar and pseudoscalar interaction for which

$$
\Delta_{\mathscr{O}} = -\frac{L(x_\perp)\mathscr{O}_1}{2}, \quad \Delta_{\mathscr{G}} = \frac{C(x_\perp)}{2}\mathscr{G}_1 = \frac{C(x_\perp)}{2}. \qquad (136)
$$

This combination is important not only as part of the Fierz transformed annihilation structure of electrodynamics but also for phenomenological studies of the two-nucleon problem. This particular case is especially simple since virtually all of the commutators and anticommutators in (135) vanish with the exception of the anticommutators that involve the m_i factors. These combine to give

$$
\begin{aligned}
&\mathrm{ch}(\Delta_{-})(2\,\mathrm{sh}(\Delta_L)\mathrm{sh}(\Delta_C)m_1\theta_{51} + 2\,\mathrm{sh}(\Delta_L)\\
&\quad\times\mathrm{ch}(\Delta_{\mathscr{G}})m_2\theta_{52}) + \mathrm{sh}(\Delta_{-})(2\,\mathrm{sh}(\Delta_L)\mathrm{sh}(\Delta_C)m_2\theta_{52}\\
&\quad + 2\,\mathrm{sh}(\Delta_L)\mathrm{ch}(\Delta_C)m_1\theta_{51})\\
&= 2\,\mathrm{sh}^2(\Delta_L)m_1\theta_{51} + 2\,\mathrm{sh}(\Delta_{\mathscr{S}})\mathrm{ch}(\Delta_L)m_2\theta_{52}. \qquad (137)
\end{aligned}
$$

Thus, in this case,

$$
\begin{aligned}
\mathscr{S}_1 &= \theta_1\cdot p + \epsilon_1\theta_1\cdot\hat{P} + m_1\,\mathrm{ch}(\mathscr{O}_1 L)\theta_{51}\\
&\quad - m_2\,\mathrm{sh}(\mathscr{O}_1 L)\theta_{52}\\
&\quad + i\theta_2\cdot(\partial/2)\,(C - L\mathscr{O}_1) \qquad (138)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathscr{S}_2 &= -\theta_1\cdot p + \epsilon_2\theta_2\cdot\hat{P} + m_2\,\mathrm{ch}(\mathscr{O}_1 L)\theta_{52}\\
&\quad - m_1\,\mathrm{sh}(\mathscr{O}_1 L)\theta_{51}\\
&\quad - i\theta_1\cdot(\partial/2)\,(C - L\mathscr{O}_1). \qquad (139)
\end{aligned}
$$

VII. CONCLUSION

Two different points of view about the significance of relativistic interactions determine the ways in which they appear in two-body constraint equations. One view is that relativistic potentials appear as constituent potentials in the constraint equations governing each particle. The other is that relativistic potentials originate as "system interactions" in forms dictated by the two-body system. In our treatment of the system of two spin one-half particles and that of Sazdjian, these points of view survive as our emphasis on interaction via "external potentials" versus Sazdjian's emphasis on a single-system interaction for the pair. However, in our work, these points of view are actually complementary—both realized in a single structure. Their complementary nature arises from two requirements that the two-body system ought to satisfy. Mathematical compatibility of the two constituent equations requires that in reality there be only one independent potential in the system (relativistic version of Newton's third law), while the demand that the interactions permit the description to degenerate to the usual Dirac equation in a static external potential when one particle has infinite mass restricts the acceptable dependences of the system potential on masses and c.m. energies. Satisfaction of the second requirement is most easily dealt with when the constraint equations are in "external potential" form.

For two spinless particles, these two requirements lead to a hyperbolic dependence of constituent potentials on the underlying system potential—a hyperbolic realization of Newton's third law [see Eqs. (30) and (31) and (35) and (36)]. When both particles possess spin one-half, satisfaction of these two requirements would seem to be more difficult. However, as we showed in our treatment of the world scalar interaction between two spin one-half particles, the presence of a special structure in spin space—recoil terms dictated by (one-body) supersymmetries—takes care of the spin-dependent complications and reduces these requirements to those of the spinless case (hyperbolic solution of the third law). Recently, Sazdjian found that satisfaction of one requirement—compatibility—is almost trivial no matter what the form of the system potential. His constraints are a pair of point-dependent "superpositions" of free-particle Dirac operators. In order to satisfy the requirement of compatibility he found a structure in spin space to take the place of our supersymmetries. This structure reduces to ours for forms of the interaction whose matrix coefficients are roots of unity (i.e., for the scalar, pseudoscalar, and timelike vec-

tor and pseudovector interactions) but not for those whose matrix coefficients are not (i.e., the spacelike vector and pseudovector interactions, the polar and axial tensor interactions).

In the present work, we have shown how Sazdjian's spin structure may be altered to reunite the system and constituent forms of the two-body Dirac equations for general relativistic interactions. The central ingredient is that the point-dependent superposition of free Dirac operators itself be hyperbolic. Then, as we have shown for eight covariant interactions, the equations are not only compatible (as were those of Sazdjian), but also possess forms in which the interaction appears as a system of "external potentials" (see the eight sets of equations in Sec. V). Furthermore, all of these potentials appear as minimal substitutions on masses and momenta. Thus this procedure generates the minimal substitutions for scalar and vector interactions along with new minimal substitutions for the pseudovector and tensor interactions. Realized in this fashion, the algebraic spin structure really does play the role for more general interactions that was played by our supersymmetry for the scalar, reducing the compatibility condition to the satisfaction of the third law for spinless particles in constituent form.

The fact that our superpositions depend nonlinearly on the underlying system interaction not only makes the "external potential" description possible but also simplifies the interaction dependence of the quantum-mechanical norm as well. When such equations are used as a quantum-mechanical transform of the field-theoretic Bethe–Salpeter equation,[5,17-18] they simplify the physical interpretation of interaction structures generated by a single photon exchange. This hyperbolic structure leads directly to an effective "external potential" form of interaction, which serves as a nonperturbative extrapolation of the perturbative quantum-mechanical transform[6] of the Bethe–Salpeter equation.

ACKNOWLEDGMENTS

We would like to thank Dr. R. L. Becker and Dr. C. Y. Wong for very helpful discussions and for motivating the generalization of our earlier work.

The work of one of the authors (HWC) was supported, in part, by the National Science Foundation.

APPENDIX A: COMPATIBLE DIRAC OPERATORS FOR THE POLAR PART OF THE TENSOR INTERACTION

For this interaction, $\Delta = \frac{1}{2}\mathcal{O}_4\mathcal{F}$. Now, since $\mathcal{O}_4 = \mathcal{E}_1 - 2\mathcal{R}$, we find that

$$\begin{aligned}\mathrm{ch}(\Delta) &= \mathrm{ch}\left(\frac{\mathcal{E}_1\mathcal{F}}{2} - \mathcal{R}\mathcal{F}\right)\\ &= \mathrm{ch}\left(\frac{\mathcal{F}}{2}\right)\mathrm{ch}(\mathcal{F}) - \mathcal{R}\,\mathrm{sh}\left(\frac{\mathcal{F}}{2}\right)\mathrm{sh}(\mathcal{F})\\ &= \mathrm{ch}^3\left(\frac{\mathcal{F}}{2}\right) + \frac{1}{2}\mathcal{O}_4\,\mathrm{sh}\left(\frac{\mathcal{F}}{2}\right)\mathrm{sh}(\mathcal{F}),\end{aligned} \tag{A1}$$

$$\begin{aligned}\mathrm{sh}(\Delta) &= \mathrm{sh}\left(\frac{\mathcal{E}_1}{2}\mathcal{F} - \mathcal{R}\mathcal{F}\right)\\ &= \mathrm{sh}\left(\frac{\mathcal{F}}{2}\right)\mathrm{ch}(\mathcal{F}) - \mathcal{R}\,\mathrm{ch}\left(\frac{\mathcal{F}}{2}\right)\mathrm{sh}(\mathcal{F})\\ &= \mathrm{sh}^3\left(\frac{\mathcal{F}}{2}\right) + \frac{1}{2}\mathcal{O}_4\,\mathrm{ch}\left(\frac{\mathcal{F}}{2}\right)\mathrm{sh}(\mathcal{F}).\end{aligned} \tag{A2}$$

In addition to the commutators given in (81) and (82) for the spacelike vector, we need the anticommutators

$$[\mathcal{O}_4,\theta_{i1}^{\mu}]_+ = -2\theta_{j1}^{\mu}\mathcal{E}_2 = -2\mathcal{E}_2\theta_{j1}^{\mu}, \tag{A3}$$

$$[\mathcal{O}_4,\theta_i\cdot\hat{P}]_+ = 2\mathcal{O}_4\theta_i\cdot\hat{P}, \quad [\mathcal{O}_{4i}\theta_{5i}]_+ = 2\mathcal{O}_4\theta_{5i}. \tag{A4}$$

With their aid, we find that

$$\begin{aligned}&[\mathrm{ch}(\Delta),\theta_{i1}^{\mu}]_-\\ &\quad = -\mathrm{sh}(\mathcal{F}/2)\mathrm{sh}(\mathcal{F})(\theta_{i1}^{\mu}\mathcal{O}_4 + \theta_{j1}^{\mu}\mathcal{E}_2)\\ &\quad = \mathrm{sh}(\mathcal{F}/2)\mathrm{sh}(\mathcal{F})(\mathcal{O}_4\theta_{i1}^{\mu} + \mathcal{E}_2\theta_{j1}^{\mu}),\end{aligned} \tag{A5}$$

$$[\mathrm{ch}(\Delta),\theta_i\cdot\hat{P}]_- = \left[\mathrm{ch}\left(\frac{\Delta}{2}\right),\theta_{5i}\right]_- = 0, \tag{A6}$$

$$\begin{aligned}[\mathrm{sh}(\Delta),\theta_{i1}^{\mu}]_+ &= 2\,\mathrm{sh}^3(\mathcal{F}/2)\theta_{i1}^{\mu}\\ &\quad - \mathrm{ch}(\mathcal{F}/2)\mathrm{sh}(\mathcal{F})\theta_{j1}^{\mu}\mathcal{E}_2\\ &= 2\,\mathrm{sh}^3(\mathcal{F}/2)\theta_{i1}^{\mu}\\ &\quad - \mathrm{ch}(\mathcal{F}/2)\mathrm{sh}(\mathcal{F})\mathcal{E}_2\theta_{j1}^{\mu},\end{aligned} \tag{A7}$$

$$[\mathrm{sh}(\Delta),\theta_i\cdot\hat{P}]_+ = 2\,\mathrm{sh}(\Delta)\theta_i\cdot\hat{P},$$

$$[\mathrm{sh}(\Delta),\theta_{51}]_+ = 2\,\mathrm{sh}(\Delta)\theta_{5i}. \tag{A8}$$

After using these brackets to evaluate (63) and performing the indicated multiplications by using (88)–(90) and

$$\mathcal{O}_3\theta_{i1}^{\mu} + \mathcal{E}_2\theta_{i1}^{\mu} + \mathcal{O}_4\theta_{i1}^{\mu} + \theta_{2i}^{\mu} = 0, \tag{A9}$$

$$\theta_{2i}^{\mu}\mathcal{O}_3 + \theta_{2i}^{\mu}\mathcal{E}_2 + \theta_{i1}^{\mu}\mathcal{O}_4 + \theta_{i1}^{\mu} = 0, \tag{A10}$$

we obtain

$$\begin{aligned}&-[\mathrm{ch}(\Delta),\theta_{i1}^{\mu}]_-\,\mathrm{sh}(\Delta) - [\mathrm{sh}(\Delta),\theta_{i1}^{\mu}]_+\,\mathrm{ch}(\Delta)\\ &\quad = 2\,\mathrm{sh}(\mathcal{F}\mathcal{E}_2/2)\mathrm{ch}(\mathcal{F}\mathcal{E}_2/2)\theta_{2i}^{\mu},\end{aligned} \tag{A11}$$

$$\begin{aligned}&[\mathrm{ch}(\Delta),\theta_{2i}^{\mu}]_-\,\mathrm{sh}(\Delta) + [\mathrm{sh}(\Delta),\theta_{2i}^{\mu}]_+\,\mathrm{sh}(\Delta)\\ &\quad = 2\,\mathrm{sh}^2(\mathcal{F}\mathcal{E}_2/2)\theta_{2i}^{\mu},\end{aligned} \tag{A12}$$

$$\begin{aligned}&\mathrm{ch}(\Delta)[\theta_{i1}^{\mu},\mathrm{ch}(\Delta)]_- + \mathrm{sh}(\Delta)[\theta_{2i}^{\mu},\mathrm{sh}(\Delta)]_+\\ &\quad = 2\,\mathrm{sh}^2(\mathcal{F}\mathcal{E}_2/2)\theta_{i1}^{\mu},\end{aligned} \tag{A13}$$

$$\begin{aligned}&-\mathrm{ch}(\Delta)[\theta_{2i}^{\mu},\mathrm{sh}(\Delta)]_+ - \mathrm{sh}(\Delta)[\theta_{2i}^{\mu},\mathrm{ch}(\ )]_-\\ &\quad = 2\,\mathrm{sh}(\mathcal{F}\mathcal{E}_2/2)\mathrm{ch}(\mathcal{F}\mathcal{E}_2/2)\theta_{i1}^{\mu},\end{aligned} \tag{A14}$$

which ultimately lead to Eqs. (95a) and (95b) given in the text.

APPENDIX B: GENERALIZATION OF THE COMPATIBLE DIRAC OPERATORS FOR SEPARATE POLAR (63) AND AXIAL (102) INTERACTIONS INTO ADDITIVE POLAR AND AXIAL INTERACTIONS

We begin with the identity

$$\mathrm{ch}(\Delta_+) = \mathrm{ch}(\Delta_{\mathcal{P}})\mathrm{ch}(\Delta_{\mathcal{S}}) + \mathrm{sh}(\Delta_{\mathcal{P}})\mathrm{sh}(\Delta_{\mathcal{S}}),$$

$$\mathrm{sh}(\Delta_+) = \mathrm{sh}(\Delta_{\mathcal{P}})\mathrm{ch}(\Delta_{\mathcal{S}}) + \mathrm{ch}(\Delta_{\mathcal{P}})\mathrm{sh}(\Delta_{\mathcal{S}}).$$

The even or odd character of the functions and their respec-

tive arguments dictate whether the $\mathscr{S}_{i0}$ form commutators or anticommutators as they pass through $\mathrm{ch}(\Delta_+)$ and $\mathrm{sh}(\Delta_+)$ in (50a') and (50b'). To evaluate (50a'), we use

$$S_{10}\,\mathrm{ch}(\Delta_+) = \mathrm{ch}(\Delta_-)\mathscr{S}_{10} + [\mathscr{S}_{10},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_- + [\mathscr{S}_{10},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+, \tag{B1}$$

$$\mathscr{S}_{20}\,\mathrm{sh}(\Delta_+) = -\mathrm{sh}(\Delta_-)\mathscr{S}_{20} + [\mathscr{S}_{20},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+ + [\mathscr{S}_{20},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_-. \tag{B2}$$

Thus, when we use a similar expression to evaluate (50b'), we find

$$\begin{aligned}\mathscr{S}_1 = {}& \mathrm{ch}^2(\Delta_-)\mathscr{S}_{10} + \mathrm{ch}(\Delta_-)([\mathscr{S}_{10},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_- \\ & + [\mathscr{S}_{10},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+ + [\mathscr{S}_{20},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+ \\ & + [\mathscr{S}_{20},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_-) + \mathrm{sh}(\Delta_-)([\mathscr{S}_{20},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_- + [\mathscr{S}_{20},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+ \\ & + [\mathscr{S}_{10},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+ + [\mathscr{S}_{10},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_- - \mathrm{sh}^2(\Delta_-)\mathscr{S}_{10}),\end{aligned} \tag{B3}$$

accompanied by a similar expression for $\mathscr{S}_2$. Just as we did for polar and axial interactions alone, we isolate the derivative part of the interaction for this combination. We find

$$[\theta_1\cdot p,\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_- + [\theta_1\cdot p,\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+ = -i\theta_1\cdot\partial(\Delta_+)\mathrm{sh}(\Delta_+) + [\theta^\mu_{1\perp},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_- p_\mu + [\theta^\mu_{1\perp},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+ p_\mu, \tag{B4}$$

$$[-\theta_2\cdot p,\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+ + [-\theta_2\cdot p,\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_- = i\theta_2\cdot\partial(\Delta_+)\mathrm{ch}(\Delta_+) - [\theta^\mu_{2\perp},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+ p_\mu - [\theta^\mu_{2\perp},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_- p_\mu, \tag{B5}$$

$$[-\theta_2\cdot p,\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_- + [\theta_2\cdot p,\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+ = i\theta_2\cdot\partial(\Delta_+)\mathrm{sh}(\Delta_+) - [\theta^\mu_{2\perp},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_- p_\mu - [\theta^\mu_{2\perp},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_+ p_\mu, \tag{B6}$$

$$[\theta_1\cdot p,\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+ + [\theta_1\cdot p,\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_- = -i\theta_1\cdot\partial(\Delta_+)\mathrm{ch}(\Delta_+) + [\theta^\mu_{1\perp},\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E})]_- + [\theta^\mu_{1\perp},\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E})]_+. \tag{B7}$$

The derivative part of (B3) is

$$\begin{aligned}&\mathrm{ch}(\Delta_-)(-i\theta_1\cdot\partial(\Delta_+)\mathrm{sh}(\Delta_+) + i\theta_2\cdot\partial(\Delta_+)\mathrm{ch}(\Delta_+)) + \mathrm{sh}(\Delta_-)(i\theta_2\cdot\partial(\Delta_+)\mathrm{sh}(\Delta_+) - i\theta_1\cdot\partial(\Delta_+)\mathrm{ch}(\Delta_+)) \\ &= i\theta_2\cdot\partial(\Delta_+) - i([\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E}),\theta^\mu_{1\perp}]_-\mathrm{sh}(\Delta_+) - [\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E}),\theta^\mu_{1\perp}]_+\mathrm{sh}(\Delta_+) \\ &\quad - [\mathrm{ch}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E}),\theta^\mu_{2\perp}]_-\mathrm{ch}(\Delta_+) \\ &\quad + [\mathrm{sh}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E}),\theta^\mu_{2\perp}]_+\mathrm{ch}(\Delta_+) - [\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E}),\theta^\mu_{2\perp}]_-\mathrm{sh}(\Delta_+) + [\mathrm{sh}(\Delta_\mathscr{E})\mathrm{ch}(\Delta_\mathscr{O}),\theta^\mu_{2\perp}]_-\mathrm{sh}(\Delta_+) \\ &\quad + [\mathrm{sh}(\Delta_\mathscr{O})\mathrm{ch}(\Delta_\mathscr{E}),\theta^\mu_{1\perp}]_+\mathrm{ch}(\Delta_+) - [\mathrm{ch}(\Delta_\mathscr{O})\mathrm{sh}(\Delta_\mathscr{E}),\theta^\mu_{1\perp}]_-\mathrm{ch}(\Delta_+))\partial_\mu(\Delta_+).\end{aligned}$$

Notice how $\Delta_- \to \Delta_+$ when it passes through θ_i, allowing us to use the identity $\mathrm{ch}^2(\Delta_-) - \mathrm{sh}^2(\Delta_-) = 1$. Thus, when we collect all terms, we find Eq. (135) given in the text.

[1] P. Van Alstine and H. W. Crater, J. Math. Phys. **23**, 1697 (1982).
[2] H. W. Crater and P. Van Alstine, Ann. Phys. (NY) **148**, 57 (1983).
[3] H. W. Crater and P. Van Alstine, Phys. Rev. D15 **36**, 3007 (1987).
[4] H. W. Crater and P. Van Alstine, Phys. Rev. Lett. **53**, 1577 (1984).
[5] H. W. Crater and P. Van Alstine, Phys. Rev. D1 **37**, 1982 (1988).
[6] H. Sazdjian, Phys. Rev. D1 **33**, 3401 (1986); Phys. Rev. D15 **34**, 189 (1986).
[7] H. Sazdjian, Phys. Lett. B **180**, 146 (1986).
[8] H. Sazdjian, J. Math. Phys. **29**, 1620 (1988).
[9] I. T. Todorov, in *Properties of the Fundamental Interactions*, edited by A. Zichichi (Editrice Compositori, Bologna, 1973), Vol. 9, Part C, pp. 953–979; Phys. Rev. D **3**, 2351 (1971).
[10] I. T. Todorov, "Dynamics of Relativistic Point Particles as a Problem with Constraints," Dubna Joint Institute for Nuclear Research No. E2-10175 (1976).
[11] P. Van Alstine, "Relativistic Two Body Problem from Singular Actions," Ph.D. dissertation, Yale University, 1976; M. Kalb and P. Van Alstine, Yale Reports Nos. C00-3075-146 (1976) and C00-3075-156 (1976).
[12] A. Komar, Phys. Rev. D **18**, 1887 (1978).
[13] M. King and F. Rohrlich, Phys. Rev. Lett. **44**, 62 (1980).
[14] L. P. Horowitz and F. Rohrlich, Phys. Rev. D **24**, 1528 (1981).
[15] F. A. Berezin and M. S. Marinov, JETP Lett. **21**, 678 (1975); Ann. Phys. (NY) **104**, 336 (1977).
[16] P. Van Alstine and H. W. Crater, Phys. Rev. D **15**, 1037 (1986).
[17] H. Sazdjian, Phys. Lett. **56** B, 381 (1985).
[18] H. Sazdjian, J. Math. Phys. **28**, 2618 (1987).

A new scheme for enumerating connected diagrams on the lattice

Yongxin Song[a] and Yuanben Dai
Center of Theoretical Physics, CCAST (World Laboratory), P. O. Box 8730, Beijing, People's Republic of China and Institute of Theoretical Physics, P. O. Box 2735, Beijing 100080, People's Republic of China

(Received 4 January 1990; accepted for publication 4 April 1990)

In this article, a scheme based on recursive generation of connected diagrams is proposed for enumerating the connected diagrams in various expansions for lattice field theories. This scheme is rigorous and general, and can be applied to the analytical methods in various lattice field theories. Some notions and results of the modern graph theory in mathematics are used. As examples, the diagram theoretic factors that occur in the average link energy of the lattice chiral field model and in the external source parameter of the pure lattice gauge field model are calculated.

I. INTRODUCTION

The lattice field theory has developed into an important field for both particle physics and statistical mechanics.[1] It makes it possible to make nonperturbative calculations for physical properties of a strongly interacting system. With the appearance of the discrete mechanics and the discrete quantum mechanics,[2] the interest in lattice field theories is increasing.

In Lagrangian formulation of the lattice field theory, the measured value of any physical quantity is expressed as the mathematical expectation of the corresponding observable computed with the Boltzman weight $\exp[S(U)]$. Therefore, to evaluate approximately the infinite multiple integrals evolved becomes a central problem of lattice field theories. In the numerical Monte Carlo[3] method, the simulation can be performed only on a finite lattice. In a variety of analytical methods, the infinite multiple integrals are expanded into infinite series of finite multiple integrals. One hopes to calculate systematically corrections to some order of approximation in the expansion.[4]

In either different physical problems or different analytical methods, these diagrams differ much in dimension, shape, and criterion of the classification. Also, the number of lattice diagrams increases drastically with the order of correction. The result contains a lot of diagrams differing in topological structure. Therefore, the above problems of the lattice diagrammatic theory are already far beyond the attainment of modern graph theory.[5] It is necessary that a general scheme for enumerating exactly the lattice diagrams be devised starting from fundamental mathematical principles. The present article is devoted to this problem.

This paper is organized as follows. Section II brings up problems of the diagrammatic theory in lattice field theories and introduce necessary definitions and notations. Section III proves some fundamental theorems and proposes a standard program for solving exactly these problems. Sections IV and V apply this standard program to the analytical calculation of the diagram theoretic factors in the lattice chiral field model and the pure lattice gauge field model, respectively. The concluding remarks are given in Sec. VI.

II. THE ORIGIN OF PROBLEMS

To solve exactly the problems of the lattice diagrammatic theory mentioned above, we need to clearly understand its physical origin, and introduce some definitions and notations. A lattice always contains various lattice constituents, sites, links, plaquettes, multiangularities, and so on. On quite a few types of the lattice constituents, we place some physical quantities. For example, a field variable is defined on each site for matter fields, on each link for gauge fields, Wilson's action is defined on all of plaquettes. Field variables play a special role in configuration integrals, for they are integral variables. In order to fix the notation, we write each field variable as U_μ, where the Greek subscript μ indicates both species of the field and its location (the position and the direction) on the space-time lattice, and the collection of U_μ as a set U. The lattice constituent on which the field variable U_μ is placed is called a lattice element.

We want to present illustratively the integral

$$\bar{Y} = \int DU\, Y(U'), \tag{2.1}$$

of the local observable $Y(U')$ over the configuration variables, where DU is the corresponding measure normalized to 1, and can be factorized as the infinite product $\Pi_\mu\, dU_\mu$. "Local" here means that U' is a finite set of field variables.

The observable $Y(U')$ is assumed to take the form

$$Y = \sum_i N_i Y_i, \tag{2.2}$$

where Y_i is a monomial of field variables and the integer N_i is the number of Y_i terms contained in Y. The diagrammatical representations for U_μ and Y_i are denoted by $d(U_\mu)$ and $d(Y_i)$, respectively. The $d(Y_i)$ is obtained by drawing $d(U_\mu)$ on the corresponding lattice element in a figure of space-time lattice for each factor U_μ appearing in the monomial Y_i. The diagram is labeled by lattice locations of its elements. Notice that the constant coefficient and group theoretical factors of Y_i are neglected in this diagrammatical representation. Therefore, the integral $\bar{Y}$ is represented by the union of a set of labeled diagrams denoted by $D(Y)$,

[a] Also at Department of Physics, Sichuan University, Chengdu, People's Republic of China.

 0022-2488/90/082015-10$03.00

$$D(Y)=\bigcup_i N_i d(Y_i)\,. \tag{2.3}$$

Here, $D(Y)$ can be written schematically as

$$D(Y)=\bigcup_{\{n_\mu\}} N_{\{n_\mu\}}\prod_\mu d^{\,n_\mu}(U_\mu)\,, \tag{2.4}$$

if $Y_i=\text{const}\times\Pi_\mu(U_\mu)^{n_\mu}$. In (2.4) Π_μ means the collection of $d(U_\mu)$'s appearing in the diagram specified by the integers n_μ.

We introduce an algebraic expression for the diagram set $D(Y)$. Consider an isomorphic mapping f_1 from diagrams to direct products of vectors with integer components. Let $\hat{U}_\mu$ be the unit vector corresponding to the field variable U_μ and let f_1 be defined by

$$\begin{aligned} &f_1 d(U_\mu)\equiv a(U_\mu)=\hat{U}_\mu\,,\\ &f_1\prod_\mu d(U_\mu)\equiv a\left(\prod_\mu U_\mu\right)=\prod_\mu \hat{U}_\mu\,,\\ &f_1\bigcup_i N_i d(Y_i)\equiv f_1\left(\sum_i N_i Y_i\right)=\sum_i N_i a(Y_i)\,. \end{aligned} \tag{2.5}$$

In the above formula, Π_μ denotes the direct product of vectors $\hat{U}_\mu$. Therefore, the algebraic expression of $D(Y)$ in (2.4) is

$$A(Y)=\sum_i N_i a(Y_i)=\sum_{\{n_\mu\}} N_{\{n_\mu\}}\prod_\mu[\hat{U}_\mu]^{n_\mu}\,. \tag{2.6}$$

The product of two algebraic expressions $A(Y_1)$ and $A(Y_2)$ is defined in the usual sense of the direct product of tensor spaces. Evidently,

$$A(Y_1)A(Y_2)=f_1[D(Y_1)\sqcap D(Y_2)]=A(Y_1Y_2)\,. \tag{2.7}$$

This corresponds to the union of the set of diagrams each one of which is composed of two subdiagrams from the two sets $D(Y_1)$ and $D(Y_2)$, respectively.

The algebraic expression is very useful. The one-to-one correspondence between $A(Y)$ and $D(Y)$ makes it possible to carry out the program of enumeration and other operations on diagrams by analytical calculations.

The diagrams having the same topological structure but differ in labels of lattice locations are said to be equivalent. Geometrically, these equivalent diagrams can be brought to one another through appropriate symmetric operations. Configuration integrals corresponding to equivalent diagrams are equal in value. The number of the equivalent diagrams is called the diagram theoretic factor, and these equivalent diagrams may be represented by any one of them without label. The problem of grouping the terms contained in the configuration integral in classes changes into the classification of the diagrams in terms of topological structure, and the problem of calculating the number of terms in each class changes into the enumeration of the equivalent diagrams. This is just what the lattice diagrammatic theory needs to cope with.

Now let us turn to the physics of the lattice field theory. We are interested in the path integrals of the partition function

$$Z=\int DU\exp[S(U)]\,, \tag{2.8}$$

and the expectation value

$$\langle X\rangle=Z^{-1}\int DU\,X(U^r)\exp[S(U)] \tag{2.9}$$

of the local observable $X(U^r)$ computed with the Boltzmann weight $\exp[S(U)]$, where $S(U)$ is the action governing the dynamics. The action has the following properties: It is an infinite sum of the form

$$S(U)=\sum s_i(U^i)\,, \tag{2.10}$$

where each term $s_i(U^i)$ is a monomial describing the coupling among the field variables in the finite subset U^i. Hence s_i is defined on the lattice constituent composed of the elements that correspond to the field variables involved in $s_i(U^i)$. Here, the Latin subscript i indicates both the lattice location (the position and the orientation) of this lattice constituent and the particular monomial among those defined on the same lattice constituent.

Due to translation and rotation invariance of the theory, the functional form of s_i is usually independent of the location of the lattice constituent.

Due to the important role and the special form of the action, we call the diagram representation of S_i the basic diagram. To be slightly more general, we shall denote it by $d(b_i)$ instead of $d(s_i)$. The union of the set of all $d(b_i)$ is denoted by $D(B)$. Correspondingly, the analytical expression for $d(b_i)$ and $D(B)$ are denoted by b_i and B, respectively. The b_i can be written as the out product of $\hat{U}_\mu$'s corresponding to U_μ's contained in s_i:

$$A(S)=\sum_i b_i=B\,. \tag{2.11}$$

To evaluate the infinite multiple integral, we have to adopt various methods of successive approximations. For example, to expand the Boltzmann weight in powers of the action

$$\begin{aligned} \exp[S(U)]&=\prod_i\exp[s_i(U_i)]\\ &=\prod_i\left[1+s_i(U_i)+\tfrac{1}{2}s_i^2(U_i)+\cdots\right] \end{aligned} \tag{2.12}$$

and to integrate term by term over the configuration variables. This is a standard methods of the strong-coupling (high-temperature) expansion. Afterward, the k th-order correction of the expectation value means the sum of terms involving the factor $\Pi_i s_i^{k_i}(U)$, where $\{k_i\}$ satisfies the relation $\Sigma_i k_i=k$.

The diagrammatical representation corresponding to the k th correction of $\langle X\rangle$ is called the k th-order correction diagram. The k th-order correction diagram set in the strong coupling expansions can be plotted out with the rules described above. The coefficients of Taylor's expansion that do not affect diagrammatical enumeration are temporarily neglected just as the convention treatment does. Thus, we have

$$D(X^{(k)})=d(X)\sqcap D(B^k)=D(X^{(k-1)})\sqcap D(B)\,, \tag{2.13}$$

where $D(B^k)$ is the union of the set of k basic diagrams,

$$D(B^k)=\overset{k}{\sqcap}D(B)=D(B^{k-1})\sqcap D(B)\,. \tag{2.14}$$

That is to say, the k th-order correction diagram set $D(X^{(k)})$ is the direct product of $D(X)$ and k sets of basic diagram $D(B^k)$. The analytical expression of the k th-order correction diagram set can be written as

$$A(X^{(k)}) = A(X)\hat{B}^k = A(X^{(k-1)})\hat{B}. \tag{2.15}$$

This expression underlies the analytical calculation of the diagram theoretic factors of $\langle X \rangle$.

The direct evaluation of the expansion in powers of the action is not very useful in the actual computation of the series. In order to avoid the enumeration of disconnected diagrams, let us expand the free energy of the system in powers of the action:

$$F = -\ln Z = \sum_n \frac{1}{n!} \overline{S}_c^n, \tag{2.16}$$

where

$$\overline{S}_c' = \overline{S}, \quad \overline{S}_c^2 = \overline{S}^2 - (\overline{S})^2,$$
$$\overline{S}_c^3 = \overline{S}^3 - 3\overline{S}_c^2 S - (\overline{S})^3, \tag{2.17}$$

which is the so-called comulant expansion.[6] This method is extensively used in statistical mechanics and lattice gauge theory. The following treatment is restricted to connected diagrams.

To decouple the infinite multiple integral, some trial action in the form of the generalized mean field is used in the variational method of Lagrangian formulation. This trial action simulates the physical behavior of ordering in the region of weak coupling. Many results can be obtained with this method. Furthermore, making the cumulant expansion of the difference between the real action and the trial action, we can collect advantages of the two methods mentioned above, and can give good results in good agreement with the Monte Carlo method. Also in this method, so called the variational-cumulant expansion method,[7] the classification and the enumeration of connected diagrams are needful.

