Additive isometries on a quaternionic Hilbert space

C. S. Sharma and D. F. Almeida
Mathematical Research Unit, University of London, Birkbeck College, 43 Gordon Square, London, WC1H 0PD, England

(Received 26 July 1989; accepted for publication 20 December 1989)

A systematic study of additive isometries on a quaternionic Hilbert space is presented. A number of new results describing the properties of such operators are proved. The work culminates in the first mathematical proof of Wigner's theorem for quaternionic Hilbert spaces of dimension other than 2 which asserts that any operator which preserves the absolute value of the inner product on a quaternionic Hilbert space is equivalent, in the sense of differing pointwise by a mere phase factor, to a linear isometry. A complete and concise description of the exceptional situation in a two-dimensional quaternionic Hilbert space is given.

I. INTRODUCTION

This work continues the fundamental study of additive operators on a quaternionic Hilbert space undertaken in our earlier works.[1-3] The relevance of quaternionic vector spaces in celestial mechanics was recently demonstrated by Vivarelli,[4] in quantum mechanics by Horwitz and Biedenharn[5] and Adler[6] and in relativity by Rocher[7] and Sharma.[8] Homomorphisms on any linear space are, of course, linear, but because of the noncommutativity of quaternions, linear operators on a quaternionic Hilbert space do not form an algebra although they do form a ring. In many quantum mechanical applications it is desirable to have an algebra of operators and it was shown in Ref. 3 that the smallest algebra of operators which contains the ring of bounded linear operators is the algebra of bounded additive operators. It is for this reason that additivity takes the central role in the study of operators on a quaternionic space: Even in complex space study of additivity is fruitful as has been demonstrated in numerous works (see Ref. 2 where further references will be found). Isometries are among the simplest operators and, therefore, study of additive isometries comes to the forefront as soon as the basic properties of additive operators have been defined.[3] We prove a number of new results on additive isometries but our final result—Wigner's theorem for quaternionic Hilbert spaces—is not new. It was correctly stated by Bargmann,[9] who also gave a "proof," which was a remarkable achievement in view of the very limited number of tools available to him. Our own proof is a simple generalization of our proof[10] for the complex case which in the opionion of the referee of that work was the first "mathematical" proof of the theorem: We hope the same adjective applies to our proof in the present case.

In Sec. II we establish our notation and state with indications of proofs several elementary lemmas which we use in the later sections. In Sec. III we state and prove all our main results the last of which is Wigner's theorem. In Sec. IV we give a concise and complete description of the exceptional situation in the two-dimensional case and make a few concluding remarks indicating the direction in which our work will proceed next.

II. FORMALITIES

We denote the fields of real and complex numbers by $\mathbb{R}$ and $\mathbb{C}$, respectively, and the skew field of quaternionic numbers by $\mathbb{H}$. Elementary properties of quaternions are described in Ref. 1. We state briefly the properties we are going to need in this work. Quaternions form a normed associative division algebra over $\mathbb{R}$ and are best described with the help of three distinct linearly independent abstract square roots of -1, which are denoted by symbols i,j, and k and whose products are defined by

$$i^2 = j^2 = k^2 = -1, \tag{2.1}$$

$$ij = -ji = k, \tag{2.2}$$

$$jk = -kj = i, \tag{2.3}$$

$$ki = -ik = j. \tag{2.4}$$

It is easy to verify that $\mathbb{H}$ is a four-dimensional vector space over $\mathbb{R}$ where $1, i, j$, and k are members of a basis. Thus any $\gamma \in \mathbb{H}$ has a unique representation as

$$\gamma = \gamma_0 + \gamma_1 i + \gamma_2 j + \gamma_3 k, \tag{2.5}$$

with $\gamma_0, \gamma_1, \gamma_2, \gamma_3 \in \mathbb{R}$. Quaternionic conjugation is defined by

$$1^* = 1, \quad i^* = -i, \quad j^* = -j, \quad k^* = -k. \tag{2.6}$$

It is easy to verify that $\mathbb{H}$ is a normed algebra with the norm defined by

$$\|\gamma\| = (\gamma^*\gamma)^{1/2}. \tag{2.7}$$

In addition to the axioms of the norm, the norm satisfies, as in the complex case,

$$\|\gamma_1\gamma_2\| = \|\gamma_1\|\|\gamma_2\|. \tag{2.8}$$

In the context of a quaternionic Hilbert space, to avoid confusion we shall refer to the norms of quaternions as moduli though we shall continue to use the same notation for the norm.

Some simple properties of quaternionic numbers that we need are collected together as the following lemma.

 0022-2488/90/051035-07$03.00

Lemma 2.1: Let $\gamma\in\mathbb{H}$. Then

(i) $\gamma = 0 \Leftrightarrow \gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = 0$,

where $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are as in (2.5).

(ii) If the imaginary square root of -1 in $\mathbb{C}$ is identified with the quaternionic i then γ has a unique representation as

$$\gamma = c_0 + c_1 j, \tag{2.9}$$

with $c_0, c_1\in\mathbb{C}$ and further

$$i\gamma i^* = -i\gamma i = c_0 - c_1 j. \tag{2.10}$$

Furthermore, in the preceding assertion i can be replaced by j (resp. k) and j by k (resp. i).

$$\text{(iii)} \sum_{\tau = i,j,k} \tau\gamma\tau = -\gamma - 2\gamma^*. \tag{2.11}$$

(iv) Let ρ be any other quaternion, then

$$\|1 + \gamma\rho\| = \|1 + \rho\gamma\|. \tag{2.12}$$

Proof: (i) Follows from the fact that $1, i, j, k$ form a basis in $\mathbb{H}$ as a vector space over $\mathbb{R}$ and (ii) and (iii) follow as a result of straightforward calculations using Eqs. (2.1)–(2.6). If either ρ or γ is zero the validity of (iv) is self-evident but if neither is zero, (iv) follows from the following simple calculation:

$$\begin{aligned}\|1 + \gamma\rho\| &= \|\gamma\|\|\gamma^{-1} + \rho\| = \|\gamma\|\|1 + \rho\gamma\|\|\gamma^{-1}\| \\ &= \|1 + \rho\gamma\|. \end{aligned}\tag{2.13}$$

□

Let $\mathscr{H}$ be a vector space over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. We define a positive definite Hermitian form on $\mathscr{H}$ by

$$\langle,\rangle : \mathscr{H}\times\mathscr{H} \to \mathbb{F},$$

$$\langle pu, qv\rangle = p\langle u,v\rangle q^*, \tag{2.14}$$

$$\langle u + v, w\rangle = \langle u,w\rangle + \langle v,w\rangle, \tag{2.15}$$

$$\langle u,v\rangle^* = \langle v,u\rangle, \tag{2.16}$$

$$\langle u,u\rangle = 0 \text{ only if } u = 0, \tag{2.17}$$

where $p^* = p$ if $\mathbb{F}$ is real, $p^* =$ *complex conjugate* of p if $\mathbb{F}$ is complex, and $p^* =$ *quaternionic conjugate* of p if $\mathbb{F}$ is quaternionic.

Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be Hilbert spaces over $\mathbb{F}$. We say that a map $A : \mathscr{H}_1 \to \mathscr{H}_2$ is *additive* if and only if for all $u, v\in\mathscr{H}_1$

$$A(u + v) = A(u) + A(v). \tag{2.18}$$

If, in addition, the map A satisfies

$$A(pu) = pA(u) \tag{2.19}$$

for all $p\in\mathbb{F}$ and all $u\in\mathscr{H}_1$, then it is called *linear*.

If, on the other hand, A satisfies

$$A(pu) = p^*A(u), \tag{2.20}$$

for all $p\in\mathbb{F}$ and all $u\in\mathscr{H}_1$, then it is called *semilinear*. It was shown in Ref. 2 that coordinate-free semilinear maps according to this particular definition do not exist for the quaternionic case. In a quaternionic Hilbert space it is necessary to define three different kinds of semilinearities called *i-*, *j-*, and *k-semilinearity* thus: A i-semilinear map A from $\mathscr{H}_1$ to $\mathscr{H}_2$ is an additive map that satisfies

$$A(ru) = rAu, \quad A(iu) = iAu, \quad A(ju) = -jAu,$$

$$A(ku) = -kAu, \tag{2.21}$$

for all $r\in\mathbb{R}$ and all $u\in\mathscr{H}_1$ and j-, and k-semilinearities have analogous definitions. It was shown by Coulson[11] that any additive map A can be written as a sum of four maps A_0, A_1, A_2, and A_3, where A_0 is linear and A_1, A_2, and A_3 are, respectively, i-, j-, and k-semilinear and that this decomposition is unique. In this decomposition A_0, A_1, A_2, and A_3 are given by

$$A_0u = \tfrac{1}{4}[Au - iA(iu) - jA(ju) - kA(ku)], \tag{2.22a}$$

$$A_1u = \tfrac{1}{4}[Au - iA(iu) + jA(ju) + kA(ku)], \tag{2.22b}$$

$$A_2u = \tfrac{1}{4}[Au + iA(iu) - jA(ju) + kA(ku)], \tag{2.22c}$$

$$A_3u = \tfrac{1}{4}[Au + iA(iu) + jA(ju) - kA(ku)]. \tag{2.22d}$$

Here we have a slight generalization of a similar decomposition given in Ref. 5.

Let p be a quaternion of unit modulus. An additive map A from a quaternionic Hilbert space $\mathscr{H}_1$ to another quaternionic Hilbert space $\mathscr{H}_2$ is said to be p-semilinear if it satisfies for every $u\in\mathscr{H}_1$ and every $\alpha\in\mathbb{H}$

$$A(\alpha u) = p^*\alpha pAu. \tag{2.23}$$

It is evident that the various semilinearities defined above are special cases of this general definition.

As was correctly shown in Ref. 2, a coordinate-free definition implied in (2.20) leads to a contradiction, but given a basis $\{u_i\}$, not necessarily orthonormal, we can define a semilinear operator A analogous to that implied by (2.20) by

$$A\left(\sum_i \alpha_i u_i\right) = \sum_i \alpha_i^* A(u_i). \tag{2.24}$$

Coordinate-dependent objects usually do not play a fundamental role in either the mathematical or the physical development of a theory, though, there are, of course, important exceptions the most familiar of which is the K operator of Wigner[12] (though, of course, even K is coordinate independent in the Hilbert space of square integrable complex-valued functions on an Euclidean manifold). Furthermore the correspondence $\alpha \mapsto \alpha^*$ in the quaternionic case is an anti-isomorphism (anti because it is product reversing) rather than an isomorphism: for this reason we call the operator defined in (2.24) an *antilinear operator*. Again we can have coordinate-dependent definitions of i-, j-, and k-antilinear operators similar to those defined by (2.21) and analogous relations defined relative to a basis and more generally for a quaternion p of unit modulus a *p-antilinear operator* is an additive operator that satisfies

$$A\left(\sum_i \alpha_i u_i\right) = \sum_i p^*\alpha_i^* pA(u_i). \tag{2.25}$$

It was proved by Pian and Sharma[13] that in the complex case every additive continuous map from $\mathscr{H}_1$ to $\mathscr{H}_2$ is a direct sum of a linear and a semilinear continuous map from $\mathscr{H}_1$ to $\mathscr{H}_2$ (Ref. 13 was in the more general context of Banach spaces of which Hilbert spaces are particular cases). The corresponding decomposition in the present case is given by Eqs. (22a)–(22d).

A map from a set $\mathscr{S}$ to itself will be called an operator on $\mathscr{S}$.

Let A be an operator or a matrix. A is said to be *involutive*[14] if and only if

$$A^2 = I. \tag{2.26}$$

Let $\mathscr{A}$ be an algebra. An *involution* $*$ on $\mathscr{A}$ is an involutive operator on $\mathscr{A}$ which takes A to A^* and satisfies the following properties.

(i) The operator $*$ is a homomorphism of the additive group in the algebra, that is,

$$(A+B)^* = A^* + B^*, \tag{2.27}$$

for all $A, B \in \mathscr{A}$,

(ii) The operator $*$ is product reversing, that is,

$$(AB)^* = B^* A^*, \tag{2.28}$$

and being involutive, of course, means that it satisfies

$$A^{**} = A, \tag{2.29}$$

for all $A \in \mathscr{A}$. This definition is a generalization of the one given by Rudin.[15] Here, unlike Rudin, we do not require $*$ to be semilinear but we require it to be additive.

It is now known[3] that if A is a bounded p-semilinear operator on a quaternionic Hilbert space $\mathscr{H}$, then its adjoint A^* exists, is p^*-semilinear and satisfies

$$\langle Ax, y\rangle = p^* \langle x, A^* y\rangle p, \tag{2.30}$$

and this includes the case when A is linear which corresponds to $p = 1$.

Let $\mathscr{H}$ be a Hilbert space over any field. Two operators A and B on $\mathscr{H}$ are said to be Wigner *equivalent*[10] if their images at each point in $\mathscr{H}$ differ merely by a phase factor, that is, for each $x \in \mathscr{H}$, $Ax = \alpha Bx$, where α has unit norm. It is trivially easy to verify that Wigner equivalence is an equivalence relation.

We now state some basic lemmas with brief indication of their proofs which we shall need in the later sections.

Lemma 2.2: Every inner-product preserving map on a Hilbert space is necessarily linear. (Note that such a map need not be surjective.)

Proof: Let A be an inner-product preserving map on a Hilbert space. The lemma follows by a simple verification of

$$\|A(\alpha x + \beta y) - \alpha Ax - \beta Ay\| = 0, \tag{2.31}$$

where x and y are any pair of vectors and α and β are any pair of scalars. □

Lemma 2.3: The real algebra of quaternions is isomorphic with the real algebra generated by complex 2×2 matrices with entries α and $\bar{\alpha}$ as the diagonal elements and β and $-\bar{\beta}$ as the off-diagonal elements where α and β are any two complex numbers.

Proof: See Ref. 1. (An easy way of obtaining this result is to identify the quaternionic i, j, and k with $i_c\sigma_z$, $i_c\sigma_y$, and $i_c\sigma_x$, respectively, where the σ's are Pauli matrices and i_c is the complex squareroot of -1 and is identified with the quaternionic i.) □

Lemma 2.4: The group of quaternions of unit modulus [called Sp(1)] is isomorphic with SU(2).

Proof: An elementary consequence of the identification made in Lemma 2.3. □

Lemma 2.5: The algebra SO(3) is isomorphic with $SU(2)/\mathscr{C}SU(2)$, where $\mathscr{C}SU(2)$ is the center of SU(2).

Proof: It follows by verifying that the transformation

$$\begin{bmatrix} -z' & x'+iy' \\ x'-iy' & z' \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{bmatrix} \begin{bmatrix} -z & x+iy \\ x-iy & z \end{bmatrix} \begin{bmatrix} \bar{\alpha} & -\beta \\ \bar{\beta} & \alpha \end{bmatrix} \tag{2.32}$$

corresponds to the rotation of Euclidean coordinates (x,y,z) in $\mathbb{R}^3$ by Euler angles (θ,ϕ,ψ) as defined in Ref. 12 if

$$\alpha = e^{-i(\theta+\psi)/2}\cos(\phi/2) \tag{2.33}$$

and

$$\beta = -e^{-i(\theta-\psi)/2}\sin(\phi/2), \tag{2.34}$$

and that the correspondence is a bijection if the transforming unitary matrices and their negatives (obtained by reversing the sign of each entry) are regarded to be the same in the sense which is equivalent to taking the quotient by the center. □

It is easy to see that the transformation $\alpha \mapsto \gamma\alpha\gamma^{-1}$ (resp. $\gamma\alpha^*\gamma^{-1}$), where γ is a fixed nonzero quaternion and α is any quaternion is an automorphism (resp. antiautomorphism) that preserves the norm and the center of the algebra of the quaternions. An automorphism (resp. antiautomorphism) that can be written in this form is called an *inner automorphism* (resp. *inner antiautomorphism*). Having defined an inner automorphism we are in a position to assert our next lemma.