III. THE SOLVING SCHEME

In the previous section we have talked about connected diagrams. Since the terminology sometimes causes confusion, and also because we want to cope with general cases, we need to have an abstract definition of connectivity. In various expansions mentioned above, connectivity of two connected parts of a diagram refers to overlapping of sets of field variables corresponding to them in the multiple integral. This can be clearly seen, for example, in (2.17). To define connectivity of the diagram made of several connected parts, first of all, we have to give the definition of adjacency. Noting that any diagram is in fact a finite subset of $D(U)$, we say d_i and d_j are adjacent if and only if there exists a nonempty subset $D(U^{ij})$ of $D(U)$ in both d_i and d_j, i.e., the intersection of d_i and d_j is not the empty set ϕ,

$$d_i \cap d_j = D(U^{ij}) \quad (\phi \neq D(U^{ij}) \subset D(U)). \tag{3.1}$$

The diagram is said to be a connected diagram if and only if any two of its connected parts are connected through successive adjacent parts. We can make a homomorphic mapping of a lattice diagram $d = \Pi_i d_i$ into a graph g in the modern graph theory: $d \to g$ under which each connected part d_i of d maps to a point i, and points i and j are adjacent if and only if their preimages are adjacent. Thus, g is a $(p,.)$ graph, where p is the number of the connected parts of its preimage. The connectivity of diagrams defined above implies that d is a connected diagram if and only if its homomorphic g is a connected graph in the sense of the graph theory. Therefore, we can make use of notions and results in this branch of mathematics.

It is clear from (2.10)–(2.14) that the basic diagram is the building block of the diagrams in various expansions and describes the coupling among field variables, hence it is regarded as an elementary connected part. Therefore, the connectivity of diagrams in various expansions is precisely defined by the above general definition of connectivity. Whether the diagram for the local observable X, $d(X)$ is considered as a connected part depends on one wants to include or exclude the diagrams in which different parts of it are not connected through basic diagrams. We shall consider mainly the case in which $d(X)$ can be considered as a connected part.

We want to construct successively the connected diagrams occurring in cumulant expansions. Similar to the cut point of a connected graph in the modern graph theory, we define the cut basic diagram of a connected diagram. For this purpose, we introduce the homomorphic mapping f_2: $d(X^{(k)}) \to g(X^{(k)})$. Under f_2, each connected part $d_i(X)$ of $d(X) = \Pi_i d_i(X)$ maps into a labeled point $X_i(1)$, the n_ith-multiple basic diagram $d^{n_i}(b_i)$ into a labeled point $b_i(n_i)$, and a pair of these labeled points is adjacent if and only if their preimages are adjacent. Therefore, the homomorphic $g(X^{(k)})$ of $d(X^{(k)})$ under f_2 is a $(p+q,.)$ graph that is labeled by the lattice locations and the multiplicities of all the connected parts of its preimage $d(X^{(k)})$, where q and p are the numbers of the connected parts of $d(X)$ and the lattice locations occupied by the basic diagrams of $d(X^{(k)})$, respectively. If $b_i(1)$ is a cut point of $g(X^{(k)})$, then $d^1(b_i)$ is called a cut basic diagram of $d(X^{(k)})$. "Cut basic diagram" here is a new concept introduced by us. If any one of its cut basic diagrams is removed, a connected diagram becomes disconnected.

From (2.10) we know any $d_c(X^{(k)})$ in the connected k th-order correction diagram set $D_c(X^{(k)})$ of $\langle X \rangle$ can be constructed from the $(k-1)$th-order correction diagram set $D(X^{(k-1)})$ by adding one basic diagram. That is to say, the number $n[d_c(X^{(k)})]$ of the diagram $d_c(X^{(k)})$ in the cumulant expansion is correctly given by the number of $d_c(X^{(K)})$ appearing in the set

$$D(X^{(k-1)}) \sqcap D(B). \tag{3.2}$$

Let $n_c[d_c(X^{(k)})]$ be the number of $d_c(X^{(k)})$ that can be constructed from the connected $(k-1)$th degree correction set $D_c(X^{(k-1)})$ by adding one basic diagram, i.e., the number of $d_c(X^{(k)})$ appearing in the set:

$$D_c(X^{(k-1)}) \sqcap D(B). \tag{3.3}$$

The relation between $n[d_c(X^{(k)})]$ and $n_c[d_c(X^{(k)})]$ is given by the following theorem.

Theorem 1: If c is the number of cut basic diagrams of

the connected k th-order correction diagram $d_c(X^{(k)})$ of $\langle X\rangle$, then

$$n[d_c(X^{(k)})] = [k/(k-c)]n_c[d_c(X^{(k)})]. \quad (3.4)$$

Proof: A given $d_c(X^{(k)})$ belonging to the connected k th-order correction diagram set $D_c(X^{(k)})$ of $\langle X\rangle$ can be expressed algebraically by $a(X^{(k)}) = a(X)\,\Pi_i\, b_i^{k_i}$, where the set $\{k_i\}$ is determined by $d_c(X^{(k)})$ and satisfies the condition $\Sigma_i\, k_i = k$. To find the algebraic expression of all $(k-1)$th-order correction diagrams that can be used to construct $d_c(X^{(k)})$ with respect to each functional argument b_i;

$$\sum_i \frac{\delta a(X^{(k)})}{\delta b_i} = \sum_i k_i a(X) b_i^{k_i-1} \prod_{j\neq i} b_j^{k_j}. \quad (3.5)$$

Letting f_1 be the isomorphic mapping from a diagram to its algebraic expression defined above, we have

$$f_1^{-1}\left[a(X) b_i^{k_i-1} \prod_{j\neq i} b_j^{k_j}\right] \in \begin{cases} D(X^{(k-1)}) & (k_i>0), \\ D(X^{(k-1)})/D_c(X^{(k-1)}) & (k_i=1), \end{cases} \quad (3.6)$$

from the definition of the cut basic diagram. Therefore, the number of terms on the right-hand side of (3.5) (the term with a coefficient k_i is counted k_i times,) is $n[d_c(X^{(k)})]$ while that with $k_i = 1$ terms excluded is just $n_c[d_c(X^{(k)})]$. Thus, we obtain Eq. (3.4).

The following conclusion can be obtained from the first theorem. The diagram $d_c(X^{(k)})$ of $D_c(X^{(k)})$ actually appearing in the expansion of $\langle X\rangle$ ($n[d_c(X^{(k)})]\neq 0$) can be completely constructed and exactly enumerated from $D_c(X^{(k-1)})$ if and only if $n_c[d_c(X^{(k)})]\neq 0$. In other words, the number of the cut basic diagrams is less than the number of the basic diagrams, i.e.,

$$c < k. \quad (3.7)$$

This leads to the restriction on the diagram representation $d(X)$ of the local observable X given by the second theorem.

Theorem 2: If $d(X)$ is a connected diagram, then any $d_c(X^{(k)})$ ($k\geqslant 1$) contains at least one basic diagram that is not a cut basic diagram.

Proof: Under the mapping f_2 defined above, $d(X^{(k)})$ maps into a graph $(p+q,.)$. When $d(X)$ is a connected diagram, $q=1$. In the language of the homomorph of $d(X^{(k)})$ under f_2, the second theorem is just the theorem of the graph theory: Any nontrivial connected graph contains at least two points that are not the cut points. Here a trivial graph (1,0) corresponds to the connected diagram $d(X)$, the connected zeroth-order approximation diagram $d(X^{(0)})$.

The above two theorems lay a solid foundation for constructing completely and enumerating precisely the connected diagram set of each order of corrections in the expansion of $\langle X\rangle$ from that of the order lower by 1, if the studied local observable corresponds to a connected diagram. So, we are in the position to design a realizable scheme for constructing successively the connected diagram set.

The operator $P(U_\mu)$ defined by the rule

$$P(U_\mu)A(U^t) = \sum_{U_\mu\in U^t} \delta_{\mu\nu} A(U^t), \quad (3.8)$$

(here, $\delta_{\mu\nu}$ is the Kronecker delta symbol) is idempotent, and can be taken as a projection operator. The terms depending on U_μ are chosen from $A(U^t)$ by the act of $P(U_\mu)$ upon $A(U^t)$. The equation

$$P(U_\mu)A(U^t) = 0, \quad (3.9)$$

implies that $A(U^t)$ is independent of U_μ, i.e., $U_\mu \bar{\in} U^t$. Therefore, we can give and prove the following theorem.

Theorem 3: If $d_c(a) = f_1^{-1}a(d)$ is a connected diagram and $D_c(A) = f_1^{-1}A(D_c)$ is a set of connected diagrams, then

$$f_1^{-1}[a(d_c)P(U_\mu)A(D_c)]_{P(U_\mu)a(d_c)\neq 0} = \{[d_c(a)\sqcap D_c(A)]_c | d(U_\mu)\in d_c(a)\cap D_c(A)\}, \quad (3.10)$$

i.e., $a(d_c)P(U_\mu)A(D_c)|_{P(U_\mu)a(d_c)\neq 0}$ expresses algebraically the connected diagram set made of $d_c(a)$ and $D_c(A)$ that share at least U_μ.

Proof: When $P(U_\mu)a(d_c)\neq 0$ indicates that $d_c(U)$ contains U_μ, and the act of $P(U_\mu)$ upon $A(D_c)$ chooses the diagrams containing U_μ from $D_c(A)$. Thus, $d_c(a)$ shares at least U_μ with $D_c(A)$. Both $d_c(a)$ and $D_c(A)$ can be regarded as connected parts in the determination of the connectivity of diagrams made of them, for they are all connected diagrams. So that, from the general definition of the connectivity we know that the lhs of Eq. (3.10) is the set of connected diagrams made of $d_c(a)$ and $D_c(A)$.

In the same way, the operator

$$P(U^r) = \prod_{U_\mu\in U^r} P(U_\mu), \quad (3.11)$$

is also idempotent and can be taken as a projection operator. The act of $P(U^r)$ upon $A_c(D_c)$ results in an algebraic expression of diagrams that contain at least the r subset $\hat{U}^r = \{\hat{U}_{\mu_1}, \hat{U}_{\mu_2}\cdots\}$ of $\hat{U}$. Thus, $[a(d_c)P(U^r)A(D_c)]_{P(U^r)A(D_c)\neq 0}$ algebraically expresses diagrams that belong to the connected diagram set $[d_c(a)\Pi D_c(A)]_c$ and satisfy $D(U^r)\subset d_c(a)\cap D_c(A)$, i.e.,

$$f_1^{-1}[a(d_c)P(U^r)A(D_c)]_{P(U^r)a(d_c)\neq 0} = \{[d_c(a)\sqcap D_c(A)]_c | D(U^r)\subset d_c(a)\cap D_c(A)\}. \quad (3.12)$$

Here we have to emphasize the following fact. The r.h.s. of Eq. (3.12) is the set of connected diagrams that satisfy $D(U^r)\subset d_c(a)\cap D_c(A)$ not $D(U^r) = d_c(a)\cap D_c(A)$. That is to say, besides every element of $D(U^r)$, some elements not belonging to $D(U^r)$ are also shared by $d_c(a)$ and $D_c(A)$. Hence, a lot of diagrams differing in topological structure are contained in the rhs of Eq. (3.12), and need to be separated. This aim can be reached by means of the sieve method in combinatorial mathematics.[8]

Theorem 4: If $d_c(a) = f_1^{-1}a(d_c)$ is a connected diagram and $D_c(A) = f_1^{-1}A(D_c)$ is a set of connected diagrams, and U^r is a r subset of $\hat{U}(a) = \{U_\mu | P(U_\mu) a_c(d_c)\neq 0\}$, then

$$a(d_c) \sum_{\substack{\hat{U}(a) \\ \hat{U}^j \subset \hat{U}(a)/\hat{U}^r}} (-)^j P(U^j) P(U^r) A(D_c)$$

$$= f_1\{[d_c(a) \sqcap D_c(A)]_c | D(U^r) = d_c(a) \cap D_c(A)\}, \tag{3.13}$$

(where j is the number of factors U_μ in U^j,) expresses analytically the connected diagram set made of $d_c(a)$ and $D_c(A)$ that share only $D(U^r)$.

Proof: The formula (3.12) is just a specific form of the principle of inclusion and exclusion in combinatorial mathematics for our problem.

The above four theorems are enough to establish a standard program for constructing successively, classifying topologically, and enumerating exactly the connected diagram set, provided that the local observable corresponds to a connected diagram set. To find the diagram theoretic factors of cumulant expansions, proceed as follows.

(i) In Eq. (3.13), taking $D_c(A)$ to be the basic diagram set $D(B)$, $d_c(a)$ one of the connected $(k-1)$th-order correction diagrams $d_c(X^{(k-1)})$ of $\langle X \rangle$, $D(U^r)$ each subset of the elements belonging to the diagram $d_c(X^{(k-1)})$, and defining the projection operators according to Eqs. (3.8) and (3.11), we obtain all the connected k th-order correction diagrams $[d_c(X^{(k-1)}) \sqcap D(B)]_c$ that are constructed from $d_c(X^{(k-1)})$ and classified by the elements shared by $d_c(X^{(k-1)})$ and $D(B)$. When $d_c(X^{(k-1)})$ runs over all elements of the connected $(k-1)$th-order correction diagram set $D_c(X^{(k-1)})$, and the terms that are of the same form but constructed from different elements of $D_c(X^{(k-1)})$ are incorporated, all the connected k th-order correction diagrams that can be constructed from $D_c(X^{(k-1)})$ are worked out.

(ii) To take the contribution of disconnected $(k-1)$th-order correction diagrams into account, for each connected k th-order correction diagram $d_c(X^{(k)})$ obtained in step (i), we read out the number of the cut basic diagrams c and give each $d_c(X^{(k)})$ a weight factor $k/(k-c)$ according to Theorem I.

(iii) We group diagrams obtained above into classes in topological structure. Diagrams with the same topological structure can be transformed to each other just by changing the labels for lattice locations, we draw a diagram without label that represents an equivalent class. We make the sum of weight factors $k/(k-c)$ of all diagrams in each class. The correct diagram theoretic factors are obtained.

It is a problem to apply the above program to the case that $d(X)$ is not a connected diagram, for the diagrams satisfying $c = p$ are omitted. At this point, we can construct the connected diagrams by the above program starting from any connected part of $d(X)$. Then, we choose the diagrams containing all the diagrammatical representation for field variable belonging to $d(X)$ by means of the projection operator.

So far we have neglected the property of group theoretic factors in our discussion. Diagrams that can be brought to each other by transformations of corresponding field variables in the internal symmetry group of the theory should be considered as equivalent also.

IV. THE LATTICE CHIRAL FIELD MODEL

In the lattice chiral field models, the chiral field variable U_n is a given representation of the gauge group, and defined on each site $S_n = (n_1, n_2, ..., n_d)$ of the d-dimensional space-time lattice.

When calculating the diagram theoretic factors of this model, we shall represent the field variable U_n in the integral with a point $d(U_n)$ on the site s_n. Adopting the simple hypercubic Euclidean lattice, and regarding the lattice spacing a as 1, we can take the coordinates of sites $n_1, n_2, ..., n_d$ to be all integers. The action S of the SU(N) lattice chiral field model is of the form

$$S(U) = \frac{\beta}{2N} \sum_{n,\mu} \mathrm{Tr}\,(U_n U^\dagger_{n+\hat{\mu}} + \mathrm{h.c.}), \tag{4.1}$$

where $\beta = 1/g^2$, g is the coupling constant of the chiral fields, $\hat{\mu}$ the unit vector on the x axis direction, and h.c. the Hermitian conjugate of the former term. Throwing away the constant $(\beta/2N)\cdot\frac{1}{2}$ and the group structure, which have no effect on the diagrammatic enumeration, we can straightforwardly read out the algebraic expression

$$B = \sum_{n,\mu} \hat{L}_{n,n+\hat{\mu}}, \quad \mu = \pm 1, \pm 2, ..., \pm d, \tag{4.2}$$

of the basic diagram set

$$D(B) = \{d(L_{n,n+\hat{\mu}})\}. \tag{4.3}$$

The quantity

$$\hat{L}_{n,n+\hat{\mu}} = \hat{U}_n \hat{U}_{n+\hat{\mu}} = \hat{U}_{n+\hat{\mu}} \hat{U}_n = \hat{L}_{n+\hat{\mu},n}, \tag{4.4}$$

determined by the disorder pair of field variables at neighboring sites $(s_n, s_{n+\hat{\mu}})$ is represented diagrammatically by a line segment $d(L_{n,n+\hat{\mu}})$ on the link $L_{n,n+\hat{\mu}}$ connecting the site s_n with the sites $s_{n+\hat{\mu}}$. In the calculation of the group theoretic factors, $L_{n,n+\hat{\mu}}$ contributes a factor $(\beta/2N) \times \mathrm{Tr}(U_n U^\dagger_{n+\hat{\mu}} + \mathrm{h.c.})$ into the integrand of the configuration integral.

In the study of the lattice chiral field model, the average link energy is customarily considered to be the order parameter, by which we detect the phase transition of the system. We desire to compute the mathematical expectation of the local observable

$$X(U^r) = (\beta/2N)\mathrm{Tr}(U_0 U_{\hat{1}}^\dagger + \mathrm{h.c.}). \tag{4.5}$$

In the same way, the effect of the group structure and the constant $\beta/2N$ can be neglected temporarily in the calculation of the diagrammatic factors. So, the diagram set of the zeroth-order approximation

$$D(X^{(0)}) = d(X) = d(L_{0,\hat{1}}), \tag{4.6}$$

can be analytically expressed as

$$A(X^{(0)}) = a(X) = \hat{L}_{0,\hat{1}}, \tag{4.7}$$

and the connected diagram set of the k th-order correction of $\langle X \rangle$ becomes the set of the connected diagrams made of k links and the link $d(L_{0,\hat{1}})$. Our task is just the classification and the enumeration of these diagrams.

Let us introduce the projection operators $P(U_n)$ on $d(U_n)$ at the site s_n and $P(U^r)$ on the given r subset $D(U^r)$ on the site set $\{s_n\}$ in terms of the rule (3.8) and (3.11). It is easy to verify that the projection operators have the follow-

ing properties: (a) $P(U_n)B=\Sigma_\mu \hat{L}_{n,n+\hat{\mu}}$. (b) If there is a site $s_{\hat{i}}$ satisfying $s_{\hat{i}}\in L_{n,n+\hat{\mu}}$ and $d(s_m,s_{\hat{i}})>1$, then $P(U_m)\hat{L}_{n,n+\hat{\mu}}=0$. (c) If there are two sites s_m and s_n satisfying s_m, $s_n\in S^r$ and $d(s_m,s_n)>1$ then $P(U^r)B=0$. Here, the notation $s_{\hat{i}}\in L_{n,n+\hat{\mu}}$ means that the site $s_{\hat{i}}$ is one of the end points of the link $L_{n,n+\hat{\mu}}$, and the distance between the sites s_m and s_n is defined as below:

$$d(s_m,s_n)=\sum_\mu |m_\mu-n_\mu| . \tag{4.8}$$

The above properties of the projection operators make it convenient to evaluate the quantities $P(U^r)B$.

Now we are in the position to construct the connected diagram set in the order by order manner. There is only one diagram $d(L_{0,\hat{i}})$ in the zeroth-order approximation diagram set that has only two points at the sites s_0 and $s_{\hat{i}}$. Thus, it is sufficient to calculate the quantities $P(U_0)B$, $P(U_{\hat{i}})B$ and $P(U_0,U_{\hat{i}})B$ for constructing the first-order correction diagram set. By virtue of the sieve method, we find out the algebraic expression of the connected first-order correction diagram set

$$A(X^{(1)})=\sum_{m=1}^{3} A_{1,m} , \tag{4.9}$$

where

$$\begin{aligned}
A_{1,1}&=\hat{L}_{0,\hat{i}}P(U_0,U_{\hat{i}})B=\hat{L}^2_{0,\hat{i}} ,\\
A_{1,2}&=\hat{L}_{0,\hat{i}}[P(U_0)-P(U_0,U_{\hat{i}})]B\\
&=\hat{L}_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}_{0,\hat{\mu}} ,\\
A_{1,3}&=\hat{L}_{0,\hat{i}}[P(U_{\hat{i}})-P(U_0,U_{\hat{i}})]B\\
&=\hat{L}_{0,\hat{i}}\sum_{\mu\neq -1}\hat{L}_{\hat{i},\hat{i}+\hat{\mu}} .
\end{aligned} \tag{4.10}$$

Now, we start to construct the connected second-order correction diagram set by using the standard program proposed in the preceding section. We need to start from every diagram in the connected first-order diagram set (4.9). Besides that, the points at sites s_0 and $s_{\hat{i}}$ are the elements of every diagram. Points at the sites $s_{\hat{\mu}}$ ($\mu\neq 1$) and $s_{\hat{i}+\hat{\mu}}$ ($\mu\neq -1$) are also the ones of the diagrams $D(A_{1,2})$ and $D(A_{1,3})$, respectively. Thus, we must calculate the quantities $P(U^r)B$ where the corresponding set of sites is $s^r=\{s_0\}$, $\{s_{\hat{i}}\}$, $\{s_{\hat{\mu}}\}$, $\{s_{\hat{i}+\hat{\mu}}\}$, $\{s_0,s_{\hat{i}}\}$, $\{s_0,s_{\hat{\mu}}\}$, $\{s_0,s_{\hat{i}+\hat{\mu}}\}$, $\{s_{\hat{i}},s_{\hat{\mu}}\}$, $\{s_{\hat{i}},s_{\hat{i}+\hat{\mu}}\}$, $\{s_0,s_{\hat{i}},s_{\hat{\mu}}\}$, $\{s_0,s_{\hat{i}},s_{\hat{i}+\hat{\mu}}\}$. We can obtain all of the connected second-order correction diagrams by means of the principle of inclusion and exclusion (3.13). In addition, when the order of correction k is higher than 1, it is necessary to supply the weight factor $k/(k-c)$ for the diagram containing c cut basic diagrams according to Theorem 1. From the definition given in preceding section, we can determine the value of c of each obtained diagram. In this way, it is found that the diagrams $d(L_{0,\hat{i}}\ \Sigma_{\mu\neq 1}L_{0,\hat{\mu}}\ \Sigma_{\nu\neq -\mu}L_{\hat{\mu},\hat{\mu}+\hat{\nu}})$ and $d(L_{0,\hat{i}}\ \Sigma_{\mu\neq -1}L_{\hat{i},\hat{i}+\hat{\mu}}\Sigma_{\nu\neq -\mu}L_{\hat{i}+\hat{\mu},\hat{i}+\hat{\mu}+\hat{\nu}})$ are of $k=2$, and contain the cut basic diagrams $d(L_0,\hat{\mu})$ and $d(L_{\hat{i},\hat{i}+\hat{\mu}})$, respectively, hence we have to give them the weight factor 2. Incorporating the same diagrams constructed from different diagrams of the first-order correction, we have worked out the analytical expression of the connected second-order correction diagram set

$$A(X^{(2)})=\sum_{m=1}^{10} A_{2,m} , \tag{4.11}$$

where

$$\begin{aligned}
A_{2,1}&=A_{1,1}P(U_0,U_{\hat{i}})B=\hat{L}^3_{0,\hat{i}} ,\\
A_{2,2}&=A_{1,1}[P(U_0)-P(U_0,U_{\hat{i}})]B+A_{1,2}[P(U_0,U_{\hat{i}})\\
&\quad -P(U_0,U_{\hat{i}},U_{\hat{\mu}})]B\\
&=2\hat{L}^2_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}_{0,\hat{\mu}} ,\\
A_{2,3}&=A_{1,1}[P(U_{\hat{i}})-P(U_0,U_{\hat{i}})]B\\
&\quad +A_{1,3}[P(U_0,U_{\hat{i}})-P(U_0,U_{\hat{i}},U_{\hat{i}+\hat{\mu}})]B\\
&=2\hat{L}^2_{0,\hat{i}}\sum_{\mu\neq -1}\hat{L}_{\hat{i},\hat{i}+\hat{\mu}} ,\\
A_{2,4}&=A_{1,2}[P(U_0,U_{\hat{\mu}})-P(U_0,U_{\hat{i}},U_{\hat{\mu}})]B\\
&=\hat{L}_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}^2_{0,\hat{\mu}} ,\\
A_{2,5}&=A_{1,3}[P(U_{\hat{i}},U_{\hat{i}+\hat{\mu}})-P(U_0,U_{\hat{i}},U_{\hat{i}+\hat{\mu}})]B\\
&=\hat{L}_{0,\hat{i}}\sum_{\mu\neq -1}\hat{L}^2_{\hat{i},\hat{i}+\hat{\mu}} ,\\
A_{2,6}&=A_{1,2}[P(U_0)-P(U_0,U_{\hat{i}})-P(U_0,U_{\hat{\mu}})\\
&\quad +P(U_0,U_{\hat{i}},U_{\hat{\mu}})]B\\
&=\hat{L}_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}_{0,\hat{\mu}}\sum_{\substack{\nu\neq 1\\ \nu\neq\mu}}\hat{L}_{0,\hat{\nu}} ,\\
A_{2,7}&=A_{1,3}[P(U_{\hat{i}})-P(U_0,U_{\hat{i}})-P(U_{\hat{i}},U_{\hat{i}+\hat{\mu}})\\
&\quad +P(U_0,U_{\hat{i}},U_{\hat{i}+\hat{\mu}})]B\\
&=\hat{L}_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}_{\hat{i},\hat{i}+\hat{\mu}}\sum_{\substack{\nu\neq -1\\ \nu\neq\mu}}\hat{L}_{\hat{i},\hat{i}+\hat{\nu}} ,\\
A_{2,9}&=A_{1,2}[P(U_{\hat{\mu}})-P(U_0,U_{\hat{\mu}})-P(U_{\hat{i}},U_{\hat{\mu}})\\
&\quad +P(U_0,U_{\hat{i}},U_{\hat{\mu}})]B+A_{1,3}[P(U_0)-P(U_0,U_{\hat{i}})\\
&\quad -P(U_0,U_{\hat{i}+\hat{\mu}})+P(U_0,U_{\hat{i}},U_{\hat{i}+\hat{\mu}})]B\\
&=2\hat{L}_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}_{0,\hat{\mu}}\sum_{\nu\neq -1}\hat{L}_{\hat{i},\hat{i}+\hat{\nu}} ,\\
A_{2,9}&=2A_{1,2}[P(U_{\hat{\mu}})-P(U_0,U_{\hat{\mu}})-P(U_{\hat{i}},U_{\hat{\mu}})\\
&\quad +P(U_0,U_{\hat{i}},U_{\hat{\mu}})]B\\
&=2\hat{L}_{0,\hat{i}}\sum_{\mu\neq 1}\hat{L}_{0,\hat{\mu}}\sum_{\nu\neq -\mu}\hat{L}_{\hat{\mu},\hat{\mu}+\hat{\nu}} ,\\
A_{2,10}&=2A_{1,3}[P(U_{\hat{i}+\hat{\mu}})-P(U_0,U_{\hat{i}+\hat{\mu}})-P(U_{\hat{i}},U_{\hat{i}+\hat{\mu}})\\
&\quad +P(U_0,U_{\hat{i}},U_{\hat{i}+\hat{\mu}})]B\\
&=2\hat{L}_{0,\hat{i}}\sum_{\mu\neq -1}\hat{L}_{\hat{i},\hat{i}+\hat{\mu}}\sum_{\nu\neq -\mu}\hat{L}_{\hat{i}+\hat{\mu},\hat{i}+\hat{\mu}+\hat{\nu}} .
\end{aligned}$$

Without doubt, the same method can be applied to the construction of the connected higher-order correction diagram set. However, with the increase of the order k, the diagrams belong to the connected $(k-1)$th-order correc-

tion diagram set, and the sites contained in every diagram increases rapidly in number. Therefore, the amount of computation becomes very large.

In this example, the diagram $d(X)$ corresponding to the local observable $X(U^r)$ represents the same group theoretic factor as the basic diagram $d\ (L_{0,\hat{1}})$. So, a lot of diagrams in formulas (4.10) and (4.12) are actually equal in value and can be added together. Taking the third step of standard program, we finish without difficulty the classification and the enumeration of the obtained diagrams, and can give the result in Table I.

V. THE LATTICE GAUGE FIELD MODEL

To check and use our scheme, as the second example, the diagram theoretic factors occuring in the cumulant expansion of the pure lattice gauge field model are calculated analytically.

In the lattice gauge field theory, the gauge field variable $U_{n,\mu}$ is defined on each link $\mathbf{L}_{n,\mu}$ that is represented diagrammatically by the unit vector $d(\mathbf{L}_{n,\mu})$ from site s_n to site $s_{n+\hat{\mu}}$ called the directed link.

Following Wilson, we take the action of the SU(N) pure lattice gauge field as the form

$$S(U) = \frac{\beta}{2N} \sum_{\substack{n,\mu,\nu \\ \mu>\nu}} \mathrm{Tr}(U_{n,\mu} U_{n+\hat{\mu},\nu} U^{+}_{n+\hat{\nu},\mu} U^{+}_{n,\nu} + \mathrm{h.c.}). \tag{5.1}$$

Throwing away the factor $\beta/2N \times \frac{1}{8}$ and the group structure of field variables, we immediately write down the algebraic expression

$$B = \sum_{\substack{n,\mu,\nu \\ \mu>\nu}} \hat{P}(n,\mu,\nu), \tag{5.2}$$

of the basic diagram set

$$D(B) = \{d(P_{n,\mu,\nu})\}, \tag{5.3}$$

where

$$\hat{P}(n,\mu,\nu) = \hat{\mathbf{P}}(n,\mu,\nu) + \hat{\mathbf{P}}(n,\nu,\mu), \tag{5.4}$$

$$\hat{\mathbf{P}}(n,\mu,\nu) = \hat{U}_{n,\mu} \hat{U}_{n+\hat{\mu},\nu} \hat{U}_{n+\hat{\mu}+\hat{\nu},-\mu} \hat{U}_{n+\hat{\nu},-\nu}, \tag{5.5}$$

and only $\hat{\mu},\hat{\nu}$ in positive direction are taken in the summation $\bar{\Sigma}$. The quantity $\hat{\mathbf{p}}(n,\mu,\nu)$ is represented illustratively by the diagram $d(\mathbf{p}_{n,\mu,\nu})$ made of four directed links $d(\mathbf{L}_{n,\mu})$, $d(\mathbf{L}_{n+\hat{\mu},\nu})$, $d(\mathbf{L}_{n+\hat{\mu}+\hat{\nu},-\mu})$, and $d(\mathbf{L}_{n+\hat{\nu},-\nu})$. The diagram $d(\mathbf{p}_{n,\mu,\nu})$ is the boundary of a oriented plaquette, hence called the oriented plaquette. The diagram $d(p_{n,\mu,\nu})$ of $p(n,\mu,\nu)$ is the union of two plaquettes having identical position and opposite orientation. For simplicity, $d(p_{n,\mu,\nu})$ is drawn as the diagram made of four links without direction $d(L_{n,n+\hat{\mu}})$, $d(L_{n+\hat{\mu},n+\hat{\mu}+\hat{\nu}})$, $d(L_{n+\hat{\nu},n+\hat{\mu}+\hat{\nu}})$, and $d(L_{n,n+\hat{\nu}})$ and called the plaquette.

In quite a few analytical methods in the Lagrangian formalism, the mathematical expectation $\langle X \rangle$ of any local observable $X(U^r)$ is calculated by means of various expansions and the external source technique. Use is made of the Legendre transformation from the single link integral

$$Z(J) = \int DU \exp(\mathrm{Tr}\, JU^{+} + \mathrm{h.c.}), \tag{5.6}$$

which can be evaluated exactly. Without any question, the convergence can be accelerated by a better choice of the external source parameter J. The value of J is determined by various methods, such as the self-consistent condition in the old-fashioned mean field theory, the system of the saddle point equations in the saddle point method, and minimizing the linear main part of the free energy in variational methods. These methods are often used, and have been argued by

TABLE I. The diagram theoretic factors occurring in the average link of the d-dimensional lattice chiral field model.

order	diagrams and their enumerations
0	1
1	1 ; $2r_1$
2	1 ; $6r_1$; $4r_1 r_2$; $6r_1^2$

$r_1 = 2d-1$, $r_2 = d-1$.

either mathematical or physical reasons. However, we hope to find a method by which the external source parameter can be determined more exactly. For this purpose, computing the functional derivative of the functional integral

$$I=\int DU\exp\left[\sum_{n,\mu}(\mathrm{Tr}\,J_{n,\mu}U^{+}_{n,\mu}+\mathrm{h.c.})-S\right], \tag{5.7}$$

with respect to the functional argument $U^{+}_{n,\mu}$, we have the following Ward identity:

$$J_{n,\mu}=\left\langle\frac{\delta S}{\delta U^{+}_{n,\mu}}\exp(-S)\right\rangle_J\Big/\langle\exp(-S)\rangle_J, \tag{5.8}$$

where the notation

$$\langle X\rangle_J=\left[\sqcap_{n,\mu}Z(J_{n,\mu})\right]^{-1}\int DU\,X(U')\times\exp\left[\sum_{n,\mu}\mathrm{Tr}(J_{n,\mu}U^{+}_{n,\mu}+\mathrm{h.c.})\right], \tag{5.9}$$

represents the expectation value of the local observable $X(U')$ computed with the Boltzmann weight $\exp[\Sigma_{n,\mu}\mathrm{Tr}(J_{n,\mu}U^{+}_{n,\mu}+\mathrm{h.c.})]$.

In fact, the Eq. (5.8) is the generalization of the self-consistent condition in the mean field theory. If the algebraic expressions of both $\langle(\delta S/\delta U^{+}_{n,\mu})\exp(-S)\rangle_J$ and $\langle\exp(-S)\rangle_J$ were evaluated exactly, the Ward indentity (5.8) would be considered as a precise system of equations satisfied by the parameters $J_{n,\mu}$. Therefore, we adopt the successively approximate method of expanding the exponential function $\exp(-S)$ into the power series of the action S. Since the group theoretic factors occurring in expansions can be calculated by means of the external source technique, as soon as the diagram theoretic factors are derived correctly, we can work out the algebraic expressions of both $\langle(\delta S/\delta U^{+}_{n,\mu})\exp(-S)\rangle_J$ and $\langle\exp(-S)\rangle_J$, and obtain the system of equations determining the parameters $\{J_{n,\mu}\}$ in successive approximations. It is obvious that the zeroth order approximation of this method is equivalent to the saddle point method and the variational method in respect of determining the value of the parameter J.[9] There is no problem to apply the scheme of determining J by Ward identity to statistical mechanics, but it might conflict with Elitzur's theorem[10] when applied to the lattice gauge field theory. In spite of the difficulty, we can still take it as an example of our scheme for enumerating connected diagrams, in which $d(X)$ is different from the basic diagram in shape and property.

Now let us calculate the diagram theoretic factors occurring in the expansion of $\langle(\delta S/\delta U^{+}_{n,\mu})\exp(-S)\rangle_J$. Since the minus sign before the action S can be absorbed into the expansion coefficients, the basic diagram set of this expansion is identical with the one of the cumulant expansions. For Wilson's action, due to the unitarity of the gauge group, the local observable determining the parameter $J_{0,1}$ is given by the formula

$$X(U')=\frac{\delta S}{\delta U^{+}_{0,1}}=\frac{\beta}{zN}\sum_{\alpha\,(|\alpha|\neq 1)}(U_{0,\alpha}U_{\hat{\alpha},1}U^{+}_{\hat{1},\alpha})^{T}, \tag{5.10}$$

where T means the transpose of a matrix. We can immediately read out the algebraically expression

$$A(X)=\sum_{\alpha\,(|\alpha|\neq 1)}\Pi(0,\alpha,1), \tag{5.11}$$

of the diagram set corresponding to $X(U')$. Here

$$\Pi(0,\alpha,1)=\hat{\mathbf{L}}_{0,\alpha}\hat{\mathbf{L}}_{\hat{\alpha},1}\hat{\mathbf{L}}_{\hat{1}+\hat{\alpha},-\alpha} \tag{5.12}$$

is represented illustratively by the labeled diagram composed of three directed links $d(\mathbf{L}_{0,\alpha})$, $d(\mathbf{L}_{\hat{\alpha},1})$ and $d(\mathbf{L}_{\hat{1}+\hat{\alpha},-\alpha})$ and contributes the factor $(\beta/2N)(U_{0,\alpha}U_{\hat{\alpha},1}U^{+}_{\hat{1},\alpha})^{T}$ into the integrand of the configuration integral. Equation (5.11) expresses algebraically the connected zeroth-order approximation diagram set of $\langle X\rangle$.

In order to construct the connected diagram set of higher order correction, we introduce the projection operator $p(\mathbf{L}_{n,\mu})$ on the directed link $d(\mathbf{L}_{n,\mu})$ and the one $P(\mathbf{L}^{r})$ on given r links without direction $d(L^{r})$ according to rules (3.8) ad (3.11), respectively, and define the one $P(L_{n,n+\hat{\mu}})$ on the link without direction $d(L_{n,n+\hat{\mu}})$ as the form

$$P(L_{n,n+\hat{\mu}})=\tfrac{1}{2}[P(\mathbf{L}_{n,\mu})+P(\mathbf{L}_{n+\hat{\mu},-\mu})]. \tag{5.13}$$

In the computation and the simplification of the diagrams constructed successively, the following properties are useful:

(i)

$$\begin{aligned}\hat{p}(n,\mu,\nu)&=\hat{p}(n+\hat{\mu},\nu,-\mu)=\hat{p}(n+\hat{\mu}+\hat{\nu},-\mu,-\nu)\\&=\hat{p}(n+\hat{\nu},-\nu,\mu)=\hat{p}(n,\nu,\mu)=\hat{p}(n+\hat{\nu},\mu,-\nu)\\&=\hat{p}(n+\hat{\mu}+\hat{\nu},-\nu,-\mu)=\hat{p}(n+\hat{\mu},-\mu,\nu).\end{aligned}$$

(ii) $P(L_{n,n+\hat{\mu}})B=\overline{\sum_{\alpha\,(|\alpha|\neq|\mu|)}}\hat{p}(n,\mu,\alpha)$.

(iii) If there exist two sites s_i and s_j satisfying $s_i\in P(n,\mu,\nu)$, $S_j\in L_{m,m+\hat{\alpha}}$ and $d(s_i,s_j)>3$, then $P(L_{m,m+\hat{\alpha}})\hat{P}(n,\mu,\nu)=0$.

(iv) If there are two sites s_i and s_j satisfying s_i, $s_j\in L^{r}$ and $d(s_i,s_j)>3$, then $P(L^{r})B=0$.

By using the standard program given in Sec. III, we can construct, classify, and enumerate all of connected diagrams. The result obained is listed in Table II.

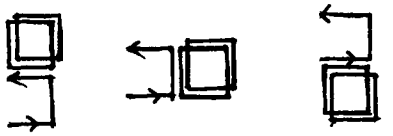
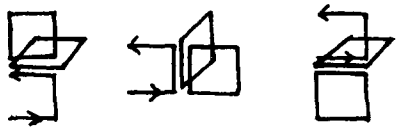
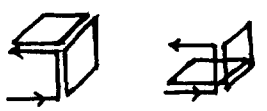
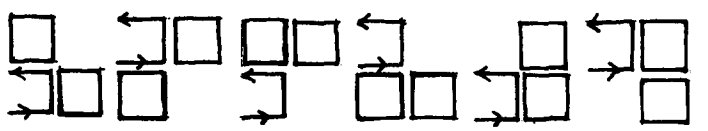
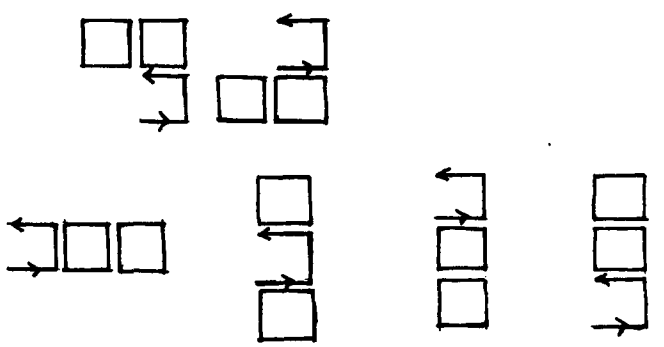
In general, these diagrams are inequivalent, because $\Pi(0,\alpha,1)$ is a directed diagram. Since the property of group theoretic factors is neglected in our discussion, for some specific gauge groups it might happen that topologically different diagrams correspond to the same group theoretic factor and can be incorporated.

VI. CONCLUDING REMARKS

Since the formulation used in our method is rigorous and construction, classification, and enumeration of diagrams are performed analytically, the results obtained with this method are rigorous and quite general. The discussion is restricted to the diagram theoretic factors, and the result can be applied to a variety of models with different group theoretic factors. The dependence on the number of space-time dimensions is automatically absorbed in results, hence the standard program is applicable to models with any space-time dimension without any trouble.

The basic diagrams describing the coupling among field variables can be composed of elements differing in number and location on the lattice, hence can have different shapes. Furthermore, the basic diagram set, as the illustrative repre-

TABLE II. The diagram theoretic factors occurring in the external source parameter of the pure lattice gauge field model with Wilson action.

order	diagrams and their enumerations
0	$2r_2$ 
1	$2r_2$ 
	$2r_2r_3$ 
2	$2r_2$ 
	$4r_2r_3$ 
	$2r_2r_3$ 
	$4r_2r_3r_4$ 
	$8r_2r_4$ 
	$4r_2(2r_3r_4+1)$ 
	$4r_2r_3^2$ 

$\gamma_2 = d-1$, $\gamma_3 = 2d-3$, $\gamma_4 = d-2$.

sentation of action $S(U)$, can be the union of several basic diagram sets with different shapes. So that, our standard program is suitable for a variety of physical models having different forms of action, such as the model with both nearest-neighbour and next-nearest interactions.

The "action $S(U)$" in Sec. III is only the quantity, in which the Boltzmann weight is expanded as the power series. It may be an effective action instead of the fundamental one. This enlarges the range of application of our standard program. For example, when the integral over fermion field variables and the trace of Dirac matrices have been performed, the lattice theory with fermion is equivalent to a pure lattice gauge field model with an effective action in which some terms are added.[11] We can calculate the diagram theoretic factors with our standard program, if the basic diagram set contain all of the diagrams corresponding to additional terms. What is more, $S(U)$ may be the quantity introduced for the convenience of mathematical treatment. So, the standard program can be applied to the variational-cumulant expansion, as long as the basic diagram set representing the trial action is added.

It appears that our scheme can be applied to cumulant expansions only. However, the basic formulation and the projection-sieve scheme, as the result of Theorems 3 and 4 is general and applicable to a variety of kinds of problems in which the classification and the enumeration of lattice diagrams are needed.

While our scheme has the advantage of rigor and universality, it has also the disadvantage of tedium of calculation. However, it can be used as the base of computer programs for enumeration of diagrams appearing in higher orders in the expansion. Furthermore, it may be used for proving general relations and theorems in diagram analysis in lattice field theories. Considering the property of the group theoretic factors at first and incorporating equivalent diagrams of the $(k-1)$th-degree correction before constructing the k th-degree correction diagram set, the amount of computation can be reduced. However, the improvement of the scheme is still needful, specially, when the diagrams corresponding to the local observable are disconnected.

[1]K. Wilson, Phys. Rev. D **10**, 2445 (1974).
[2]R. Friedberg and T. D. Lee, CU-TH-254.
[3]See, e.g., J. B. Kogut *et al.*, *Lattice Gauge Theory Using Parallel Processors*, edited by X.-Y. Li *et al.* (Gordon and Breach, New York, 1987).
[4]J.-M. Drouffe and J.-B. Zuber, Phys. Rep. C **102**, 1 (1983); and further references therein.
[5]F. Harary, *Graph Theory* (Addison-Wesley, London, 1969); F. Harary and E. M. Palmer, *Graphic Enumeration* (Academic, New York, 1973).
[6]E. Meeron, J. Chem. Phys. **27**, 1238 (1957).
[7]X.-H. He, D.-C. Xian, and Y.-X.Song, Phys. Lett. B **153**, 417 (1985); X.-T. Zheng, T.-L. Chen, and D.-C. Zian, Phys. Lett. B **154**, 166 (1985).
[8]C. Ko and W.-D. Wei, *Combinatorial Theory*, (Science and Technology Press, Beijing, 1984), Vol. 1.
[9]D.-F. Chen, Y.-X. Song, X.-T. Zheng, T.-L. Chen, X.-H. He, and D.-C. Xian, Phys. Energ. Fort. Phys. Nucl. **10**, 32 (1986).
[10]S. Elitzur, Phys. Rev. D **12**, 3978 (1975).
[11]H. Hamber, E. Marinari, G. Parisi, and C. Rebbi, Phys. Lett. B **124**, 99 (1983); C.-M. Wu and P.-Y. Zhao, Commun. Theor. Phys. **7**, 253 (1987).

U(2) extension of modified Bessel function with two indices

N. Tanimura
Institut für Theoretische Physik, Justus-Liebig-Universität Giessen, D-6300 Giessen, West Germany

O. Tanimura
Araki 1160, Fukumitsu, Toyama, Japan

(Received 1 December 1989; accepted for publication 14 March 1990)

A new set of functions that are given by the coefficients of the character expansion of the single-link action in the U(2) lattice gauge theory is studied. The set relies on the introduction of the character of the mixed term of the normal and adjoint U(2) variables. From the Schwinger–Dyson variational method the recursion relations among the functions are derived. From the relations the asymptotic behaviors of the function at the inverse coupling $\to 0$ and ∞ are derived. Through the combination of the recursion relations and the differentiation the linear differential equation of the fourth order is derived. The properties of the solution are discussed.

I. INTRODUCTION

It has been shown by Brower *et al.*[1] that the partition function of the single-link action in the U(N) lattice gauge theory[2] is systematically described by the combination of the modified Bessel functions. These functions correspond to the coefficients of the character of the unit dimension in the expansion of the single-link action.[3] Inversely, it appears interesting to study systematically the coefficient functions of all the characters at a given unitary group.

In previous papers[4,5] we have established the properties of coefficient functions in SU(3) lattice gauge theory: the recursion relations among the functions, the expansion with respect to the inverse coupling, the asymptotic behaviors of the function, and the differential equation that the function obeys. This set of functions is regarded as the SU(3) extension of the modified Bessel function.

In this group we investigate systematically the coefficient functions of the U(2) group. We first represent the exponentiated single-link action in terms of the characters and give the general expansion formula with respect to the inverse coupling β for the coefficient function by direct group integration. We then derive the recursion relations from the Schwinger–Dyson variational method[6] and determine the asymptotic behaviors at $\beta\to 0$ and $\beta\to\infty$. By the combination of the recursion relation and differentiation we determine the differential equation of the fourth order and discuss the properties of the solutions.

II. COEFFICIENTS OF U(2) CHARACTER EXPANSION

We start with the character expansion of the exponentiated single-link action for the U(2) link variable U. Let $\chi_{\lambda\mu}(U)$ be the usual character of the U(2) group, where λ and μ denote the numbers of the first and second rows of the Young tableau, respectively. The Young tableau is described in terms of $\mathrm{tr}(U)$ in Ref. 3. We further introduce the character for the mixed term of U and $U^\dagger$, which is defined by a negative value of μ such as $\chi_{\lambda-|\mu|}(U)$ (the general expression is given in Appendix A). Then we may write the character expansion as follows:

$$\begin{aligned}\exp[\beta\,\mathrm{tr}(U+U^\dagger)] &= C_{00}(\beta)+\sum_{\lambda=1}^{\infty}\sum_{\mu=0}^{\lambda} d_{\lambda\mu}C_{\lambda\mu}(\beta)\\ &\quad\times[\chi_{\lambda\mu}(U)+\phi^*_{\lambda\mu}(U)]\\ &\quad+\sum_{\lambda=1}^{\infty}\Big\{d_{\lambda-\lambda}C_{\lambda-\lambda}(\beta)\chi_{\lambda-\lambda}(U)\\ &\quad+\sum_{\mu=1}^{\lambda-1} d_{\lambda-\mu}C_{\lambda-\mu}(\beta)\\ &\quad\times[\chi_{\lambda-\mu}(U)+\chi^*_{\lambda-\mu}(U)]\Big\},\end{aligned}\tag{2.1}$$

with

$$d_{\lambda\mu}=\lambda-\mu+1\quad(\lambda\geqslant\mu\geqslant-\lambda),\tag{2.2}$$

where the symbol $d_{\lambda\mu}$ denotes the dimension. Here $C_{\lambda\mu}(\beta)$ is the real polynomial of the real variable β. The symmetry property of $\chi_{\lambda-\mu}(U)=\chi_{\mu-\lambda}(U)$ (see Appendix A) assures the symmetry of $C_{\lambda-\mu}(\beta)=C_{\mu-\lambda}(\beta)$ $(\mu>0)$. For example, $\chi_{1-1}(U)$ is given by

$$\chi_{1-1}(U)=\mathrm{tr}(U)\mathrm{tr}(U^\dagger)-1.\tag{2.3}$$

The orthogonality relation reads as

$$\int dU\,\chi^*_{\lambda'\mu'}(U)\chi_{\lambda\mu}(U)=\delta_{\lambda\lambda'}\delta_{\mu\mu'},\tag{2.4}$$

where $\mu,\mu'<0$ or $\geqslant 0$. With the use of (2.4) the character expansion (2.1) is inverted to be

$$\begin{aligned}C_{\lambda\mu}(\beta)=\frac{1}{d_{\lambda\mu}}\int dU\,\chi^*_{\lambda\mu}(U)\\ \times\exp[\beta\,\mathrm{tr}(U+U^\dagger)]\quad(\lambda\geqslant\mu\geqslant-\lambda),\end{aligned}\tag{2.5}$$

where the integral is taken over the invariant measure for the group element U.

Let us take $(\lambda,\mu)=(0,0)$, $(1,0)$, $(1,1)$, and $(1,-1)$ for simple examples. Substituting the respective expressions into (2.5) and performing the group integration, we find

$$C_{00}(\beta)=Z_{0,0}(\beta),\tag{2.6}$$

$$C_{10}(\beta)=\beta\sum_{k=0}^{1}\binom{1}{k}\beta^{2k}Z_{1-k,k}(\beta),\tag{2.7}$$

 0022-2488/90/082025-08$03.00

$$C_{11}(\beta)=\beta^2\sum_{s=0}^{1}\frac{2}{(2-s)!}\binom{1}{s}\sum_{k=0}^{1}\binom{2-2s}{k}\times\beta^{2k}Z_{2-2s-k,k+s}(\beta), \tag{2.8}$$

$$C_{1-1}(\beta)=\frac{1}{2}C_{11}(\beta)+\frac{1}{2}\beta^2\sum_{k=0}^{2}\binom{2}{k}\times\beta^{2k}Z_{2-k,k}(\beta), \tag{2.9}$$

with

$$Z_{k,r}(\beta)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{2^n\beta^{2n+4m}}{(n+2m+k+2r+1)!(m+r)!n!m!}, \tag{2.10}$$

where $C_{00}(\beta)$ is the single-link partition function and is given by Eriksson *et al.*[7] The successive evaluations of the $C_{\lambda\mu}(\beta)$'s lead us to the following general formula for $\mu\geqslant 0$:

$$C_{\lambda\mu}(\beta)=\beta^{\lambda+\mu}\sum_{s=0}^{\mu}\frac{(\lambda+1)!}{(\lambda-s+1)!}\binom{\mu}{s}\sum_{k=0}^{\lambda+\mu-2s}\binom{\lambda+\mu-2s}{k}\times\beta^{2k}Z_{\lambda+\mu-2s-k,k+s}(\beta) \tag{2.11}$$

and for $\mu<0$ we find, with setting $\nu=-\mu$,

$$C_{\lambda-\nu}(\beta)=\frac{\lambda!\nu!}{(\lambda+\nu)!}\left[C_{\lambda+\nu,0}(\beta)+\sum_{r=1}^{\nu}\frac{(\lambda+\nu)!(\lambda+\nu-2r+1)}{r!(\lambda+\nu+1-r)!}C_{\lambda+\nu-r,r}(\beta)\right]. \tag{2.12}$$

We note that the β series begins with the power of $\lambda+|\mu|$ in $C_{\lambda\mu}(\beta)$ $(\lambda\geqslant\mu\geqslant-\lambda)$. In fact, (2.11) becomes, at $\beta\to 0$,

$$C_{\lambda\mu}(\beta)\sim\beta^{\lambda+\mu}\sum_{s=0}^{\mu}\frac{(\lambda+1)!}{(\lambda-s+1)!}\binom{\mu}{s}\frac{2}{(\lambda+\mu+1)!s!}\sim\beta^{\lambda+\mu}\frac{1}{(\lambda+1)!\mu!}, \tag{2.13}$$

where we have used the relevant mathematical formula with respect to the two-term coefficient. On the other hand, (2.12) at $\beta\to 0$ is shown to be

$$\begin{aligned}C_{\lambda-\nu}(\beta)&\sim\frac{\lambda!\nu!}{(\lambda+\nu)!}\beta^{\lambda+\nu}\left[\frac{1}{(\lambda+\nu+1)!}+\sum_{r=1}^{\nu}\frac{(\lambda+\nu)!(\lambda+\nu-2r+1)}{[(\lambda+\nu+1-r)!]^2(r!)^2}\right]\\&\sim\frac{\lambda!\nu!(\lambda+\nu)^2}{(\lambda+\nu)!(\lambda+\nu+1)!}\beta^{\lambda+\nu}\times\left[1+\frac{(\lambda+\nu+1)(\lambda+\nu-3)}{2^2}+\cdots\right]\\&\sim\beta^{\lambda+\mu}\frac{1}{(\lambda+\nu+1)\lambda!\mu!},\end{aligned} \tag{2.14}$$

where we have used (2.13) and bracketed the first two terms successively.

III. THE SCHWINGER–DYSON EQUATION

We derive the Schwinger–Dyson equation[6] for the function containing the multilink variables. Here we consider only the case of $\mu\geqslant 0$ since the operation between the characters is invariant under the transformation of $\mu\to-\mu$, as shown in Appendix B.

We first define the generating function of the single-link ensemble average by

$$\langle f(U)\rangle=\int dU f(U)\exp[\mathrm{tr}(UJ+J^\dagger U^\dagger)]|_{J=\beta\sigma^0}, \tag{3.1}$$

where $f(U)$ is a polynomial of the single-link variable U and $U^\dagger$. The symbol J stands for an element of the group GL(2,C). The symbol σ^0 is the 2×2 unit matrix. Then $C_{\lambda\mu}(\beta)$ in (2.5) is written as

$$C_{\lambda\mu}(\beta)=(1/d_{\lambda\mu})\langle\chi^*_{\lambda\mu}(U)\rangle. \tag{3.2}$$

We consider the following variation:

$$U\to U^\epsilon U=\exp(i\epsilon\sigma^\alpha)U, \tag{3.3}$$

where ϵ denotes the infinitesimal parameter, σ^r ($r=1$, 2, and 3) denotes the Pauli matrix, and $\alpha\in\{0,1,2,3\}$. Here σ^α has the following properties:

$$\sigma^\alpha_{nm}\sigma^\alpha_{kr}=2\delta_{nr}\delta_{mk},\quad \mathrm{tr}(\sigma^\alpha\sigma^\beta)=2\delta_{\alpha\beta}. \tag{3.4}$$

Under the variation (3.3) Eq. (3.1) is transformed to

$$\begin{aligned}\langle f(U)\rangle&=\int d(U^\epsilon U)f(U^\epsilon U)\times\exp[\mathrm{tr}(U^\epsilon UJ+J^\dagger U^\dagger U^{\epsilon\dagger})]|_{J=\beta\sigma^0}\\&=\int dU f((1+i\epsilon\sigma^\alpha)U)\times\exp[\mathrm{tr}\{(1+i\epsilon\sigma^\alpha)UJ+\text{c.c.}\}]|_{J=\beta\sigma^0},\end{aligned} \tag{3.5}$$

which leads to a variational equation. Explicitly we set

$$f(U)=\mathrm{tr}(\sigma^\alpha U^n)[\mathrm{tr}(U)]^k. \tag{3.6}$$

Then with the use of (3.4) we obtain

$$\begin{aligned}&2\langle\mathrm{tr}(U^n)U^k\rangle+\sum_{r=1}^{n-1}\langle\mathrm{tr}(U^r)\mathrm{tr}(U^{n-r})U^k\rangle\\&\quad+k\langle\mathrm{tr}(U^{n+1})u^{k-1}\rangle+\beta[\langle\mathrm{tr}(U^{n+1})u^k\rangle\\&\quad+\langle\mathrm{tr}(U^{n+1})u^k\rangle]=0,\end{aligned} \tag{3.7}$$

where for simplicity we have used

$$u=\mathrm{tr}(U). \tag{3.8}$$

Equation (3.7) is the Schwinger–Dyson equation for the multilink variable in the U(2) group.

IV. RECURSION RELATION

A. Derivation of the recursion relations

We express the Schwinger–Dyson equation (3.7) in $C_{\lambda\mu}(\beta)$. Here we consider only the case of $\mu\geqslant 0$ because of the symmetrization under the replacement of $\mu\to-\mu$ (Appendix B). Hence we make use of the explicit formula of the characters shown in Ref. 3. For simplicity we omit the argument β of $C_{\lambda\mu}(\beta)$ below.

We take a simple example of $n=1$ and $k=0$ in (3.7). Equation (3.7) is then written in terms of the characters in Ref. 3 as

$$2\langle\chi_{10}\rangle+\beta[\langle\chi_{20}\rangle-\langle\chi_{11}\rangle-2\langle\chi_{00}\rangle]=0. \tag{4.1}$$

By substituting (3.2) into (4.1), we obtain one of the relations among the $C_{\lambda\mu}$'s:

$$4C_{10}+\beta[3C_{20}-C_{11}-2C_{00}]=0. \tag{4.2}$$

In general (3.7) leads to some combination of simple relations. We regard (3.7) as one of the simultaneous equations for the $C_{\lambda\mu}$'s without the β factor. For example, we take two sets of parameters of $(n,k)=(1,1)$ and (2,0) in (3.7). Then we have two simultaneous equations for C_{20} and C_{11}: These are solved and written in terms of the $C_{\lambda\mu}$'s with the β factor as

$$C_{11}=-\beta[C_{21}-C_{10}], \tag{4.3}$$

$$9C_{20}=-\beta[4C_{30}-C_{21}-3C_{10}]. \tag{4.4}$$

Likewise, with successive variations of n and k in (3.7) we obtain a series of relations for the $C_{\lambda\mu}$'s. From these relations we can find regularities with respect to λ and μ: They are attributed to two recursion relations

$$[(\lambda-\mu+1)/\beta\,]C_{\lambda\mu}-C_{\lambda,\mu+1}+C_{\lambda+1,\mu}-C_{\lambda-1,\mu}+C_{\lambda,\mu-1}=0, \tag{4.5}$$

$$\lambda C_{\lambda,\mu+1}-(\mu-1)C_{\lambda+1,\mu}+\mu C_{\lambda-1,\mu}-(\lambda+1)C_{\lambda,\mu-1}=0. \tag{4.6}$$

where $\lambda\geqslant 1$ is assumed in both formulas. As shown in Appendix A, the character is extended to the character containing the mixed terms of U and $U^{\dagger}$ by making use of the negative value of μ. Hence $\lambda-1\geqslant\mu\geqslant-\lambda+1$ holds in both (4.5) and (4.6). For later use we consider two variants that are given by the linear combinations (4.5)$\times(\lambda+1)$ + (4.6) and (4.5)$\times\lambda$ + (4.6), respectively,

$$[(\lambda+1)(\lambda-\mu+1)/\beta\,]C_{\lambda\mu}-C_{\lambda,\mu+1}+(\lambda-\mu+2)C_{\lambda+1,\mu}-(\lambda-\mu+1)C_{\lambda-1,\mu}=0, \tag{4.7}$$

$$[\lambda(\lambda-\mu+1)/\beta\,]C_{\lambda\mu}+(\lambda-\mu+1)C_{\lambda+1,\mu}-(\lambda-\mu)C_{\lambda-1,\mu}-C_{\lambda,\mu-1}=0. \tag{4.8}$$

It is of key importance that both (4.5) and (4.6) with the replacement $\mu\to-\mu$ are in symmetry with respect to the exchange of λ and μ.

For $\mu<0$ we take an example of $(\lambda,\mu)=(1,-1)$. By using the symmetry of $C_{\lambda,-|\mu|}=C_{|\mu|,-\lambda}$, (4.5) is written as

$$3C_{1-1}+\beta[2C_{2-1}-2C_{10}]=0. \tag{4.9}$$

B. Behavior of $C_{\lambda\mu}(\beta)$ at $\beta\to 0$

We check the behavior of $C_{\lambda\mu}(\beta)$ at $\beta\to 0$ from the recursion relations. Let us first consider the case of $\mu\geqslant 0$. We start with the knowledge of $C_{00}\sim 1$ and the assumption in (4.7):

$$-C_{\lambda,\mu+1}+(\lambda-\mu+2)C_{\lambda+1,\mu}=O(\beta^{\lambda+\mu+1}). \tag{4.10}$$

With this assumption and from the successive variation of λ and μ we find that $C_{\lambda\mu}(\beta)\sim O(\beta^{\lambda+\mu})$. Equation (4.7) then appears at $\beta\to 0$:

$$C_{\lambda\mu}(\beta)\sim[1/(\lambda+1)]\beta C_{\lambda-1,\beta}(\beta)\sim\{1/[(\lambda+1)\lambda\cdots(\mu+2)]\}\beta^{\lambda-\mu}C_{\mu\mu}(\beta). \tag{4.11}$$

Setting $\lambda=\mu$ in (4.8), we find, with the use of (4.10),

$$C_{\mu\mu}(\beta)\sim(1/\mu)\beta C_{\mu,\mu-1}(\beta)\sim[1/\mu(\mu+1)]\beta^{2}C_{\mu-1,\mu-1}(\beta)\sim[1/\mu!(\mu+1)!]\beta^{2\mu}C_{00}(\beta). \tag{4.12}$$

Combining (4.11) and (4.12), we obtain

$$C_{\lambda\mu}(\beta)\sim[1/\mu!(\lambda+1)!]\beta^{\lambda+\mu}, \tag{4.13}$$

which coincides with (2.12) and satisfies the assumption of (4.10).

For $\mu<0$ we assume in (4.8) with $\nu=-\mu$ that

$$(\lambda+\nu+1)C_{\lambda+1,-\nu}-C_{\lambda,-\nu-1}=O(\beta^{\lambda+\nu+1}). \tag{4.14}$$

Then (4.8) reads as

$$C_{\lambda,-\nu}(\beta)\sim[(\lambda+\nu)/\lambda(\lambda+\nu+1)]\beta C_{\lambda-1,-\nu}(\beta)\sim\nu!(2\nu+1)!/\lambda!(2\nu)!(\lambda+\nu+1)\times\beta^{\lambda-\nu}C_{\nu,-\nu}(\beta). \tag{4.15}$$

Setting $\lambda=\nu$ in (4.15) and using $C_{\nu-1,-\nu}=C_{\nu,-(\nu-1)}$, we find

$$C_{\nu,-\nu}(\beta)\sim[2\nu/\nu(2\nu+1)]\beta C_{\nu,-(\nu-1)}(\beta)\sim[(2\nu)!/\nu!\nu!(2\nu+1)!]\beta^{2\nu}C_{00}(\beta). \tag{4.16}$$

From (4.15) and (4.16) we find

$$C_{\lambda,-\nu}(\beta)\sim[1/(\lambda+\nu+1)\lambda!\nu!]\beta^{\lambda+\nu}, \tag{4.17}$$

which coincides with (2.13) and again satisfies the assumption of (4.10). The results of (4.13) and (4.17) give a confirmation of the recursion relations (4.5) and (4.6).

C. Asymptotic behavior at $\beta\to\infty$

Here we show that recursion relations determine the asymptotic behavior of $C_{\lambda\mu}(\beta)$, starting with the knowledge of $C_{00}(\beta)$ at $\beta\to\infty$. Following Brower *et al.*,[1] the single-link partition function $C_{00}(\beta)$ is given by

$$C_{00}(\beta)=\left.\frac{\beta' I_0(2\beta)I_1(2\beta')-\beta I_1(2\beta)I_0(2\beta')}{\beta'^{2}-\beta^{2}}\right|_{\beta'=\beta}, \tag{4.18}$$

where β^{2} and β'^{2} correspond to the eigenvalues of $JJ^{\dagger}$ in Ref. 1. One can check that (4.18) has the same β series as given by (2.6) with (2.10). Here $I_k(z)$ denotes the modified Bessel function, which has the asymptotic behavior at $z\to\infty$:

$$I_n(z)\sim(e^{z}/\sqrt{2\pi z})[1-(4n^{2}-1)/8z]. \tag{4.19}$$

The substitution of (4.19) into (4.18) generates, at $\beta\to\infty$,

$$C_{00}(\beta)\sim(e^{4\beta}/8\pi\beta^{2})[1+1/8\beta\,]. \tag{4.20}$$

We further need the asymptotic behaviors of $C_{10}(\beta)$ and $C_{11}(\beta)$ at the beginning of the recursion relations, which are given by, respectively,

$$C_{10}(\beta)=\frac{1}{4}\frac{d}{d\beta}C_{00}(\beta)\sim\frac{e^{4\beta}}{8\pi\beta^2}\left[1-\frac{3}{8\beta}\right], \quad (4.21)$$

$$C_{11}(\beta)=\frac{1}{2}\frac{d}{d\beta}C_{10}(\beta)+\frac{3}{\beta}C_{10}(\beta)-C_{00}(\beta)$$
$$\sim\frac{e^{4\beta}}{8\pi\beta^2}\left[1-\frac{3}{8\beta}\right]. \quad (4.22)$$

Here we have used the relation of differentiation shown in (5.3) below. Equation (4.21) is given directly from (5.3) with $\lambda=\mu=0$ and (4.22) is given from the combination of (4.2) and (5.3) with $(\lambda,\mu)=(1.0)$. Here we define the normalized value for the simplicity of the derivation:

$$A_{\lambda\mu}(\beta)=C_{\lambda\mu}(\beta)/C_{00}(\beta). \quad (4.23)$$

From (4.20)–(4.22) we find $A_{10}(\beta)$ and $A_{11}(\beta)$ at $\beta\to\infty$:

$$A_{10}(\beta)\sim A_{11}(\beta)\sim 1-1/2\beta. \quad (4.24)$$

From the recursion relations (4.5) and (4.6) we can evaluate any $A_{\lambda\mu}(\beta)$ at $\beta\to\infty$ with the use of (4.24). For example, (4.3) generates

$$A_{21}(\beta)=A_{10}(\beta)-(1/\beta)A_{11}(\beta)\sim 1-3/2\beta. \quad (4.25)$$

Through the successive variations of λ and μ in the recursion relations we easily find the general formula of the asymptotic behavior of $A_{\lambda\mu}(\beta)$:

$$A_{\lambda\mu}(\beta)\sim 1-(1/4\beta)[\lambda(\lambda+1)+\mu(\mu-1)]. \quad (4.26)$$

We prove formula (4.26) for $\mu\geqslant 0$ by the induction method in the following. We first assume that (4.26) holds for $\lambda+\mu\leqslant n$. Let $\lambda+\mu$ be equal to n. When (4.26) is substituted into (4.8) and divided by $C_{00}(\beta)$ apart from $C_{\lambda+1,\mu}$, we then find

$$A_{\lambda+1,\mu}(\beta)\sim 1-(1/4\beta)[(\lambda+2)(\lambda+1)+\mu(\mu-1)]. \quad (4.27)$$

Since (4.27) holds for all λ and μ satisfying $\lambda+\mu=n$, the validities of (4.27) are brought down successively to those of the $A_{\lambda\mu}(\beta)$'s with $\lambda+\mu=2$ and 1, which are apparent and seen in (4.2)–(4.4). This establishes the proposition.