Lemma 2.6: All norm-preserving automorphisms of the quaternions are inner automorphisms.

Proof: An automorphism evidently takes the center into the center and it is obvious that the only norm-preserving automorphism of $\mathbb{R}$ (which is the center in this case) is the identity transformation. Thus any norm-preserving automorphism of the quaternions leaves the center fixed pointwise. Regard the quaternions as a four-dimensional real space in which the imaginary part of the quaternions are isomorphic with $\mathbb{R}^3$. The norm-preserving automorphisms of $\mathbb{R}^3$ are members of SO(3) which as we have just seen is isomorphic with the group of unit quaternions and the correspondence is through a transformation which makes all these rotations into the inner automorphisms of the quaternions. □

The proof we have given depends on the proof of Lemma 2.5 which depends on some lengthy though straightforward computations. There is another way of looking at things which gets to the result avoiding the computations.

Alternative proof of Lemma 2.6: Note first that the transformation

$$\begin{bmatrix} \alpha_1' & \alpha_2' \\ \alpha_3' & \alpha_4' \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \tag{2.35}$$

is a linear transformation of the coordinates $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ of a vector in $\mathbb{C}^4$. It is very simple to verify that there are 16 linearly independent transformations of this type and the dimension of the space of linear transformations from $\mathbb{C}^4$ to $\mathbb{C}^4$ is also 16, hence every linear transformation on $\mathbb{C}^4$ can be written as a linear combination of terms in this form. It is not difficult to prove that all automorphisms that preserve the value of $\alpha_1\alpha_4 - \alpha_2\alpha_3$ can be written as a single term in this form. In a representation of an automorphism the two transforming matrices must each be invertible and since the automorphism takes the unit element of the quaternions which is represented by a unit matrix into itself it will have to be of the

form $p\alpha p^{-1}$. Finally it must take a quaternion α of a unit modulus (that is, $\alpha\alpha^* = 1$) into a quaternion of unit modulus, hence

$$(p\alpha p^{-1})^* = (p\alpha p^{-1})^{-1}, \tag{2.36a}$$

which is the same thing as

$$p\alpha^{-1}p^{-1} = p^{*-1}\alpha^{-1}p^* \tag{2.36b}$$

or

$$p^*p\alpha^{-1} = \alpha^{-1}p^*p. \tag{2.37}$$

Thus p^*p is a 2×2 matrix that commutes with every unitary matrix among them the matrices $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$. The requirement that p^*p commutes with the first of these matrices leads to the conclusion that its off-diagonal elements are zero and the requirement that it commutes with the second matrix leads to the conclusion that its diagonal elements are equal. This proves that p is a complex constant times a unitary matrix, but a unitary p and a multiple of p by a complex constant clearly correspond to the same automorphism. Thus every automorphism of $\mathbb{H}$ is an inner automorphism. □

This result is, of course, a particular case of the well-known theorem in ring theory which asserts that if A is an automorphism of a simple algebra which is finite-dimensional over its center and if every point of the center is a fixed point of A then A is an inner automorphism.[16] Our method can also be used to prove the analogous result: *All norm-preserving antiautomorphisms of* $\mathbb{H}$ *are inner antiautomorphisms.*

We now come to our last basic lemma.

Lemma 2.7: Let $f:\mathbb{H}\to\mathbb{H}$ be a map which preserves inner products on $\mathbb{H}$ regarded as an Euclidean four-dimensional space and which satisfies $f(1) = 1$. Then f is either an inner automorphism or an inner antiautomorphism on $\mathbb{H}$.

Proof: Remembering that when $\mathbb{H}$ is regarded as a four-dimensional Euclidean space 1, i, j, and k satisfying Eqs. (2.1)–(2.4) are regarded as an orthonormal basis, by Lemma 2.2 f is real-linear and since it preserves inner products and satisfies $f(1) = 1$, it leaves real quaternions fixed and transforms the imaginary space of quaternions in such a way that orthogonal vectors go into orthogonal ones. We know geometrically that only such transformations in three dimension are space inversion and rotation. A pure rotation will take i, j, and k into orthogonal vectors which continue to satisfy Eqs. (2.1)–(2.4) and an immediate consequence is that products are preserved in which case our map is an inner automorphism by the preceding Lemma. If rotation is combined with space inversion, products are reversed and we have an inner antiautomorphism. □

Corollary 2.7.1: Let f: $\mathbb{H}\to\mathbb{H}$ be a norm-preserving map with 1 as a fixed point and satisfying

$$\|1 + \alpha\beta^*\| = \|1 + f(a)(f(\beta))^*\|, \tag{2.38}$$

then f is either an inner automorphism or an inner antiautomorphism on $\mathbb{H}$.

Proof: Follows from the preceding lemma by noting that if $\mathbb{H}$ is regarded as a four-dimensional Euclidean space, then the inner product of two quaternions α and β is $\mathrm{Re}(\alpha\beta^*)$. □

Lemma 2.8: Let A be an operator on a Hilbert space $\mathscr{H}$ over any field whose restriction to each one-dimensional subspace of $\mathscr{H}$ is additive and which for any pair of mutually perpendicular vectors x and y in $\mathscr{H}$ satisfies

$$A(x + y) = Ax + Ay, \tag{2.39}$$

then A is additive on $\mathscr{H}$.

Proof: In view of what is given additivity has to be verified only for a linearly independent pair of vectors which are not mutually perpendicular and since the sum of any two such vectors can be written as a sum of two mutually perpendicular vectors verification is trivial with the help of what is given. □

III. THE MAIN RESULTS

Proposition 3.1: Let A be a p-semilinear isometry from a quaternionic Hilbert space $\mathscr{H}$ to itself.

Then

$$\langle Ax,Ay\rangle = p^*\langle x,y\rangle p. \tag{3.1}$$

Further $A^*A = I$ and $AA^* = P$, where I is the identity map on $\mathscr{H}$ and P is an orthogonal projection on $P(\mathscr{H})$. Furthermore if A is surjective then $P = I$ and pA is unitary.

Proof: Consider first the case when $p = 1$, in which case A is linear. Since A preserves the norms of vectors, by considering the norms of Ax, Ay, and $A(x + y)$, where x and y are arbitrary vectors in $\mathscr{H}$, we easily deduce

$$\langle Ax,Ay\rangle + \langle Ay,Ax\rangle = \langle x,y\rangle + \langle y,x\rangle. \tag{3.2}$$

Replacing y, in turn, by iy, jy, and ky, we get after simple rearrangement

$$\langle Ax,Ay\rangle + i\langle Ay,Ax\rangle i = \langle x,y\rangle + i\langle y,x\rangle i, \tag{3.3}$$

and two similar equations with i replaced by j and k, respectively. Adding together the three equations thus obtained and taking into account (2.11) gives us

$$\langle Ax,Ay\rangle - \langle Ay,Ax\rangle = \langle x,y\rangle - \langle y,x\rangle. \tag{3.4}$$

By adding the preceding equation to (3.2) we get

$$\langle Ax,Ay\rangle = \langle x,y\rangle. \tag{3.5}$$

In the general case when p is any quaternion of unit length, it follows from (2.23) that pA is linear and by using (3.5) we are able to establish immediately the validity of (3.1).

By using (2.30) we have in all cases

$$\langle A^*Ax,y\rangle = \langle x,y\rangle. \tag{3.6}$$

This implies that $A^*A = I$. Then AA^* is clearly linear, self-adjoint, and idempotent so that it is an orthogonal projection on its image. The rest of the assertion is obvious. □

We note that (3.1) is also valid if A is a p-semilinear isometry from one quaternionic Hilbert space to another. Further if A is a linear isometry from a quaternionic Hilbert space $\mathscr{H}$ to itself then every unit scalar multiple of A (when we multiply a linear operator with a unit scalar p^* we get a p-semilinear operator) is an additive operator on $\mathscr{H}$ which preserves the modulus of the inner product. We prove that all additive operators on $\mathscr{H}$ that preserve the modulus of the inner product arise in this way provided the dimension of $\mathscr{H}$ is at least 2. Before we prove this we provide counterexamples in the one-dimensional case: here we can identify $\mathscr{H}$ with $\mathbb{H}$ and any inner antiautomorphism is additive and preserves the modulus of the inner product and we will see in the corollary of the next proposition that such an operator can not be a scalar multiple of a linear isometry.

Proposition 3.2: Let $\mathscr{H}$ be a quaternionic Hilbert space of dimension greater than 1. An operator preserving the modulus of the inner product on $\mathscr{H}$ cannot be a p-antilinear operator.

Proof: Suppose an operator A preserves the modulus of the inner product and is p antilinear with respect to a normalized basis. Since the dimension of $\mathscr{H}$ is at least 2, we can certainly find two distinct members u and v in this basis so that they are both unit vectors and linearly independent. We shall establish a contradiction. Consider

$$\begin{aligned} 0 &= \|\langle u+iv, ku-jv\rangle\| = \|\langle A(u+iv), A(ku-jv)\rangle\| \\ &= \|\langle Au,Au\rangle p^*kp - \langle Au,Av\rangle p^*jp \\ &\quad - p^*ip\langle Av,Au\rangle p^*kp + p^*ip\langle Av,Av\rangle p^*jp\| \\ &= \|2p^*kp + \langle Au,Av\rangle p^*ipp^*kp - p^*ip\langle Av,Au\rangle p^*kp\|. \end{aligned} \tag{3.7}$$

This implies

$$\begin{aligned} 2 &= p^*ip\langle Av,Au\rangle + \langle Au,Av\rangle p^*i^*p \\ &= 2\,\mathrm{Re}\, p^*ip\langle Av,Au\rangle. \end{aligned} \tag{3.8}$$

Hence

$$\begin{aligned} 1 \leqslant \|p^*ip\langle Av,Au\rangle\| &= \|\langle Av,Au\rangle\| \\ &= \|\langle v,u\rangle\| \leqslant \|v\|\|u\| = 1. \end{aligned} \tag{3.9}$$

This is possible only if

$$\|\langle v,u\rangle\| = \|v\|\|u\|, \tag{3.10}$$

or, in other words, only if u and v are linearly dependent (or parallel). This contradiction proves our assertion. □

Corollary 3.2.1: An inner antiautomorphism cannot be a scalar multiple of an inner automorphism.

Proof: We first prove that no inner antiautomorphism can be represented as an inner automorphism. To see this, suppose that the inner antiautomorphism induced by the quaternion p (that is, $\alpha \mapsto p\alpha^*p^*$) is equal to the inner automorphism induced by the quaternion q (that is, $\alpha \mapsto q\alpha q^*$) where without loss of generality we assume that p and q have unit moduli. Then for any $\alpha \in \mathbb{H}$, we have

$$p\alpha^*p^* = q\alpha q^* \tag{3.11}$$

or

$$\alpha^* = p^*q\alpha q^*p. \tag{3.12}$$

If we now take $\alpha = k$, we have

$$-k = p^*qkq^*p = p^*qijq^*p = p^*qiq^*pp^*qjq^*p = i^*j^* = k. \tag{3.13}$$

This contradiction proves that no isomorphism can be both an inner automorphism and an inner antiautomorphism. Finally, since both preserve the real quaternions we do not really have a choice of a scalar multiple because the only scalar which multiplied to a real number leaves that number unchanged is 1. □

Proposition 3.3: Let A be an additive operator on a quaternionic Hilbert space $\mathscr{H}$ of dimension greater than 1 that preserves the modulus of the inner product of any pair of vectors. Then A is a p-semilinear isometry.

Proof: Since A preserves orthogonality and norms we have for any $x \in \mathscr{H}$ and any $\alpha \in \mathbb{H}$

$$A(\alpha x) = \alpha_x Ax, \tag{3.14}$$

where $\|\alpha_x\| = \|\alpha\|$. We note that for every $x \in \mathscr{H}$, $1_x = 1$. Let x and y be mutually orthogonal unit vectors. Then

$$\begin{aligned} &\|1 + \alpha\beta^*\| \\ &\quad = \|\langle x+\alpha y, x+\beta y\rangle\| = \|\langle A(x+\alpha y), A(x+\beta y)\rangle\| \\ &\quad = \|1 + \alpha_y\beta_y^*\|. \end{aligned} \tag{3.15}$$

By Corollary 2.7.1 the correspondence $\alpha \mapsto \alpha_y$ is either an inner automorphism or inner antiautomorphism of $\mathbb{H}$. Let z be any unit vector perpendicular to y, then

$$\begin{aligned} 0 = \|\langle \alpha z + y, z - \alpha^* y\rangle\| &= \|\langle A(\alpha z + y), A(z - \alpha^* y)\rangle\| \\ &= \|\alpha_z - \alpha_y\|, \end{aligned} \tag{3.16}$$

for all $\alpha \in \mathbb{H}$. Hence $\alpha_z = \alpha_y$. Thus $\alpha \mapsto \alpha_y$ is the same function for all members of an orthonormal basis. Proposition 3.2 rules out the possibility that this function is an inner antiautomorphism. Hence there exists a unit quaternion p such that

$$A\alpha x = p^*\alpha p Ax, \tag{3.17}$$

which shows that A is a p-semilinear isometry. □

Proposition 3.4: Let T be an operator on a quaternionic Hilbert space $\mathscr{H}$ that satisfies

$$\langle Tx, Ty\rangle = p^*\langle x,y\rangle p, \quad \forall (x,y) \in \mathscr{H} \times \mathscr{H}, \tag{3.18}$$

for some unit quaternion p. Then T is a p-semilinear isometry.

Proof: Follows by observing that pT preserves the inner product and therefore by Lemma 2.2 is linear and therefore, $T = p^*pT$ is p-semilinear which evidently preserves the norm and is an isometry. □

Proposition 3.5: Let $\mathscr{H}$ be a two-dimensional quaternionic Hilbert space. Then there exist operators on $\mathscr{H}$ which are not Wigner equivalent to a linear isometry.