Formula (4.26) also holds for $\mu<0$, which can be proved in the same manner as shown above. Combining (4.26) with (4.20) we obtain the asymptotic behavior of $C_{\lambda\mu}(\beta)$:

$$C_{\lambda\mu}(\beta)\sim(e^{4\beta}/8\pi\beta^2)[1-(1/8\beta)\{2\lambda(\lambda+1)+2\mu(\mu-1)-1\}]. \quad (4.28)$$

For simplicity, here we have calculated only the β^{-1} term. Since we can find the complete β^{-1} expansion for $C_{00}(\beta)$ from (4.18), we can also calculate the β^{-1} expansion for the $C_{\lambda\mu}(\beta)$'s with the use of the recursion relations in the same way as shown above.

D. The U(*N*) differential equation by Brower *et al.*[1]

We interpret the general differential equation for the U(N) group derived by Brower *et al.*[1] in terms of our language. The U(N) differential equation reads as[1]

$$\left[\frac{1}{N}\sum_k x_k\frac{\partial}{\partial x_k}+\frac{1}{N^2}\sum_k x_k^2\frac{\partial^2}{\partial x_k^2}+\frac{1}{N^2}\sum_{k\neq s}\frac{x_k x_s}{x_k-x_s}\times\left(\frac{\partial}{\partial x_k}-\frac{\partial}{\partial x_s}\right)-\frac{1}{N^2}\sum_k x_k\right]\zeta(x_k)=0, \quad (4.29)$$

where x_k is the k th eigenvalue of $JJ^\dagger$ and $\zeta(x_k)$ is the partition function. First we investigate the equation for U(2). Since our expansion (2.1) is based on the assumption that $JJ^\dagger$ has the equal eigenvalues of β^2, we set $x_1=x_2=\beta^2$ after the differentiation in (4.29) with $N=2$. Since $\zeta(x_k)$ is the partition function with the different eigenvalues, it becomes equal to $C_{00}(\beta)$ in the case of the equal eigenvalues:

$$\zeta(x_1=x_2=\beta^2)=C_{00}(\beta)=Z_{0,0}(\beta), \quad (4.30)$$

where we have used (2.6) with (2.10). For the differentiation of $\zeta(x_k)$ we find

$$\frac{\partial}{\partial x_k}\zeta(x_1=x_2=\beta^2)=Z_{1,0}(\beta)+\beta^2 Z_{0,1}(\beta)=\frac{C_{10}(\beta)}{\beta}, \quad (4.31)$$

where we have used (2.7) and set $x_1=x_2=\beta^2$ after the differentiation. Likewise, we see the following correspondences:

$$\frac{\partial^2}{\partial x_k^2}\zeta(x_1=x_2=\beta^2)=\frac{C_{20}(\beta)}{\beta^2}, \quad (4.32)$$

$$\frac{x_1x_2}{x_1-x_2}\left(\frac{\partial}{\partial x_1}-\frac{\partial}{\partial x_2}\right)\zeta(x_1=x_2=\beta^2)=\beta^2[C_{20}(\beta)-C_{11}(\beta)], \quad (4.33)$$

where we have used (2.11). Substituting (4.30)–(4.33) into (4.29), we obtain

$$4C_{10}+\beta[3C_{20}-C_{11}-2C_{00}]=0,$$

which is (4.2). That is, (4.29) with the equal eigenvalues for U(2) is reduced to the first equation of the recursion relation in U(2).

We next calculate the first equation of the recursion relation for the U(N) group, which should correspond to (4.29) with the equal eigenvalues. Let us write the Schwinger–Dyson equation for $f(U)=\mathrm{tr}(\lambda^\alpha U)$, where λ^α is the U(N) Gell-Mann matrix and satisfies the same equations as (3.4). Then the variation (3.5) with $\sigma^\alpha\to\lambda^\alpha$ leads to the following:

$$N\langle\mathrm{tr}(U)\rangle+\beta[\langle\mathrm{tr}(U^2)\rangle-N\langle 1\rangle]=0. \quad (4.34)$$

Here we introduce the character $\chi_{n_1n_2\cdots n_N}$ with the dimension $d_{n_1n_2\cdots n_N}$ in U(N), where n_r denotes the number of boxes in the rth row of the Young tableau. The coefficient function $C_{n_1n_2\cdots n_N}$ is given by

$$C_{n_1n_2\cdots n_N}=(d_{n_1n_2\cdots n_N})^{-1}\langle\chi^*_{n_1n_2\cdots n_N}\rangle. \quad (4.35)$$

Substituting the general expression for the character given in Ref. 3 into (4.35), we obtain the first equation of the recursion relation in U(N):

$$2NC_{10\cdots0}(\beta)+\beta[(N+1)C_{20\cdots0}(\beta)-(N-1)C_{110\cdots0}(\beta)-2C_{0\cdots0}(\beta)]=0. \quad (4.36)$$

Relation (4.36) is equivalent to (4.29) with the equal eigenvalues $x_k=\beta^2$ for all k.

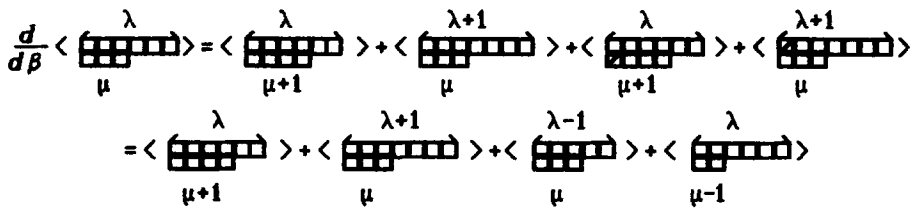
V. DIFFERENTIATION

We examine the relation containing the differentiation of $C_{\lambda\mu}(\beta)$ with respect to β. When operating $d/d\beta$ on (3.1), we find

$$\frac{d}{d\beta}\langle f(U)\rangle = \int dU f(U)[\mathrm{tr}(U)+\mathrm{tr}(U^\dagger)] \times \exp[\mathrm{tr}(UJ+J^\dagger U^\dagger)]|_{J=\beta\sigma''}. \quad (5.1)$$

This means that the differentiation of the average of the multilink variable is equal to the multiplication of $\mathrm{tr}(U)+\mathrm{tr}(U^\dagger)$ on the multilink variable. Thus when one sets $f(U)$ explicitly, e.g., to be (3.6), and performs the integral in (5.1), one can obtain the differentiation of $C_{\lambda\mu}(\beta)$ in the same manner as in Sec. IV A.

However, we can calculate the differentiation more directly by making use of the Young tableau. As shown in Appendix A, the multiplications of $\mathrm{tr}(U)$ and $\mathrm{tr}(U^\dagger)$ correspond to the additions of □ and its adjoint ⊡ to the original Young tableau, respectively. From the rule defined in Appendix A we illustrate $d/d\beta\,\langle\chi_{\lambda\mu}(U)\rangle$ schematically as



(5.2)

which corresponds to

$$\frac{d}{d\beta}\langle\chi_{\lambda\mu}\rangle = \langle\chi_{\lambda,\mu+1}\rangle + \langle\chi_{\lambda+1,\mu}\rangle + \langle\chi_{\lambda-1,\mu}\rangle + \langle\chi_{\lambda,\mu-1}\rangle. \quad (5.3)$$

With the use of (3.2) we find

$$(\lambda-\mu+1)\frac{d}{d\beta}C_{\lambda\mu} = (\lambda-\mu)[C_{\lambda,\mu+1}+C_{\lambda-1,\mu}] + (\lambda-\mu+2)[C_{\lambda+1,\mu}+C_{\lambda,\mu-1}]. \quad (5.4)$$

Equation (5.4) holds also for $\mu<0$ since the multiplication between the Young tableaux is invariant under the change of $\mu\to-\mu$, as shown in Appendix B. Therefore, (5.4) holds in the region of $\lambda\geqslant\mu\geqslant-\lambda$.

VI. DIFFERENTIAL EQUATION

A. Derivation

Let (λ,μ) be the two-dimensional index space. In the differentiation (5.4) and the symmetrical recursion relation (4.5) the point (λ,μ) is associated with the adjacent four points $(\lambda+1,\mu)$, $(\lambda,\mu+1)$, $(\lambda-1,\mu)$, and $(\lambda,\mu-1)$, while in the recursion relation (4.6) the same four points are associated with one another without (λ,μ). This indicates that (λ,μ) is located at the center in the index space in Eqs. (5.4), (4.5), and (4.6): We call it the center.

We will derive the linear differential equation for $C_{\lambda\mu}(\beta)$ from Eqs. (5.4), (4.5), and (4.6). Therefore, we should eliminate all C_{nm}'s other than $C_{\lambda\mu}$. In order to match the number of the unknown C_{nm}'s with Eqs. (5.4), (4.5), and (4.6) we introduce (4.5), (4.6), and (5.4) for the following four centers: (λ,μ), $(\lambda-1,\mu)$, $(\lambda,\mu+1)$, and $(\lambda+1,\mu+1)$: This choice of centers is one possible way to do this. Here we assume that the centers (n,m) lie in the region of $n\geqslant m\geqslant 0$. Hence the total number of simultaneous equations is $3\times4=12$.

On the other hand, the resulting unknown C_{nm}'s are given by

$C_{n,\mu-1}$, $(\lambda-1\leqslant n\leqslant\lambda)$, $C_{n,\mu}$, $(\lambda-2\leqslant n\leqslant\lambda+1)$,
$C_{n,\mu+1}$, $(\lambda-1\leqslant n\leqslant\lambda+2)$, $C_{n,\mu+2}$, $(\lambda\leqslant n\leqslant\lambda+1)$.

It follows that the number of C_{nm}'s is equal to 12. Since these C_{nm}'s are coupled to one another in the simultaneous differential equations, one can in principle solve these equations. We eliminate 11 C_{nm}'s $(\neq C_{\lambda\mu})$ from the simultaneous equations in the conventional manner, which consists of some arithmetic calculations and differentiations with respect to β. After some tedious calculations we find the following differential equation of the fourth order:

$$\left\{\frac{d^4}{d\beta^4}+\frac{8}{\beta}\frac{d^3}{d\beta^3}+\left[\frac{-2X+13}{\beta^2}-16\right]\times\frac{d^2}{d\beta^2}-\left[\frac{6X+2d_{\lambda\mu}^2-1}{\beta^3}+\frac{64}{\beta}\right]\times\frac{d}{d\beta}+\left[\frac{2X(d_{\lambda\mu}^2-1)-(d_{\lambda\mu}^2-1)^2}{\beta^4}-\frac{32}{\beta^2}\right]\right\}\times\mathscr{C}_{\lambda\mu}(\beta)=0, \quad (6.1)$$

with

$$X=\lambda(\lambda+1)+\mu(\mu-1), \quad (6.2)$$

where we have replaced the expression of the function $C_{\lambda\mu}(\beta)$ by $\mathscr{C}_{\lambda\mu}(\beta)$ since the solution of this differential equation contains not only $C_{\lambda\mu}(\beta)$, but also the other unknown solutions. For simplicity we write (6.1) as

$$\sum_{n=0}^{4}\left(\sum_{s=0}^{1}\beta^{n+2s-4}f_{ns}(\lambda,\mu)\right)\frac{d^n}{d\beta^n}\mathscr{C}_{\lambda\mu}(\beta)=0. \quad (6.3)$$

Here we note that $f_{ns}(\lambda,-\mu)=f_{ns}(\mu,-\lambda)$ and $f_{n1}(\mu,\lambda)=0$ for $n=3$ and 4.

For the spectral value of λ or μ the number of coupled equations is reduced on account of the properties of $C_{\lambda,-\mu}=C_{\mu,-\lambda}$ $(\mu\geqslant0)$. Then the differential equation (6.3) is reduced to a lower order equation. For example, let us take $\lambda=\mu=0$. Then we find

$$\left\{\frac{d^3}{d\beta^3}+\frac{5}{\beta}\frac{d^2}{d\beta^2}+\left[\frac{3}{\beta^2}-16\right]\frac{d}{d\beta}-\frac{16}{\beta}\right\}\mathscr{C}_{00}(\beta)=0. \quad (6.4)$$

Equation (6.4) satisfies the general differential equation (6.1) by multiplying the following operator:

$$\left[\frac{d}{d\beta}+\frac{3}{\beta}\right][\text{lhs of (6.4)}]=[\text{lhs of (6.1)}]|_{\lambda=\mu=0}. \tag{6.5}$$

B. Expansion with respect to β

We solve the differential equation (6.1) by means of the expansion method with respect to β. We set

$$\mathscr{C}_{\lambda\mu}(\beta)=\sum_{p=0}^{\infty}a_p\beta^{p+\xi}. \tag{6.6}$$

Here ξ is the unknown to be determined from the indicial equation

$$\begin{aligned}F(\xi)&\equiv\sum_{n=0}^{4}f_{n0}(\lambda,\mu)\frac{\xi!}{(\xi-n)!}\\&=[\xi-(\lambda+\mu)][\xi-(\lambda-\mu)][\xi+\lambda+\mu]\\&\quad\times[\xi+\lambda-\mu+2]=0.\end{aligned} \tag{6.7}$$

We thus have four integer solutions: $\lambda+\mu$, $\lambda-\mu$, $-(\lambda+\mu)$, and $-(\lambda-\mu+2)$. We take only one solution $\xi=\lambda+\mu$ for $\mu\geqslant 0$ or $\xi=\lambda-\mu$ for $\mu<0$, which corresponds to the coefficient of character expansion $C_{\lambda\mu}(\beta)$. Then (6.6) with $\xi=\lambda+|\mu|$ shows that the β series of this solution starts with $\beta^{\lambda+|\mu|}$.

We next calculate the coefficient a_p's for both $\mu\geqslant 0$ and $\mu<0$. Substituting (6.6) into (6.1) and then setting the coefficient of each power of β to be zero, we then obtain the following iterative relation:

$$\begin{aligned}a_p&=-\frac{a_{2p-2}}{F(\lambda+\mu+2p)}\\&\quad\times\sum_{n=0}^{2}f_{n1}(\lambda,\mu)\frac{(\lambda+|\mu|+2p)!}{(\lambda+|\mu|+2p-n)!}.\end{aligned} \tag{6.8}$$

We can solve the iterative equation (6.8) as

$$\begin{aligned}a_p&=(-)^p\prod_{k=1}^{p}\left[\frac{1}{F(\lambda+\mu+2k)}\right.\\&\quad\left.\times\sum_{n=0}^{2}f_{n1}(\lambda,\mu)\frac{(\lambda+|\mu|+2k)!}{(\lambda+|\mu|+2k-n)!}\right]a_0.\end{aligned} \tag{6.9}$$

Here we note that a_0 is not determined from the differential equation, but is a multiplicative constant. When we set a_0 to be the coefficient of (4.13) or (4.17), this solution is equivalent to $C_{\lambda\mu}(\beta)$:

$$\begin{aligned}\mathscr{C}_{\lambda\mu}(\beta)&=\sum_{p=0}^{\infty}(-)^p\prod_{k=1}^{p}\left[\frac{1}{F(\lambda+\mu+2k)}\sum_{n=0}^{2}f_{n1}(\lambda,\mu)\right.\\&\quad\left.\times\frac{(\lambda+|\mu|+2k)!}{(\lambda+|\mu|+2k-n)!}\right]\beta^{\lambda+|\mu|}a_0,\end{aligned} \tag{6.10}$$

with

$$a_0=\begin{cases}\dfrac{1}{\mu!(\lambda+1)!}, & \mu\geqslant 0,\\[2ex] \dfrac{1}{(\lambda+\nu+1)\lambda!\nu!}, & \mu<0.\end{cases} \tag{6.11}$$

It can be numerically checked that the β series of $C_{\lambda\mu}(\beta)$ given by (2.12) is equivalent to (6.10). Here we did not treat solutions other than $C_{\lambda\mu}(\beta)$ since we focus on the general form of the differential equation.

C. Asymptotic behavior of the solution

The resulting differential equation enables us to study the asymptotic behavior of the solution. We have seen in Sec. IV C that $C_{\lambda\mu}(\beta)$ at $\beta\to\infty$ behaves like (4.28). This means that (4.28) satisfies (6.1) at $\beta\to\infty$ and can really be checked. Thus one can expand the solution in terms of $1/\beta$ as

$$\mathscr{C}_{\lambda\mu}(\beta)=\frac{e^{4\beta}}{\beta^2}\sum_{p=0}^{\infty}b_p\beta^{-p}, \tag{6.12}$$

where $b_0=\frac{1}{8}\pi$. Substituting (6.12) into (6.1) and then setting the coefficient of each inverse power of β to be zero, we find the following iterative equation:

$$b_p=-\frac{\displaystyle\sum_{k=q}^{p-1}b_k\left\{\sum_{n=0}^{4}\sum_{s=0}^{1}\binom{n}{n+2s-k+p-3}\left[\prod_{r=1}^{n+2s-k+p-3}(-k-2-r+1)\right]4^{-2s+k-p+3}f_{ns}(\lambda,\mu)\right\}}{\displaystyle\sum_{n=0}^{4}\sum_{s=0}^{1}\binom{n}{n+2s-3}\left[\prod_{r=1}^{n+2s-3}(-p-2-r+1)\right]4^{-2s+3}f_{ns}(\lambda,\mu)} \tag{6.13}$$

with

$$q=\begin{cases}0, & p\leqslant 3,\\ p-3, & p>3.\end{cases} \tag{6.14}$$

Specifically, for $p=1$ we have

$$b_1=-[(2X-1)/8]b_0=-[(2X-1)/64\pi], \tag{6.15}$$

which is equivalent to the result given by (4.28). This suggests the self-consistence of our treatment.

VII. SUMMARIES

We have studied the coefficient of the extended character expansion of the exponentiated single-link action in the U(2) group, in which we have introduced the character of the mixed term of U and $U^\dagger$. Finally, we have reached the compact set of mathematical functions with two group indices, i.e., a Bessel-type function associated with the U(2) group.

We have derived the recursion relations, the relation of the differentiation, the β-series expansion, the asymptotic behavior at $\beta \to 0$ and $\beta \to \infty$, and the differential equation of the fourth order.

The resulting differential equation is contrasted with that of the corresponding functions in the U(1) and SU(2) groups, i.e., the second-order differential equation of the modified Bessel function. The differentiation of $I_k(2\beta)$ [the coefficient function corresponding to the U(1) or SU(2) group] generates the two functions with the adjacent integer suffices $k-1$ and $k+1$, while in U(2) the differentiation leads to the connection of four functions with adjacent indices in the symmetrized equation (5.4). The same fact also holds for the point of the recursion relation. It may following that the fourth-order differential equation is derived in the U(2) group, while in U(1) or SU(2) the second-order differential equation is derived. In view of this we regard the function $C_{\lambda\mu}(\beta)$ as the U(2) extension of the modified Bessel function.

We have shown that the differential equation for the general U(N) partition function derived by Brower *et al.*[1] is attributed to the first equation of the recursion relation for the coefficient function of the U(N) group in the case of the equal eigenvalues.

APPENDIX A: CHARACTERS OF MIXED TERMS OF U AND U^+

Here we define $\chi_{\lambda,-\nu}(U)$ $(\nu>0)$ and represent it with the Young tableau. First we connect $\mathrm{tr}(U^\dagger)$ to a slashed tableau:

$$\mathrm{tr}(U_{\lambda\mu}) \to \square.$$

The property of $UU^\dagger = 1$ assures symmetry with respect to λ and ν for $\chi_{\lambda,-\nu}(U)$. Therefore, $\chi_{\lambda,-\nu}(U)$ is represented by means of the single-row tableau in the following diagram:

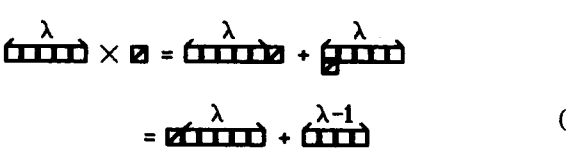
$\chi_{\lambda,-\nu}(U) \to$  (A1)

The rules of the multiplication between the characters containing the slashed tableaux are the following:

(i) The same multiplication rule holds as in the conventional rule,

(ii) the slashed tableau is movable freely in the same row and if a normal tableau and a slashed one lie in the same column, they are both dropped out.

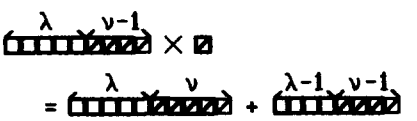
Let us take the simple example of the $\chi_{\lambda 0} \times \mathrm{tr}(U^\dagger)$, as in the following diagram:

 (A2)

It follows that

$$\chi_{\lambda 0} \times \mathrm{tr}(U^\dagger) = \chi_{\lambda,-1} + \chi_{\lambda-1,0}. \tag{A3}$$

Let us consider the multiplication of $\chi_{\lambda,-(\nu-1)}$ $(\nu>0)$ and $\mathrm{tr}(U^\dagger)$: It is represented in the Young tableau with the use of rule (ii) in the diagram

 (A4)

which corresponds to

$$\chi_{\lambda,-(\nu-1)} \times \mathrm{tr}(U^\dagger) = \chi_{\lambda,-\nu} + \chi_{\lambda-1,-(\nu-1)}. \tag{A5}$$

To calculate $\chi_{\lambda,-\nu}$, we make use of formula (A5) successively as follows:

$$\begin{aligned}\chi_{\lambda,-\nu} &= \chi_{\lambda,-(\nu-1)}\,\mathrm{tr}(U^\dagger) - \chi_{\lambda-1,-(\nu-1)} \\ &= \chi_{\lambda,-(\nu-2)}[\mathrm{tr}(U^\dagger)]^2 - 2\chi_{\lambda-1,-(\nu-2)} \\ &\quad \times \mathrm{tr}(U^\dagger) + \chi_{\lambda-2,-(\nu-2)} \\ &= \sum_{s=0}^{\nu} (-)^s \binom{\nu}{s} [\mathrm{tr}(U^\dagger)]^{\nu-s} \chi_{\lambda-s,0}. \end{aligned} \tag{A6}$$

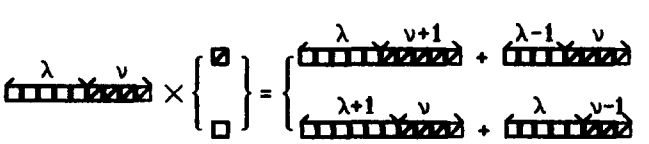
Thus one can evaluate the $\chi_{\lambda,-\mu}(U)$ from the known character[3] $\chi_{\lambda' 0}(U)$ $(\lambda' \leqslant \lambda)$. The dimension is given by setting $U=1$ in (A.6):

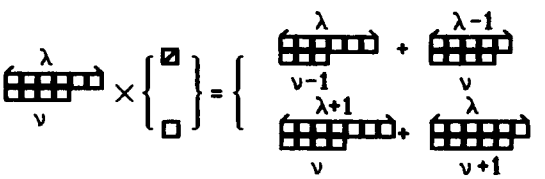
$$\begin{aligned}\chi_{\lambda,-\nu}(1) &= \sum_{s=0}^{\nu} (-)^s \binom{\nu}{s} 2^{\nu-s}(\lambda - s + 1) \\ &= \lambda + \nu + 1, \end{aligned} \tag{A7}$$

which satisfies (2.2). One can also check the orthonormal property of $\chi_{\lambda,-\nu}(U)$ by the mathematical induction method.

APPENDIX B: SYMMETRY PROPERTY OF CHARACTER

We show the symmetry property between $\chi_{\lambda\nu}(U)$ and $\chi_{\lambda,-\nu}(U)$ $(\nu>0)$. We first note that (2.2) also holds for $\chi_{\lambda,-\nu}$, as shown in (A7). We calculate the multiplications for $\chi_{\lambda,-\nu}(U)$ and $\chi_{\lambda\nu}(U)$ in the following diagrams:

 (B1)

 (B2)

which correspond, respectively, to

$$\chi_{\lambda,-\nu} \times \begin{Bmatrix} \chi^*_{10} \\ \chi_{10} \end{Bmatrix} = \begin{Bmatrix} \chi_{\lambda,-\nu-1} + \chi_{\lambda-1,-\nu} \\ \chi_{\lambda+1,-\nu} + \chi_{\lambda,-\nu+1} \end{Bmatrix}, \tag{B3}$$

$$\chi_{\lambda,\nu} \times \begin{Bmatrix} \chi^*_{10} \\ \chi_{10} \end{Bmatrix} = \begin{Bmatrix} \chi_{\lambda,\nu-1} + \chi_{\lambda-1,\nu} \\ \chi_{\lambda+1,\nu} + \chi_{\lambda,\nu+1} \end{Bmatrix}. \tag{B4}$$

Comparing (B3) with (B4), we find that the former is given by the replacement of $\nu \to -\nu$ in the latter. This establishes

the symmetry property of $\chi_{\lambda\nu}$ and $\chi_{\lambda,-\nu}$ in the multiplication of the characters.

[1]R. C. Brower, P. Rossi, and C.-I. Tang, Phys. Rev. D **23**, 992 (1981).
[2]K. Wilson, Phys. Rev. D **10**, 2445 (1974).
[3]I. Bars, J. Math. Phys. **21**, 2678 (1980).
[4]N. Tanimura and O. Tanimura, J. Math. Phys. **31**, 721 (1990).
[5]N. Tanimura and O. Tanimura, J. Math. Phys., submitted.
[6]D. Foester, Phys. Lett. B **87**, 87 (1978); T. Eguchi, Phys. Lett. B **87**, 91 (1979); R. C. Brower and M. Nauenberg, Nucl. Phys. A **180**, 221 (1981); S. S. Xue, T. C. Hsien, and C. M. Wu, Phys. Lett. B **175**, 341 (1987); N. Tanimura, Phys. Lett. B **221**, 55 (1989).
[7]K. E. Eriksson, N. Svartholm, and B. S. Skagerstam, J. Math. Phys. **22**, 2276 (1981).

On a Lie algebra of vector fields on a complex torus

Murray R. Bremner
Mathematical Sciences Research Institute, 1000 Centennial Drive, Berkeley, California 94720

(Received 24 October 1989; accepted for publication 21 March 1990)

This paper discusses a modification of the Krichever–Novikov Lie algebra of vector fields on the torus in which the two specified points where singularities can occur are points of order 1 and 2 rather than points of infinite order.

I. INTRODUCTION

Krichever and Novikov[1] initiated the study of Lie algebras of meromorphic vector fields, on compact Riemann surfaces of arbitrary genus, which are holomorphic except that poles are allowed at two specified points. In genus zero, the surface is the Riemann sphere, and one conventionally takes the two points to be 0 and ∞, thus obtaining the (centerless) Virasoro algebra, which has been studied in great detail. In genus 1, one represents a complex torus Σ as the quotient of the complex plane $\mathbf{C}$ by a lattice

$$\Lambda := \mathbf{Z}(2\omega_1) \oplus \mathbf{Z}(2\omega_2) \quad (\omega_1,\omega_2 \in \mathbf{C}, \quad \operatorname{Im} \omega_1/\omega_2 > 0),$$

chooses a point z_0 in the fundamental parallelogram, and sets $P_\pm := \pm z_0$. (We shall write z for both a point in $\mathbf{C}$ and its image in Σ.) Using the theory of elliptic functions, together with the Riemann–Roch theorem, one determines that the Lie algebra in this case has a basis $E_i := e_i(z)(d/dz)$, $i \in \mathbf{Z} + \frac{1}{2}$, where $e_i(z)$ is defined using the Weierstrass σ function:

$$e_i(z) := \sigma^{i-1/2}(z - z_0)\sigma(z + 2iz_0)/\sigma^{i+1/2}(z + z_0). \tag{1}$$

Thus $e_i(z)$ has a zero of order $i - \frac{1}{2}$ at P_+, a pole of order $i + \frac{1}{2}$ at P_-, and (as required by double periodicity) another zero, this time simple, at $-2iz_0$. The structure constants of this Lie algebra can then be expressed in terms of the Weierstrass ζ function.

One sees easily from (1) that the choice of z_0 is not completely free: For example, if z_0 is a torsion point, say $nz_0 = 0$, then $e_{n-1/2}(z) = e_{n+1/2}(z)$. There is no redundancy of this type in (1) if and only if z_0 is a point of infinite order.

However, one can take both P_+ and P_- to be torsion points, but the resulting Lie algebra will not have the form described by Krichever and Novikov. The simplest example occurs for $P_+ = 0, P_- = \omega_1 + \omega_2$; this is the case that will be considered in the remainder of this paper.

II. BASIS AND COMMUTATION RELATIONS

Write $\mathscr{K}$ for the Lie algebra of meromorphic vector fields on $\Sigma := \mathbf{C}/\Lambda$, which are holomorphic on $\Sigma \setminus \{P_\pm\}$. Let $\wp(z)$ be the Weierstrass $\wp$ function for the lattice Λ, and set $\omega := \omega_1 + \omega_2$, $\rho := \wp(\omega)$. Now $\wp(z) - \rho$ has a double pole at $z = 0$, a double zero at $z = \omega$ [since $\wp'(\omega) = 0$], and no other zeros or poles.[2] Thus $\mathscr{K}$ must include the vector fields

$$K_a := k_a(z)\frac{d}{dz}, \quad k_a(z) := (\wp(z) - \rho)^a, \quad a \in \mathbf{Z}. \tag{2}$$

Calculating commutators of these gives

$$[K_a, K_b] = (b - a)(\wp(z) - \rho)^{a+b-1}\wp'(z)\frac{d}{dz},$$

and thus $\mathscr{K}$ must also contain

$$J_a := j_a(z)\frac{d}{dz}, \quad j_a(z) := (\wp(z) - \rho)^{a-1}\wp'(z), \quad a \in \mathbf{Z}. \tag{3}$$

Proposition: The vector fields J_a and K_a for $a \in \mathbf{Z}$ form a basis of $\mathscr{K}$.

Proof: On Σ the vector field d/dz has no zeros or poles. Thus we have a vector space isomorphism from $\mathscr{F}$, the field of meromorphic functions on Σ, holomorphic on $\Sigma \setminus \{P_\pm\}$, to $\mathscr{K}$, given by $f(z) \mapsto f(z)(d/dz)$. So it suffices to show that the functions $j_a(z)$, $k_a(z)$ for $a \in \mathbf{Z}$, form a basis of $\mathscr{F}$. This is an immediate corollary of the Riemann–Roch theorem,[3] as follows. Consider the positive divisor $D := cP_- + dP_+$ $(c,d \in \mathbf{Z}_{>0})$ on Σ, and let $\mathscr{L}(D)$ be the finite-dimensional vector space of meromorphic functions $f(z)$ on Σ such that $\operatorname{div}(f) \geqslant -D$, together with the zero function. Then $\dim \mathscr{L}(D) = c + d$. One checks that $k_a(z)$ has a pole of order $2a$ at $z = 0$, a zero of order $2a$ at $z = \rho$, and no other zeros or poles, and that $j_a(z)$ has a pole of order $2a + 1$ at $z = 0$, a zero of order $2a - 1$ at $z = \rho$, and simple zeros at $z = \omega_1$ and $z = \omega_2$. Thus

$$\{k_{-[d/2]},\ldots,k_0 \equiv 1,\ldots,k_{[c/2]}\} \cup \{j_{-[(d-1)/2]},\ldots j_0 \cdots j_{[(c-1)/2]}\},$$

is a set of $c + d$ linearly independent functions $f(z)$ satisfying $\operatorname{div} f \geqslant -D$. Q.E.D.

The $\wp$ function satisfies the differential equation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3, \quad g_i = g_i(\omega_1,\omega_2), \tag{4}$$

which by differentiation gives

$$\wp''(z) = 6\wp(z)^2 - \tfrac{1}{2}g_2. \tag{5}$$

We also have the addition formula

$$\wp(z_1 + z_2) = \frac{1}{4}\left(\frac{\wp'(z_1) - \wp'_1(z_2)}{\wp(z_1) - \wp(z_2)}\right)^2 - \wp(z_1) - \wp(z_2), \quad z_1 \neq z_2. \tag{6}$$

Proposition: The commutation relations among the J_a and K_a are as follows:

 0022-2488/90/082033-02$03.00

$$[J_a,J_b] = (b-a)((12\rho^2 - g_2)J_{a+b-1} + 12\rho J_{a+b} + 4J_{a+b+1}), \tag{7}$$

$$[K_a,K_b] = (b-a)J_{a+b}, \tag{8}$$

$$[J_a,K_b] = (b-a+\tfrac{1}{2})(12\rho^2 - g_2)K_{a+b-1} + (b-a)12\rho K_{a+b} + 4(b-a-\tfrac{1}{2})K_{a+b+1}. \tag{9}$$

Proof: We only give the details for (9); (7) and (8) are simpler. Using (4) and (5) we have

$$\begin{aligned}
[J_a,K_b] &= (\wp(z)-\rho)^{a-1}\wp'(z)\frac{d}{dz}(\wp(z)-\rho)^b\frac{d}{dz} - (\wp(z)-\rho)^b\frac{d}{dz}(\wp(z)-\rho)^{a-1}\wp'(z)\frac{d}{dz} \\
&= (b-a+1)(\wp(z)-\rho)^{a+b-2}\wp'(z)^2\frac{d}{dz} - (\wp(z)-\rho)^{a+b-1}\wp''(z)\frac{d}{dz} \\
&= (b-a+1)(\wp(z)-\rho)^{a+b-2}(4(\wp(z)-\rho)^3 + 12\rho(\wp(z)-\rho)^2 + (12\rho^2-g_2)(\wp(z)-\rho))\frac{d}{dz} \\
&\quad - (\wp(z)-\rho)^{a+b-1}\left(6(\wp(z)-\rho)^2 + 12\rho(\wp(z)-\rho) + \left(6\rho^2 - \frac{1}{2}g_2\right)\right)\frac{d}{dz} \\
&= 4(b-a-\tfrac{1}{2})(\wp(z)-\rho)^{a+b+1}\frac{d}{dz} + 12\rho(b-a)(\wp(z)-\rho)^{a+b}\frac{d}{dz} \\
&\quad + (12\rho^2-g_2)(b-a+\tfrac{1}{2})(\wp(z)-\rho)^{a+b-1}\frac{d}{dz},
\end{aligned}$$

which gives (9). Q.E.D.

The automorphism of Σ given by $z \mapsto z+\omega$ interchanges P_+ and P_-, and induces an automorphism of $\mathscr{K}$, namely $\sigma: f(z)(d/dz) \mapsto f(z+\omega)(d/dz)$.

Proposition: The automorphism σ takes the explicit form $\sigma(J_a) = -\beta^{-2a}J_{-a}$, $\sigma(K_a) = \beta^{-2a}K_{-a}$, where β is a square root of $1/(3\rho^2 - \frac{1}{4}g_2)$.

Proof: Note that for $f(X) := 4X^3 - g_2X - g_3$, we have $f'(X) = 4(3X^2 - \frac{1}{4}g_2)$. By assumption $f(X)$ has distinct roots, and so $f(\rho) = 0$ implies $f'(\rho) \neq 0$. Thus $3\rho^2 - \frac{1}{4}g_2 \neq 0$ and β is well defined. Using (4) and (6) we obtain

$$\begin{aligned}
\wp(z+\omega) - \rho &= \frac{1}{4}\frac{\wp'(z)^2}{(\wp(z)-\rho)^2} - \wp(z) - 2\rho \\
&= \frac{1}{4}\frac{1}{(\wp(z)-\rho)^2}((12\rho^2 - g_2) \\
&\quad \times(\wp(z)-\rho) + (4\rho^3 - g_2\rho - g_3)) \\
&= \beta^{-2}\frac{1}{(\wp(z)-\rho)}.
\end{aligned}$$

Differentiating gives

$$\wp'(z+\omega) = -\beta^{-2}[\wp'(z)/(\wp(z)-\rho)^2].$$

Thus

$$\begin{aligned}
\sigma(J_a) &= (\wp(z+\omega)-\rho)^{a-1}\wp'(z+\omega)\frac{d}{dz} \\
&= -\beta^{-2a}(\wp(z)-\rho)^{-a-1}\wp'(z)\frac{d}{dz} \\
&= -\beta^{-2a}J_{-a}.
\end{aligned}$$

The calculation of $\sigma(K_a)$ is similar. Q.E.D.

Thus if we define $J'_a := \beta^a J_a$, $K'_a := \beta^a K_a$, then σ takes the more convenient form $\sigma(J'_a) = -J'_{-a}$, $\sigma(K'_a) = K'_{-a}$. If we further define $J''_a := -\frac{1}{4}\beta J'_a$, $K''_a := -\frac{1}{2}\beta K'_a$, we still have $\sigma(J''_a) = -J''_a$, $\sigma(K''_a) = K''_{-a}$, and the commutation relations take a very simple and symmetric form. Here, we omit the double-prime superscripts, and set $\gamma := 3\rho\beta$:

$$[J_a,J_b] = (a-b)(J_{a+b-1} + \gamma J_{a+b} + J_{a+b+1}), \tag{10a}$$

$$[K_a,K_b] = (a-b)\beta J_{a+b}, \tag{10b}$$

$$[J_a,K_b] = (a-b-\tfrac{1}{2})K_{a+b-1} + (a-b)\gamma K_{a+b} + (a-b+\tfrac{1}{2})K_{a+b+1}. \tag{10c}$$

ACKNOWLEDGMENTS

I thank the Mathematical Sciences Research Institute for its hospitality.

This work was supported by NSF Grant DMS-85-05550.

[1] I. M. Krichever and S. P. Novikov "Algebras of Virasoro type, Riemann surfaces, and structures of the theory of solitons," Funct. Anal. Appl. **21**, 126 (1987).

[2] D. Husemöller, *Elliptic Curves* (Springer-Verlag, New York, 1987).

[3] J. Silverman, *The Arithmetic of Elliptic Curves* (Springer-Verlag, New York, 1986).

A class of integral transforms associated with the Majorana representation of SL(2,C)

Debabrata Basu and T. K. Kar
Department of Physics, Indian Institute of Technology, Kharagpur 721 302, West Bengal, India

(Received 21 March 1989; accepted for publication 14 February 1990)

It is shown that the boson operators for the Majorana representation of SL(2,C) realized as hyperdifferential operators in the Bargmann Hilbert space of analytic functions yield, on exponentiation, a parametrized continuum of integral transforms. Each value of the group parameters yields an integral transform pair. The formula for the inversion of the transform is obtained simply by replacing the group element by its inverse.

I. INTRODUCTION

The two Majorana wave equations[1] are based on two special unitary representations of the homogeneous Lorentz group or its spinor group SL(2,C). In customary notation these two unitary irreducible representations (UIR's) are denoted by $D^{(0,1/2)}$ and $D^{(1/2,0)}$. The former is a member of the supplementary series of UIR's and the latter belongs to the principal series. The infinitesimal generators of these representations are contained as a subset in the Dirac realization[2] of the "remarkable" representation of the 3 + 2 de Sitter group, as well as in the Biedenharn and van Dam construction[3] of a special representation of the symplectic group Sp(2,2) in terms of boson creation and annihilation operators. Dirac operators were the natural consequence of a special simplifying feature of the representation of the group when a particular generator has only integral eigenvalues. On the other hand, the Biedenharn and van Dam construction[3] was motivated by the necessity of choosing a relative orientation for the Minkowski versus spin space in the application of the method of Galilean subdynamics developed in connection with their interpretation of the Dirac positive energy wave equation.[4,5] The boson realization of the de Sitter generators has also been introduced by van Dam *et al.*[6,7] in their attempt to incorporate some of the internal structure of particles in an elegant algebraic framework. This structure was shown to exhibit group theoretically defined discrete mass–spin relationships that are stable for any given value of a timelike four-momentum. These operators were also used in the formulation of a constrained relativistic Hamiltonian dynamics for composite systems.

The Majorana representation $D^{(0,1/2)} \oplus D^{(1/2,0)}$ plays a very special role in the construction of the infinite-dimensional UIR's of SL(2,C) by the boson operator technique.[8] It is indeed the basic representation generated by the Dirac–Biedenharn–van Dam–Mukunda[2-7] operators of SL(2,C), just as the metaplectic representation is the basic representation generated by the Holman–Biedenharn–Moshinsky–Quesne[8] operators of SL(2,R). The boson creation and annihilation operators are conveniently represented by operators defined in a certain Hilbert space of analytic functions $B(C)$. This Hilbert space has been introduced and studied in detail by Bargmann[9] for a finite number of boson operators and by Segal[10] for an infinite number of operators. The basic advantage of this procedure is that one obtains explicit expressions not only for the infinitesimal generators, but also for their exponential, the unitary operator of the representation itself. In a previous paper[11] we have shown that the use of this Hilbert space as the carrier space of the UIR's of SL(2,R) leads to a parametrized continuum of integral transforms mapping $B(C)$ onto itself. Each value of the group parameters yields an integral transform pair.

It is the object of the present paper to extend this analysis to the unitary representations of SL(2,C). The boson operators for the Majorana representations are constructed in Bargmann's Hilbert space $B(C_2)$, which consists of entire analytic functions $f(z_1, z_2)$ of two complex variables z_1 and z_2. As noted by Dirac,[4] in this realization the generators of the space rotation are first-order operators, while the Lorentz boosts are operators of second order. This simplicity is a special feature of the Bargmann realization of SL(2,C). The action of a finite element of the group on an element of $B(C_2)$ yields a parametrized continuum of integral transforms mapping $B(C_2)$ onto itself. The integral kernel for the inversion of this transform is obtained simply by replacing the group element by its inverse. Our method of exponentiation of the generators is based on an adaption of the Barut and Raczka analysis[12] of the "heat equation" on a Lie group and analytic vectors. Following Ref. 11 we first factorize the unitary operator of the representation into an appropriate Baker–Campbell–Hausdorff formula by using a theorem due to Wilcox.[13] The successive application of the operator factors on an element of $B(C_2)$ yields the integral transform pair.

II. FINITE ELEMENT OF THE GROUP AND THE ASSOCIATED INTEGRAL TRANSFORM

To make this paper self-contained we first describe some basic properties of the Bargmann Hilbert space $B(C_2)$, where C_2 is the two-dimensional complex Euclidean space. The elements of $B(C_2)$ are entire analytic functions $f(z_1, z_2)$ having a finite norm according to the scalar product

$$(f,g) = \int \overline{f(z_1,z_2)}\, g(z_1,z_2)\, d\mu(z_1)\, d\mu(z_2), \tag{2.1}$$

where $d\mu(z)$ is the Gaussian measure

$$d\mu(z) = (e^{-|z|^2}/\pi)\, d^2z, \quad d^2z = dx\, dy, \quad z = x + iy. \tag{2.2}$$

The scalar product satisfies

 0022-2488/90/082035-05$03.00

$$(z_i f,g) = \left(f,\frac{\partial}{\partial z_i} g\right), \quad i = 1,2. \tag{2.3}$$

A complete orthonormal set in $B(C_2)$ is given by the powers

$$u_{n_1 n_2}(z_1,z_2) = (z_1^{n_1} z_2^{n_2})/(n_1! n_2!)^{1/2}, \quad n_1,n_2 = 0,1,2,\ldots . \tag{2.4}$$

We now introduce the principal vectors $e_{z_1 z_2}$, which are bounded linear functionals in $B(C_2)$ satisfying

$$f(z_1,z_2) = (e_{z_1 z_2}, f). \tag{2.5}$$

The explicit form of the principal vector is given by

$$e_{z_1 z_2}(\xi_1,\xi_2) = e^{\bar{z}_1 \xi_1 + \bar{z}_2 \xi_2},$$

so that Eq. (2.5) reads as

$$f(z_1,z_2) = \int e^{z_1 \bar{\xi}_1 + z_2 \bar{\xi}_2} f(\xi_1,\xi_2) d\mu(\xi_1) d\mu(\xi_2). \tag{2.6}$$

The principal vectors will play a crucial role in our analysis.

The group SL(2,C) consists of the complex unimodular matrices

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det g = ad - bc = 1. \tag{2.7}$$

The group possesses six generators J_i, F_i $(i = 1,2,3)$, where J_i are the generators of space rotation and F_i those of pure Lorentz transformation. The Lie algebra of SL(2,C) is defined by the commutation relations

$$[J_i,J_j] = i\epsilon_{ijk} J_k, \quad [J_i,F_j] = i\epsilon_{ijk} F_k,$$
$$[F_i,F_j] = -i\epsilon_{ijk} J_k, \tag{2.8}$$

and has two Casimir operators, both of second degree, in the generators

$$C_1 = \mathbf{J}^2 - \mathbf{F}^2, \quad C_2 = \mathbf{J}\cdot\mathbf{F}. \tag{2.9}$$

In a UIR of SL(2,C) the six generators would be represented by Hermitian operators and C_1, C_2 would be represented by real numbers. The values of C_1 and C_2 can be written as

$$C_1 = j_0^2 + \sigma^2 - 1, \quad C_2 = j_0 \sigma. \tag{2.10}$$

The UIR's can be classified into two families.

(i) For the family of the principal series $D^{(j_0,\sigma)}$, j_0 is an integer or half-integer and σ is pure imaginary:

$$\sigma = i\rho, \quad -\infty < \rho < \infty.$$

(ii) For the family of the supplementary series $D^{(0,\sigma)}$, $j_0 = 0$ and σ is a real number lying in the interval $0 < \sigma < 1$.

To construct a unitary representation of the group in $B(C_2)$ we introduce the Dirac–Biedenharn–van Dam–Mukunda solution of the commutation relations.[14]

$$J_1 = \frac{1}{2}\left(z_1 \frac{\partial}{\partial z_2} + z_2 \frac{\partial}{\partial z_1}\right), \quad J_2 = -\frac{i}{2}\left(z_1 \frac{\partial}{\partial z_2} - z_2 \frac{\partial}{\partial z_1}\right),$$
$$J_3 = \frac{1}{2}\left(z_1 \frac{\partial}{\partial z_1} - z_2 \frac{\partial}{\partial z_2}\right),$$
$$F_1 = -\frac{1}{4}\left(z_1^2 - z_2^2 + \frac{\partial^2}{\partial z_1^2} - \frac{\partial^2}{\partial z_2^2}\right), \tag{2.11}$$
$$F_2 = \frac{i}{4}\left(z_1^2 + z_2^2 - \frac{\partial^2}{\partial z_1^2} - \frac{\partial^2}{\partial z_2^2}\right),$$
$$F_3 = \frac{1}{2}\left(z_1 z_2 + \frac{\partial^2}{\partial z_1 \partial z_2}\right).$$

Equation (2.3) ensures that the operators (2.11) are Hermitian under the scalar product (2.1).

Explicit calculation with the generators (2.11) yields

$$C_1 = j_0^2 + \sigma^2 - 1 = -\tfrac{3}{4}, \quad C_2 = j_0 \sigma = 0; \tag{2.12}$$

this is possible if

$$\sigma = 0, \quad j_0 = \tfrac{1}{2},$$

which corresponds to a representation of the principal series $D^{(1/2,0)}$ or if

$$j_0 = 0, \quad \sigma = \tfrac{1}{2},$$

which corresponds to a UIR belonging to the supplementary series $D^{(0,1/2)}$. Therefore, the representation generated by the above operators is the direct sum

$$D = D^{(1/2,0)} \oplus D^{(0,1/2)}.$$

A finite element of the group is obtained by exponentiating the operators (2.11). For this we introduce the canonical factorization[15]

$$g = u' \epsilon u, \tag{2.13}$$

where u' is the full SU(2) matrix

$$u' = \begin{pmatrix} \alpha' & \beta' \\ -\bar{\beta}' & \bar{\alpha}' \end{pmatrix} = \begin{pmatrix} e^{i(\phi' + \Psi')/2} \cos(\theta'/2) & e^{i(\phi' - \Psi')/2} \sin(\theta'/2) \\ -e^{-i(\phi' - \Psi')/2} \sin(\theta'/2) & e^{-i(\phi' + \Psi')/2} \cos(\theta'/2) \end{pmatrix}. \tag{2.14}$$

Here u is the truncated SU(2) matrix

$$u = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$$
$$= \begin{pmatrix} \cos(\theta/2) e^{i\phi/2} & \sin(\theta/2) e^{-i\theta/2} \\ -\sin(\theta/2) e^{i\phi/2} & \cos(\theta/2) e^{-i\phi/2} \end{pmatrix} \tag{2.15}$$

and ϵ is the boost

$$\epsilon = \begin{pmatrix} e^{-\tau/2} & 0 \\ 0 & e^{\tau/2} \end{pmatrix}. \tag{2.16}$$

The SL(2,C) matrix (2.7) can be related to the elements of u, ϵ, and u' by

$$a = e^{-\tau/2} \alpha' \alpha - e^{\tau/2} \beta' \bar{\beta}, \quad b = e^{-\tau/2} \alpha' \beta + e^{\tau/2} \beta' \bar{\alpha},$$
$$c = -e^{-\tau/2} \bar{\beta}' \alpha - e^{\tau/2} \bar{\alpha}' \bar{\beta}, \quad d = -e^{-\tau/2} \bar{\beta}' \beta + e^{\tau/2} \bar{\alpha}' \bar{\alpha}. \tag{2.17}$$

In what follows we shall need the inverse relation, namely, we have to express the elements of u, ϵ, and u' in terms of a, b, c, d, and their complex conjugates.

To calculate the elements of the truncated SU(2) matrix u in terms of a, b, c, d, etc. we introduce the positive Hermitian matrix

$$G = g^\dagger g = \begin{pmatrix} |a|^2 + |c|^2 & \bar{a}b + \bar{c}d \\ a\bar{b} + c\bar{d} & |b|^2 + |d|^2 \end{pmatrix} = \begin{pmatrix} A & B \\ \bar{B} & D \end{pmatrix}, \tag{2.18}$$

where

$$A = |a|^2 + |c|^2 \geqslant 1, \quad B = \bar{a}b + \bar{c}d, \quad D = |b|^2 + |d|^2 \geqslant 1. \tag{2.19}$$

The decomposition (2.13) yields

$$G = u^\dagger \epsilon u'^\dagger u' \epsilon u = u^\dagger \epsilon^2 u. \tag{2.20}$$

Thus

$$\epsilon^2 = \begin{pmatrix} e^{-\tau} & 0 \\ 0 & e^{\tau} \end{pmatrix} = uGu^\dagger. \tag{2.21}$$

Therefore, the truncated SU(2) matrix u diagonalizes the Hermitian matrix G and the eigenvectors of G constitute the columns of $u^\dagger$. The eigenvectors of G satisfying

$$G\xi = \lambda\xi, \tag{2.22}$$

where ξ is the column vector

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \tag{2.23}$$

can be easily obtained. The eigenvalues are given by

$$\begin{aligned} \lambda_1 &= e^{-\tau} = \sigma/2 - (\sigma^2/4 - 1)^{1/2} \leqslant 1, \\ \lambda_2 &= e^{\tau} = \sigma/2 + (\sigma^2/4 - 1)^{1/2} \geqslant 1, \end{aligned} \tag{2.24a}$$

where

$$\sigma = |a|^2 + |b|^2 + |c|^2 + |d|^2 \geqslant 2. \tag{2.24b}$$

The eigenvector corresponding to $\lambda = \lambda_1$ is given by

$$\xi = (\lambda_2 - \lambda_1)^{-1/2} \begin{pmatrix} (D - \lambda_1)^{1/2} \\ -(A - \lambda_1)^{1/2} \quad \bar{B}/|B| \end{pmatrix} \tag{2.25}$$

and that corresponding to $\lambda = \lambda_2$ is given by

$$\eta = (\lambda_2 - \lambda_1)^{-1/2} \begin{pmatrix} (A - \lambda_1)^{1/2} \\ (D - \lambda_1)^{1/2} \quad \bar{B}/|B| \end{pmatrix}. \tag{2.26}$$

From (2.25) and (2.26) we easily obtain (after a trivial adjustment of phase) the truncated SU(2) matrix

$$u = (\lambda_2 - \lambda_1)^{-1/2} \times \begin{pmatrix} [(D - \lambda_1)\bar{B}/|B|]^{1/2} & -[(A - \lambda_1)B/|B|]^{1/2} \\ [(A - \lambda_1)\bar{B}/|B|]^{1/2} & [(D - \lambda_1)B/|B|]^{1/2} \end{pmatrix}. \tag{2.27}$$

We therefore obtain

$$\begin{aligned} \alpha &= \left[\frac{(D - \lambda_1)\bar{B}}{(\lambda_2 - \lambda_1)|B|}\right]^{1/2}, \\ \beta &= -\left[\frac{(A - \lambda_1)B}{(\lambda_2 - \lambda_1)|B|}\right]^{1/2}, \end{aligned} \tag{2.28}$$

where A, B, D, λ_1, and λ_2 are given by Eqs. (2.19) and (2.24).

To determine the elements of u' we analogously introduce

$$G' = gg^\dagger = \begin{pmatrix} A' & B' \\ \bar{B}' & D' \end{pmatrix}, \tag{2.29}$$

where

$$A' = |a|^2 + |b|^2 \geqslant 1, \quad B' = a\,\bar{c} + b\bar{d}, \quad D' = |c|^2 + |d|^2 \geqslant 1. \tag{2.30}$$

If we now introduce the "reduced" SU(2) matrix u_r,

$$u' = u_r \begin{pmatrix} e^{i\Psi'/2} & 0 \\ 0 & e^{-i\Psi'/2} \end{pmatrix}, \tag{2.31}$$

where

$$u_r = \begin{pmatrix} \alpha_r & \beta_r \\ -\bar{\beta}_r & \bar{\alpha}_r \end{pmatrix}, \tag{2.32a}$$

$$\alpha_r = e^{i\phi'/2}\cos(\theta'/2), \quad \beta_r = e^{i\phi'/2}\sin(\theta'/2), \tag{2.32b}$$

we obtain

$$\epsilon^2 = u_r^\dagger G' u_r. \tag{2.33}$$

Thus the reduced SU(2) matrix diagonalizes the positive Hermitian matrix $G' = gg^\dagger$ and the eigenvectors of G' will therefore constitute the columns of u_r. Proceeding in the same way as before we obtain

$$\alpha_r = \left[\frac{(D' - \lambda_1)B'}{(\lambda_2 - \lambda_1)|B'|}\right]^{1/2}, \quad \beta_r = \left[\frac{(A' - \lambda_1)B'}{(\lambda_2 - \lambda_1)|B'|}\right]^{1/2}. \tag{2.34}$$

The remaining parameter Ψ' can now be calculated from Eq. (2.17). However, we shall not give the explicit form of $\exp(i\Psi'/2)$ in terms of a, b, c, d because only α_r, β_r will enter explicitly into the calculation.

A finite element of the group according to the canonical decomposition (2.13) is given by

$$T_g = T_{u'} T_\epsilon T_u \tag{2.35}$$

where

$$\begin{aligned} T_{u'} &= e^{i\phi' J_3} e^{i\theta' J_2} e^{i\Psi' J_3}, \\ T_\epsilon &= e^{i\tau F_3}, \quad T_u = e^{i\theta J_2} e^{i\phi J_3}. \end{aligned} \tag{2.36}$$

The action of T_u on an arbitrary element $f(z_1, z_2)$ is simple:

$$T_u f(z_1, z_2) = f(\alpha z_1 - \bar{\beta} z_2, \beta z_1 + \bar{\alpha} z_2).$$

To determine the action of the Lorentz boost T_ϵ we express the generator F_3 in a different form by introducing

$$z = (z_1 + z_2)/\sqrt{2}, \quad z' = (z_1 - z_2)/\sqrt{2}; \tag{2.37a}$$

thus

$$F_3 = \frac{1}{4}\left(z^2 - z'^2 + \frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial z'^2}\right). \tag{2.37b}$$

We now proceed to obtain a Baker–Campbell–Hausdorff formula for exponential operators of the form

$$e^{\alpha(\partial^2/\partial z^2) + \beta z^2} \tag{2.38}$$

by using a theorem due to Wilcox.[8]

Let P and Q be any two operators satisfying the commutation relation

$$[P,Q]=cI, \tag{2.39}$$

where c is a complex number. Thus P and Q may be the annihilation and creation operators, momentum and coordinate operators, etc. Then using the Wilcox theorem one can show[11]

$$e^{\alpha P^2+\beta Q^2}=J^{-1/2}e^{\beta Q^2}e^{(1/c)\ln(cG+1)QP}e^{\Lambda P^2}, \tag{2.40}$$

where

$$\alpha^{-1}A=\beta^{-1}\beta=(\lambda J)^{-1}\sinh\lambda,$$
$$G=c^{-1}(J^{-1}-1),\quad J=\cosh\lambda,\quad \lambda=(-4\alpha\beta)^{1/2}. \tag{2.41}$$

Thus

$$T_\epsilon=\left(\cosh\left(\frac{\tau}{2}\right)\right)^{-1}\exp\left[\left(\frac{i}{2}\right)\tanh\left(\frac{\tau}{2}\right)(z^2-z'^2)\right]$$
$$\times\exp\left[\ln\operatorname{sech}\left(\frac{\tau}{2}\right)\left(z\frac{\partial}{\partial z}+z'\frac{\partial}{\partial z'}\right)\right]$$
$$\times\exp\left[\frac{i}{2}\tanh\left(\frac{\tau}{2}\right)\left(\frac{\partial^2}{\partial z^2}-\frac{\partial^2}{\partial z'^2}\right)\right]. \tag{2.42}$$

To determine the action of the operator (2.42) on $f(\alpha z_1-\bar\beta z_2,\beta z_1+\bar\alpha z_2)$ we use the fundamental property of the principal vector as given by Eq. (2.6):

$$f(\alpha z_1-\bar\beta z_2,\beta z_1+\bar\alpha z_2)$$
$$=\int\exp\left[(\alpha z_1-\bar\beta z_2)\bar\xi_1+(\beta z_1+\bar\alpha z_2)\bar\xi_2\right]$$
$$\times f(\xi_1,\xi_2)d\mu(\xi_1)d\mu(\xi_2). \tag{2.43}$$

Expressing the rhs of Eq. (2.43) in terms of the variables z, z' as defined by Eq. (2.37) we have

$$T_w f=\int e^{zu+z'v}f(\xi_1,\xi_2)d\mu(\xi_1)d\mu(\xi_2), \tag{2.44a}$$

where

$$u=(1/\sqrt{2})[(\alpha\bar\xi_1+\beta\bar\xi_2)+(\bar\alpha\bar\xi_2-\bar\beta\bar\xi_1)],$$
$$v=(1/\sqrt{2}[(\alpha\bar\xi_1+\beta\bar\xi_2)-(\bar\alpha\bar\xi_2-\bar\beta\bar\xi_1)]. \tag{2.44b}$$

We first operate Eqs. (2.44) with the second-order operator appearing on the extreme right of Eq. (2.42). Thus

$$\exp\left[(i/2)\tanh(\tau/2)\left(\frac{\partial^2}{\partial z^2}-\frac{\partial^2}{\partial z'^2}\right)\right]T_w f$$
$$=\int\exp\left[\left(\frac{i}{2}\right)\tanh\left(\frac{\tau}{2}\right)(u^2-v^2)+zu+z'v\right]$$
$$\times f(\xi_1,\xi_2)d\mu(\xi_1)d\mu(\xi_2). \tag{2.45}$$

Applying the remaining factors in (2.42) successively and writing the resulting expression in terms of the original variables z_1, z_2 we have

$$T_\epsilon T_w f=\left(\operatorname{sech}\left(\frac{\tau}{2}\right)\right)\int K(z_1,z_2;\xi_1,\xi_2)$$
$$\times f(\xi_1,\xi_2)d\mu(\xi_1)d\mu(\xi_2), \tag{2.46a}$$

where

$$K(z_1,z_2;\xi_1,\xi_2)=\exp\Big(i\operatorname{sech}(\tau/2)\Big[z_1z_2\sinh\left(\frac{\tau}{2}\right)$$
$$+\sinh(\tau/2)(\bar\alpha\bar\xi_2-\bar\beta\bar\xi_1)(\alpha\bar\xi_1+\beta\bar\xi_2)$$
$$-iz_1(\alpha\bar\xi_1+\beta\bar\xi_2)-iz_2(\bar\alpha\bar\xi_2-\bar\beta\bar\xi_1)\Big]\Big). \tag{2.46b}$$

Finally, applying the operator $T_{u'}$; setting

$$u_g(z_1,z_2)=[T_g f](z_1,z_2); \tag{2.47}$$

and expressing the parameters in terms of a, b, c, d, we obtain

$$u_g(z_1,z_2)=\int A_g(z_1,z_2;\xi_1,\xi_2)f(\xi_1,\xi_2)d\mu(\xi_1)d\mu(\xi_2), \tag{2.48}$$

where

$$A_g(z_1,z_2;\xi_1,\xi_2)$$
$$=2(|a|^2+|b|^2+|c|^2+|d|^2+2)^{-1/2}\exp(i[(a\bar c+b\bar d)z_1^2-(\bar a c+\bar b d)z_2^2+(|c|^2+|d|^2-|a|^2-|b|^2)z_1z_2$$
$$+(a\bar b+c\bar d)\bar\xi_1^2-(\bar a b+\bar c d)\bar\xi_2^2+(|b|^2+|d|^2-|a|^2-|c|^2)\bar\xi_1\bar\xi_2-2iz_1\{(\bar d+a)\bar\xi_1+(b-\bar c)\bar\xi_2\}$$
$$-2iz_2\{(d+\bar a)\bar\xi_2-(\bar b-c)\bar\xi_1\}]/(|a|^2+|b|^2+|c|^2+|d|^2+2)). \tag{2.49}$$

The formula for the inversion of the transform follows at once from

$$f(\xi_1,\xi_2)=[T_{g^{-1}}\quad u_g](\xi_1,\xi_2). \tag{2.50}$$

Since

$$g^{-1}=\begin{pmatrix}d & -b\\ -c & a\end{pmatrix} \tag{2.51}$$

and

$$A_{g^{-1}}(\xi_1,\xi_2;z_1,z_2)=\overline{A_g(z_1,z_2;\xi_1,\xi_2)}, \tag{2.52}$$

the inversion formula is given by

$$f(\xi_1,\xi_2)=\int\overline{A_g(z_1,z_2;\xi_1,\xi_2)}\,u_g(z_1,z_2)d\mu(z_1)d\mu(z_2). \tag{2.53}$$

Equations (2.48) and (2.53) constitute an integral transform pair for each allowed value of the group parameters a, b, c, d.

For the SU(1,1) subgroup, namely, for $a=\alpha$, $d=\bar\alpha$, $b=\beta$, $c=\bar\beta$ we obtain the two-dimensional version of the SU(1,1) transform considered in Ref. II. A simple special case for SL(2,C) not contained in SU(1,1) is obtained by the choice

$$g=\begin{pmatrix}1 & -2i\\ 0 & 1\end{pmatrix}, \tag{2.54}$$

which yields the following transform pair:

$$\begin{aligned} u(z_1,z_2) &= \int K(z_1,z_2;\xi_1,\xi_2) f(\xi_1,\xi_2) \\ &\quad \times d\mu(\xi_1) d\mu(\xi_2), \\ f(\xi_1,\xi_2) &= \int \overline{K(z_1,z_2;\xi_1,\xi_2)} u(z_1,z_2) \\ &\quad \times d\mu(z_1) d\mu(z_2), \end{aligned} \tag{2.55}$$

where

$$\begin{aligned} &K(z_1,z_2;\xi_1,\xi_2) \\ &\quad = (1/\sqrt{2})\exp\{\tfrac{1}{4}[(z_1+z_2-i\bar{\xi}_1-i\bar{\xi}_2)^2 \\ &\qquad + 2(1+i)(\bar{\xi}_1\bar{\xi}_2 - z_1z_2 + z_1\bar{\xi}_1 + z_2\bar{\xi}_2)]\}. \end{aligned} \tag{2.56}$$

The Plancherel formula for the transform pair is obtained from the unitarity of the representation

$$(f_1,f_2) = (T_g f_1, T_g f_2) = (u_{1g}, u_{2g}), \tag{2.57}$$

which yields

$$\begin{aligned} &\int \overline{f_1(\xi_1,\xi_2)} f_2(\xi_1,\xi_2) d\mu(\xi_1) d\mu(\xi_2) \\ &\quad = \int \overline{u_{1g}(z_1,z_2)}\, u_{2g}(z_1,z_2) d\mu(z_1) d\mu(z_2). \end{aligned} \tag{2.58}$$

ACKNOWLEDGMENT

The authors wish to thank Professor G. P. Sastry for helpful discussions.