Proof: Let $\{u,v\}$ be an orthonormal basis in $\mathscr{H}$. Let U be any unitary operator on $\mathscr{H}$, then $\{Uu, Uv\}$ is also an orthonormal basis in $\mathscr{H}$. Let p be any nonreal unit quaternion. We first write an arbitrary vector w in the form $w = r\alpha u + \beta v$, where r is real, α is a unit quaternion, and β is any quaternion. If r is zero or α is real we define $\hat{U}$ by

$$\hat{U}w = p^*(r\alpha)^*pUu + p^*\beta^*pUv, \tag{3.19}$$

which is straightforward p-antilinearity. When r is nonzero and α is nonreal, we define $\hat{U}$ by first writing w as

$$w = r\alpha u + \alpha\alpha^*\beta v, \tag{3.20}$$

and then by the following relation which looks like p antilinearity but is not because it fails to reverse the product of α and $\alpha^*\beta$:

$$\hat{U}w = p^*(r\alpha)^*pUu + p^*\alpha^*pp^*(\alpha^*\beta)^*pUv. \tag{3.21}$$

With the help of Lemma 2.1 (iv) it is easily verified that $\hat{U}$ preserves the moduli of inner products. The only linear isometry which takes u into Uu and v into Uv is U. We shall show that $\hat{U}$ cannot be Wigner equivalent to U. Suppose the contrary is true, that is for some unit quaternion γ depending on both α and β

$$\hat{U}w = \gamma Uw. \tag{3.22}$$

Substitution in (3.21) with the help of (3.20) gives

$$\gamma = p^*\alpha^* p\alpha^*, \tag{3.23}$$

which shows that γ is independent of β and substituting this in the coefficient of Uv gives

$$\alpha^*\beta = p^*(\alpha^*\beta)^* p. \tag{3.24}$$

Since $\alpha^*\beta$ as β ranges over $\mathbb{H}$ is merely a different enumeration (or permutation) of members of $\mathbb{H}$, we have established a contradiction with the help of Corollary 3.2.1. □

Proposition 3.6: Let T be an operator on a quaternionic Hilbert space $\mathscr{H}$ of dimension other than 2 that preserves the modulus of the inner product, then T is Wigner equivalent to a linear isometry A. If in addition T is surjective, then A is unitary.

Proof: In dimension 1 any operator that preserves the modulus of the inner product and therefore the norm is clearly Wigner equivalent to the identity operator.

Suppose now that the dimension of $\mathscr{H}$ is at least 3. Given T with the stated properties we shall construct a linear isometry U which is Wigner equivalent to T. Take any unit vector x in $\mathscr{H}$, let $\mathscr{P}$ be the one-dimensional space spanned by x. We shall first define an operator A on $(x+\mathscr{P}^\perp)\cup\mathscr{P}^\perp$, and later we shall use A to define U. We start by defining

$$Ax = Tx. \tag{3.25}$$

For any unit vector $y\in\mathscr{P}^\perp$ and any quaternion α we define the images of A at $x+\alpha y$ and αy to be unit scalar multiples of the corresponding images of T in such a way that

$$\langle Ax, A(x+\alpha y)\rangle = 1 \tag{3.26}$$

and

$$\langle A(x+\alpha y), A(\alpha y)\rangle = \|\alpha\|^2. \tag{3.27}$$

Since in the domain in which A has been defined thus far, it is Wigner equivalent to T, it too preserves the moduli of the inner products. It is evident that any operator which preserves the modulus of the inner product preserves norms and perpendicularity of vectors: this fact together with the defining equations imply that

$$A(x+\alpha y) = Ax + A(\alpha y) = Ax + \alpha_y Ay, \tag{3.28}$$

where

$$\|\alpha_y\| = \|\alpha\|. \tag{3.29}$$

For a fixed vector $y\in\mathscr{P}^\perp$ and α, $\hat{\alpha}\in\mathbb{H}$, we have

$$\|1+\alpha\hat{\alpha}^*\| = \|\langle A(x+\alpha y), A(x+\hat{\alpha}y)\rangle\| = \|1+\alpha_y\hat{\alpha}_y^*\|. \tag{3.30}$$

This together with the fact that (3.28) implies that $1_y = 1$ leads us to conclude with the help of Corollary 2.7.1 that the mapping $\alpha\mapsto\alpha_y$ is either an inner automorphism or an inner antiautomorphism of $\mathbb{H}$; in either case A restricted to the one-dimensional subspace spanned by y is additive. We shall prove next that A restricted to $\mathscr{P}^\perp$ is additive which by Proposition 3.3 will rule out the possibility that the mapping $\alpha\mapsto\alpha_y$ is an inner antiautomorphism of $\mathbb{H}$.

Next let y and z be any pair of mutually perpendicular unit vectors in $\mathscr{P}^\perp$ and let α and β be any pair of quaternions. Remembering that $\alpha y+\beta z$ divided by its norm is a unit vector in $\mathscr{P}^\perp$, the equation corresponding to (3.28) for $A(x+\alpha y+\beta z)$ reads, after again taking into account that A preserves norms and perpendicularity of vectors:

$$A(x+\alpha y+\beta z) = Ax + A(\alpha y\,\beta z) = Ax + \alpha' Ay + \beta' Az, \tag{3.31}$$

where α' and β' have the same moduli as α and β, respectively. Taking the absolute values of the inner product of the two sides of (3.28) with the corresponding sides of (3.31) gives

$$\|1+\alpha\alpha^*\| = \|1+\alpha_y\alpha'^*\|, \tag{3.32}$$

which evidently implies

$$\alpha' = \alpha_y, \tag{3.33}$$

and similarly

$$\beta' = \beta_z, \tag{3.34}$$

so that

$$A(\alpha y+\beta z) = A(\alpha y) + A(\beta z), \tag{3.35}$$

which implies that the restriction of A to $\mathscr{P}^\perp$ satisfies the requirements of Lemma 2.8 and is, therefore, additive. Hence from Proposition 3.3 it now follows that the restriction of A to $\mathscr{P}^\perp$ is a p-semilinear map for some unit quaternion p and thus pA is a linear isometry which is Wigner equivalent to A and, therefore, to T also. We take U to be pA extended by linearity to the whole of $\mathscr{H}$. Since we know that the values of U differ from those of T $(x+\mathscr{P}^\perp)\cup\mathscr{P}^\perp\cup\mathscr{P}$ by a phase factor only, all that remains to be done is to check that $U(\alpha x+\beta y)$ differs from $T(\alpha x+\beta y)$ by a phase factor only: here $y\in\mathscr{P}^\perp$ and α and β are an arbitrary pair of nonzero quaternions. By the linearity of U and the known fact that on $x+\mathscr{P}^\perp U$ and T differ by a phase factor only, we have

$$\begin{aligned} T(\alpha x+\beta y) &= T(\alpha(x+\alpha^{-1}\beta y)) = \hat{\alpha}T(x+\alpha^{-1}\beta y) \\ &= \hat{\alpha}\gamma U(x+\alpha^{-1}\beta y) \\ &= \hat{\alpha}\gamma\alpha^{-1}U(\alpha x+\beta y), \end{aligned} \tag{3.36}$$

where γ has unit modulus and α and $\hat{\alpha}$ have the same moduli. Thus T is Wigner equivalent to the linear isometry U. The proof of the rest of the assertion is obvious. □

In this connection it remains only to remark that U is Wigner equivalent to $-U$, as was first pointed out by Bargmann.[9] Consequently U in the proposition above is unique up to a sign.

IV. DISCUSSION

The most remarkable thing about Wigner's theorem for quaternionic Hilbert spaces is that for such spaces dimension 2 is exceptional. The first counter example was given by Bargmann.[9] Our proof of Wigner's theorem enables us to systematically describe all possible counter examples: Bargmann's and our own among them. Let T be an operator on a quaternionic Hilbert space $\mathscr{H}$ of dimension 2 which preserves the modulus of the inner product, but is not Wigner equivalent to any unitary operator on $\mathscr{H}$. We shall determine all possible forms of T. Let $\{x,y\}$ be an orthonormal basis in $\mathscr{H}$, then since T preserves orthonormality $\{Tx,Ty\}$ is also an orthonormal basis. Clearly there is a unitary operator U which takes $\{x,y\}$ to $\{Tx,Ty\}$. Proceeding as in the proof of Wigner's theorem we define on operator A which is

Wigner equivalent to T and which demonstrates the various possibilities. We set $Ax = Tx = Ux$ and define A on $(x + \mathscr{P}^1) \cup \mathscr{P}^1$. As in the proof of Wigner's theorem it follows that A thus defined induces either an inner automorphism or an inner antiautomorphism. If A were inducing an inner automorphism, then as in the last part of the proof it will follow that A is Wigner equivalent to a unitary operator. Thus A must be a p-antilinear operator on $(x + \mathscr{P}^1) \cup \mathscr{P}^1$. We have seen that such an A cannot be Wigner equivalent to either a linear of antilinear operator, in the latter case because we have seen that antilinear operators do not preserve the moduli of inner products unless the dimension of the space is 1. On $\mathscr{P}$ all maps to $T(\mathscr{P})$ that preserve the moduli of inner products are Wigner equivalent to each other. For considering the remaining points of $\mathscr{H}$, that is those not in $\mathscr{P} \cup \mathscr{P}^1 \cup (x + \mathscr{P}^1)$, these can be written in the form $\alpha x + \beta y$ with α neither zero nor 1 and β nonzero. We can rewrite these vectors in the form $\alpha(x + \alpha^{-1}\beta y)$ and from the argument familiar from the proof of Wigner's theorem we can write for any definition of A on these points which preserves its Wigner equivalence to T

$$A(\alpha(x + \alpha^{-1}\beta y)) = \gamma T(\alpha(x + \alpha^{-1}\beta y)) = \gamma\alpha' T(x + \alpha^{-1}\beta y) = \alpha'' A(x + \alpha^{-1}\beta y), \tag{4.1}$$

where γ has unit modulus and α, α', and α'' have the same moduli. We can choose the correspondence $\alpha \mapsto \alpha''$ to be any inner automorphism or any inner antiautomorphism on $\mathbb{H}$. The interesting point to note is that even if we choose this correspondence to be the same as the one which gives us a p-antilinear operator on $(x + \mathscr{P}^1) \cup \mathscr{P}^1$, we do not get p antilinearity on the whole space. To see this, consider the action of A defined by these prescriptions on $\alpha x + \beta y (\alpha \neq 0)$, clearly

$$A(\alpha x + \beta y) = p^*\alpha^* pAx + p^*\alpha^*\beta^*\alpha^{*-1} pAy,$$

which is not the same as $p^*\alpha^* pAx + p^*\beta^* pAy$ that one would expect if A were p antilinear on the whole space. Finally we can compose any such A with any unitary transformation either on the left or on the right to get another operator which preserves the modulus of the inner product without being Wigner equivalent to any unitary transformation. It is possible to systematically classify all these possibilities and study how they are related to each other, but at the present time there does not seem to be sufficient motivation for undertaking such a study. We will, however, point out that the counterexamples of Bargmann[9] and of this paper are covered by these possibilities as they must be because the possibilities above are exhaustive. Finally we point out that, as is easy to see, each counterexample gives a discontinuous operator (to see this, take U in the example of Proposition 3.5 to be the identity operator on $\mathscr{H}$ and set $p = \alpha = i$ and $\beta = j$.

The vector jv can also be written as $\lim_{r\to 0} r\alpha u + \alpha\alpha^* jv$. Hence if $\hat{U}$ were continuous, $\hat{U}(jv)$ can be evaluated by either (3.19) or (3.21), but one of them gives the value jv while the other gives $-jv$. The demonstration for the general case is no harder.). Thus we could have used the lack of continuity to prove that these operators are not Wigner equivalent to any linear operator. However, there is no need to use analysis where one can reach one's goal with algebra alone!

We hope that our proof of Wigner's theorem makes the theorem as obvious to everyone as it was to Wigner and we hope and believe that it is mathematically more rigorous as well as more elementary than Bargmann's.[9] Apart from a new counterexample and a new proof we have presented a number of results that are technically new but we wish to put greater emphasis on the newness of our methods rather than of our results. Even the methods are not all that new but they are designed in such a way that they bring rigorous mathematical methods with the use of only simple and easy concepts.

In our future work we shall continue to develop the mathematics of operators on a quaternionic Hilbert space, but time seems ripe for a reappraisal of the theory of time reversal operator as well as charge conjugation and parity in the light of all the experience gained thus far and to see how much mathematical rigour can be brought into the theory of these operators. Work is in progress and will be reported in due course.

ACKNOWLEDGEMENTS

We are grateful to Professor H. Van Dam for some enlightening information about the earlier proofs of Wigner theorem and for encouragement. It is a pleasure to acknowledge our indebtedness to the referee who has made numerous suggestions for improvement of our presentation and helped us to avoid a number of ambiguities, though we continue to claim title absolute to all the shortcomings which have survived his scrutiny.

[1] C. S. Sharma and T. J. Coulson, J. Math. Phys. **28**, 1941 (1987).
[2] C. S. Sharma, J. Math. Phys. **29**, 1069 (1988).
[3] C. S. Sharma and D. F. Almeida, J. Math. Phys. **30**, 369 (1989).
[4] M. D. Vivarelli, Celestial Mech. **29**, 45 (1983); **36**, 45 (1985).
[5] L. P. Horwitz and L. C. Biedenharn, Ann. Phys. (NY) **157**, 432 (1984); A. Razon, L. P. Horwitz, and L. C. Biedenharn, J. Math. Phys. **30**, 59 (1989).
[6] S. Adler, Commun. Math. Phys. **104**, 611 (1986).
[7] E. Y. Rocher, J. Math. Phys. **13**, 1919 (1972).
[8] C. S. Sharma, Nuovo Cimento B **103**, 431 (1989).
[9] V. Bargmann, J. Math. Phys. **5**, 862 (1964).
[10] C. S. Sharma and D. F. Almeida, Ann. Phys. (NY) **197**, 300 (1990).
[11] T. J. Coulson, Ph.D. thesis, London University, 1986.
[12] E. P. Wigner, *Group Theory* (Academic, New York, 1959).
[13] J. Pian and C. S. Sharma, Int. J. Theor. Phys. **22**, 107 (1983).
[14] P. M. Cohn, *Algebra I* (Wiley, Chichester, 1982).
[15] W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973).
[16] T. W. Hungerford, *Algebra* (Holt, Rinehart, and Winston, New York, 1974).

Some examples of the algebra of flows

S. K. Bose
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Received 10 May 1989; accepted for publication 7 January 1990)

The algebra of the group of smooth maps from a manifold M to a compact simple Lie group G is studied for two cases. The first is when M is the double coset $SO(d,R)\backslash SO(d+1,R)/SO(d,R)$, the corresponding maps are those from a d sphere to G that are invariant under left translations by elements from $SO(d,R)$. In the second example, M is a two-dimensional torus. The problem of central extension of these algebras is solved. For the first example, no central extension is possible. For the second, the number of independent central extensions is infinite.

I. INTRODUCTION

Let G be a Lie group, g its Lie algebra over the field F, and M a smooth C^∞ manifold. Homotopically trivial smooth maps from M to G constitute a group G^M under pointwise multiplication. This group has the natural structure of an infinite-dimensional Lie group. The corresponding Lie algebra g^M, the algebra of flows, consists of C^∞ functions $M \to g$ with pointwise commutator. That is to say, the algebra of flows is definable by the structure $g \otimes C^\infty(M)$ with the commutator given by the formula

$$[g_1 \otimes a_1, g_2 \otimes a_2] = [g_1, g_2] \otimes a_1 a_2,$$

where $g_1, g_2 \in g$ and $a_1, a_2 \in C^\infty(M)$ and the commutative algebra $C^\infty(M)$ is defined over the same field F as is g. When M is a circle S^1, g^{S^1} is the loop algebra—the quotient of the Kac–Moody algebra[1,2] by its center. When M is a d-dimensional sphere S^d, the corresponding sphere algebras are of interest. Two special cases of these, namely, $d = 2$ and $d = 3$, and G compact and simple, have been studied recently.[3,4]

To analyze the structure of an algebra of flows, one chooses a basis. This is provided by a suitable choice of the basis of g and a choice of a complete set of functions on M. For the loop algebra, a complete set of functions is provided by Z^n, n = integer, here the complex number Z parametrizes the circle $Z = e^{i\theta}$. For the cases $M = S^2$ and $M = S^3$, the complete set of functions are, respectively, the spherical harmonics $Y_{lm}(\theta, \phi)$ and the rotation matrices $D^l_{mm'}(\alpha, \beta, \gamma)$. Here, α, β, and γ are the Euler angles. The problem of understanding the structure of an algebra of flows, in its essence, is the problem of displaying the *algebraic* structure of the corresponding complete set of functions in a compact, tractable, and reasonably simple manner.