[1] E. Majorana, Nuovo Cimento **9**, 335 (1932); see, also, I. M. Gel'fand and A. M. Yaglom, Zh. Eksp. Teor. Fiz. **18**, 703 (1948).
[2] P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963).
[3] L. C. Biedenharn and H. van Dam, Phys. Rev. D **9**, 471 (1974).
[4] P. A. M. Dirac, Proc. R. Soc. London Ser. A **322**, 435 (1971); **328**, 1 (1972).
[5] L. C. Biedenharn, M. Y. Han, and H. van Dam, Phys. Rev. D **8**, 1735 (1973).
[6] H. van Dam and L. C. Biedenharn, Phys. Lett. **62B**, 190 (1976).
[7] N. Mukunda, H. van Dam, and L. C. Biedenharn, Phys. Rev. D **22**, 1938 (1980).
[8] See Refs. 2–7; see, also, W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **39**, 1 (1966); Y. Nambu, Phys. Rev. **160**, 1171 (1967); A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938 (1970); M. Moshinsky and C. Quesne, *ibid.* **12**, 1172, 1780 (1971).
[9] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961); Commun. Pure Appl. Math. **20**, 1 (1967); V. Bargmann, *in Analytic Methods in Mathematical Physics*, edited by P. Gilber and R. G. Newton (Gordon and Breach, New York, 1970).
[10] I. E. Segal, Illinois J. Math. **6**, 500 (1962).
[11] D. Basu, J. Math. Phys. **30**, 1 (1989).
[12] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN, Polish Scientific, Warsaw, 1980), p. 358.
[13] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).
[14] These operators are, in fact, a rotated form of the operators of Dirac, Biedenharn, van Dam, and Mukunda.
[15] M. A. Naimark, *Linear Representation of the Lorentz Group* (Pergamon, Oxford, 1964), p. 164.

Superconformal algebra of meromorphic vector fields with multipoles on a super Riemann sphere

Shikun Wang
CCAST (World Laboratory), P.O. Box 8730, Beijing 100080, People's Republic of China, and Institute of Applied Mathematics, Academia Sinica, P. O. Box 2734, Beijing 100080, People's Republic of China
Kaiwen Xu
Physics Department, Zhejiang University, Hangzhou, People's Republic of China

(Received 6 December 1989; accepted for publication 21 March 1990)

The superconformal algebras of meromorphic vector fields with multipoles and the relevant Abelian differential of the third kind on a super Riemann sphere were constructed. The algebra includes two Ramond sectors as subalgebras.

I. INTRODUCTION

The Virasoro algebra and super Virasoro algebra play very important roles in the string theory, conformal field theory, the superstring theory, the superconformal field theory, and other physical theories with conformal or superconformal symmetry.[1-5] It is well known that the Virasoro algebra is isomorphic to the central extension algebra of the algebra of meromorphic vector fields with two poles on the sphere S^2.

In recent papers,[6-10] the algebras of meromorphic vector fields with multipoles on Riemann surfaces have been constructed and the central extension version of the algebra has been investigated for the Riemann sphere S^2. In this paper, we will concentrate upon the corresponding super version of the algebra of the meromorphic vector field with multipoles on the Riemann sphere. First, we will construct the superconformal algebra of meromorphic vector fields with multipoles on a super Riemann sphere. Second, we will construct an Abelian differential of the third kind and use it to introduce the concept of Euclidean time and to give a picture of the interaction of superstrings.

II. THEORY

According to the point of view of Giddings and Nelson,[11] the super Riemann surface is a complex supermanifold with a superconformal structure. The class of superconformal structure is the class of supercomplex in which the transition functions are analytic and satisfy the following condition:

$$D_{\theta_\alpha} Z_\beta = \theta_\beta D_{\theta_\alpha} \theta_\beta, \tag{1}$$

where

$$Z_\beta = f_{\beta\alpha}(Z_\alpha \theta_\alpha), \quad \theta_\beta = \phi_{\beta\alpha}(Z_\alpha, \theta_\alpha),$$
$$D_{\theta_\alpha} = \frac{\partial}{\partial \theta_\alpha} + \theta_\alpha \frac{\partial}{\partial Z_\alpha}.$$

We choose N different points P_i $(i = 1,2,...,N)$ with local coordinate $w_i(P_i) = 0$, $\theta_i(P_i) = 0$ on a super Riemann sphere. Here,

$$z = w_1 = w_2 + z_2 = \cdots = w_{N-1} + z_{N-1} = -1/w_N,$$
$$\frac{1}{z}\theta = \frac{1}{w_1}\theta_1 = \frac{1}{w_2 + z_2}\theta_2 = \cdots = \frac{1}{w_{N-1} + z_{N-1}}\theta_{N-1} = \theta_N. \tag{2}$$

It is clear that these gluing conditions satisfy (1).

By use of $\lambda = -1$ and $\lambda = -\frac{1}{2}$ differentials with condition (2), we can introduce a basis H^i_n and G^i_r $(i = 1,2,...,N-1)$ of superfields with multipoles on a super Riemann sphere, which, in neighborhoods of the points P_i, will have the following form:

$$H^1_n = w_1^{1-n}\frac{\partial}{\partial w_1} + \frac{1}{2}(1-n)w_1^{-n}\theta_1\frac{\partial}{\partial \theta_1}$$
$$= (z - z_1)^{1-n}\frac{\partial}{\partial z} + \frac{1}{2}(1-n)(z-z_1)^{-n}\theta\frac{\partial}{\partial \theta},$$
$$\text{here } z_1 = 0, \quad n \in Z; \tag{3}$$

$$H^i_n = w_i^{1-n}\frac{\partial}{\partial w_i} + \frac{1}{2}(1-n)w_i^{-n}\theta_i\frac{\partial}{\partial \theta_i}$$
$$= (z - z_i)^{1-n}\frac{\partial}{\partial z} + \frac{1}{2}(1-n)(z-z_i)^{-n}\theta\frac{\partial}{\partial \theta},$$
$$\text{here } n \geqslant 2 \quad (i = 2,3,...,N-1); \tag{4}$$

$$G^1_r = w_i^{1-n}\left(\frac{\partial}{\partial \theta_1} - \theta_1\frac{\partial}{\partial w_1}\right)$$
$$= (z - z_1)^{1-n}\left(\frac{\partial}{\partial \theta} - \theta\frac{\partial}{\partial z}\right), \quad \text{here } n \in Z; \tag{5}$$

$$G^i_r = w_i^{1-n}\left(\frac{\partial}{\partial \theta_i} - \theta_i\frac{\partial}{\partial w_i}\right)$$
$$= (z - z_i)^{1-n}\left(\frac{\partial}{\partial \theta} - \theta\frac{\partial}{\partial z}\right),$$
$$\text{here } n \geqslant 2 \quad (i = 2,3,...,N-1). \tag{6}$$

With respect to the basis H^i_n, G^i_r, we can get the following algebra relations:

 0022-2488/90/082040-02$03.00

$$[H^i_n,H^i_m] = (n-m)H^i_{n+m}, \tag{7}$$

$$[H^i_n,H^j_m] = \sum_{\alpha=2}^{n+1}(2n-\alpha)X^{ij}_{nm}(\alpha)H^i_\alpha - \sum_{\alpha=2}^{m+1}(2m-\alpha)\,X^{ji}_{mn}(\alpha)H^j_\alpha, \quad i\neq j, \tag{8}$$

$$\{G^i_r,G^i_s\} = -2H^i_{r+s-1}, \tag{9}$$

$$\{G^i_r,G^j_s\} = \sum_{\alpha=2}^{r+1}(-2)Y^{ij}_{rs}(\alpha)H^i_{\alpha-1} + \sum_{\alpha=2}^{s+1}(-2)Y^{ji}_{sr}(\alpha)H^j_{\alpha-1}, \quad i\neq j, \tag{10}$$

$$[H^i_n,G^i_s] = (\tfrac{1}{2}n+\tfrac{1}{2}-s)G^i_{n+s}, \tag{11}$$

$$[H^i_n,G^j_s] = \sum_{\alpha=2}^{n+1}\left(\frac{3}{2}n+\frac{1}{2}-\alpha\right)X^{ij}_{ns}(\alpha)G^i_\alpha - \sum_{\alpha=2}^{s+1}\left(\frac{3}{2}s+\frac{1}{2}-\frac{\alpha}{2}\right)\times X^{ji}_{ns}(\alpha)G^j_\alpha, \quad i\neq j, \tag{12}$$

where $i,j = 1,2,...,N-1$,

$$X^{ij}_{nm}(\alpha) = \begin{cases}\binom{1-m}{n+1-\alpha}(z_i-z_j)^{\alpha-m-n}, & \text{when } n\geqslant 2,\\ (1-\delta_{2-\alpha,0})\binom{1-m}{\alpha-m-n}(z_i-z_j)^{\alpha-m-n}, & \\ \text{when } i=1 \text{ and } n\leqslant 1,\end{cases} \tag{13}$$

$$Y^{ij}_{rs}(\alpha) = \begin{cases}(1-\delta_{2-\alpha,0})\binom{1-s}{r+1-\alpha}(z_i-z_j)^{\alpha-r-s}, & \\ \text{when } r\geqslant 2,\\ \binom{1-s}{\alpha-m-n}(z_i-z_j)^{\alpha-r-s}, & \\ \text{when } i=1 \text{ and } r\leqslant 1.\end{cases} \tag{14}$$

$$\binom{1-m}{n+1-\alpha} = \begin{cases}\dfrac{(1-m)(-m)(-m-1)\cdots(1-m-n+\alpha)}{(n+1-\alpha)!},\\ 0, \text{ when } n+1-\alpha<0 \text{ or } n+1-\alpha>1-m>0.\end{cases}$$

In Ref. 4, a $(P/2,0)$-type holomorphic superdifferential is defined as

$$d\theta^p\,\phi(z,\theta) + d\theta^{p-1}\eta D_\theta\phi(z,\theta); \quad \eta = (dz+\theta\,d\theta). \tag{15}$$

From that we define an Abelian differential of the third kind only with three poles, for simplicity, on a super Riemann sphere as

$$W = [(z-2)/z(z-1)]dz + \partial_z f(z)\theta\,dz + f(z)d\theta. \tag{16}$$

Using the Abelian differential and following the case of two poles, we will now introduce the concept of Euclidean time on a super Riemann sphere and give it a simple explanation. Choose an arbitrary point P_0 different from P_i $(i = 1,2,...,N)$, and define a function

$$\tau(P) = \mathrm{Re}\int_{P_0}^{P} w. \tag{17}$$

Obviously, the function is univalent. If a superfield defined on a super Riemann surface is also written as $\varphi(t,\theta,\sigma)$, here $\tau = t+\theta$, and satisfies the canonical equation:

$$\partial_i\varphi(t,\theta,\sigma) = [H,\varphi(t,\theta,\sigma)], \quad \partial_\theta\varphi(t,\theta,\sigma) = [Q,\varphi(t,\theta,\sigma)]. \tag{18}$$

where H is a Hamiltonian and Q is a supersymmetry charge.

Then the level curves of the function $\tau(p), C_\tau\{p\in S^2,\tau(p)=\tau\}$, describe the position and interacting picture of a superstring at the time t. When $t\to-\infty$, C_τ will become a small circle around $z=0$, and when $t\to\infty$, C_τ will become two small circles around $z=1,\ \infty$, respectively. So with evolution of the time from $-\infty$ to $+\infty$, the superstring enters at $z=0$ for $t\to-\infty$ and can split in two at some time and somewhere, then exits at $z=1,\ \infty$ for $t\to+\infty$.

The result obtained above is only one kind of super case, Ramond sector. For the Neven–Schwartz (NS) sector of multipoles, a similar result will be obtained if such a way to expand the polynomial of the NS sector can be found.

ACKNOWLEDGMENTS

We wish to thank Professor H. Y. Guo for his suggestion, Dr. Z. Qin for his hospitality during our visiting IPT, and we also wish to express our sincere thanks to Professor Zamolodchikov, Professor Vafa, and Professor Sen for valuable discussions.

This work was supported, in part, by the National Natural Science Foundation of China.

[1] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, Nucl. Phys. B **241**, 333 (1984).
[2] D. Friedan, Z. Qiu, and S. H. Shenker, in *Proceedings of the Conference on Vertex Operator in Mathematics and Physics*, November 1983, edited by J. Lepowsky (Springer-Verlag, New York, 1984); Phys. Rev. Lett. **52**, 1575 (1984).
[3] D. Friedan, S. H. Shenker, and E. Maránec, Nucl. Phys. B **271**, 93 (1986).
[4] E. Verlinde, Nucl. Phys. B **300**, 360 (1988).
[5] L. Alvarez Gaumé, C. Gomez, G. Moore, and C. Vafa, Nucl. Phys B **303**, 455 (1988).
[6] H. Y. Gao, J. S. Na, J. M. Shen, S. K. Wang, and Q. H. Yu, J. Phys. A. Math. Gen. **23**, 379 (1990).
[7] H. Y. Guo, J. S. Na, J. M. Shen, S. K. Wang, and Q. H. Yu, Commun. Theor. Phys. **12**, 349 (1989).
[8] R. Dick, Lett. Math. Phys. **18**, 255 (1989).
[9] M. Schlichenmaier, Krichever–Novikov algebras for more than two points, preprint Manusk. Fak. u. Inf. Mannheim 97-1989, to appear to Lett. Math. Phys.
[10] M. Schlichenmaier, Krichever–Novikov algebras for more than two points: explicit generators, preprint Manusk. Fak. u. Inf. Mannheim 101-1989, to appear to Lett. Math. Phys.
[11] S. B. Giddings and P. Nelson, Comm. Math. Phys. **116**, 607 (1988).

Relativistic relaxation models for a simple gas

A. Majorana
Dipartimento di Matematica, viale A. Doria, 6, 95125 Catania, Italy

(Received 29 December 1988; accepted for publication 28 March 1990)

A general relativistic time-relaxation model is introduced and compared with the relativistic Boltzmann equation. It is proved that the model has the same features (formal properties of linear operator, positive transport coefficients, etc.) of the Boltzmann equation. It is also shown that the phase speed of infinitesimal disturbances, propagating in a gas described by our model, is less than the speed of light.

I. INTRODUCTION

The study of relativistic kinetic theory has recently received a new impetus, partly on account of developments in plasma physics. One of the major shortcomings in dealing with the relativistic Boltzmann equation is the complicated structure of the collision integral.[1] Therefore, alternative simpler expressions have been proposed for the collision term, just as in the classical case, by Marle,[2] Anderson and Witting,[3] Cercignani,[4] Cercignani and Majorana.[5,6]

The model proposed by Marle is

$$p^\alpha \frac{\partial f}{\partial x^\alpha} = -\frac{m}{\tau_0}(f-F), \tag{1}$$

where $f(x,p)$ is the distribution function, which is a function of the space-time coordinates x^α and the energy-momentum vector p^α. Here τ_0 is the collision time depending only on space-time cordinates, m is the rest mass of a particle, and F a local Jüttner function.

In the classical limit Eq. (1) corresponds to the nonrelativistic Bhatnager–Gross–Krook (BGK) model;[7] however, as pointed out by Anderson and Witting,[3] in the extreme relativistic limit, the results for transport coefficients with the Marle formulation[2] differ functionally from the results one calculates with the relativistic Grad moment method. In a previous paper with Cercignani[5] we gave an analysis for the propagation of infinitesimal disturbances using the Marle model. Our results do not exhibit any unphysical features, but lead to the unpleasant circumstance that no true discrete spectrum for the wave propagation exists even in the limit of very small frequencies, when one would expect to find the analogous behavior of sound, shear, and heat waves described by relativistic continuum theories. However, an asymptotic analysis has been used to show that such a discrete spectrum exists in an asymptotic sense and is in agreement with physical expectations. The resulting asymptotic expansion agrees with the results obtained by De Groot *et al.*[1] by a direct formal expansion of the solution of the linearized Boltzmann equation.

Anderson and Witting[3] proposed the following model:

$$p^\alpha \frac{\partial f}{\partial x^\alpha} = -\frac{u^\alpha p_\alpha}{\tau_1}(f-F), \tag{2}$$

where u^α is the dimensionless (with respect to the light velocity c) hydrodynamic four-velocity and τ_1 is a parameter depending only on $x=(x^\alpha)$. Anderson and Witting calculated explicitly the transport coefficients according to Eq. (2) and showed that the results agree with the values obtained using the relativistic Grad moment method. In Ref. 6 the propagation of the infinitesimal waves according to the linearized form of Eq. (2) is studied. The dispersion relation has plausible solutions describing thermal, sound, and shear waves for frequencies not exceeding certain critical values of the order of the collision frequency. This is in agreement with the results obtained in the classical case.[8,9] The analytical continuation of the dispersion relation beyond the critical values leads to a continuation of the curves giving the phase speed and attenuation rates as functions of the frequency. However, the phase speed exceeds the speed of light for very high frequencies. This indicates that the analytically continued solutions do not share the property of the true solutions of the Boltzmann equation, as proved by Cercignani.[10]

In this paper we study a general kinetic model which includes the previous models. We analyze in detail the linearized equation and show that it satisfies the same formal properties of the Boltzmann equation.[1,10] In particular, it is proved that the transport coefficients are positive and the phase speed of infinitesimal disturbances is less than the speed of light.

II. BASIC FEATURES OF THE RELATIVISTIC BOLTZMANN EQUATION

Our attention will be confined throughout to a simple gas, i.e., a system of material particles all having the same proper mass m and four-momentum $p=(p^\alpha)$, with $p^\alpha p_\alpha = -m^2c^2$. We neglect quantum effects and regard binary collisions as the only interactions.

Since we employ the metric $g^{\alpha\beta}=\mathrm{diag}(-1,1,1,1)$, a future pointing timelike vector v^α is characterized by $v^\alpha v_\alpha < 0$ and $v^0 > 0$. In our notation Greek indices run from 0–3.

The evolution of the gas can be described by the relativistic Boltzmann equation

$$p^\alpha \frac{\partial f}{\partial x^\alpha} = \frac{1}{2}\int (f'f'_* - ff_*)\, W(p,p_*|p',p'_*)\, \omega_*\omega'\omega'_*, \tag{3}$$

where $\omega = dp^1\, dp^2\, dp^3/p^0$ is the volume element; the transition rate W has the form

$$W(p,p_*|p',p'_*) = \sigma s \delta^{(4)}(p+p_*-p'-p'_*), \tag{4}$$

with $s = 2m^2c^2 - 2p^\alpha_* p_\alpha$; and σ is the differential cross sec-

 0022-2488/90/082042-05$03.00

tion depending on s and the scattering angle.

To derive the linearized form of the relativistic Boltzmann equation, one assumes that

$$f = \Phi(x,p)(1+\phi(x,p)), \tag{5}$$

where

$$\Phi = [1/(2\pi\hbar)^3]\exp\{[\mu(x) + cp^\alpha u_\alpha(x)]/k_B T(x)\}, \tag{6}$$

is a local equilibrium function and ϕ is a perturbation function. Usually one identifies the five parameters μ, u_α, and T with the Gibbs function per particle, the dimensionless hydrodynamic four-velocity, and the temperature of the gas, respectively. In other words, one imposes the conditions of fit

$$\int \Phi\phi p^\alpha \omega = 0, \quad \int \Phi\phi (p^\alpha u_\alpha)^2 \omega = 0 \tag{7}$$

according to the Eckart definition of dynamical quantities. Of course, one can choose the Landau–Lifshitz description by changing conditions (7).

The first-order perturbation function ϕ has to be determined from the integral equation

$$p^\alpha \frac{\partial}{\partial x^\alpha}[\Phi(1+\phi)] = \frac{1}{2}\int \Phi\Phi_*(\phi' + \phi'_* - \phi - \phi_*) W\omega_*\omega'\omega'_*. \tag{8}$$

In a hydrodynamical context one derives thermodynamical equations multiplying the transport equation (8) by various combinations of the powers of p^α; then one integrates over momentum space. Therefore, if g is a generic function of p^α only, we have

$$\frac{\partial}{\partial x^\alpha}\int p^\alpha \Phi(1+\phi) g\omega = \frac{1}{2}\int \Phi\Phi_*(\phi' + \phi'_* - \phi - \phi_*) Wg\omega\omega_*\omega'\omega'_* \tag{9}$$

$$= -n^2\sigma(T)[\phi,g], \tag{10}$$

where $n = n(x^\alpha)$ is the particle density, $\sigma(T)$ is a characteristic cross section depending on the absolute temperature T, and $[,]$ indicates the bilinear symmetric functional defined by Eqs. (9) and (10).

We recall two important properties of the bracket $[,]$:

(i) if g is a linear function of p^α, then $[\phi,g] = 0 \ \forall\phi$;

(ii) $[g,g] \geqslant 0 \ \forall g$.

Properties (i) and (ii) are used to obtain the transport coefficients of the Eckart equations. Indeed, one defines the volume viscosity ζ, the heat conductivity λ, and the shear viscosity η as

$$\zeta = [k_B T/c\sigma(T)]\{[5-3\gamma)zG - 3\gamma]^2/A^{22}\}, \tag{11}$$

$$\lambda = [3ck_B/\sigma(T)]\{[\gamma/(\gamma-1)]^2/B^{11}\}, \tag{12}$$

$$\eta = 10[k_B T/c\sigma(T)](z^2G^2/C^{00}), \tag{13}$$

where γ is the ratio of the heat capacities per particle,

$$G = G(z) = K_3(z)/K_2(z)$$

is the Synge function,[11] $z = mc^2/k_B T$, and $K_n(z)$ is the nth modified Bessel function of the second kind. Furthermore,

$$A^{22} = [\tau^2,\tau^2], \tag{14}$$

$$B^{11} = h_{\alpha\beta}[\tau\pi^\alpha,\tau\pi^\beta], \tag{15}$$

$$C^{00} = [\langle\pi^\alpha\pi^\beta\rangle,\langle\pi_\alpha\pi_\beta\rangle], \tag{16}$$

with

$$\pi^\alpha = cp^\alpha/(k_B T), \quad \tau = -\pi^\alpha u_\alpha, \tag{17}$$

$$h^{\alpha\beta} = u^\alpha u^\beta + g^{\alpha\beta}, \tag{18}$$

where the angular brackets denote the symmetric, purely spatial, trace-free part of the enclosed tensor, i.e.,

$$\langle A_{\alpha\beta}\rangle = \tfrac{1}{2}h^\lambda_\alpha h^\mu_\beta (A_{\lambda\mu} + A_{\mu\lambda} - \tfrac{2}{3}h_{\lambda\mu}h^{\rho\sigma}A_{\rho\sigma}).$$

We recall that the transport coefficients are positive because the functional $[,]$ is positive semidefinite.

III. RELATIVISTIC TIME-RELAXATION MODELS

The BGK model replaces the collision operator of the Boltzmann equation with an operator which has a linear appearance. This latter operator is proportional to the difference between the distribution function and a local equilibrium function. The coefficient of proportionality is called the collision frequency. If $\bar\nu$ denotes the collision frequency, we note that

$$\nu = \bar\nu/n\sigma(T)k_B T$$

is a dimensionless scalar quantity. It is reasonable to assume that $\bar\nu$ depends only on the thermodynamic variables n, T, u^α, and the momentum vector p^α. Now, since ν is a scalar dimensionless function, it can depend only on the dimensionless scalar variables τ and z. Therefore, the transport equation is written as

$$p^\alpha \frac{\partial f}{\partial x^\alpha} = \frac{n}{c}\sigma(T)k_B T\nu(\tau,z)(F-f). \tag{19}$$

Two different choices of ν reproduce the models proposed by Marle[2] and Anderson and Witting.[3]

The function F is a local Jüttner function which contains five unknown functions. These parameters are determined by requiring that the conservation of energy-momentum and particle-number flux are given by Eq. (19). We therefore have the following five conditions:

$$\int \nu(F-f)g_k\omega = 0, \quad (g_k) = (1,p^\alpha). \tag{20}$$

The set of equation (19) and (20) gives the relativistic version of the BGK model.

A linear model can be derived by assuming that f is given by Eq. (5). The corresponding linear transport equation is written as

$$p^\alpha \frac{\partial}{\partial x^\alpha}[\Phi(1+\phi)] = \frac{n}{c}\sigma(T)k_B T\nu\Phi(a + b_\mu\pi^\mu - \phi), \tag{21}$$

where the five parameters a and b_μ (which derive from the five parameters of F) must satisfy the linear equations

$$\int \nu(\tau,z)\Phi(a + b_\mu\pi^\mu - \phi)g_k\omega, \quad (g_k) = (1,p^\alpha). \tag{22}$$

To solve Eqs. (22) with respect to a and b_μ, it is useful to define

$$J(\psi) = \frac{1}{4\pi m^2 c^2 K_2(z)} \int \nu(\tau,z)\exp(-\tau)\psi\omega. \tag{23}$$

We note that the functional J has the same dimension of ψ and $\psi \geqslant 0$ implies $J(\psi) \geqslant 0$.

Let

$$A = J(1), \quad B = J(\tau), \quad C = J(\tau^2), \tag{24}$$

and, taking into account that

$$(n/c)\sigma(T)k_B T\nu(\tau,z)\Phi = [n^2\sigma(T)/4\pi m^2 c^2 K_2(z)]\nu(\tau,z)\exp(-\tau),$$

Eqs. (22) become

$$\begin{aligned} &Aa + Bb_\mu u^\mu = J(\phi), \\ &Bau^\alpha + \left[Cu^\alpha u^\mu + \tfrac{1}{3}(C - z^2 A)h^{\alpha\mu}\right]b_\mu = J(\pi^\alpha\phi). \end{aligned} \tag{25}$$

Explicit expressions for a and b^α are obtained by multiplying properly the second equation in (25) by u_α and h^β_α; one finds

$$a = [1/(AC - B^2)][CJ(\phi) - BJ(\phi\tau)], \tag{26}$$

$$b^\beta = [1/(AC - B^2)][BJ(\phi) - AJ(\phi\tau)]u^\beta + [3/(C - z^2 A)]h^\beta_\alpha J(\phi\pi^\alpha), \tag{27}$$

where, as proved in the Appendix, $AC - B^2$ and $C - z^2 A$ are positive functions. By inserting Eqs. (26) and (27) into Eq. (21), we obtain the following equation for ϕ:

$$p^\alpha \frac{\partial}{\partial x^\alpha}[\Phi(1+\phi)] = \frac{n}{c}\sigma(T)k_B T\nu\Phi\left(\frac{1}{AC - B^2}[(C - \tau B)J(\phi) + (A\tau - B)J(\phi\tau)] + \frac{3}{C - z^2 A}h_{\lambda\mu}\pi^\lambda J(\phi\pi^\mu) - \phi\right). \tag{28}$$

Denoting by $\mathcal{L}(\phi)$ the rhs of Eq. (28), the linearized model assumes the form

$$\begin{aligned} &p^\alpha \frac{\partial}{\partial x^\alpha}[\Phi(1+\phi)] = \mathcal{L}(\phi), \\ &\int \Phi\phi p^\alpha \omega = 0, \quad \int \Phi\phi(p^\alpha u_\alpha)^2\omega = 0. \end{aligned} \tag{29}$$

We remark that Eqs. (29) are not completely linear because the parameters T and u^α appear inside the function Φ. Since, also, $\mathcal{L}$ depends on n, T, and u^α, we write, when necessary, $\mathcal{L} = \mathcal{L}(\phi;n,T,u^\alpha)$.

Analogous to the Boltzmann equation, one can write moment equations, multiplying the transport equation (21) by a polynomial function. We now have

$$\frac{\partial}{\partial x^\alpha}\int p^\alpha\Phi(1+\phi)g\omega = \frac{n}{c}\sigma(T)k_B T\int \nu\Phi(a + b_\mu\pi^\mu - \phi)g\omega = n^2\sigma(T)[aJ(g) + b_\mu J(\pi^\mu g) - J(\phi g)],$$

which, introducing the functional

$$L(\phi,g) = J(\phi g) - aJ(g) - b_\mu J(\pi^\mu g) \tag{30}$$

and related with $\mathcal{L}(\phi)$ by

$$\int \mathcal{L}(\phi)g\omega = -n^2\sigma(T)L(\phi,g)$$

assumes the form

$$\frac{\partial}{\partial x^\alpha}\int p^\alpha\Phi(1+\phi)g\omega = -n^2\sigma(T)L(\phi,g). \tag{31}$$

Equation (31) coincides with Eq. (10), replacing $L(\phi,g)$ with $[\phi,g]$.

Inserting Eqs. (26) and (27) into Eq. (30), we obtain

$$\begin{aligned} L(\phi,g) = J(\phi g) &+ [1/(AC - B^2)][BJ(\phi\tau)J(g) \\ &+ BJ(\phi)J(g\tau) - CJ(\phi)J(g) - AJ(\phi\tau)J(g\tau)] \\ &- [3/(C - z^2 A)]h_{\alpha\beta}J(\pi^\alpha\phi)J(\pi^\beta g). \end{aligned} \tag{32}$$

Using expression (32) one can easy verify that the following properties hold:

(i) L is a bilinear symmetric functional;

(ii) if ϕ is a linear function of p^α, then $L(\phi,g)$ is zero identically for every g.

We note that the function g is an argument of the linear operators [see Eqs. (31) and (32)] and then, without loss of generality, we can assume that g is a scalar function. Defining

$$\begin{aligned} \Lambda(g) = L(g,g) &= J(g^2) + [1/(AC - B^2)][2BJ(g\tau)J(g) - CJ^2(g) \\ &- AJ^2(g\tau)] - [3/(C - z^2 A)]h_{\alpha\beta}J(\pi^\alpha g)J(\pi^\beta g) \end{aligned} \tag{33}$$

we can prove the following theorem.

Theorem 1: We find that $\Lambda(g) \geqslant 0$ and $\Lambda(g) = 0$ if and only if g is a linear function of p^α.

Proof: Let us consider a linear function of p^α, depending on g, defined by

$$\begin{aligned} S(g) = [1/(AC - B^2)]\{&CJ(g) - BJ(g\tau) \\ &+ [AJ(g\tau) - BJ(g)]\tau\} \\ &+ [3/(C - z^2 A)]h_{\alpha\beta}\pi^\alpha J(\pi^\beta g) \end{aligned} \tag{34}$$

and $g_S = g - S(g)$. It is easy to verify that

$$J(g_S) = 0, \quad J(g_S\pi^\mu) = 0 \tag{35}$$

for every g. Since by the previous properties (i) and (ii),

$$\Lambda(g) = L(g,g) = L[g_S + S(g), g_S + S(g)] = L(g_S,g_S) = \Lambda(g_S),$$

it follows, using Eqs. (33) and (35), that

$$\Lambda(g) = \Lambda(g_S) = J(g_S^2) \geqslant 0. \tag{36}$$

Thus the first part of Theorem 1 is proved. Now, to complete the proof, it remains to be shown that $\Lambda(g) = 0$ if and only if g is a linear function of p^α.

If $\Lambda(g) = 0$ Eq. (36) implies $J(g_S^2) = 0$. Since $g_S^2 \geqslant 0$ and J is defined by Eq. (23), then $g_S = 0$ and hence (by definition of g_S) $g = S(g)$, i.e., g is a linear function of p^α. Conversely, if g is linear, it is easy to verify that $S(g) = g$; hence $g_S = 0$ and $\Lambda(g) = 0$. ∎

Another important property of the function Λ is established by the following theorem.

Theorem 2: We find that $\Lambda(g)$ is a convex function.

Proof: We have to prove that

$$\Lambda[\lambda g+(1-\lambda)h]\leqslant\lambda\Lambda(g)+(1-\lambda)\Lambda(h),$$

$$\forall g,h;\forall\lambda\in[0,1].$$

To this aim it is enough to show the identity

$$\Lambda[\lambda g+(1-\lambda)h]-\lambda\Lambda(g)-(1-\lambda)\Lambda(h)$$
$$=-\lambda(1-\lambda)\Lambda(g-h)$$

because $\Lambda(g-h)$ is non-negative. Taking into account the bilinear and symmetric properties of L, one obtains

$$\begin{aligned}&\Lambda[\lambda g+(1-\lambda)h]-\lambda\Lambda(g)-(1-\lambda)\Lambda(h)\\&=\lambda^2L(g,g)+2\lambda(1-\lambda)L(g,h)\\&\quad+(1-\lambda)^2L(g,g)-\lambda L(g,g)-(1-\lambda)L(h,h)\\&=-\lambda(1-\lambda)[L(g,g)-2L(g,h)+L(h,h)]\\&=-\lambda(1-\lambda)L(g-h,g-h)\\&=-\lambda(1-\lambda)\Lambda(g-h).\end{aligned}$$

■

Thus we have shown that the linear operator $L(,)$ has the same formal properties of $[,]$ and hence that the transport coefficients derived by the relaxation transport equation are positive. Of course, the relaxation transport coefficients can differ from those obtained from the Boltzmann equation; however, one can choose properly the collision frequency ν to fit them as best as possible.

By straightforward calculation using $L(,)$ instead of $[,]$, one finds

$$A^{22}=E+[1/(AC-B^2)](2BCD-AD^2-C^3), \tag{37}$$

$$B^{11}=E-z^2C-[1/(C-z^2A)](D-z^2B)^2, \tag{38}$$

$$C^{00}=\tfrac{2}{3}(E-2z^2C+z^4A), \tag{39}$$

where

$$D=J(\tau^3),\quad E=J(\tau^4). \tag{40}$$

We report explicit results for two special cases of ν.

(i) $\nu=\nu_0$ (constant with respect to p^α):

$$A=(1/z)\nu_0(G-4/z),\quad B=\nu_0,\quad C=\nu_0(zG-1),$$
$$D=\nu_0z(z+3G),\quad E=\nu_0z[2z+(z^2+15)G].$$

(ii) $\nu=\nu_1\tau$ (ν_1 constant with respect to p^α):

$$A=\nu_1,\quad B=\nu_1(zG-1),\quad C=\nu_1z(z+3G),$$
$$D=\nu_1z[2z+(z^2+15)G],\quad E=\nu_1z(z^2+15)(6G+z),$$

where $G=G(z)$ is the Synge function.

IV. PROPAGATION OF INFINITESIMAL DISTURBANCES

We study the evolution of an infinitesimal disturbance, traveling in a medium otherwise in equilibrium, according to the relaxation model.

Denoting by μ_0, T_0, and u_0^α the space-time independent parameters of the distribution function Φ^{eq}, which describes the equilibrium state to be disturbed, we let

$$\mu(x)=\mu_0+\delta\mu(x), \tag{41}$$

$$T(x)=T_0+\delta T(x), \tag{42}$$

$$u^\alpha(x)=u_0^\alpha+\delta u^\alpha(x). \tag{43}$$

Expanding the local equilibrium function Φ, we obtain

$$\Phi=\Phi^{eq}[1+\chi(x^\alpha,p^\alpha)]+\text{nonlinear terms}, \tag{44}$$

where

$$\chi=(1/k_BT_0)[\delta\mu(x)-(1/T_0)(\mu_0+cp_\alpha u_0^\alpha)$$
$$\times\delta T(x)+cp_\alpha\,\delta u^\alpha(x)]. \tag{45}$$

Since

$$\Phi(1+\phi)=\Phi^{eq}(1+\phi+\chi)+\text{nonlinear terms},$$
$$\Phi\phi=\Phi^{eq}\phi+\text{nonlinear terms},$$

defining

$$h=\phi+\chi \tag{46}$$

we have

$$\mathscr{L}(\phi;n,T,u^\alpha)=\mathscr{L}(\phi;n_0,T_0,u_0^\alpha)+\text{nonlinear terms}=\mathscr{L}(h-\chi;n_0,T_0,u_0^\alpha)+\text{nonlinear terms}$$
$$=\mathscr{L}(h;n_0,T_0,u_0^\alpha)-\mathscr{L}(\chi;n_0,T_0,u_0^\alpha)+\text{nonlinear terms}=\mathscr{L}(h;n_0,T_0,u_0^\alpha)+\text{nonlinear terms}.$$

The above equation follows from the fact that $\mathscr{L}(\chi;n_0,T_0,u_0^\alpha)=0$ because χ is a linear function of p^α.

As a consequence of the previous relations, Eqs. (29) become

$$p^\alpha\frac{\partial h}{\partial x^\alpha}=\frac{1}{\Phi^{eq}}\mathscr{L}(h;n_0,T_0,u_0^\alpha),$$

$$\int\Phi^{eq}(h-\chi)p^\alpha\omega=0,\quad\int\Phi^{eq}(h-\chi)(p_\alpha u_0^\alpha)^2\omega=0. \tag{47}$$

The set of the six above equations contains the six unknowns h, $\delta\mu(x)$, $\delta T(x)$, and δu_0^α. However, the first equation in (47) contains only the variable h and then can be solved independently.

Let

$$\mathscr{L}_0(h)=(1/\Phi^{eq})\mathscr{L}(h;n_0,T_0,u_0^\alpha). \tag{48}$$

Denoting by

$$(g,h)=\int\bar{g}h\Phi^{eq}\omega \tag{49}$$

the scalar product in the Hilbert space $H=\mathscr{L}^2(\Phi^{eq}\omega)$ if h is real and belongs to H, we have

$$(\mathscr{L}_0(h),h)=-n^2\sigma(T)L(h,h)=-n^2\sigma(T)\Lambda(h)\leqslant0. \tag{50}$$

We look for solutions of Eq. $(47)_1$ of the kind

$$h=\rho(p^\alpha)\exp(ik_\beta x^\beta), \tag{51}$$

where k^β is the wave vector. Since we consider forced waves in an unbound medium, we take the real (positive) frequency as given and the complex wave vector as unknown.

Inserting Eq. (51) into Eq. (47), we obtain

$$ik_\alpha p^\alpha h = \mathcal{L}_0(h). \tag{52}$$

Assuming, without loss of generality, $k^0 = \omega(>0)$, $k^i = (k^1,0,0)$, $k^1 = \omega(a+ib)$, Eq. (52) becomes

$$[-i\omega p^0 + i\omega(a+ib)p^1]\rho = \mathcal{L}_0(\rho). \tag{53}$$

Then

$$-i\omega(p^0\rho,\rho) + i\omega(a+ib)(p^1\rho,\rho) = (\mathcal{L}_0(\rho),\rho) \leqslant 0.$$

Since $(\mathcal{L}_0(\rho),\rho)$ is real, we obtain

$$(p^0\rho,\rho) - a(p^1\rho,\rho) = 0, \tag{54}$$

$$-b(p^1\rho,\rho) \leqslant 0. \tag{55}$$

Therefore, the phase speed $v_{\rm ph}$ is

$$v_{\rm ph} = c/|a| = c[|(p^1\rho,\rho)|/(p^0\rho,\rho)] < c$$

because $|p^1| < |p^0|$ almost everywhere. This result shows that disturbances propagate with a velocity less than the light speed as one aspect in the relativistic context. Inequality (55) means that waves damp when propagating. The same result was obtained by Cercignani[10] using the Boltzmann equation.

The case when h does not belong to H can be analyzed in exactly the same way as in Ref. 10.

Explicit results of wave propagation were obtained by Cercignani and Majorana[5,6] using the Marle and Anderson and Witting models. It was shown in particular that the phase speed is less then c when the frequency ω does not exceed a critical frequency ω_c (depending on the model and T_0). When $\omega > \omega_c$, the original dispersion relation does not admit solutions, but, considering the analytically continued equation, one can again obtain solutions, which are in some way a continuation of the true roots. The continued solutions are meaningless when ω is greater than a frequency $\omega_f > \omega_c$ because the phase speed reaches the light velocity.

The present results state that the relaxation model introduced here is consistent with the relativistic framework, at least for a wave propagation with not too high a frequency. If $\omega > \omega_c$ one cannot look for solutions of the type given by Eq. (51), but could consider a boundary value problem associated with Eq. (47).

ACKNOWLEDGMENT

This research was supported by Ministero della Pubblica Istruzione of Italy.

APPENDIX: THE POSITIVITY OF $AC-B^2$ AND $C-z^2A$

We prove that $AC - B^2 > 0$ and $C > z^2A$. In fact, if λ is a real number, we have

$$A\lambda^2 + 2B\lambda + C = \lambda^2 J(1) + 2\lambda J(\tau) + J(\tau^2)$$
$$= J(\lambda^2 + 2\lambda\tau + \tau^2) = J[(\lambda + \tau)^2] > 0$$

and hence $AC - B^2$ is positive.

The second inequality follows from

$$C - z^2A = J(\tau^2 - z^2)$$
$$= (c/k_{\rm B}T)^2 J[(p^\alpha u_\alpha)^2 - m^2c^2] > 0.$$

[1] S. R. de Groot, W. A. van Leeuwen, and Ch. G. van Weert, *Relativistic Kinetic Theory* (North-Holland, Amsterdam, 1980).
[2] C. Marle, Ann. Inst. Henri Poincaré **10A**, 67, 127 (1969).
[3] J. L. Anderson and H. R. Witting, Physica **74**, 466 (1974).
[4] C. Cercignani, in *Atti del 5° Congresso di Relativita' Generale e Fisica della Gravitazione*, Catania, Italy, September 1982.
[5] C. Cercignani and A. Majorana, Meccanica **19**, 175 (1984).
[6] C. Cercignani and A. Majorana, Phys. Fluids **28**, 1673 (1985).
[7] P. L. Bhatnagar, E. P. Gross, and M. Krook, Phys. Rev. **94**, 511 (1954).
[8] L. Sirovich and J. K. Thurber, J. Acoust. Soc. Am. **37**, 329 (1965).
[9] C. Cercignani and A. Majorana, J. Appl. Math. Phys. (ZAMP) **36**, 699 (1985).
[10] C. Cercignani, Phys. Rev. Lett. **50**, 1122 (1983).
[11] J. L. Synge, *The Relativistic Gas* (North-Holland, Amsterdam, 1957).

Effective conductivity and average polarizability of random polycrystals

Marco Avellaneda and Oscar Bruno[a)]
Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, New York 10012

(Received 2 October 1989; accepted for publication 28 March 1990)

A third-order expansion for the effective thermal conductivity tensor K* of anisotropic polycrystalline cell materials is derived. The coefficients of the expansion are given in terms of the average polarizability tensor, a nondimensional quantity determined from the grain shape and crystallographic orientation distributions independent of other details of the microgeometry such as two (or more) particle correlation functions. Explicit numerical results for a wide variety of microgeometries made of ellipsoidal cells are obtained. This calculation uses a new method that exploits the symmetry properties of the effective conductivity tensor of a cell material as a function of the single-crystal conductivities.

I. INTRODUCTION

Single-phase polycrystals are assemblages of homogeneous grains or crystallites with arbitrary shapes, packed together so as to fill all space. The three principal thermal conductivities are the same in all grains, while the crystallographic orientations may vary randomly from one grain to another. A wide variety of crystalline solids, including most metals, are commonly found in polycrystalline form. According to Fourier's law of heat conduction, the temperature $T(x)$ and heat flux density $\mathbf{q}(x)$ satisfy

$$\mathbf{q}(x) = -\mathsf{K}(x)\cdot\nabla T(x) \tag{1}$$

and

$$\nabla\cdot q(x) = 0, \tag{2}$$

where $\mathsf{K}(x)$ is the local thermal conductivity tensor. We can write this tensor in the form

$$\mathsf{K}(x) = \sum_i \chi i(x)\,{}^t\mathsf{R}_i\mathsf{K}^0\mathsf{R}_i, \tag{3}$$

where $\chi_i(x) = 1$ if x is in grain i and $\chi_i(x) = 0$ otherwise;

$$\mathsf{K}^{(0)} = \mathrm{diag}(k_1,k_2,k_3), \tag{4}$$

where k_j, $1\leqslant j\leqslant 3$ are the principal conductivities of the crystalline phase and ${}^t\mathsf{R}_i\mathsf{K}^{(0)}\mathsf{R}_i$ is the conductivity in grain i, $i = 1,2,3,...$, where R_i are suitable rotations. In (2), it is implicity understood that the normal component of $\mathbf{q}(x)$ is continuous across grain boundaries. We assume for simplicity that the medium occupies all space and that $\mathsf{K}(x)$ is ergodic. Solutions of (1) and (2) can then be decomposed into the average and fluctuating components

$$\nabla T(x) = \langle\nabla T\rangle + (\nabla T)'(x),$$
$$\mathbf{q}(x) = \langle\mathbf{q}\rangle + \mathbf{q}'(x). \tag{5}$$

The effective conductivity tensor K* is defined by the relation

$$\langle\mathbf{q}\rangle = -\mathsf{K}^*\cdot\langle\nabla T\rangle \tag{6}$$

as $\langle\nabla T\rangle$ varies over all three-vectors. Physcially, K* represents the conductivity of an equivalent homogeneous material and can be determined experimentally by direct or indirect measurements on samples which are large compared to the typical grain size.[1]

The theoretical determination of K* is made difficult by the complex character of Eqs. (1) and (2) for realistic microgeometries. Also, for real materials, only limited information is available on the microgeometry, usually through measurements of correlation functions or the grain shape distribution. Considerable work has been done to estimate K* starting with Voigt,[2] Reuss,[3] and Wiener,[4] who derived upper and lower bounds on the effective conductivity k^* of isotropic aggregates:

$$k_R \equiv 3/(1/k_1 + 1/k_2 + 1/k_3) \leqslant k^* \leqslant k_V$$
$$\equiv \tfrac{1}{3}(k_1 + k_2 + k_3). \tag{7}$$

Recently, this interval was narrowed to

$$k_S \leqslant k^* \leqslant k_V, \tag{8}$$

where k_S is the positive root of

$$k_S^3 + (k_1 + k_2 + k_3)k_S^2 - 4k_1k_2k_3 = 0 \tag{9}$$

and the new interval was shown to be, in a sense, optimal.[5] However, the size of the interval (8) is too wide to reasonably determine k^* even for moderately anisotropic crystals. More efficient estimates on the effective conductivity that incorporate additional statistical information were derived in 1969 and 1970 by Molyneux,[6,7] who obtained a perturbation expansion for k^* valid to third order in $\delta k_i = k_i - 1$, $1\leqslant i\leqslant 3$, and related bounds. This expansion incorporates the two- and three-point correlation functions of the material. The statistical information needed as input in such expressions must then be determined experimentally, e.g., by image processing techniques.[8,9]

We will focus here on a statistical model introduced by Miller,[10] for which there exist estimates on k^* which are significantly narrower than the $k_s - k_V$ bounds and yet more explicit than the Molyneux bounds. These estimates require as input the particle shape distribution and crystallographic orientation distribution within grains. Specifically, we will assume that (i) the crystallographic orientations of different grains are statistically independent and (ii) the crystallographic orientation within each grain is distributed randomly in such a way that for any rotation R, the six conductivities

$${}^t\mathsf{R}\cdot\mathrm{diag}(k_i,k_j,k_l)\cdot\mathsf{R}, \tag{10}$$

obtained by permutation of the indices i, j, k, arise with equal

[a)] Present address: School of Mathematics, University of Minnesota, Minneapolis, MN 55455.

 0022-2488/90/082047-10$03.00

probability. Assumption (ii) is satisfied, for instance, if the crystallographic orientation is uniformly distributed (and then *all* orientations are equally likely). Such assumptions determine the correlation functions to some extent, but not completely. For instance, the correlation length is of the order of one particle diameter. For isotropic aggregates of this type, Hashin and Shtrikman[11] and Walpole[12] showed that

$$k^* = k_V - \frac{1}{9}\sum_{i=1}^{3}(k_i - k_V)^2 + o(\delta k)^2 \tag{11}$$

and

$$k_1\frac{4k_1^2 + 8k_1k_2 + 8k_1k_3 + 7k_2k_3}{16k_1^2 + 5k_1k_2 + 5k_1k_3 + k_2k_3} \leqslant k^* \leqslant k_3\frac{4k_3^2 + 8k_3k_2 + 8k_1k_3 + 7k_1k_2}{16k_3^2 + 5k_3k_2 + 5k_3k_1 + k_1k_2} \tag{12}$$

if $k_1 \leqslant k_2 \leqslant k_3$.

Subsequently, Willemse and Caspers[13] calculated the corresponding third-order expansion for k^* by assuming uniformly distributed crystallographic orientations. This expansion depends on a single parameter related to the particle shape distribution.

In this paper, we introduce a new approach for estimating the effective conductivity which fully exploits the symmetry properties implied by the statistical assumptions (i) and (ii). Our main result is a third-order Taylor expansion of the anisotropic tensor $\mathsf{K}^*(k_1,k_2,k_3)$ in δk_i. The coefficients of the expansion depend on the *average polarizability tensor* of the aggregate, a quantity determined only from the grain shape and crystallographic orientation distributions and independent from other details of the microgeometry. We recover the results of Hashin and Shtrikman,[11] Walpole,[12] and Willemse and Caspers[13] as a special case when K^* is isotropic (see Appendix B).

II. THE MAIN RESULT: RELATION BETWEEN CONDUCTIVITY AND AVERAGE POLARIZABILITY

Consider the Taylor series expansion of $\mathsf{K}^* = \mathsf{K}^*(k_1,k_2,k_3)$ near $k_i = 1$:

$$\begin{aligned}\mathsf{K}^*(k_1,k_2,k_3) = \mathsf{I} &+ \sum_{i=1}^{3}\frac{\partial}{\partial k_i}\mathsf{K}^*(1,1,1)(\delta k_i) \\ &+ \frac{1}{2}\sum_{i,j=1}^{3}\frac{\partial^2}{\partial k_i\,\partial k_j}\mathsf{K}^*(1,1,1)(\delta k_i)(\delta k_j) \\ &+ \frac{1}{6}\sum_{i,j=1}^{3}\frac{\partial^3}{\partial k_i\,\partial k_j\,\partial k_l}\mathsf{K}^*(1,1,1)(\delta k_i) \\ &\times(\delta k_j)(\delta k_l) + \cdots .\end{aligned} \tag{13}$$

From first principles,

$$\mathsf{K}^*(\lambda k_1,\lambda k_2,\lambda k_3) = \lambda\mathsf{K}^*(k_1,k_2,k_3). \tag{14}$$

Also, we have

$$\mathsf{K}^*(k_1,k_2,k_3) = \mathsf{K}^*(k_i,k_j,k_l), \quad i\neq j\neq l \tag{15}$$

since our statistical assumptions ensure that k_1, k_2, k_3 are interchangeable. Differentiating K^* at $k_i = 1$ and using (14) and (15), we obtain the relations

$$\frac{\partial}{\partial k_i}\mathsf{K}^*(1,1,1) = \frac{1}{3}\mathsf{I}, \quad 1\leqslant i\leqslant 3, \tag{16}$$

$$\frac{\partial^2}{\partial k_1^2}\mathsf{K}^*(1,1,1) = \frac{\partial^2}{\partial k_2^2}\mathsf{K}^*(1,1,1) = \frac{\partial^2}{\partial k_3^2}\mathsf{K}^*(1,1,1) \equiv \mathsf{A}, \tag{17}$$

$$\frac{\partial^2}{\partial k_i\,\partial k_j}\mathsf{K}^*(1,1,1) = -\frac{1}{2}\frac{\partial^2}{\partial k_1^2}\mathsf{K}^*(1,1,1), \quad i\neq j, \tag{18}$$

$$\frac{\partial^2}{\partial k_1^3}\mathsf{K}^*(1,1,1) = \frac{\partial^3}{\partial k_2^3}\mathsf{K}^*(1,1,1) = \frac{\partial^3}{\partial k_3^3}\mathsf{K}^*(1,1,1) \equiv \mathsf{B}, \tag{19}$$

$$\frac{\partial^3}{\partial k_i^2\,\partial k_j}\mathsf{K}^*(1,1,1) = -\frac{1}{2}\frac{\partial^3}{\partial k_1^3}\mathsf{K}^*(1,1,1) - \frac{1}{2}\frac{\partial^2}{\partial k_1^2}\mathsf{K}^*(1,1,1), \quad i\neq j, \tag{20}$$

and

$$\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\mathsf{K}^*(1,1,1) = \frac{\partial^3}{\partial k_1^3}\mathsf{K}^*(1,1,1) + \frac{3}{2}\frac{\partial^2}{\partial k_1^2}\mathsf{K}^*(1,1,1). \tag{21}$$

Equation (20) was obtained by differentiating the relation $\mathsf{K}^*(k_1,k_1,1) = k_1\mathsf{K}^*(1,1,k_1^{-1})$ and using (19), while Eq. (21) was obtained using the identity $\mathsf{K}^*(k_1,k_2,k_3) = k_3\mathsf{K}^*(k_1/k_3,k_2/k_3,1)$ together with (19) and (20). Thus we see that the third-order Taylor expansion (13) depends only on the two tensors of microgeometric parameters A and B. For later use in comparison with other work, we observe that this expansion can be rewritten in the form

$$\begin{aligned}K^*(k_1,k_2,k_3) = k_V\Bigg[\mathsf{I} &+ \frac{3}{4}\mathsf{A}\sum_{i=1}^{3}\left(\frac{k_i}{k_V} - 1\right)^2 \\ &+ \left(\frac{3}{4}\mathsf{A} + \frac{3}{4}\mathsf{B}\right)\sum_{i=1}^{3}\left(\frac{k_i}{k_V} - 1\right)^3\Bigg] \\ &+ o\left(\left|\frac{k_i}{k_V} - 1\right|\right)^3,\end{aligned} \tag{22}$$

with $k_V = \frac{1}{3}(k_1 + k_2 + k_3)$. To further determine the tensors A and B, we introduce the averge polarizability tensor. Consider first a configuration consisting of a single grain G with the conductivity $\mathsf{K}_G = {}^t\mathsf{R}\mathsf{K}^{(0)}\mathsf{R}$ embedded in an isotropic medium with unit conductivity. Given an arbitrary three-vector $\mathbf{e}$, denote by $T_{\mathbf{e}}^G(x)$ the temperature field which solves the corresponding field equations and such that

$\lim_{|x|\to\infty} \nabla T_e^G(x) = \mathbf{e}$. The single-grain polarizability tensor $\mathscr{L}_G$ is defined by the relation

$$ {}^t\mathbf{e}\cdot\mathscr{L}_G\cdot\mathbf{e} = \frac{1}{\mathrm{Vol}(G)}\int_G {}^t\nabla T_e^G(x)\cdot(\mathsf{K}_G - I)\cdot\mathbf{e}\,dx. \quad (23) $$

The average polarizability tensor of the polycrystal is defined by

$$ \mathscr{L} = \langle \mathscr{L}_G \rangle, \quad (24) $$

where the average is taken over the grain shape and crystallographic orientation distributions. Physcially, $\mathscr{L}$ represents the first-order correction to the value of the effective conductivity of an isotropic medium of unit conductivity containing an embedded dilute array of grains randomly sampled from the polycrystal (see Fig. 1). This tensor can be explicitly computed for several microgeometries (see Sec. III).

The identities relating A, B, and the average polarization tensor $\mathscr{L}$ for general grain shapes and orientations are

$$ \mathsf{A} = \frac{2}{3}\left[\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) - \frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1)\right] \quad (25) $$

and

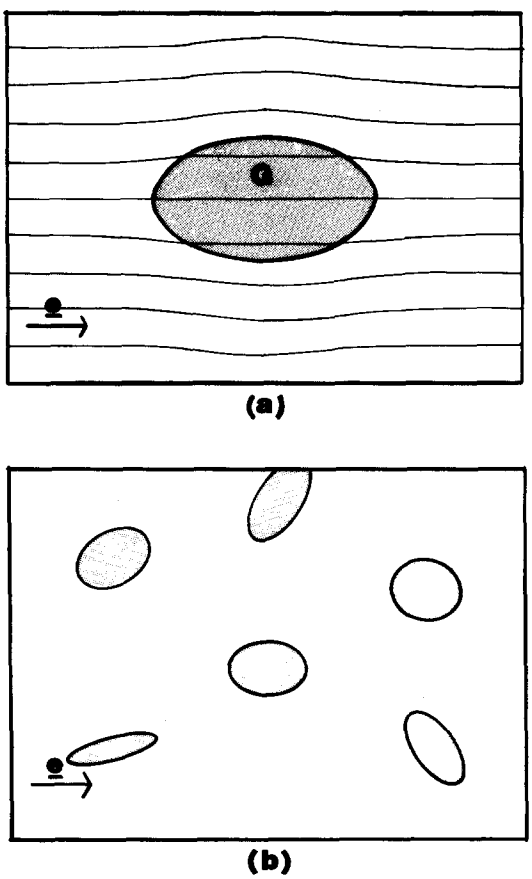


FIG. 1. (a) A schematic view of the single-grain configuration and the corresponding field lines of ∇T_e^G. The conductivity is equal to K_G inside the grain and I outside. (b) The average polarizability $\mathscr{L} = \langle \mathscr{L}_G \rangle$ satisfies $\mathsf{K}_{\mathrm{eff}}^* = \mathsf{I} + f\mathscr{L} + o(f)$, where $\mathsf{K}_{\mathrm{eff}}^*$ is the effective conductivity of a configuration consisting of a dilute array of randomly selected grains occupying a volume fraction f and surrounded by a medium with the conductivity $\mathsf{K} = \mathsf{I}$.

$$ \mathsf{B} = \frac{2}{9}\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) - \frac{2}{3}\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1) + \frac{4}{9}\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\mathscr{L}(1,1,1) - \frac{2}{3}\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) + \frac{2}{3}\frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1). \quad (26) $$

Formulas (25) and (26) constitute the main new result of this paper: They are derived in Secs. IV–VI.

We point out that there exists a general linear relationship between the coefficients of the classical perturbation expansion in the conductivity contrast of a cell material and the expansion in powers of the volume fraction for the conductivity of a uniform reference medium containing a dilute dispersion of randomly selected cells from the composite. Such a relation was first systematically derived in Bruno's thesis,[14] although some special cases of it already appeared in the work of Miller[10] and Elsayed.[15] The present results consist of an application of this general principle to single-phase polycrystals. It was shown that the computation of fourth-order and higher-order expansions requires the input of quantities related to the second- (and higher) order corrections to the conductivity of a uniform medium containing a dilute array of cells.[14,15] Such corrections necessarily involve input of quantities related to interactions of two or more particles.

III. AGGREGATES OF ELLIPSOIDAL GRAINS

According to a classical result in potential theory going back to Maxwell,[16] the field $\nabla T_e^G(x)$ is uniform within the grain G if G is an ellipsoid. This allows us to obtain an explicit formula for $\mathscr{L}_G$. Thus for aggregates consisting of ellipsoidal grains—not necessarily with the same shapes or orientations—the average polarizability $\mathscr{L} = \langle \mathscr{L}_G \rangle$ can be determined explicitly.

Fixing a Cartesian reference frame, $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$, $\mathbf{e}^{(3)}$, the shape and position of a single ellipsoidal grain is determined by three unit vectors $\mathbf{n}^{(1)}$, $\mathbf{n}^{(2)}$, $\mathbf{n}^{(3)}$ in the directions of its semiaxes and by the corresponding depolarization factors L_i, $1 \leqslant i \leqslant 3$, which satisfy $0 \leqslant L_i \leqslant 1$, $\Sigma_i L_i = 1$. If the position of the vector $\mathbf{n}^{(i)}$ is specified by

$$ \mathbf{n}^{(i)} = {}^t\mathsf{O}\mathbf{e}^{(i)}, \quad (27) $$

where O is a rotation matrix, and the conductivity within the grain is $\mathsf{K}_G = {}^t\mathsf{R}\mathsf{K}^{(0)}\mathsf{R}$, it can be shown that

$$ \mathscr{L}_G = \delta\mathsf{K}_G(I + {}^t\mathsf{OLO}\,\delta\mathsf{K}_G)^{-1}, \quad (28) $$

where

$$ \delta\mathsf{K}_G = {}^t\mathsf{R}(\mathsf{K}^{(0)} - \mathsf{I})\mathsf{R} = {}^t\mathsf{R}\,\delta\mathsf{K}^{(0)}\mathsf{R} \quad (29) $$

and

$$ \mathsf{L} = \mathrm{diag}(L_1, L_2, L_3). \quad (30) $$

Thus denoting by $d\mu(\mathsf{L},\mathsf{O},\mathsf{R})$ the joint grain shape–crystallographic distribution orientation, we have, from (24) and (28),

$$\mathscr{L} = \int {}^{t}\mathsf{R}\,\delta\mathsf{K}^{(0)}\mathsf{R}(\mathsf{I} + {}^{t}\mathsf{O}\mathsf{L}\mathsf{O}{}^{t}\mathsf{R}\,\delta\mathsf{K}^{(0)}\mathsf{R})^{-1}\,d\mu(\mathsf{L},\mathsf{O},\mathsf{R}). \tag{31}$$

Computing the second and third partial derivatives of $\mathscr{L}$ at $k_i = 1$ and substituting their expressions into (25) and (26), we obtain an explicit third-order expansion for aggregates of ellipsoidal grains satisfying assumptions (i) and (ii). This calculation is straightforward, although rather lengthy (see Appendix C). When the orientation distribution within each grain is uniformly distributed, the corresponding expressions for A and B are

$$\mathsf{A} = -\frac{2}{45}\int {}^{t}\mathsf{O}\mathsf{L}\mathsf{O}\,d\mu - \frac{2}{15}\mathsf{I} \tag{32}$$

and

$$\mathsf{B} = \frac{32}{315}\int {}^{t}\mathsf{O}\mathsf{L}^2\mathsf{O}\,d\mu - \frac{2}{105}\int \operatorname{tr}\mathsf{L}^2\,d\mu + \frac{2}{315}\int {}^{t}\mathsf{O}\mathsf{L}\mathsf{O}\,d\mu - \frac{4}{21}\mathsf{I}. \tag{33}$$

For instance, if the grains are oriented uniformly in all directions,

$$\mathsf{A} = -(4/27)\mathsf{I} \tag{34}$$

and

$$\mathsf{B} = \left[\tfrac{2}{135}\langle L_1^2 + L_2^2 + L_3^2\rangle + \tfrac{26}{135}\right]\mathsf{I}, \tag{35}$$

where $\langle L_1^2 + L_2^2 + L_3^2\rangle$ denotes averaging over all grain shapes. For spherical grains, we obtain, with $L_1 = L_2 = L_3 = \frac{1}{3}$,

$$\mathsf{B}_S = \tfrac{16}{81}\mathsf{I}. \tag{36}$$

In the limiting cases of randomly oriented lamellar particles ($L_1 = 1$, $L_2 = L_3 = 0$) and needle-like particles ($L_1 = L_2 = \frac{1}{2}$, $L_3 = 0$) we obtain, respectively,

$$\mathsf{B}_L = \tfrac{28}{135}\mathsf{I}, \quad \mathsf{B}_N = \tfrac{53}{270}\mathsf{I}. \tag{37}$$

IV. AN ASSOCIATED *N*-PHASE POLYCRYSTAL

To incorporate into the calculation the statistical independence of the crystallographic orientations of different cells, we introduce a mathematical artifice. Let N be a large positive integer. We assign randomly an integer between 1 and N to each grain. In this way, we obtain N statistically equivalent sets of grains, each one formed by those cells with the same assigned number. Next, we replace the single-crystal conductivity matrix K^0 by

$$\mathsf{K}^{(j)} = \operatorname{diag}(x_j, y_j, z_j) \tag{38}$$

in all grains of the set with the assigned number j, $1 \leqslant j \leqslant N$, where (x_j, y_j, z_j) are N triples of positive numbers. The local conductivity tensor of this N-phase polycrystal is

$$\mathsf{K}_N(x) = \sum_{j=1}^{N}\left[\sum_{\text{grain } i \text{ has assigned number } j} \chi_i(x)\,{}^{t}\mathsf{R}_i\mathsf{K}^{(j)}\mathsf{R}_i\right]. \tag{39}$$

Let

$$\mathsf{K}_N^* = \mathsf{K}_N^*[(x_1,y_1,z_1),\ldots,(x_N,y_N,z_N)] \tag{40}$$

denote the corresponding effective conductivity, viewed as a function of the $3N$ variables (x_i, y_i, z_i), $i \leqslant N$. Clearly, by construction,

$$\mathsf{K}_N^*[(x_1,y_1,z_1),\ldots,(x_i,y_i,z_i),\ldots,(x_j,y_j,z_j),\ldots,(x_N,y_N,z_N)] = \mathsf{K}^*[(x_1,y_1,z_1),\ldots,(x_j,y_j,z_j),\ldots,(x_i,y_i,z_i),\ldots,(x_N,y_N,z_N)] \tag{41}$$

for all $i \neq j$. Moreover, by assumption (ii) on the crystallographic orientation distributions,

$$\mathsf{K}_N^*[(x_1,y_1,z_1),\ldots,(x_i,y_i,z_i),\ldots,(x_N,y_N,z_N)] = \mathsf{K}_N^*[(x_1,y_1,z_1),\ldots,(y_i,x_i,z_i),\ldots,(x_N,y_N,z_N)], = \mathsf{K}_N^*[(x_1,y_1,z_1),\ldots,(x_i,z_i,y_i),\ldots,(x_N,y_N,z_N)], \quad \text{etc.} \tag{42}$$

The effective conductivity of the original single-phase polycrystal satisfies

$$\mathsf{K}^*[(k_1,k_2,k_3)] = \mathsf{K}_N^*[(k_1,k_2,k_3),\ldots,(k_1,k_2,k_3),\ldots,(k_1,k_2,k_3)]. \tag{43}$$

We shall assume first, for simplicity, that there are finitely many grains per unit volume. This assumption is removed in Appendix A. Under this condition, each of the N phases is dilute.[10] Hence

$$\mathsf{K}_N^*[(x_1,y_1,z_1),(1,1,1),\ldots,(1,1,1)] = \mathsf{I} + (1/N)\mathscr{L}(x_1,y_1,z_1) + o(1/N). \tag{44}$$

Equation (44) together with (43), provide a useful mathematical relation between the average polarizability and the conductivity of the polycrystal, which we exploit in the next sections.