A related problem concerns the question of central extensions of an algebra of flows. The number of independent central extensions to g^M is equal to the dimension of the second cohomology spaces $H^2(g^M)$ of algebra g^M, with coefficients in F (Ref. 5). When M is a manifold of dimension greater than one and G compact, the space $H^2(g^M)$ is infinite dimensional, as has been proved by Feigin.[6] In Ref. 4, the spaces $H^2(g^M)$ for the cases $M = S^2$, $M = S^3$, and G compact and simple, have been explicitly constructed. Thus for the case $M = S^2$, the current algebra corresponding to g^M has an anomaly term that contains an arbitrary function on the two sphere. When this function is expanded into the complete set Y_{lm} of functions, one gets the infinite number of central extensions. Similarly, for the case $M = S^3$, the anomaly term contains three arbitrary functions.

How much of the analysis of Ref. 4 could be generalized to the d-dimensional sphere S^d ($d > 3$)? For a straightforward generalization, we need to have at our disposal a complete set of functions on S^d. While this task entails no difficulty in principle, it is quite messy in practice. The functions are of d independent variables; and their algebraic structure (reduction of a product of two functions into the sum) is rather complicated, to say the least. One way to cut down on the number of variables, in fact, to be left with a single variable is to restrict oneself to a particular subspace (of the d sphere) that is realizable as a certain double coset, as follows.[7,8] Let $H = SO(d+1, R)$ and the closed subgroup $K = SO(d, R)$, the left coset space H/K is S^d, our M is the double coset $K\backslash H/K$. Functions on the left coset are realizable as functions $f(s)$ on H such that $f(st) = f(s)$ for all $t \in K$; that is, as functions on H that are invariant under right translations by elements from K. Similarly, the functions on the right coset are functions on H that are invariant under left translations by elements from K. Functions on H, $f(s)$ which are invariant under both left and right translations by elements from K, $f(tst') = f(s)$ for t, t' in K, are the functions on the double coset $K\backslash H/K$. They may also be identified with functions on S^d such that $h(t \cdot x) = h(x)$ for $t \in K$ and $x \in S^d$. A basis for the space $L^2(K\backslash H/K)$ of square-integrable (with respect to the canonical measure) functions on the double coset is provided by the so-called zonal spherical harmonics.[7,8] Embedding the sphere S^d in R^{d+1} in the canonical way, one may also identify the zonal spherical harmonics with the restriction to S^d of the homogeneous harmonic polynomials in R^{d+1}. These functions are *very* interesting, they play a central role in the theory of generalized Fourier analysis of groups.[7,8] In Sec. II, we study the algebra of flows for the case where $M = K\backslash H/K$, $H = SO(d+1, R)$, $K = SO(d, R)$, and G is compact and simple. We shall explicitly display the basic commutation relations for this Lie algebra. We shall also

 0022-2488/90/051042-05$03.00

find that this algebra does *not* possess any central extension. We shall call these the zonal sphere algebras.

The fact that the number of possible independent central extensions to g^M is infinite, whenever the dimension of M is greater than one, raises many questions. Looked at one way, it seems that the whole thing has become a bit meaningless. On the other hand, if there is, as there indeed is, an infinite number of central terms, then it is with an infinite number of central terms that one has got to learn to live. There is yet another possibility. The centrally extended algebra might possess *representations* in which *all but a finite number* of central terms vanish. In any event, it is plainly desirable to work out an example of an algebra with an infinite number of central terms which is, at the same time, simpler than the examples in Ref. 4. The simplest example of such an algebra is obtained when the manifold M is a two-dimensional torus. In Sec. III, we shall study this example.

II. ZONAL SPHERE ALGEBRAS

We wish to consider the algebra of flows corresponding to the group of smooth (C^∞) maps from M to G, where M is subspace of S^d, which is realized as the double coset $\mathrm{SO}(d)\backslash\mathrm{SO}(d+1)/\mathrm{SO}(d)$; here the d sphere S^d is identified with the left coset space $\mathrm{SO}(d+1)/\mathrm{SO}(d)$. For functions on M, a basis (a complete set of functions) is provided by the zonal spherical harmonics. The latter are identifiable[7,8] with the classical Gegenbauer polynomials $C_m^\lambda(x)$; here, m is the degree of the polynomial and the index λ is related to the dimension d of the sphere S^d via $d = 2\lambda + 1$, and $x = \cos\theta$ where θ parametrizes the double coset. Let us consider some examples. For $d = 2$, $C_m^{1/2}$ is the usual Legendre polynomial $P_m(\cos\theta)$. For $d = 3$, C_m^1 is $\sin(m+1)\theta/\sin\theta$. Recall that S^3 is the manifold of SU(2) and thus a complete set of functions on S^3 is provided by the functions that provide the set of unitary irreducible representations of SU(2) (rotation matrices). The polynomial $C_m^1(\cos\theta)$ is the corresponding *character* function. For a fixed λ, the product of two polynomials can be expanded as

$$C_n^\lambda(x)C_m^\lambda(x)=\sum_l C_{nm}^l C_l^\lambda(x), \tag{2.1}$$

where the expansion coefficients C_{nm}^l are given by

$$[2^{1-2\lambda}\pi/(\Gamma(\lambda))^2][\Gamma(l+2\lambda)/(l+\lambda)!]C_{nm}^l$$

$$=\int_{-1}^{1} C_l^\lambda(x)C_n^\lambda(x)C_m^\lambda(x)(1-x^2)^{\lambda-1/2}\,dx. \tag{2.2}$$

The integral that appears above is known.[9] We thus obtain the following expression for the coefficient C_{nm}^l: It is zero unless $l + m + n$ is even and there exists a triangle with sides l,m,n; when nonvanishing it is given by

$$C_{nm}^l=\frac{\alpha_{s-l}\alpha_{s-m}\alpha_{s-n}}{\alpha_s}\frac{l+\lambda}{s+\lambda}\frac{\Gamma(l+1)}{\Gamma(s+1)}\frac{\Gamma(s+2\lambda)}{\Gamma(l+2\lambda)}, \tag{2.3}$$

where

$$\alpha_n=\binom{n+\lambda-1}{n}=\frac{\Gamma(n+\lambda)}{\Gamma(n+1)}\frac{1}{\Gamma(\lambda)} \tag{2.4}$$

and

$$s=\tfrac{1}{2}(l+m+n). \tag{2.5}$$

We note that the coefficients C_{nm}^l are non-negative, the nonvanishing ones given by (2.3) are positive. It is also interesting to note that for the special case $\lambda = 1$, corresponding to the zonal harmonics for S^3, all the nonzero coefficients become unity, $C_{nm}^l = 1$. This is as it should be. Similarly, the case $\lambda = \frac{1}{2}$ is easily checked to reproduce the correct result for Legendre polynomials.

Let T^a, $a = 1,2,\ldots,\dim(G)$, be a basis for the Lie algebra g of G, and f^{abc} the corresponding structure constants. A basis for the algebra g^M may now be chosen as T_m^a, m is a non-negative integer, in such a way that the basic commutation relation can be calculated from the "defining representation" $T_m^a = T^a C_m^\lambda$ (the label λ is suppressed on the T_m^a, its presence is understood from the context). We thus obtain the desired commutator

$$[T_n^a, T_m^b]=f^{abc}C_{nm}^l T_l^c. \tag{2.6}$$

The consistency of the above relation is easily checked. Notice that C_{nm}^l is symmetric in the lower indices, thus antisymmetry of the commutator is guaranteed. The Jacobi identity for the double commutator is also easily checked. Finally, a summation over the repeated indices c and l in (2.6) is understood. Consider now the problem of central extension of the algebra. Toward this end we write

$$[T_n^a, T_m^b]=f^{abc}C_{nm}^l T_l^c + d_{nmj}^{ab}K^j, \tag{2.7}$$

where K^j are the central generators. To proceed further we restrict G to be compact and simple. The structure constants of G may now be taken as completely antisymmetric satisfying $f^{abc}f^{abd} = \delta^{cd}$. Following a standard argument[4] we can now, without loss of generality, set

$$d_{nmj}^{ab}=\delta^{ab}d_{nmj}. \tag{2.8}$$

The Jacobi identity for the commutator (2.7) now gives

$$C_{nm}^k d_{kl} + C_{ml}^k d_{kn} + C_{ln}^k d_{km}=0, \tag{2.9}$$

where a summation over the dummy index is understood and we have suppressed the additional index [j in (2.8)] on d_{nm}. After all, this index simply labels the linearly independent solutions to (2.9), which is precisely what we are now going to study. Equation (2.9) together with the antisymmetry property

$$d_{nm}=-d_{mn} \tag{2.10}$$

are the relations needed to determine the central terms.

We are now going to prove that the system (2.9) and (2.10) of equation possess no nontrivial solution. The range of summation over the dummy index k in (2.9) is determined by the nonvanishing property of the coefficients C_{nm}^k. Thus in the first term of (2.9) k ranges from $|n - m|$ to $n + m$ in steps of 2, in the second term of (2.9) k ranges from $|m - l|$ to $m + l$, and in the third term from $|l - n|$ to $l + n$, in steps of 2. For each summand

C^k_{nm} is now given by (2.3), and every such C^k_{nm} is *positive*. Also note $C^k_{no} = \delta_{nk}$. Now take $l = m = 0$ in (2.9) to conclude that $d_{no} = 0$ for all n. To proceed further we argue as follows. Suppose there exists a positive integer n such that we have $d_{ij} = 0$ for all $i \leqslant n - j + 1$, as j takes up positive integer values (1,2,3,...) up to $(n - 1)/2$ if n is odd and to $n/2$ if n is even. Because of antisymmetry of d_{ij} it really amounts to the statements $d_{i1} = 0$ for $i \leqslant n$, $d_{i2} = 0$ for $i \leqslant n - 1$ and so on; the sequence ends with $d_{(n+3)/2,(n-1)/2} = 0$ for n = odd or with $d_{(n+2)/2,\,n/2} = 0$ for n = even. When these conditions are satisfied we shall say that we have a null system based on n. Now we state the following.

Proposition: If there exists a null system based on n for a particular value n_0 of n, then so does a null system based on $n_0 + 1$.

Proof: Take $l = i, m = 1, n = n_0 - i + 1$ in Eq. (2.9) to obtain

$$C^k_{n_0-i+1,1}d_{ki} + C^k_{1i}\,d_{k,n_0-i+1} + C^k_{i,n_0-i+1}d_{k1} = 0. \tag{2.11}$$

Consider the third term above. In the k summation, the maximum value is $n_0 + 1$, all other values of k are lower in steps of 2. Thus only one term $k = n_0 + 1$ survives. In the second term in (2.11) the allowed values of k are $k = i - 1$ and $k = i + 1$; of these only the term $k = i + 1$ survives. Similarly, in the k summation in the first term only the one with $k = n_0 - i + 2$ is nonzero. We thus obtain

$$C^{n_0+1}_{n_0-i+1,i}\,d_{n_0+1,1} + C^{n_0-i+2}_{n_0-i+1,1}\,d_{n_0-i+2,i} - C^{i+1}_{i,1}d_{n_0-i+1,i+1} = 0. \tag{2.12}$$

In case n_0 is even, the above provides us with a system of equations $n_0/2$ is number (i runs as $i = 1,2,...,n_0/2$), the number of unknowns is also $n_0/2$, due to antisymmetry of d_{ij}. In case n_0 is odd, the above is a set of $(n_0 + 1)/2$ equations for that number of unknowns [i here runs from 1 to $(n_0 + 1)/2$)]. We now set $d_{n_0-i+2,i} = X_i$, assemble the X_i's into a column vector X, and write Eq. (2.12) as the matrix equation

$$AX = 0, \tag{2.13}$$

where A is a certain square matrix, with the following structure. The ith row of the matrix has the shape: first entry is positive, the ith entry is again positive, the $(i + 1)$th entry is negative, all other entries are zero. This means that each row, except the first and the last, has exactly three entries with the stated signs. The first row has *two* entries, first positive and second negative. The last row has two entries, the first and the last, *both positive*. The conclusion regarding the signs comes from the fact that the C's as given by (2.3), are all positive. Beyond this, no further information regarding the matrix A will be needed. Upon writing out the matrix A it becomes clear, after a little inspection (start with the lower right-hand corner, examine the cofactors and walk up) that A must be nonsingular; its determinant cannot vanish. Thus Eq. (2.3) has only the trivial solution $X = 0$. Another way of proving this result has been shown to the author by McGlinn. Fix the convention that $d_{n_0+1,1}$ is positive. Equation (2.12) with $i = 1$ now gives $d_{n_0,2}$ is positive. Using this, Eq. (2.12) now gives for $i = 2$ that $d_{n_0-1,3}$ is positive. Continuing this way and using all but the last equation in (2.12) one proves that *all* d's are positive—a conclusion that is contradicted by the very last equation in (2.12). Thus all d's must vanish. Our proposition stands proved.

We can now prove that Eqs. (2.9) and (2.10) possess no nontrivial solution. The existence of a null system based on n_0 for the cases $n_0 = 2$ and $n_0 = 3$ follows directly from Eq. (2.12). In view of our proposition, we now have the following theorem.

Theorem: No central extension exists for the zonal sphere algebras associated with a compact simple Lie group.

III. TWO TORUS ALGEBRAS

In this example M is a two-dimensional torus T^2, parametrized by the ordered pair (Z_1,Z_2) of complex numbers, each with magnitude equal to unity. The elements of the dual space $\tilde{T}^2$ are ordered pairs $N = (n_1,n_2)$ of integers. In the sequel, we shall consistently denote such ordered pairs by the corresponding capital letter. In $\tilde{T}^2$, we have an operation of component wise addition:

$$N + M = (n_1 + m_1,\, n_2 + m_2); \qquad N = (n_1,n_2), \quad M = (m_1,m_2). \tag{3.1}$$

We define the antisymmetric product ($\tilde{T}^2 \otimes \tilde{T}^2 \to \mathbf{Z}$) by the formula

$$N \wedge M = n_1 m_2 - n_2 m_1. \tag{3.2}$$

We introduce the Kronecker symbol $\delta_{MN} = \delta_{m_1n_1}\delta_{m_2n_2}$. We choose the basis for g such that the structure constants f^{abc} are completely antisymmetric (G is simple compact and Lie). We may now select a basis for our algebra of flows g^{T^2} to be given by T^a_N that obey the commutation relations

$$[T^a_N,\, T^b_M] = f^{abc}\,T^c_{N+M}. \tag{3.3}$$

Consider, now, the problem of central extension of the algebra. Toward this end we write

$$[T^a_N,\, T^b_M] = f^{abc}\,T^c_{N+M} + \delta^{ab}\,d_{NM(J)}K^j, \tag{3.4}$$

where K^j are the central generators, indexed by J. We should note that the assumed dependence of the central terms on a and b in (3.4) does not imply any loss of generality. In order for (3.4) to satisfy the Jacobi identity, the following condition must be satisfied:

$$d_{N+M,L(J)} + d_{M+L,N(J)} + d_{L+N,M(J)} = 0. \tag{3.5}$$

The above plus the requirement of antisymmetry determine the central extensions. Equation (3.5) has the solutions:

$$d_{NM(L)} = (N \wedge M)\,\delta_{N+M,L}. \tag{3.6}$$

These solutions are indexed by $L \in \tilde{T}^2$. There are two other linearly independent solutions to (3.5):

$$d_{NM(1)}=n_1\,\delta_{N+M,0};\quad d_{NM(2)}=n_2\,\delta_{N+M,0}. \tag{3.7}$$

We summarize our result as follows:

$$[T^a_N,\,T^b_M]=f^{abc}T^c_{N+M}+\delta^{ab}(N\wedge M)K^{N+M}+\delta^{ab}(n_1K^1+n_2K^2)\delta_{N+M,0}. \tag{3.8}$$

Here, K^1, K^2, and $K^L(L\in\tilde{T}^2)$ are the central charges. Before proceeding further it is useful to look at the corresponding current algebra.

Let us define the "currents":

$$T^a(\theta_1,\theta_2)=\frac{1}{(2\pi)^2}\sum_{n_1n_2}T^a_{n_1n_2}Z_1^{n_1}Z_2^{n_2};\quad Z_1=e^{i\theta_1},\ Z_2=e^{i\theta_2}. \tag{3.9}$$

From (3.8) and (3.9) we compute the current commutation relations:

$$[T^a(\theta_1,\theta_2),T^b(\theta_1',\theta_2')]=f^{abc}T^c\delta(\theta_1-\theta_1')\delta(\theta_2-\theta_2')+\delta^{ab}\Big\{\frac{\partial h}{\partial\theta_1}\delta(\theta_1-\theta_1')\delta'(\theta_2-\theta_2')-\frac{\partial h}{\partial\theta_2}\delta'(\theta_1-\theta_1')\delta(\theta_2-\theta_2')$$
$$-iK^1\delta'(\theta_1-\theta_1')\delta(\theta_2-\theta_2')-iK^2\delta(\theta_1-\theta_1')\delta'(\theta_2-\theta_2')\Big\}, \tag{3.10}$$

where

$$h(\theta_1,\theta_2)=\sum_{n_1,n_2}K^{(n_1,n_2)}Z_1^{n_1}Z_2^{n_2}. \tag{3.11}$$

Thus the anomaly term in the current algebra contains the arbitrary function $h(\theta_1,\theta_2)$. Indeed, we could have started from the other end and written the anomaly term in the current algebra in the form

$$\delta'(\theta_1-\theta_1')\delta(\theta_2-\theta_2')h_1(\theta_1,\theta_2)+\delta(\theta_1-\theta_1')\delta'(\theta_2-\theta_2')h_2(\theta_1,\theta_2).$$

The Jacobi identity would then give the condition

$$\frac{\partial h_1}{\partial\theta_1}+\frac{\partial h_2}{\partial\theta_2}=0. \tag{3.12}$$

The functions h_1, h_2 are identifiable with the components of a certain close one-form. The Hodge decomposition theorem plus the deRham cohomology of the two torus now gives the general solution of (3.12) in the form

$$h_1=-\frac{\partial h}{\partial\theta_2}+c_1,\quad h_2=\frac{\partial h}{\partial\theta_1}+c_2, \tag{3.13}$$

where c_1 and c_2 are constants, and h is an arbitrary function of θ_1 and θ_2. Expanding $h(\theta_1,\theta_2)$ in terms of the basis functions we would then, once again, obtain the commutation relation Eq. (3.8).

The algebra (3.8) contains subalgebras isomorphic to the Kac–Moody algebra. Generators T^a_N, where N is of the form $N=(n_1,0)$, obey

$$[T^a_{n_1},\,T^b_{m_1}]=f^{abc}T^c_{n_1+m_1}+\delta^{ab}n_1\delta_{n_1+m_1,0}K^1, \tag{3.14}$$

where we have put $T^a_N=T^a_{n_1}$ for $N=(n_1,0)$. Similarly, T^a_N with N of the form $N=(0,n_2)$ generate algebra

$$[T^a_{n_2},\,T^b_{m_2}]=f^{abc}T^c_{n_2+m_2}+\delta^{ab}n_2\delta_{n_2+m_2,0}K^2. \tag{3.15}$$

Note that (3.12) and (3.13) are mutually noncommuting. Yet another Kac–Moody algebra is generated by operators of the type T^a_N with $N=(n,n)$. We have

$$[T^a_n,\,T^b_m]=f^{abc}T^c_{n+m}+\delta^{ab}n\delta_{n+m,0}(K^1+K^2), \tag{3.16}$$

where $T^a_N\equiv T^a_n$ for $N=(n,n)$. These facts are sufficient to guarantee that the central charges K^1 and K^2 are quantized in the same fashion as the Kac–Moody central charge.

Is it possible to associate with the algebra (3.8) a root vector system in a finite dimensional root vector space? First, we note that we can obviously introduce a pair d_i $(i=1,2)$ of grading operators with the defining property

$$[d_i,\,T^a_N]=n_iT^a_N;\quad i=1,2, \tag{3.17}$$

plus the requirement that d_i commute with each other and with all the central charges. However, the algebra (3.8) *cannot* be extended by appending the elements d_i. One checks easily that (3.8) together with (3.17) are inconsistent with the Jacobi identifies (because of the presence of the central charges K^N, $N\in\tilde{T}^2$). Thus the answer to the question raised at the beginning of this paragraph is no. In this sense, our algebra is essentially distinct from the Kac–Moody algebras.

The algebra given by Eq. (3.8) is probably the simplest conceivable extension of the notion of a Kac–Moody algebra. This algebra was considered previously by Bars,[3] but the central extension terms were not obtained in their generality by him (the term denoted here as K^N was left out). In conclusion, we make the following remark. It is immediately apparent from the commutation relations (3.8) that there exists a class of representations in which all the central generators except two, namely K^1 and K^2, vanish (have zero eigenvalue). From this class, we can pick up the class of "highest weight representations" with the defining property

$$T^a_N|s\rangle=0, \tag{3.18}$$

whenever $N = (n_1, n_2)$ is such that *either* of the two conditions (1) $n_1 > 0, n_2 \geqslant 0$, (2) $n_1 \geqslant 0, n_2 > 0$ is fulfilled. The states $|s\rangle$ then provide a representation of g, which may be taken as irreducible. The resulting irrep of (3.8) can further be made unitary if the central charges K^1 and K^2 are quantized in the fashion of a Kac–Moody central charge:

$$2K^i/\psi^2 \in \mathbb{Z}, \quad K^i \geqslant \psi \cdot \mu_0, \quad i=1, 2, \tag{3.19}$$

where ψ is the long root of g and μ_0 the highest weight of the "vacuum" representation provided by the states $|s\rangle$.

We did not address the question of possible physical application of the mathematical structures outlined in this paper. Actually, it appears entirely feasible to construct physical models corresponding to our two-torus algebra. This problem is under current investigation and its conclusions will be reported in a future communication.

ACKNOWLEDGMENTS

The author is much indebted to Professor W. D. McGlinn for conversations and to Professor H. Azad of the Department of Mathematics, University of Islamabad, for introducing him to the concept of functions on a double coset.

[1] V. G. Kac, Funct. Anal. Appl. **1**, 328 (1967); *Infinite-dimensional Lie Algebras—An Introduction* (Birkhauser, Boston, MA, 1983).
[2] R. V. Moody, Bull. Am. Math. Soc. **73**, 217 (1967); J. Algebra **10**, 211 (1968).
[3] I. Bars, in *Vertex Operators in Mathematical Physics*, edited by J. Lepowsky, S. Mendelstam, and I. M. Singer (MSRI, Berkeley Publ. #3, Berkeley, CA, 1984), p. 373.
[4] S. K. Bose and S. A. Bruce, Notre Dame preprint, 1988.
[5] D. B. Fuks, *Cohomology of Infinite-dimensional Lie Algebras*, translated by A. B. Sosinski (Consultants Bureau, New York and London, 1986), p. 33.
[6] B. L. Feigin, Uspekhi. Math. Nauk **35**(2), 223 (1980). Translation in Russian Math. Surveys **35**(2), 239 (1980).
[7] J. Diedonneé, *Special Functions and Linear Representation of Lie Groups* (American Mathematical Society, Providence, RI, 1980).
[8] F. A. Berezin and I. M. Gelfand, Am. Math. Soc. Trans., Ser. 2 **21**, 193 (1962).
[9] G. Szegö, *Orthogonal Polynomials*, Am. Math. Soc. Colloq. Pub., Vol. XXIII (American Mathematical Society, New York, 1959), p. 387.

The Jordan–Schwinger representations of Cayley–Klein groups. I. The orthogonal groups

N. A. Gromov
Komi Scientific Centre, Academy of Sciences of the USSR, Syktyvkar, 167610, USSR

V. I. Man'ko
P. N. Lebedev Physical Institute, Academy of Sciences of the USSR, Moscow, 117333, USSR

(Received 16 May 1989; accepted for publication 25 October 1989)

The Cayley–Klein groups are defined as the groups that are obtained by the contractions and analytical continuations of the orthogonal groups. The Jordan–Schwinger representations of Cayley–Klein groups are discussed based on the mixed sets of creation and annihilation operators of boson or fermion type. The matrix elements of finite group transformations are obtained in the bases of coherent and Fock states.

I. INTRODUCTION

In 1935, Jordan[1,2] introduced the so-called Jordan mapping that is a mapping from a one-particle realization of the kinematic symmetry into field operators of either boson or fermion type. This mapping preserves the commutation relations of matrices. In 1952 Schwinger[3] introduced an original treatment of the rotation group by representing the matrix generators in terms of their bilinear forms with respect to boson annihilation and creation operators. Since this representation is equivalent to the Jordan mapping it is often called[4] the Jordan–Schwinger representation. It has been widely used to provide a treatment of representations of Lie groups. On the other hand, there is the well-developed theory of a many-body quantum system in the second quantized field formalism whose Hamiltonians are multidimensional quadratic in boson or fermion creation and annihilation operators.[5,6] The methods of this theory may be used for the calculation of matrix elements of finite transformations of Lie groups in the bases of coherent and Fock states.[7]

In the present series of papers we shall discuss the Jordan–Schwinger representation of Cayley–Klein groups, i.e., groups obtained from classical (orthogonal, special unitary, and symplectic) one's by all possible contractions and analytical continuations of group parameters. The Inonu–Wigner contractions[8] and analytical continuations are regarded on the basis of a unified description[9–13] with the help of Clifford dual numbers. We consider the Jordan–Schwinger representation of the matrix generators of groups under discussion based on either fermion or boson operators. For the groups obtained from classical one's by only contractions the set of particle operators describing the representation is pure, i.e., all members of the set are either annihilation or creation operators. However, if the groups are obtained by analytical continuations only or both continuations and contractions then the representations are based on mixed sets of annihilation and creation operators.[4] The matrix elements of the Jordan–Schwinger representation of the finite group transformations are calculated in the bases of coherent states, which were introduced by Glauber.[14] In the case of boson representations we use the important property of the coherent states, namely, that the coherent state gives a generating function for discrete Fock states. Then the matrix elements of the finite group transformations in the coherent state bases are the generating function for the matrix elements in Fock bases. The last matrix elements are expressed in terms of Hermite polynomials of several variables with zero arguments.

The outline of the content is as follows: In Sec. II we present a brief account of the (well-known) theory of quantum systems with quadratics in boson and fermion operator Hamiltonians that is adapted to the calculation of the matrix elements of the Jordan–Schwinger representations of Lie groups. In Sec. III we describe the orthogonal Cayley–Klein groups. In Sec. IV we construct the Jordan–Schwinger representation of Cayley–Klein groups of arbitrary dimension and obtain the matrix elements of the finite group transformations in the coherent basis. For the illustration of the developed formalism we regard in Sec. V some groups of low dimensions. The connections of Jordan–Schwinger representations with a stationary quantum system are briefly discussed in the concluding remarks.

II. THE SECOND QUANTIZATION METHOD AND MATRIX ELEMENTS

Let G be a group of N-dimensional matrices with the generators X_k and commutators $[X_k, X_m] = \Sigma_s c_{km}^s X_s$. Define the operators $\hat{X}_k = \Sigma_{p,q} (X_k)_{pq} \hat{a}_p^+ \hat{a}_q$, where $\hat{a}_p^+$ and $\hat{a}_q$ are the boson or fermion creation and annihilation operators, respectively, satisfying the canonical commutation relations

$$\hat{a}_i \hat{a}_k = \epsilon \hat{a}_k \hat{a}_i, \quad \hat{a}_i^+ \hat{a}_k^+ = \epsilon \hat{a}_k^+ \hat{a}_i^+, \qquad \hat{a}_i \hat{a}_k^+ - \epsilon \hat{a}_k^+ \hat{a}_i = \delta_{ik}. \tag{2.1}$$

Here $\epsilon = 1$ in the boson case and $\epsilon = -1$ in the fermion case. Then the operators $\hat{X}_k$ satisfy the commutation relations of the Lie algebra of group G and realize their Jordan–Schwinger representation. The finite group transformation operator $\hat{U}_g(\mathbf{r})$ is connected with the general element $\hat{X}(\mathbf{r}) = \Sigma_k r_k \hat{X}_k$ of the Lie algebra by the exponential map $\hat{U}_g(\mathbf{r}) = \exp(-\hat{X}(\mathbf{r}))$, where r_k are the group parameters.

The representation space H is the state space of the N-

 0022-2488/90/051047-07$03.00

dimensional quantum oscillator. We shall use here as the bases in the representation space the overcomplete family of Glauber coherent states,[14] i.e., the eigenstates of the annihilation operators $\hat{a}_k|\alpha\rangle = \alpha_k|\alpha\rangle$, $\langle\alpha|\alpha\rangle = 1$. In the boson case α_k, $k = 1,2,...,N$ are complex variables and in the fermion case α_k are Grassmann anticommutative variables.[15] A vector $|\mathbf{f}\rangle \in H$ is determined[7] by the analytic (with respect to $\boldsymbol{\alpha}^*$) function $f(\boldsymbol{\alpha}^*)$

$$f(\boldsymbol{\alpha}^*) = \exp(\tfrac{1}{2}|\boldsymbol{\alpha}|^2)\langle\boldsymbol{\alpha}|\mathbf{f}\rangle, \tag{2.2}$$

and the operator $\hat{U}_g$ by the kernel $U(\boldsymbol{\alpha}^*, \boldsymbol{\beta})$,

$$U(\boldsymbol{\alpha}^*, \boldsymbol{\beta}) = \exp(\tfrac{1}{2}|\boldsymbol{\alpha}|^2 + \tfrac{1}{2}|\boldsymbol{\beta}|^2)\langle\boldsymbol{\alpha}|\hat{U}_g|\boldsymbol{\beta}\rangle. \tag{2.3}$$

A transformed vector $|\mathbf{f}'\rangle = \hat{U}_g|\mathbf{f}\rangle$ is represented by the function

$$f'(\boldsymbol{\alpha}^*) = \int U(\boldsymbol{\alpha}^*, \boldsymbol{\beta}) f(\boldsymbol{\beta}^*) d\mu(\boldsymbol{\beta}),$$

$$d\mu(\boldsymbol{\beta}) = \pi^{-N}\exp(-|\boldsymbol{\beta}|^2) d^2\boldsymbol{\beta}, \tag{2.4}$$

where $|\boldsymbol{\alpha}|^2 = \Sigma_{k=1}^N |\alpha_k|^2$, $|\boldsymbol{\beta}|^2 = \Sigma_{k=1}^N |\beta_k|^2$, $d^2\boldsymbol{\beta} = \Pi_{k=1}^N d(\mathrm{Re}\,\beta_k)\cdot d(\mathrm{Im}\,\beta_k)$.