V. EXPRESSION FOR *A* IN TERMS OF $\partial^2\mathscr{L}$

We set

$$\xi_1 = \frac{\partial^2}{\partial x_1^2}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)], \tag{45}$$

$$\xi_2 = \frac{\partial^2}{\partial x_1\,\partial y_1} \mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)], \tag{46}$$

and

$$\xi_3 = \frac{\partial^2}{\partial x_1\,\partial x_2} \mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)], \tag{47}$$

Differentiating (43) twice at $k_i = 1$, we obtain

$$\mathsf{A} = \frac{\partial^2}{\partial k_1^2} \mathsf{K}^*(1,1,1) = N\xi_1 + N(N-1)\xi_3. \tag{48}$$

Three independent linear relations between ξ_1, ξ_2, ξ_3 and the second derivatives of $\mathscr{L}$ are obtained by differentiating the equations

$$\mathsf{K}_N^*[(z,1,1),(1,1,1),\ldots,(1,1,1)] = 1 + (1/N)\mathscr{L}(z,1,1) + o(1/N), \tag{49}$$

$$\mathsf{K}_N^*[(z,z,z),(1,1,1),\ldots,(1,1,1)] = 1 + (1/N)\mathscr{L}(z,z,z) + o(1/N), \tag{50}$$

and

$$\mathsf{K}_N^*[(z,z,z),\ldots,(z,z,z)] = z. \tag{51}$$

We obtain, respectively,

$$\xi_1 = \frac{1}{N}\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right),$$

$$3\xi_1 + 6\xi_2 = \frac{3}{N}\left[\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) + 2\frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1)\right] + o\left(\frac{1}{N}\right), \tag{52}$$

and

$$\xi_1 + 2\xi_2 + 3(N-1)\xi_3 = 0,$$

which yield

$$\xi_1 = \frac{1}{N}\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right),$$

$$\xi_2 = \frac{1}{N}\frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right), \tag{53}$$

$$\xi_3 = -\frac{1}{3N(N-1)}\left[\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) + 2\frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1)\right] + o\left(\frac{1}{N^2}\right).$$

Substitution of expressions (53) into (48) and passage to the limit as $N \to \infty$ gives the desired formula (25) for A.

VI. EXPRESSION FOR B IN TERMS OF $\partial^2\mathscr{L}$ AND $\partial^3\mathscr{L}$

Set

$$\eta_1 = \frac{\partial^3}{\partial x_1^3}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)],$$

$$\eta_2 = \frac{\partial^3}{\partial x_1^2\,\partial y_1}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)],$$

$$\eta_3 = \frac{\partial^3}{\partial x_1\,\partial y_1\,\partial z_1}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)], \tag{54}$$

$$\eta_4 = \frac{\partial^3}{\partial x_1^2\,\partial x_2}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)],$$

$$\eta_5 = \frac{\partial^3}{\partial x_1\,\partial y_1\,\partial x_2}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)],$$

and

$$\eta_6 = \frac{\partial^3}{\partial x_1\,\partial x_2\,\partial x_3}\mathsf{K}_N^*[(1,1,1),\ldots,(1,1,1)].$$

An expression for B is obtained by differentiating Eq. (43) three times:

$$\mathsf{B} = \frac{\partial^3}{\partial k_1^3}\mathsf{K}^*(1,1,1) = N\eta_1 + 3N(N-1)\eta_4 + N(N-1)(N-2)\eta_6. \tag{55}$$

Again, we derive a system of linear equations to determine the unknown tensors $\eta_1,\ldots,\eta_6$. For this purpose we will use, in addition to (49)–(51),

$$\mathsf{K}_N^*[z,z,1),\ (1,1,1),\ldots,(1,1,1)] = 1 + (1/N)\mathscr{L}(z,z,1) + o(1/N), \tag{56}$$

$$\mathsf{K}_N^*[(z,z,z),(1,1,1),\ldots,(1,1,1)] = z\mathsf{K}_N^*\left[(1,1,1),\left(\frac{1}{z},\frac{1}{z},\frac{1}{z}\right),\ldots,\left(\frac{1}{z},\frac{1}{z},\frac{1}{z}\right)\right], \tag{57}$$

and

$$\mathsf{K}_N^*[(z,1,1),\ldots,(z,1,1)] = z\mathsf{K}_N^*\left[\left(1,\frac{1}{z},\frac{1}{z}\right),\ldots,\left(1,\frac{1}{z},\frac{1}{z}\right)\right]. \tag{58}$$

Equation (56) is a low volume-fraction expansion, while (57) and (58) reflect the homogeneity of K_N^*. Differentiating (49), (56), (50), (51), (57), and (58), respectively, we obtain, after some algebraic manipulations,

$$\eta_1 = \frac{1}{N}\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right), \tag{59}$$

$$2\eta_1 + 6\eta_2 = \frac{1}{N}\left[2\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) + 6\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1)\right] + o\left(\frac{1}{N}\right), \tag{60}$$

$$3\eta + 18\eta_2 + 6\eta_3 = \frac{1}{N}\left[3\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) + 18\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1) + 6\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\mathscr{L}(1,1,1)\right] + o\left(\frac{1}{N}\right), \tag{61}$$

$$3N\eta_1 + 18N\eta_2 + 6N\eta_3 + 27N(N-1)\eta_4 + 54N(N-1)\eta_5 + 27N(N-1)(N-2)\eta_6 = 0, \tag{62}$$

$$\begin{aligned}3N\eta_1 + 18N\eta_2 + 6N\eta_3 &+ 27(N-1)(N-2)\eta_4 + 54(N-1)(N-2)\eta_5 \\ &+ 27(N-1)(N-2)(N-3)\eta_6 + 9(N-1)\xi_1 + 18(N-1)\xi_2 + 27(N-1)(N-2)\xi_3 = 0,\end{aligned} \tag{63}$$

and

$$\begin{aligned}3N\eta_1 + 6N\eta_2 &+ 15N(N-1)\eta_4 + 12N(N-1)\eta_5 + 9N(N-1)(N-2)\eta_6 \\ &+ 6N\xi_1 + 6N\xi_2 + 12N(N-1)\xi_3 = 0.\end{aligned} \tag{64}$$

Equations (59)–(61) can be solved in terms of $\partial^3\mathscr{L}$ as

$$\begin{aligned}\eta_1 &= \frac{1}{N}\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right), \\ \eta_2 &= \frac{1}{N}\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right), \\ \eta_3 &= \frac{1}{N}\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\mathscr{L}(1,1,1) + o\left(\frac{1}{N}\right).\end{aligned} \tag{65}$$

Equations (62)–(64) can be cast in the form

$$\eta_4 + 2\eta_5 + (N-2)\eta_6 = -[1/9(N-1)](\eta_1 + 6\eta_2 + 2\eta_3), \tag{66}$$

$$\eta_4 + 2\eta_5 + (N-3)\eta_6 = -[N/9(N-1)(N-2)](\eta_1 + 6\eta_2 + 2\eta_3), \tag{67}$$

and

$$\begin{aligned}5\eta_4 + 4\eta_5 + 3(N-2)\eta_6 &= -[1/(N-1)](\eta_1 + 2\eta_2) - [2/(N-1)] \\ &\times[\xi_1 + \xi_2 + 2(N-1)\xi_3],\end{aligned} \tag{68}$$

where the last equation in (52) was used to simplify the rhs of (67). Upon solving (66)–(68) in η_4, η_5, η_6 and using (65), we obtain

$$\begin{aligned}N(N-1)\eta_4 &= -\frac{2}{9}\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) + \frac{2}{9}\frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1) \\ &\quad -\frac{1}{3}\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) - \frac{2}{3}\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1) + o(1),\end{aligned} \tag{69}$$

$$\begin{aligned}N(N-1)\eta_5 &= \frac{1}{9}\frac{\partial^2}{\partial k_1^2}\mathscr{L}(1,1,1) - \frac{1}{9}\frac{\partial^2}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1) \\ &\quad -\frac{2}{3}\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1) - \frac{1}{3}\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\mathscr{L}(1,1,1) + o(1),\end{aligned} \tag{70}$$

and

$$\begin{aligned}N(N-1)(N-2)\eta_6 &= \frac{2}{9}\frac{\partial^3}{\partial k_1^3}\mathscr{L}(1,1,1) + \frac{4}{3}\frac{\partial^3}{\partial k_1^2\,\partial k_2}\mathscr{L}(1,1,1) \\ &\quad + \frac{4}{9}\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\mathscr{L}(1,1,1) + o(1).\end{aligned} \tag{71}$$

The expression for B in terms of $\partial^2\mathscr{L}$ and $\partial^3\mathscr{L}$ is obtained by substituting (65), (69), and (71) into formula (55) and taking the limit as $N \to \infty$.

ACKNOWLEDGMENTS

We are grateful to Professor G. Milton and Dr. J. Berryman for useful discussions which helped improve this paper.

This research was partially supported by a National

Science Foundation Grant No. NSF-DMS-8802739 and by the US Army Research Office through Grants Nos. ARO-DAAL-0389K0039 and ARO-DAAL-0388K0110.

APPENDIX A: REMOVAL OF THE ASSUMPTION OF FINITELY MANY GRAINS PER UNIT VOLUME

For an arbitrary grain G of finite volume, we set

$$T_{\mathbf{e}}^{G}(x) = \mathbf{e}\cdot x + T'(x), \tag{A1}$$

so that

$${}^{t}\mathbf{e}\cdot\mathscr{L}_{G}\cdot\mathbf{e} = \frac{1}{\mathrm{Vol}(G)}\int_{G} {}^{t}\nabla T'(x)\cdot\delta\mathsf{K}_{G}\cdot\mathbf{e} + {}^{t}\mathbf{e}\cdot\delta\mathsf{K}_{G}\cdot\mathbf{e}. \tag{A2}$$

The function $T'(x)$ satisfies

$$\nabla\cdot\mathsf{K}(x)\cdot\nabla T'(x) + \nabla\cdot\chi_{G}(x)\cdot\delta\mathsf{K}_{G}\cdot\mathbf{e} = 0, \tag{A3}$$

where χ_G is the characteristic function of G and $\mathsf{K}(x) = \mathsf{K}_{G}\chi_{G}(x) + \mathsf{I}(1-\chi_{G}(x))$. Multiplying both sides of Eq. (A3) by $T'(x)$ and integrating, we obtain

$$\int_{G} {}^{t}\nabla T'(x)\cdot\delta\mathsf{K}_{G}\cdot\mathbf{e}\,dx = -\int {}^{t}\nabla T'(x)\cdot\mathsf{K}(x)\cdot\nabla T'(x)dx. \tag{A4}$$

By the Cauchy–Schwartz inequality applied to (A4),

$$\left(\int |\nabla T'(x)|^{2}\,dx\right)^{1/2} \leqslant \frac{\mathrm{Vol}(G)^{1/2}}{\min(k_1,1)}|\delta\mathsf{K}_{G}\cdot\mathbf{e}|. \tag{A5}$$

Hence, ${}^{t}\mathbf{e}\cdot\mathscr{L}_{G}\cdot\mathbf{e}$ can be estimated as follows:

$$\begin{aligned}|{}^{t}\mathbf{e}\cdot\mathscr{L}_{G}\cdot\mathbf{e}| &\leqslant \frac{1}{\mathrm{Vol}(G)}\left(\int|\nabla T'(x)|^{2}dx\right)^{1/2} \\ &\quad\cdot\mathrm{Vol}(G)^{1/2}|\delta\mathsf{K}_{G}\mathbf{e}| + {}^{t}\mathbf{e}\cdot\delta\mathsf{K}_{G}\cdot\mathbf{e} \\ &\leqslant \frac{1}{\min(k_1,1)}|\delta\mathsf{K}_{G}\mathbf{e}|^{2} + {}^{t}\mathbf{e}\cdot\delta\mathsf{K}_{G}\cdot\mathbf{e}.\end{aligned} \tag{A6}$$

Equation (A6) implies that for $|\delta k_i| < \frac{1}{2}$, any entry of the tensor $\mathscr{L}_G$ satisfies

$$|\mathscr{L}_{G}^{pq}(k_1,k_2,k_3)| \leqslant 1/4\,\min[k_1,1] + \tfrac{1}{2} \leqslant 1, \quad 1\leqslant p,q\leqslant 3. \tag{A7}$$

Thus *the polarizability tensor of a grain of conductivities* k_i, $1\leqslant i\leqslant 3$ *with* $|\delta k_i| < \frac{1}{2}$ *is uniformly bounded, independent of the shape of the grain.* Since $\mathscr{L}_G$ is an analytic function of k_1, k_2, k_3 [because T' satisfies the elliptic equation (A3)], similar uniform shape-independent bounds hold for all the derivatives of $\mathscr{L}_G$ at $k_i = 1$.

Using this fact, we now remove the assumption that the polycrystal have finitely many grains per unit volume. Given an arbitrary grain configuration, we can construct a modified polycrystal by removing all grains of maximum diameter $\leqslant\epsilon$ (where ϵ is a small number) and replacing the resulting free space by new grains, so that the new aggregate has finitely many grains per unit volume. Note that the volume fraction of material in the original configuration, which has been removed and subsequently filled with new grains, tends to zero as $\epsilon\to 0$. Denote by $\mathscr{L}$ and $\tilde{\mathscr{L}}$ the average polarizability tensors corresponding to the original and modified configurations, respectively. The tensors $\mathscr{L}$, $\tilde{\mathscr{L}}$, and their derivatives are averages of the corresponding single-grain quantities, $\mathscr{L}_G$, and their derivatives over the grain shape and crystallographic orientation distributions. We conclude from the above analysis that

$$\begin{aligned}&\max_{|\alpha|<3}\|D_{k}^{\alpha}\mathscr{L}(1,1,1) - D_{k}^{\alpha}\tilde{\mathscr{L}}(1,1,1)\| \\ &\quad\leqslant C_1\times(\text{volume removed}),\end{aligned} \tag{A8}$$

where C_1 is a numerical constant.

On the other hand, the conductivity tensor $\mathsf{K}^{*}(k_1,k_2,k_3)$ is known to be analytic in k_1, k_2, k_3 at $k_i = 1$. Also, denoting the conductivity of the modified configuration by $\tilde{\mathsf{K}}^{*}$, it is known that

$$\|\mathsf{K}^{*} - \tilde{\mathsf{K}}^{*}\| \leqslant C_2(\text{volume removed})^{\nu}, \tag{A9}$$

where C_2 and ν are positive constants which depend on $\max_i k_i$ and $\min_i k_i$.[17] From this estimate and the analyticity of K^{*} and $\tilde{\mathsf{K}}^{*}$, we conclude that

$$\begin{aligned}&\max_{|\alpha|<3}\|D_{k}^{\alpha}\mathsf{K}^{*}(1,1,1) - D_{k}^{\alpha}\tilde{\mathsf{K}}^{*}(1,1,1)\| \\ &\quad\leqslant C_3\times(\text{volume removed})^{\nu},\end{aligned} \tag{A10}$$

where C_3 is a numerical constant. Since the modified configuration has only finitely many grains per unit volume, the basic identies (25) and (26) relating $\partial^{2}\tilde{\mathsf{K}}^{*}$ and $\partial^{3}\tilde{\mathsf{K}}^{*}$ to the derivatives of $\tilde{\mathscr{L}}$ are now valid. To extend such identities to the conductivity and polarizability of the original configuration, we use the estimates (A8) and (A10). Accordingly,

$$\begin{aligned}&\left|\mathsf{A} - \frac{2}{3}\left[\frac{\partial^{2}}{\partial k_1^{2}}\mathscr{L}(1,1,1) - \frac{\partial^{2}}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1)\right]\right| \\ &\quad\leqslant|\mathsf{A}-\tilde{\mathsf{A}}| + \frac{2}{3}\left|\frac{\partial^{2}}{\partial k_1^{2}}\mathscr{L}(1,1,1) - \frac{\partial^{2}}{\partial k_1^{2}}\tilde{\mathscr{L}}(1,1,1)\right| \\ &\qquad+\frac{2}{3}\left|\frac{\partial^{2}}{\partial k_1\,\partial k_2}\mathscr{L}(1,1,1) - \frac{\partial^{2}}{\partial k_1\,\partial k_2}\tilde{\mathscr{L}}(1,1,1)\right| \\ &\quad\leqslant C_3\,(\text{volume removed})^{\nu} + \frac{4C_1}{3}(\text{volume removed}).\end{aligned} \tag{A11}$$

Letting ϵ tend to zero, the volume removed tends to zero and we recover identity (25) for an arbitrary configuration. The approximation argument for B is entirely similar.

APPENDIX B: CALCULATION OF TRACE A AND APPLICATION TO ISOTROPIC POLYCRYSTALS

The partial derivatives of the single-grain polarizability $\mathscr{L}_G(k_1,k_2,k_3)$ of an arbitrary bounded grain G can be computed formally by differentiating both sides of (A2). This involves the differentiation of the function $T'(x)$ in the parameters k_i at $k_i = 1$, $1\leqslant i\leqslant 3$. In this appendix we show how such a calculation leads to the fact that tr A is independent of the grain shape and orientation distributions. We set

$$T_{i}(x) = \frac{\partial}{\partial k_i}\bigg|_{\mathbf{k}=(1,1,1)} T'(x) \tag{B1}$$

and note that differentiating Eq. (A3) once at $k_i = 1$, this function satisfies

$$\Delta T_i + ({}^t\mathbf{n}_i \cdot \mathbf{e}) \nabla \cdot \chi_G \mathbf{n}_i = 0, \quad 1 < i \leqslant 3, \tag{B2}$$

where $\mathbf{n}$ is a unit vector in the direction of the principal conductivity k_i. On the other hand, differentiating (A2) twice at $k_i = 1$, one obtains

$$\begin{aligned} {}^t\mathbf{e} \cdot \frac{\partial^2}{\partial k_i \, \partial k_j} \mathscr{L}(1,1,1) \cdot \mathbf{e} \\ = \frac{1}{\mathrm{Vol}(G)} \int_G ({}^t\nabla T_i(x) \cdot \mathbf{n}_j)({}^t\mathbf{n}_j \cdot \mathbf{e}) \\ + \frac{1}{\mathrm{Vol}(G)} \int_G ({}^t\nabla T_j(x) \cdot \mathbf{n}_i)({}^t\mathbf{n}_i \cdot \mathbf{e}). \end{aligned} \tag{B3}$$

We deduce from (B2) that formally,

$$\nabla T_i(x) = \int_{\mathbf{R}^3} \Gamma(x-y) \cdot \mathbf{n}_i ({}^t\mathbf{n}_i \cdot \mathbf{e}) \chi_G(y) dy, \tag{B4}$$

where

$$\Gamma_{pq}(x) = \frac{\partial^2}{\partial x_p \, \partial x_q} \frac{1}{4\pi|x|} = \frac{1}{4\pi|x|^3} \left(\frac{3x_i x_j}{|x|^2} - \delta_{ij} \right)$$

is the kernel of the integral operator $(\partial^2/\partial x^p \, \partial x^q)(-\Delta)^{-1}$. The Fourier transform of $\Gamma(x)$ is

$$\hat{\Gamma}_{pq}(\xi) = \int_{\mathbf{R}^3} \Gamma_{pq}(x) e^{ix\xi} \, dx = -\frac{\xi_p \xi_q}{|\xi|^2}.$$

From (B4), we conclude that the Fourier transform of $\nabla T_i(x)$ is

$$\nabla T_i(\xi) = -\left[({}^t\mathbf{n}_i \cdot \mathbf{e})({}^t\xi \cdot n_i) \hat{\chi}_G(\xi)/|\xi|^2 \right] \xi. \tag{B5}$$

Substituting expression (B5) into (B3) and using the Plancherel identity, we obtain

$$\begin{aligned} {}^t\mathbf{e} \cdot \frac{\partial^2}{\partial k_i \, \partial k_j} \mathscr{L}_G(1,1,1) \cdot \mathbf{e} \\ = -\frac{2}{\mathrm{Vol}(G)} \int_{\mathbf{R}^3} \frac{({}^t\mathbf{n}_i \cdot \mathbf{e})({}^t\mathbf{n}_j \cdot \mathbf{e})({}^t\mathbf{n}_i \cdot \xi)({}^t\mathbf{n}_j \cdot \xi)}{|\xi|^2} \\ \times |\hat{\chi}_G(\xi)|^2 \, d\xi. \end{aligned} \tag{B6}$$

The trace of $\partial^2/\partial k_i \, \partial k_j \, \mathscr{L}_G(1,1,1)$ is computed by varying $\mathbf{e}$ over an orthonormal basis in (B6) and summing. Accordingly,

$$\mathrm{tr} \frac{\partial^2}{\partial k_i \, \partial k_j} \mathscr{L}_G(1,1,1) = 0, \quad \text{if } i \neq j$$

and

$$\begin{aligned} \mathrm{tr} \frac{\partial^2}{\partial k_i^2} \mathscr{L}_G(1,1,1) \\ = -\frac{2}{3 \, \mathrm{Vol}(G)} \int_{\mathbf{R}^3} \frac{({}^t n_i \cdot \xi)^2}{|\xi|^2} |\hat{\chi}_G(\xi)|^2 \, d\xi. \end{aligned} \tag{B7}$$

Since the three principal conductivities k_1, k_2, k_3 are interchangeable,

$$\langle ({}^t n_i \cdot \xi)^2 / |\xi|^2 \rangle = \tfrac{1}{3}, \quad \xi \in \mathbb{R}^3, \quad 1 \leqslant i \leqslant 3,$$

where the brackets indicate averaging over the orientation distribution of the conductivity in grain G. Hence, the average polarizability tensor $\mathscr{L}$ satisfies

$$\begin{aligned} \mathrm{tr} \frac{\partial^2}{\partial k_i^2} \mathscr{L}(1,1,1) &= -\left\langle \frac{2}{3 \, \mathrm{Vol}(G)} \int_{\mathbf{R}^3} |\hat{\chi}_G(\xi)|^2 \, d\xi \right\rangle \\ &= -\left\langle \frac{2}{3} \frac{1}{\mathrm{Vol}(G)} \cdot \mathrm{Vol}(G) \right\rangle \\ &= -\frac{2}{3}. \end{aligned}$$

From the basic formula for (25) (for A) we conclude that

$$\mathrm{tr}\, \mathsf{A} = -\tfrac{4}{3}.$$

Substituting this result in the Taylor expansion (22) [for $\mathsf{K}^*(k_1,k_2,k_3)$], we obtain

$$\frac{1}{3} \mathrm{tr}\, \mathsf{K}^*(k_1,k_2,k_3) = k_V \left[1 - \frac{1}{9} \sum_{i=1}^{3} \left(\frac{k_i}{k_V} - 1 \right)^2 \right] + o(\delta k_i^2).$$

In particular, we recover (11) for isotropic configurations. A similar calculation, which we omit, yields an expression for the trace of B in terms of multiple integrals depending on the average grain shape, which allows us to recover the third-order expansion for isotropic polycrystals of Willemse and Caspers.[13]

APPENDIX C: CALCULATIONS FOR AGGREGATES OF ELLIPSOIDAL GRAINS

We present an outline of the calculation of the tensors of the microgeometric parameters A and B for anisotropic aggregates of ellipsoids. It is convenient to consider first a single ellipsoid G with the depolarization factors L_1, L_2, L_3 and the uniaxial or isotropic conductivity tensor K_G. We shall consider three different cases: (i) $k_1 = z$, $k_2 = k_3 = 1$; (ii) $k_1 = k_2 = z$, $k_3 = 1$; and (iii) $k_1 = k_2 = k_3 = z$.

Case (i): $k_1 = 2$, $k_2 = k_3 = 1$.

Denoting by $\mathbf{n}$ a unit vector in the axial (z) direction, we have

$$\mathsf{K}_G = I + (z-1)\mathbf{n} \otimes \mathbf{n}, \tag{C1}$$

where $(\mathbf{n} \otimes \mathbf{n})_{ij} = n_i n_j$. We expand the corresponding single-grain polarizability $\mathscr{L}_G$ given by (28) in powers of $(z-1)$ to obtain

$$\frac{d^2}{dz^2} \mathscr{L}_G(z,1,1)|_{z=1} = -2 \left(\sum_{i=1}^{3} L_i n_i^2 \right) \mathbf{n} \otimes \mathbf{n} \tag{C2}$$

and

$$\frac{d^3}{dz^3} \mathscr{L}_G(z,1,1)|_{z=1} = 6 \left(\sum_{i=1}^{3} L_i n_i^2 \right)^2 \mathbf{n} \otimes \mathbf{n}, \tag{C3}$$

where n_i are the coordinates of $\mathbf{n}$ in a frame of principal axes of G. The second and third derivatives of the average polarizability $\mathscr{L}$ of an array of ellipsoidal grains with semiaxes parallel to those of G are obtained from (C2) and (C3) by averaging over the distributions of $\mathbf{n}$ and L_1, L_2, L_3. From (C1) the distribution of $\mathbf{n}$ should be invariant under the reflections $\mathbf{n} \to (-\mathbf{n})$. In particular, odd moments of n_i, $1 \leqslant i \leqslant 3$ vanish. We find that in the above reference frame,

$$a \equiv \frac{d^2}{dz^2}\,\mathscr{L}^{pp}_G(z,1,1)|_{z=1} = -2\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right) n_p^2\right\rangle \tag{C4}$$

and

$$c \equiv \frac{d^3}{dz^3}\,\mathscr{L}^{pp}_G(z,1,1)|_{z=1} = 6\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right)^2 n_p^2\right\rangle, \tag{C5}$$

while the off-diagonal entries $\partial^2 \mathscr{L}^{pq}$, $\partial^3 \mathscr{L}^{pq}$ with $p \neq q$ are equal to zero.

Case (ii): $k_1 = k_2 = z$, $k_3 = 1$.

The conductivity within the grain is

$$\mathsf{K}_G = \mathsf{I} + (z-1)(\mathsf{I} - \mathbf{n} \otimes \mathbf{n}), \tag{C6}$$

where $\mathbf{n}$ is a unit vector in the axial direction. Substituting (C6) into (28) (with $\mathsf{O} = \mathsf{I}$) and expanding in powers of $(z-1)$, we obtain formulas for $\partial^2/\mathscr{L}_G$ and $\partial^3/\mathscr{L}_G$, which, as in case (i), can be averaged over the distributions of $\mathbf{n}$ and $\mathbf{L}$. In this way, we obtain the following formulas for the second and third derivatives of the average polarizability of an array of ellipsoids with parallel principal axes:

$$b \equiv \frac{d^2}{dz^2}\,\mathscr{L}^{pp}(z,z,1)|_{z=1}$$

$$= -\frac{2}{3}\langle L_p\rangle - 2\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right) n_p^2\right\rangle \tag{C7}$$

and

$$d \equiv \frac{d^3}{dz^3}\,\mathscr{L}^{pp}(z,z,1)|_{z=1} = 6\left\langle\left(\sum_{i=1}^{3} L_i^2 n_i^2\right) n_p^2\right\rangle$$

$$+ 12\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right) L_p n_p^2\right\rangle - 6\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right)^2 n_p^2\right\rangle. \tag{C8}$$

Case (iii): $k_1 = k_2 = k_3 = z$.

In this case $\mathsf{K}_G = z\mathsf{I}$ and using (28) and averaging, we obtain immediately

$$\frac{d^2}{dz^2}\,\mathscr{L}^{pp}(z,z,z)|_{z=1} = -2\langle L_p\rangle \tag{C9}$$

and

$$e \equiv \frac{d^3}{dz^3}\,\mathscr{L}^{pp}(z,z,z)|_{z=1} = 6\langle L_p^2\rangle. \tag{C10}$$

We use formulas (C4) through (C10) to compute the partial derivatives of $\mathscr{L}(k_1,k_2,k_3)$ at $k_i = 1$. Accordingly,

$$\frac{\partial^2}{\partial k_1^2}\,\mathscr{L}^{pp}(1,1,1) = a,$$

$$\frac{\partial^2}{\partial k_1\,\partial k_2}\,\mathscr{L}^{pp}(1,1,1) = \frac{1}{2}(b-2a),$$

$$\frac{\partial^3}{\partial k_1^3}\,\mathscr{L}^{pp}(1,1,1) = c,$$

$$\frac{\partial^3}{\partial k_1^2\,\partial k_2}\,\mathscr{L}^{pp}(1,1,1) = \frac{1}{6}(d-2c),$$

and

$$\frac{\partial^3}{\partial k_1\,\partial k_2\,\partial k_3}\,\mathscr{L}^{pp}(1,1,1) = \frac{1}{6}(e+3c-3d). \tag{C11}$$

We substitute expressions (C11) into the basic formulas (25) and (26) to obtain

$$\mathsf{A}^{pp} = \frac{4}{3}a - \frac{1}{3}b = -2\left\langle\left(\sum_i L_i n_i^2\right) n_p^2\right\rangle + \frac{2}{9}\langle L_p\rangle \tag{C12}$$

and

$$\mathsf{B}^{pp} = \frac{2}{3}c - \frac{2}{27}e - \mathsf{A}^{pp} = 6\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right)^2 n_p^2\right\rangle$$

$$- 4\left\langle\left(\sum_{i=1}^{3} L_i n_i^2\right) L_p N_p^2\right\rangle + \frac{4}{9}\langle L_p^2\rangle - \frac{2}{9}\langle L_p\rangle. \tag{C13}$$

Formulas (C12) and (C13) give the values of the eigenvalues of the average polarizability of an array of ellipsoids with parallel principal axes. The expressions for A and B for general configurations of ellipsoids can be obtained by replacing in (C12) and (C13) the tensor L by $^t\mathsf{OLO}$ and averaging over the orientation field O. We restrict our attention to the case of uniformly distributed crystallographic orientations within each grain, which corresponds to $\mathbf{n}$ being uniformly distributed on the unit sphere in three-space. The relevant moments of the vector $\mathbf{n}$ are

$$\langle n_i^2\rangle = \tfrac{1}{3}, \quad \langle n_i^4\rangle = \tfrac{1}{5}, \quad \langle n_i^6\rangle = \tfrac{1}{7} \tag{C14}$$

for $1 \leqslant i \leqslant 3$,

$$\langle n_i^2 n_j^2\rangle = \tfrac{1}{15}, \quad \langle n_i^2 n_j^4\rangle = \tfrac{1}{35} \tag{C15}$$

for $i \neq j$, and

$$\langle n_1^2 n_2^2 n_3^2\rangle = \tfrac{1}{105}. \tag{C16}$$

Substituting these values into (C12) and (C13), we obtain

$$\mathsf{A}^{pp} = -\tfrac{2}{45}\langle L_p\rangle - \tfrac{2}{15} \tag{C17}$$

$$\mathsf{B}^{pp} = \tfrac{32}{315}\langle L_p^2\rangle - \tfrac{2}{105}\langle L_1^2 + L_2^2 + L_3^2\rangle + \tfrac{2}{315}\langle L_p\rangle - \tfrac{4}{21}, \tag{C18}$$

and $\mathsf{A}^{pq} = \mathsf{B}^{pq} = 0$ for $p \neq q$. Replacing L by $^t\mathsf{OLO}$ and averaging over the grain orientation distribution yields expressions (32) and (33).

[1] Y. S. Touloukian, R. W. Powell, C. Y. Ho, and P. G. Klemens, *Thermal Conductivity, I. Metallic Elements and Alloys; Thermal Conductivity II. Non-Metallic Solids, Thermodynamical Properties of Matter* (IFI/Plenum, New York, 1970).

[2] W. Voigt, *Lehrbuch der Kristallphysik* (Teubner, Leipzig, 1928), p. 410.

[3] A. Reuss, Z. Angew. Math. Mech. **9**, 49 (1929).

[4] O. Wiener, Abhandl, Math.-Phys. Kl. Königl. Sachsischen Ges. **32**, 509 (1912).

[5] M. Avellaneda, A. V. Cherkaev, K. A. Lurie, and G. W. Milton, J. Appl. Phys. **63**, 4989 (1988).

[6] J. E. Molyneux, J. Math. Phys. **10**, 912 (1969).

[7] J. E. Molyneux, J. Math. Phys. **11**, 1172 (1970).

[8] P. B. Corson, J. Appl. Phys. **45**, 3159 (1974).

[9] J. G. Berryman, J. Appl. Phys. **57**, 2374 (1985).

[10] M. N. Miller, J. Math. Phys. **10**, 1988 (1969).

[11]Z. Hashin and S. Shtrikman, Phys. Rev. **130**, 129 (1963).
[12]L. J. Walpole, J. Mech. Phys. Solids **14**, 189 (1966).
[13]M. W. M. Willemse and W. J. Caspers, J. Math. Phys. **20**, 1824 (1979).
[14]O. Bruno, "The effective conductivity of an infinitely interchangeable mixture", Ph. D. thesis, New York University, 1989.
[15]M. A. Elsayed, J. Math. Phys. **15**, 2001 (1974).
[16]J. C. Maxwell, *Electricity and Magnetism* (Clarendon, Oxford, 1873), 1st ed.; (see, for instance, J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York (1941), Chap. III.
[17]This is a consequence of the work in N. G. Meyers and A. Elcrat, Duke Math. J. **42**, 121 (1975), on the L^p integrability ($p>2$) of gradients of solutions to elliptic equations. Considerations leading to estimate (80) can be found in M. Avellaneda, Commun. Pure Appl. Math. **40**, 527 (1987), Proposition 8.