The matrix elements (or the kernel) of the finite group transformation operator $\hat{U}_g(\mathbf{r})$ are obtained by the method of motion integrals.[6] The motion invariants are built with the help of the matrix

$$\Lambda(\mathbf{r}) = \exp(\Sigma B(\mathbf{r})) \equiv \begin{pmatrix} \xi & \eta \\ \eta_1 & \xi_1 \end{pmatrix}, \tag{2.5}$$

where the matrix

$$\Sigma = \begin{pmatrix} 0 & E \\ -\epsilon E & 0 \end{pmatrix},$$

E is an N-dimensional unit matrix and the matrix $B(\mathbf{r})$ is defined by the equation

$$\hat{X}(\mathbf{r}) = \sum_k r_k \hat{X}_k = (\hat{\mathbf{a}}, \hat{\mathbf{a}}^+) B(\mathbf{r}) \begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^+ \end{pmatrix}. \tag{2.6}$$

Here $(\hat{\mathbf{a}}, \hat{\mathbf{a}}^+)$ is the row matrix, $\binom{\hat{\mathbf{a}}}{\hat{\mathbf{a}}^+}$ is the column matrix, and the product in Eq. (2.6) is the ordinary matrix product. We shall use such an agreement throughout the paper. The kernel of the operator $\hat{U}_g(\mathbf{r})$ is given by the following equation[6]:

$$\begin{aligned} U(\boldsymbol{\alpha}^*, \boldsymbol{\beta}, \mathbf{r}) &= (\det \xi)^{-\epsilon/2} \exp\left(-\tfrac{1}{2}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}) R(\mathbf{r}) \begin{pmatrix} \boldsymbol{\alpha}^* \\ \boldsymbol{\beta} \end{pmatrix}\right) \\ &= (\det \xi)^{-\epsilon/2} \exp(-\tfrac{1}{2}\boldsymbol{\alpha}^* \xi^{-1}\zeta\boldsymbol{\alpha}^* + \boldsymbol{\alpha}^*\xi^{-1}\boldsymbol{\beta} \\ &\quad + \tfrac{1}{2}\epsilon\,\boldsymbol{\beta}\eta_1\xi^{-1}\boldsymbol{\beta}), \end{aligned} \tag{2.7}$$

where the $2N$-dimensional matrix $R(\mathbf{r})$ is as follows:

$$R(\mathbf{r}) = \begin{pmatrix} \xi^{-1}\eta & -\xi^{-1} \\ -\epsilon\xi^{-1T} & -\epsilon\eta_1\xi^{-1} \end{pmatrix}. \tag{2.8}$$

In the boson case ($\epsilon = 1$) we also regard the discrete Fock states basis in the representation space H. The Fock state $|\mathbf{n}\rangle$ is the eigenstate of the particle number operator $\hat{a}_k^+ \hat{a}_k|\mathbf{n}\rangle = n_k|\mathbf{n}\rangle$, $\mathbf{n} = (n_1, n_2, ..., n_N)$ and n_k are non-negative integer numbers. We may use the important property of the coherent states, namely, that the coherent state gives the generating function for the Fock states

$$|\boldsymbol{\alpha}\rangle = \exp\left(-\frac{1}{2}|\boldsymbol{\alpha}|^2\right) \sum_{\mathbf{n}=0}^{\infty} \frac{\boldsymbol{\alpha}^{\mathbf{n}}}{(\mathbf{n}!)^{1/2}} |\mathbf{n}\rangle, \tag{2.9}$$

where

$$\mathbf{n}! = \prod_{k=1}^N n_k!, \quad \boldsymbol{\alpha}^{\mathbf{n}} = \prod_{k=1}^N \alpha_k^{n_k}$$

and it follows immediately, that the kernel (2.7) is the generating function for the matrix elements of the operator $\hat{U}_g(\mathbf{r})$ in the Fock states basis

$$U(\boldsymbol{\alpha}^*, \boldsymbol{\beta}, \mathbf{r}) = \sum_{\mathbf{m},\mathbf{n}=0}^{\infty} \frac{\boldsymbol{\alpha}^{*\mathbf{m}}\boldsymbol{\beta}^{\mathbf{n}}}{(\mathbf{m}!\mathbf{n}!)^{1/2}} \langle\mathbf{m}|\hat{U}_g(\mathbf{r})|\mathbf{n}\rangle. \tag{2.10}$$

Multidimensional Hermite polynomials $H_{\mathbf{k}}^{(R)}(\mathbf{x})$ are defined by their generating function as follows[16]:

$$\exp\left(-\frac{1}{2}\mathbf{a}R\mathbf{a} + \mathbf{a}R\mathbf{x}\right) = \sum_{\mathbf{k}=0}^{\infty} \frac{\mathbf{a}^{\mathbf{k}}}{\mathbf{k}!} H_{\mathbf{k}}^{(R)}(\mathbf{x}). \tag{2.11}$$

Then the kernel (2.10) multiplied by $(\det \xi)^{1/2}$ is the generating function for the Hermite polynomials of $2N$ zero variables

$$\begin{aligned} (\det \xi)^{1/2} U(\boldsymbol{\alpha}^*, \boldsymbol{\beta}, \mathbf{r}) &= \exp\left(-\frac{1}{2}(\boldsymbol{\alpha}^*, \boldsymbol{\beta})\right) R(\mathbf{r}) \begin{pmatrix} \boldsymbol{\alpha}^* \\ \boldsymbol{\beta} \end{pmatrix} \\ &= \sum_{\mathbf{m},\mathbf{n}=0}^{\infty} \frac{\boldsymbol{\alpha}^{*\mathbf{m}}\boldsymbol{\beta}^{\mathbf{n}}}{\mathbf{m}!\mathbf{n}!} H_{\mathbf{m},\mathbf{n}}^{(R(\mathbf{r}))}(\mathbf{0}), \end{aligned} \tag{2.12}$$

where the matrix $R(\mathbf{r})$ is given by Eq. (2.8).

III. THE ROTATION GROUPS IN CAYLEY–KLEIN SPACES

It is well known in geometry,[17] that there are 3^n n-dimensional real spaces of constant curvature. Pimenov[18] has given their unified axiomatic description and has built the transformations of the elliptic space into arbitrary space of constant curvature. In accordance with the Erlangen Program, due to F. Klein, each geometry is associated with a motion group. Then the transformations of the geometry induce the transformations of the related motion group. This idea was used to develop the method of transitions between groups,[9–13] that naturally unify both contractions and analytical continuations of groups.

Let us define the fundamental map of the Euclidean space R_{n+1} into the space $R_{n+1}(\mathbf{j})$ as follows:

$$\psi: \quad R_{n+1} \to R_{n+1}(\mathbf{j}),$$

$$\psi x_0 = x_0, \quad \psi x_k = x_k \prod_{m=1}^k j_m, \tag{3.1}$$

where $k = 1,2,...,n$; x_0, x_k are the Cartesian coordinates, $\mathbf{j} = (j_1, j_2, ..., j_n)$, parameter j_k may be equal to the real unit 1, or to the Clifford dual unit ι_k, or to the imaginary unit i.

The dual units are characterized by the following algebraic properties: each of them are not equal to zero $\iota_k \neq 0$; a different dual unit obeys the commutative law of multiplication $\iota_k \cdot \iota_m = \iota_m \cdot \iota_k \neq 0$, $k \neq m$; the product of a dual unit multiplied by itself is always equal to zero $\iota_k^2 = 0$. Division of a real or complex number by a dual unit is not defined. We assume that division of a dual unit by itself is equal to a real unit $\iota_k/\iota_k = 1$ (but not ι_k/ι_m or ι_m/ι_k, $k \neq m$, these constructions are not defined). The last property is equivalent to

the fact that the equation $a\iota_k = b\iota_k$ has only one solution $a = b$ in the real or complex number field. The dual units may be regarded as even products of Grassmann anticommutative numbers and new "b numbers" of the paper[19] are then nothing but Clifford dual numbers.

We define the $(n+1)$-dimensional real Cayley–Klein space $R_{n+1}(\mathbf{j})$ as the $(n+1)$-dimensional vector space with the following metric:

$$\mathbf{x}^2(\mathbf{j}) = x_0^2 + \sum_{k=1}^{n} x_k^2 \prod_{m=1}^{k} j_m^2. \tag{3.2}$$

Then Eq. (3.1) is the mapping of the Euclidean space into the Cayley–Klein spaces. The space of constant curvature $S_n(\mathbf{j})$ is realized on the sphere $S_n(\mathbf{j}) = \{\mathbf{x} \in R_{n+1}(\mathbf{j}) | \mathbf{x}^2(\mathbf{j}) = 1\}$ in the Cayley–Klein space. The set of parameters $\mathbf{j}$ gives all 3^n $(n+1)$-dimensional Cayley–Klein spaces or n-dimensional spaces of constant curvature. It must be emphasized that usually the spaces with identical signatures are not distinguished, i.e., the space $R_3(1,i)$ with the metric $x_0^2 + x_1^2 - x_2^2$ and the space $R_3(i,i)$ with the metric $x_0^2 - x_1^2 + x_2^2$ are the same. We have fixed the numbers of coordinate axes and for us $R_3(1,i)$ and $R_3(i,i)$ are the different spaces.

The rotations of the Cayley–Klein space $R_{n+1}(\mathbf{j})$ form the group $\mathrm{SO}_{n+1}(\mathbf{j})$, which we call the orthogonal Cayley–Klein group. The map (3.1) induces the transformation of the group SO_{n+1} into the Cayley–Klein group $\mathrm{SO}_{n+1}(\mathbf{j})$. The generators $\widetilde{X}_{\mu\nu}$ of SO_{n+1} are the infinitesimal rotations in two-dimensional planes $\{x_\mu, x_\nu\}$, $\mu = 0,1,\ldots,n-1$, $\nu = 1,2,\ldots,n$, $\mu < \nu$. The nonzero elements of the matrix $\widetilde{X}_{\mu\nu}$ are as follows: $(\widetilde{X}_{\mu\nu})_{\mu\nu} = -1$, $(\widetilde{X}_{\mu\nu})_{\nu\mu} = 1$. It is easy to obtain the induced transformation law of the generators of SO_{n+1} under the map (3.1) in the following group[9,10]:

$$X_{\mu\nu}(\mathbf{j}) = \left(\prod_{m=\mu+1}^{\nu} j_m \right) \hat{X}_{\mu\nu}(\to). \tag{3.3}$$

Here, by $\widetilde{X}_{\mu\nu}(\to)$, we denote the transformed generator $\widetilde{X}_{\mu\nu}$ with the following nonzero matrix elements:

$$(\widetilde{X}_{\mu\nu}(\to))_{\mu\nu} = \left(\prod_{m=\mu+1}^{\nu} j_m \right) (\widetilde{X}_{\mu\nu})_{\mu\nu} = -\prod_{m=\mu+1}^{\nu} j_m,$$
$$(\widetilde{X}_{\mu\nu}(\to))_{\nu\mu} = \left(\prod_{m=\mu+1}^{\nu} j_m^{-1} \right) (\widetilde{X}_{\mu\nu})_{\nu\mu} = \prod_{m=\mu+1}^{\nu} j_m^{-1}. \tag{3.4}$$

Then the transformation (3.3) gives for the nonzero matrix elements of the generator $X_{\mu\nu}(\mathbf{j})$

$$(X_{\mu\nu}(\mathbf{j}))_{\mu\nu} = -\prod_{m=\mu+1}^{\nu} j_m^2, \quad (X_{\mu\nu})_{\nu\mu} = 1, \quad \mu < \nu. \tag{3.5}$$

The generators $X_{\mu\nu}(\mathbf{j})$ satisfy the commutation relations[9,10]

$$[X_{\mu_1\nu_1}, X_{\mu_2\nu_2}] = \begin{cases} X_{\nu_1\nu_2} \prod_{m=\mu_1+1}^{\nu_1} j_m^2, & \mu_1 = \mu_2, \quad \nu_1 < \nu_2, \\ X_{\mu_1\mu_2} \prod_{m=\mu_2+1}^{\nu_2} j_m^2, & \mu_1 < \mu_2, \quad \nu_1 = \nu_2, \\ -X_{\mu_1\nu_2}, & \mu_1 < \nu_1 = \mu_2 < \nu_2 \end{cases} \tag{3.6}$$

of the group $\mathrm{SO}_{n+1}(\mathbf{j})$.

Let us observe that when some parameters j_m are equal to the dual units the transformations (3.3) are the multidimensional Inonu–Wigner contractions.[8] Indeed, $\widetilde{X}_{\mu\nu}(\to)$ are the singular transformed generators, the products $\Pi_{m=\mu+1}^{\nu} j_m$ play the role of the zero tending parameters and the resulting generators (3.5) are not singular.

The finite rotation $\Xi(\mathbf{r},\mathbf{j}) = \exp X(\mathbf{r},\mathbf{j})$ corresponds to the general element

$$X(\mathbf{r},\mathbf{j}) = \sum_{\lambda=1}^{n(n+1)/2} r_\lambda X_\lambda(\mathbf{j}), \quad r_\lambda \in R \tag{3.7}$$

of the algebra $\mathrm{so}_{n+1}(\mathbf{j})$. Here λ is in a one-to-one accordance with μ,ν, $\mu < \nu$ due to the equation

$$\lambda = \nu + \mu(n-1) - \mu(\mu-1)/2. \tag{3.8}$$

Due to the Cayley–Hamilton theorem[20] the matrix $\Xi(\mathbf{r},\mathbf{j})$ is expressed algebraically by the matrices $X^m(\mathbf{r},\mathbf{j})$, $m = 0,1,\ldots,n$. The explicit form of the finite rotations $\Xi(\mathbf{r},\mathbf{j})$ can be directly obtained for the groups of low dimensions, namely $\mathrm{SO}_2(j_1)$, $\mathrm{SO}_3(\mathbf{j})$, $\mathrm{SO}_4(\mathbf{j})$.

Combining Eqs. (3.3) and (3.7) we observe that for the imaginary values of some parameters j_k some real group parameters r_λ are imaginary ones, i.e., they are analytically continued from the real number field into the complex one. The orthogonal group SO_{n+1} is transformed by these into some pseudoorthogonal group $\mathrm{SO}(p,q)$. For the dual values of some parameters j_k some real group parameters r_λ are pure dual ones, i.e., they are continued into the dual number field and we have a contraction of the group SO_{n+1}. Thus from the viewpoint of transformations both procedures have the same nature, namely the continuation of group parameters from the real number field into the dual (contraction) or complex ones.

IV. THE JORDAN–SCHWINGER REPRESENTATIONS OF THE ORTHOGONAL CAYLEY–KLEIN GROUPS

Let us define the transformation of annihilation and creation operators induced by the map (3.1) as follows:

$$\psi\hat{\mathbf{a}} = \left(\hat{a}_0, \hat{a}_k \prod_{m=1}^{k} j_m \right), \quad \psi\hat{\mathbf{a}}^+ = \left(\hat{a}_0^+, \hat{a}_k^+ \prod_{m=1}^{k} j_m^{-1} \right), \tag{4.1}$$
$$k = 1,2,\ldots,n,$$

where ψ is identical, i.e., $\psi\hat{\mathbf{a}} = \hat{\mathbf{a}}$, $\psi\hat{\mathbf{a}}^+ = \hat{\mathbf{a}}^+$, when $j_k = 1, \iota_k$. For the imaginary values of parameters $\mathbf{j}$ we use the well-known properties of the annihilation and creation operators: $i\hat{a}_k = \hat{a}_k^+$, $i\hat{a}_k^+ = \epsilon\hat{a}_k$. Then Eq. (4.1) may be written in the form

$$\begin{pmatrix} \psi\hat{\mathbf{a}} \\ \psi\hat{\mathbf{a}}^+ \end{pmatrix} = \begin{pmatrix} \psi_1(\mathbf{j}) & -\psi_2(\mathbf{j}) \\ \epsilon\psi_2(\mathbf{j}) & \psi_1(\mathbf{j}) \end{pmatrix} \begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^+ \end{pmatrix} \equiv \Psi^{-1}(\mathbf{j}) \begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^+ \end{pmatrix}, \tag{4.2}$$

where $\psi_1(\mathbf{j})$, $\psi_2(\mathbf{j})$ are $(n+1)$-dimensional diagonal matrices with the following nonzero matrix elements: $(\psi_1(\mathbf{j}))_{00} = 1$, $(\psi_1(\mathbf{j}))_{kk} = \pm 1$, if $\Pi_{m=1}^{k} j_m = \pm b$ and b is a positive real or dual number, $(\psi_1(\mathbf{j}))_{kk} = 0$ otherwise; $(\psi_2(\mathbf{j}))_{00} = 0$, $(\psi_2(\mathbf{j}))_{kk} = 0$, if $(\psi_1(\mathbf{j}))_{kk} = \pm 1$ and $(\psi_2(\mathbf{j}))_{kk} = \mp 1$, if $\Pi_{m=1}^{k} j_m = \pm ib$. The $2(n+1)$-dimensional matrix $\Psi(\mathbf{j})$ has the property $\Psi(\mathbf{j}) = (\Psi^{-1}(\mathbf{j}))^T$. It is easily shown by direct calculations that the operators

$$\hat{X}_{\mu\nu}(\mathbf{j}) = \psi\hat{\mathbf{a}}^{+}X_{\mu\nu}(\mathbf{j})\psi\hat{\mathbf{a}} \tag{4.3}$$

satisfy the commutation relations (3.7) and hence provide the Jordan–Schwinger representation of the group $SO_{n+1}(\mathbf{j})$.

The general element of the algebra $so_{n+1}(\mathbf{j})$ in Jordan–Schwinger representation is written in the form

$$\hat{X}(\mathbf{r},\mathbf{j}) = \sum_{\lambda=1}^{n(n+1)/2} r_\lambda \hat{X}_\lambda(\mathbf{j}) = \frac{1}{2}(\psi\hat{\mathbf{a}},\psi\hat{\mathbf{a}}^{+})\Delta(\mathbf{r},\mathbf{j})\begin{pmatrix}\psi\hat{\mathbf{a}}\\ \psi\hat{\mathbf{a}}^{+}\end{pmatrix} = \frac{1}{2}(\hat{\mathbf{a}},\hat{\mathbf{a}}^{+})B(\mathbf{r},\mathbf{j})\begin{pmatrix}\hat{\mathbf{a}}\\ \hat{\mathbf{a}}^{+}\end{pmatrix}, \tag{4.4}$$

where the matrix $\Delta(\mathbf{r},\mathbf{j})$ is given by the equation

$$\Delta(\mathbf{r},\mathbf{j}) = \begin{pmatrix}0 & \epsilon X^{T}(\mathbf{r},\mathbf{j})\\ X(\mathbf{r},\mathbf{j}) & 0\end{pmatrix}. \tag{4.5}$$

The nonzero matrix elements of $X(\mathbf{r},\mathbf{j})$ are as follows:

$$(X(\mathbf{r},\mathbf{j}))_{\nu\mu} = r_\lambda, \quad (X(\mathbf{r},\mathbf{j}))_{\mu\nu} = -r_\lambda \prod_{m=\mu+1}^{\nu} j_m^2.$$

Here λ is connected with μ,ν, $\mu<\nu$ by Eq. (3.8). Using Eq. (4.2) we conclude that the matrix $B(\mathbf{r},\mathbf{j})$ is obtained from the matrix $\Delta(\mathbf{r},\mathbf{j})$ by the following transformation:

$$B(\mathbf{r},\mathbf{j}) = \Psi(\mathbf{j})\Delta(\mathbf{r},\mathbf{j})\Psi^{-1}(\mathbf{j}). \tag{4.6}$$

We regard first the Cayley–Klein groups $SO_{n+1}(\mathbf{j})$ that are obtained from SO_{n+1} only by the contractions, i.e., when the parameters j_k are equal to the real unit or to the Clifford dual units, $j_k = 1, \iota_k$, $k = 1,2,\ldots,n$. Then (4.1), (4.2) are identical transformations, $B(\mathbf{r},\mathbf{j}) = \Delta(\mathbf{r},\mathbf{j})$, and we have from Eq. (2.5)

$$\Lambda(\mathbf{r},\mathbf{j}) = \begin{pmatrix}\Xi(\mathbf{r},\mathbf{j}) & 0\\ 0 & \Xi^{T}(-\mathbf{r},\mathbf{j})\end{pmatrix}, \tag{4.7}$$

i.e., $\eta = \eta_1 = 0$, $\xi_1 = \Xi^{T}(-\mathbf{r},\mathbf{j})$, $\xi = \Xi(\mathbf{r},\mathbf{j})$, $\det\xi = \det\Xi(\mathbf{r},\mathbf{j}) = 1$, and $\xi^{-1} = \Xi^{-1}(\mathbf{r},\mathbf{j}) = \Xi(-\mathbf{r},\mathbf{j})$. Here $\Xi(\mathbf{r},\mathbf{j})$ is the finite rotation matrix of the group $SO_{n+1}(\mathbf{j})$. From Eq. (2.7) the kernel of the finite rotation operator $\hat{U}_g(\mathbf{r},\mathbf{j}) = \exp(-\hat{X}(\mathbf{r},\mathbf{j}))$ in a coherent state basis is given by

$$U(\alpha^{*},\beta,\mathbf{r},\mathbf{j}) = \exp(\alpha^{*}\Xi(-\mathbf{r},\mathbf{j})\beta). \tag{4.8}$$

In the boson case ($\epsilon = 1$) this kernel is the generating function for the Hermite polynomials $H_{\mathbf{m},\mathbf{n}}^{(R(\mathbf{r},\mathbf{j}))}(\mathbf{0})$ with the matrix $R(\mathbf{r},\mathbf{j})$ in the form

$$R(\mathbf{r},\mathbf{j}) = \begin{pmatrix}0 & -\Xi(-\mathbf{r},\mathbf{j})\\ -\Xi^{T}(-\mathbf{r},\mathbf{j}) & 0\end{pmatrix}. \tag{4.9}$$

Contractions of $SO_{n+1}(\mathbf{j})$ under dual values of some of the parameters j_k give rise to limit processes in the generating function (4.8) and hence induce a limit processes between Hermite polynomials.[21,22]

Let the Cayley–Klein groups $SO_{n+1}(\mathbf{j})$ be obtained from SO_{n+1} by both contractions and analytical continuations, i.e., $j_k = 1, \iota_k, i$, $k = 1,2,\ldots,n$. Let us introduce the new parameters $\tilde{j}_k$, $\tilde{j}_k = 1, \iota_k$ as follows: $j_k = i\tilde{j}_k$, if $j_k = i$ and $j_k = \tilde{j}_k$, if $j_k = 1, \iota_k$. The reason of such a redefinition of the parameters is to consider explicitly the analytical continuations and give the opportunity of regarding the contractions of these analytical continuated groups. The motion integrals matrix $\tilde{\Lambda}(\mathbf{r},\tilde{\mathbf{j}})$ is obtained from the matrix (4.7) as follows[22]:

$$\tilde{\Lambda}(\mathbf{r},\tilde{\mathbf{j}}) = \Psi(\mathbf{j})\Lambda(\mathbf{r},\tilde{\mathbf{j}})\Psi^{-1}(\mathbf{j}) = \begin{pmatrix}\xi & \eta\\ \eta_1 & \xi_1\end{pmatrix}. \tag{4.10}$$

Using Eqs. (4.2), (4.7), and (4.10), we have

$$\begin{aligned}
\xi &= \psi_1\Xi(\mathbf{r},\tilde{\mathbf{j}})\psi_1 + \psi_2\Xi^{T}(-\mathbf{r},\tilde{\mathbf{j}})\psi_2,\\
\xi_1 &= \psi_2\Xi(\mathbf{r},\tilde{\mathbf{j}})\psi_2 + \psi_1\Xi^{T}(-\mathbf{r},\tilde{\mathbf{j}})\psi_1,\\
\eta &= -\psi_1\Xi(\mathbf{r},\tilde{\mathbf{j}})\psi_2 + \epsilon\psi_2\Xi^{T}(-\mathbf{r},\tilde{\mathbf{j}})\psi_1,\\
\eta_1 &= -\psi_2\Xi(\mathbf{r},\tilde{\mathbf{j}})\psi_1 + \epsilon\psi_1\Xi^{T}(-\mathbf{r},\tilde{\mathbf{j}})\psi_2,
\end{aligned} \tag{4.11}$$

and by Eq. (2.7) obtain the kernel of the finite rotation operator $\hat{U}_g(\mathbf{r},\mathbf{j})$ in a coherent state basis. Note that Eq. (2.7) includes a nonlinear operation of obtaining the inverse matrix ξ^{-1}; therefore for the kernel we do not have the simple equation as Eqs. (4.6) or (4.10).

V. EXAMPLES

To show the effectiveness of the general consideration developed for the Jordan–Schwinger representation of Cayley–Klein groups in the previous sections we shall discuss some groups of low dimensions $SO_2(j_1)$, $SO_3(\mathbf{j})$, $SO_4(\mathbf{j})$ for which it is possible to obtain the explicit form of the finite rotation matrix $\Xi(\mathbf{r},\mathbf{j})$.

A. $SO_2(j_1)$ groups

The map (3.1), namely $\psi x_0 = x_0$, $\psi x_1 = j_1 x_1$, $j_1 = 1,\iota_1,i$, gives the spaces $R_2(j_1)$ with the metric $\mathbf{x}^2(j_1) = x_0^2 + j_1^2 x_1^2$. Here $R_2(1)$ is a Euclidean plane, $R_2(i)$ is the Minkowski (or hyperbolic) plane, and $R_2(\iota_1)$ is the Galilean plane. Then three groups $SO_2(j_1)$ are as follows: $SO_2(1)$ is the usual rotation group on the plane, $SO_2(i)$ is the group of (one-dimensional) Lorentz transformations, and $SO_2(\iota_1)$ is the group of (one-dimensional) Galilean transformations. Equation (3.5) gives the matrix generator of $SO_2(j_1)$ in the form

$$X_{01}(j_1) = \begin{pmatrix}0 & -j_1^2\\ 1 & 0\end{pmatrix}. \tag{5.1}$$

Then the finite rotation matrix $\Xi(r_1,j_1) = \exp(r_1 X_{01}(j_1))$ is easily obtained:

$$\Xi(r_1,j_1) = \begin{pmatrix}\cos j_1 r_1 & -j_1\sin j_1 r_1\\ -j_1^{-1}\sin j_1 r_1 & \cos j_1 r_1\end{pmatrix}. \tag{5.2}$$

A function of dual arguments is defined by its Taylor expansion, therefore $\cos\iota_1 r_1 = 1$, $\sin\iota_1 r_1 = \iota_1 r_1$, and we have

$$\Xi(r_1,\iota_1) = \begin{pmatrix}1 & 0\\ r_1 & 1\end{pmatrix},$$

i.e., the matrix of Galilean transformation.

When $j_1 = 1, \iota_1$, the operator $\hat{X}_{01}(j_1) = \hat{\mathbf{a}}^{+}X_{01}(j_1)\hat{\mathbf{a}} = \hat{a}_1^{+}\hat{a}_0 - j_1^2\hat{a}_0^{+}\hat{a}_1$ provides the Jordan–Schwinger representation of $SO_2(j_1)$ and the kernel of the finite rotation operator $\hat{U}_g(r_1,j_1) = \exp(-r_1 X_{01}(j_1))$ in a coherent state basis is given by, using (4.8) and (5.2),

$$\begin{aligned}
U(\alpha^{*},\beta,r_1,j_1) &= \exp(\alpha^{*}\Xi(-r_1,j_1)\beta)\\
&= \exp((\alpha_0^{*}\beta_0 + \alpha_1^{*}\beta_1)\cos j_1 r_1\\
&\quad - \alpha_1^{*}\beta_0 j_1^{-1}\sin j_1 r_1\\
&\quad + \alpha_0^{*}\beta_1 j_1\sin j_1 r_1).
\end{aligned} \tag{5.3}$$

In the boson case, when α_k^*, β_k are the complex variables, the expression (5.3) is the generating function for the Hermite polynomials of four zero-valued variables. We write some first polynomials

$$\begin{aligned} &H_{0,0;0,0}(r_1,j_1) = 1, \\ &H_{1,1;1,1}(r_1,j_1) = \cos^2 j_1 r_1 - \sin^2 j_1 r_1, \\ &H_{1,0;1,0}(r_1,j_1) = H_{0,1;0,1}(r_1,j_1) = \cos j_1 r_1, \\ &H_{0,1;1,0}(r_1,j_1) = -j_1^{-1} \sin j_1 r_1, \\ &H_{1,0;0,1}(r_1,j_1) = j_1 \sin j_1 r_1. \end{aligned} \tag{5.4}$$

For the Galilean group $SO_2(\iota_1)$ we have $\hat{X}_{01}(\iota_1) = \hat{a}_1^+ \hat{a}_0$ and Eq. (5.3) gives

$$U(\alpha^*, \beta, r_1, \iota_1) = \exp(\alpha_0^* \beta_0 + \alpha_1^* \beta_1 - r_1 \alpha_1^* \beta_0). \tag{5.5}$$

Then the first Hermite polynomials are

$$\begin{aligned} H_{0,0;0,0}(r_1,\iota_1) &= H_{1,1;1,1}(r_1,\iota_1) \\ &= H_{1,0;1,0}(r_1,\iota_1) \\ &= H_{0,1;0,1}(r_1,\iota_1) = 1, \\ H_{0,1;1,0}(r_1,\iota_1) &= -r_1, \quad H_{1,0;0,1}(r_1,\iota_1) = 0. \end{aligned} \tag{5.6}$$

When $j_1 = i$ we introduce the new parameter $\tilde{j}_1$ as $j_1 = i\tilde{j}_1$ and $\tilde{j}_1 = 1, \iota_1$. The case $\tilde{j}_1 = \iota_1$ corresponds to the contraction of the Lorentz group $SO_2(i)$. Equations (4.1) give $\psi\hat{\mathbf{a}} = (\hat{a}_0, \hat{a}_1^+)$, $\psi\hat{\mathbf{a}}^+ = (\hat{a}_0^+, -\epsilon\hat{a}_1)$ and the operator $\hat{X}_{01}(i\tilde{j}_1) = \psi\mathbf{a}^+ X_{01}(ij_1)\psi\mathbf{a} = j_1^2 \hat{a}_0^+ \hat{a}_1^+ - \epsilon\hat{a}_1\hat{a}_0$ provides the Jordan–Schwinger representation of $SO_2(i\tilde{j}_1)$. From Eqs. (4.2) we obtain the matrices ψ_1 and ψ_2 in the form

$$\psi_1(i) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \psi_2(i) = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.7}$$

Replacing parameter j_1 in (5.2) by $i\tilde{j}_1$ we have

$$\Xi(r_1, i\tilde{j}_1) = \begin{pmatrix} \cosh \tilde{j}_1 r_1 & \tilde{j}_1 \sinh \tilde{j}_1 r_1 \\ \tilde{j}_1^{-1} \sinh \tilde{j}_1 r_1 & \cosh \tilde{j}_1 r_1 \end{pmatrix}. \tag{5.8}$$

Then the intermediate matrix $\Lambda(r_1, i\tilde{j}_1)$ is given by Eqs. (4.7) and (5.8). Using it in Eq. (4.10), we obtain the motion integrals matrix $\tilde{\Lambda}(r_1, i\tilde{j}_1)$ of $SO_2(i\tilde{j}_1)$, namely,

$$\begin{aligned} \xi = \xi_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cosh \tilde{j}_1 r_1, \\ \eta &= \begin{pmatrix} 0 & 1 \\ \epsilon & 0 \end{pmatrix} \tilde{j}_1 \sinh \tilde{j}_1 r_1, \\ \eta_1 &= \begin{pmatrix} 0 & \epsilon \\ 1 & 0 \end{pmatrix} \tilde{j}_1^{-1} \sinh \tilde{j}_1 r_1. \end{aligned} \tag{5.9}$$

The kernel of the operator $\hat{U}_g(r_1, i\tilde{j}_1)$ of the finite Lorentz transformation is given by Eq. (2.7) and is as follows:

$$\begin{aligned} U(\alpha^*, \beta, r_1, i\tilde{j}_1) = (\cosh \tilde{j}_1 r_1)^{-\epsilon} &\exp\{(\cosh \tilde{j}_1 r_1)^{-1} \\ &\times [\alpha_0^* \beta_0 + \alpha_1^* \beta_1 + (\beta_0 \beta_1 - \tilde{j}_1^2 \alpha_0^* \alpha_1^*) \\ &\times \tilde{j}_1^{-1} \sinh \tilde{j}_1 r_1]\}. \end{aligned} \tag{5.10}$$

Under the contraction ($\tilde{j}_1 = \iota_1$) of the Lorentz group we have for the kernel

$$U(\alpha^*, \beta, r_1, i\iota_1) = \exp(\alpha_0^* \beta_0 + \alpha_1^* \beta_1 + r_1 \beta_0 \beta_1). \tag{5.11}$$

Comparing the last expression with Eq. (5.5) we conclude that they are different though it follows from Eq. (3.5) the group $SO_2(\iota_1)$ and $SO_2(i\iota_1)$ are the same, namely Galilean group. It is the particular case of the general situation[22]: if a Cayley–Klein group $SO_{n+1}(\mathbf{j})$ is obtained from SO_{n+1} by a k-dimensional contraction, i.e., if k parameters $\mathbf{j}$ are equal to the dual units, then there are 2^k different Jordan–Schwinger representations of $SO_{n+1}(\mathbf{j})$. The connections between such representations are still under investigation.

B. $SO_3(\mathbf{j})$ groups

The map (3.1), namely $\psi x_0 = x_0$, $\psi x_1 = j_1 x_1$, $\psi x_2 = j_1 j_2 x_2$, $j_1 = 1, \iota_1, i$, $j_2 = 1, \iota_2, i$, gives the nine Cayley–Klein spaces $R_3(j_1, j_2) = R_3(\mathbf{j})$ with the metric $\mathbf{x}^2(\mathbf{j}) = x_0^2 + j_1^2 x_1^2 + j_1^2 j_2^2 x_2^2$. The nine geometries of the planes of constant curvature are realized[18] on the spheres $S_2(\mathbf{j}) = \{\mathbf{x} | \mathbf{x}^2(\mathbf{j}) = 1\}$ in the spaces $R_3(\mathbf{j})$. These geometries are as follows: $S_2(1,1)$—elliptic; $S_2(\iota_1,1)$—Euclidean; $S_2(i,1)$—Lobachevski (or hyperbolic); $S_2(1,\iota_2)$—semielliptic (or co-Euclidean); $S_2(\iota_1,\iota_2)$—Galilean; $S_2(i,\iota_2)$—semihyperbolic (or co-Minkowski); $S_2(1,i)$—anti-de Sitter; $S_2(\iota_1,i)$—Minkowski; $S_2(i,i)$—de Sitter. The rotation group $SO_3(\mathbf{j})$ is isomorphic to the motion group of the geometry $S_2(\mathbf{j})$. We shall call the Cayley–Klein group $SO_3(\mathbf{j})$ by the name of appropriate geometry.

Equation (3.5) gives the matrix generators of $SO_3(\mathbf{j})$ in the form

$$X_1(\mathbf{j}) = \begin{pmatrix} 0 & -j_1^2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2(\mathbf{j}) = \begin{pmatrix} 0 & 0 & -j_1^2 j_2^2 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$X_3(\mathbf{j}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -j_2^2 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5.12}$$

The set of generators satisfy the commutation relations

$$[X_1, X_2] = j_1^2 X_3, \quad [X_2, X_3] = j_2^2 X_1, \quad [X_3, X_1] = X_2. \tag{5.13}$$

In accordance with Eq. (3.8) we denote the generators as follows: $X_1 = X_{01}$, $X_2 = X_{02}$, $X_3 = X_{12}$. To the general element

$$\begin{aligned} X(\mathbf{r},\mathbf{j}) &= \sum_{\lambda=1}^{3} r_\lambda X_\lambda(\mathbf{j}) \\ &= \begin{pmatrix} 0 & -j_1^2 r_1 & -j_1^2 j_2^2 r_2 \\ r_1 & 0 & -j_2^2 r_3 \\ r_2 & r_3 & 0 \end{pmatrix} \end{aligned} \tag{5.14}$$

of the algebra $so_3(\mathbf{j})$ corresponds to the finite rotation of the group $SO_3(\mathbf{j})$

$$\begin{aligned} \Xi(\mathbf{r},\mathbf{j}) = \exp X(\mathbf{r},\mathbf{j}) &= E \cos r + X(\mathbf{r},\mathbf{j})(\sin r/r) \\ &+ X'(\mathbf{r},\mathbf{j})[(1 - \cos r)/r^2], \end{aligned} \tag{5.15}$$

where

$$X'(\mathbf{r},\mathbf{j}) = \begin{pmatrix} j_2^2 r_3^2 & -j_1^2 j_2^2 r_2 r_3 & j_1^2 j_2^2 r_1 r_3 \\ -j_2^2 r_2 r_3 & j_1^2 j_2^2 r_2^2 & -j_1^2 j_2^2 r_1 r_2 \\ r_1 r_3 & -j_1^2 r_1 r_2 & j_1^2 r_1^2 \end{pmatrix}, \tag{5.16}$$

$$r^2(\mathbf{j}) = j_1^2 r_1^2 + j_1^2 j_2^2 r_2^2 + j_2^2 r_3^2. \tag{5.17}$$

We shall discuss only contractions of the rotation group SO_3, i.e., $j_1 = 1$, ι_1, $j_2 = 1$, ι_2. Then the transformations (4.1), (4.2) are identical and the Jordan–Schwinger representation of the generators (5.12) is given by the operators

$$\hat{X}_1 = \hat{a}_1^+ \hat{a}_0 - j_1^2 \hat{a}_0^+ \hat{a}_1, \quad \hat{X}_2 = \hat{a}_2^+ \hat{a}_0 - j_1^2 j_2^2 \hat{a}_0^+ \hat{a}_2, \tag{5.18}$$

$$\hat{X}_3 = \hat{a}_2^+ \hat{a}_1 - j_2^2 \hat{a}_1^+ \hat{a}_2,$$

that satisfy the commutation relations (5.13). The kernel of the finite rotation operator $\hat{U}_g(\mathbf{r},\mathbf{j}) = \exp(-\hat{X}(\mathbf{r},\mathbf{j}))$ in a coherent state basis is given by, using (4.8) and (5.14)–(5.17),

$$\begin{aligned} U(\alpha^*, \beta, \mathbf{r}, \mathbf{j}) &= \exp\{\alpha^* \Xi(-\mathbf{r},\mathbf{j})\, \beta\} \\ &= \exp\Big\{\cos r \sum_{k=0}^{2} \alpha_k^* \beta_k - [(\sin r)/r][r_1(\alpha_1^* \beta_0 - j_1^2 \alpha_0^* \beta_1) + r_2(\alpha_2^* \beta_0 - j_1^2 j_2^2 \alpha_0^* \beta_2) \\ &\quad + r_3(\alpha_3^* \beta_1 - j_2^2 \alpha_1^* \beta_2) + [(1-\cos r)/r^2][j_2^2 r_3^2 \alpha_0^* \beta_0 + j_1^2 j_2^2 r_2^2 \alpha_1^* \beta_1 + j_1^2 r_1^2 \alpha_2^* \beta_2 \\ &\quad - j_2^2 r_2 r_3(\alpha_1^* \beta_0 + j_1^2 \alpha_0^* \beta_1) + r_1 r_3(\alpha_2^* \beta_0 + j_1^2 j_2^2 \alpha_0^* \beta_2) - j_1^2 r_1 r_2(\alpha_2^* \beta_1 + j_2^2 \alpha_1^* \beta_2)]\Big\}. \end{aligned} \tag{5.19}$$

In the boson case Eq. (5.19) is the generating function for the Hermite polynomials of six zero-valued variables.

C. $SO_4(\mathbf{j})$ groups

The map (3.1) gives the $3^3 = 27$ Cayley–Klein spaces $R_4(\mathbf{j})$, $\mathbf{j} = (j_1 j_2 j_3)$, $j_k = 1, \iota_k, i$, $k = 1,2,3$ with the metric $\mathbf{x}^2(\mathbf{j}) = x_0^2 + j_1^2 x_1^2 + j_1^2 j_2^2 x_2^2 + j_1^2 j_2^2 j_3^2 x_3^2$. The three-dimensional spaces of constant curvature are realized on the spheres $S_3(\mathbf{j}) = \{\mathbf{x} | \mathbf{x}^2(\mathbf{j}) = 1\}$ of the unit real radius in $R_4(\mathbf{j})$. Some of these spaces are well known, for example, $S_3(i,1,1)$—Lobachevski; $S_3(\iota_1,1,1)$—Euclidean; $S_3(\iota_1,i,1)$—Minkowski; $S_3(\iota_1,\iota_2,1)$—Galilean, and some do not have special names.

The six matrix generators of $SO_4(\mathbf{j})$ are given by Eq. (3.5) as follows:

$$X_1 = X_{01} = \begin{pmatrix} 0 & -j_1^2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$X_2 = X_{02} = \begin{pmatrix} 0 & 0 & -j_1^2 j_2^2 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$X_3 = X_{03} = \begin{pmatrix} 0 & 0 & 0 & -j_1^2 j_2^2 j_3^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

$$X_4 = X_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -j_2^2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$X_5 = X_{13} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -j_2^2 j_3^2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

$$X_6 = X_{23} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -j_3^2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{5.20}$$

and due to Eq. (3.6) satisfy the commutation relations

$$\begin{aligned} &[X_1,X_2] = j_1^2 X_4, \quad [X_1,X_3] = j_1^2 X_5, \quad [X_2,X_3] = j_1^2 j_2^2 X_6, \\ &[X_1,X_4] = -X_2, \quad [X_1,X_5] = -X_3, \quad [X_2,X_6] = -X_3, \\ &[X_2,X_4] = j_2^2 X_1, \quad [X_3,X_5] = j_2^2 j_3^2 X_1, \quad [X_3,X_6] = j_3^2 X_2, \\ &[X_4,X_5] = j_2^2 X_6, \quad [X_4,X_6] = -X_5, \quad [X_5,X_6] = j_3^2 X_4. \end{aligned} \tag{5.21}$$

Let us introduce the new denominations for the group parameters, namely, $r_4 = -s_3$, $r_5 = s_2$, $r_6 = -s_1$. Then the general element of the algebra $so_4(\mathbf{j})$ is

$$\begin{aligned} X(\mathbf{r},\mathbf{s},\mathbf{j}) &= \sum_{k=1}^{3} r_k X_k(\mathbf{j}) - s_3 X_4(\mathbf{j}) + s_2 X_5(\mathbf{j}) - s_1 X_6(\mathbf{j}) \\ &= \begin{pmatrix} 0 & -j_1^2 r_1 & -j_1^2 j_2^2 r_2 & -j_1^2 j_2^2 j_3^2 r_3 \\ r_1 & 0 & j_2^2 s_3 & -j_2^2 j_3^2 s_2 \\ r_2 & -s_3 & 0 & j_3^2 s_1 \\ r_3 & s_2 & -s_1 & 0 \end{pmatrix} \end{aligned} \tag{5.22}$$

and the finite rotation matrix of the group $SO_4(\mathbf{j})$ is given by the following equation:

$$\begin{aligned} \Xi(\mathbf{r},\mathbf{s},\mathbf{j}) &= \{(r^2+s^2)^2 - 4 j_1^2 j_3^2 (\mathbf{r},\mathbf{s})^2\}^{-1/2} \\ &\quad \times (E \cdot A + X \cdot B + (\mathbf{r},\mathbf{s}) \cdot X_1 \cdot C + X^2 \cdot D), \end{aligned} \tag{5.23}$$

where E is a four-dimensional unit matrix, the matrix X is given by Eq. (5.22), the matrices X_1 and X^2 are in the form

$$X_1 = \begin{pmatrix} 0 & -j_1^2 j_3^2 s_1 & -j_1^2 j_2^2 j_3^2 s_2 & -j_1^2 j_2^2 j_3^2 s_3 \\ j_3^2 s_1 & 0 & j_1^2 j_2^2 j_3^2 r_3 & -j_1^2 j_2^2 j_3^2 r_2 \\ j_3^2 s_2 & -j_1^2 j_3^2 r_3 & 0 & j_1^2 j_3^2 r_1 \\ s_3 & j_1^2 r_2 & -j_1^2 r_1 & 0 \end{pmatrix}, \tag{5.24}$$

$$X^2=\begin{pmatrix} -r^2 & j_1^2 j_2^2(\mathbf{r}\times\mathbf{s})_1 & j_1^2 j_2^2(\mathbf{r}\times\mathbf{s})_2 & j_1^2 j_2^2 j_3^2(\mathbf{r}\times\mathbf{s})_3 \\ j_2^2(\mathbf{r}\times\mathbf{s})_1 & j_3^2 s_1^2 - j_1^2 r_1^2 - s^2 & j_2^2 j_3^2 s_1 s_2 - j_1^2 j_2^2 r_1 r_2 & j_2^2 j_3^2 s_1 s_3 - j_1^2 j_2^2 j_3^2 r_1 r_3 \\ (\mathbf{r}\times\mathbf{s})_2 & j_3^2 s_1 s_2 - j_1^2 r_1 r_2 & j_2^2 j_3^2 s_2^2 - j_1^2 j_2^2 r_2^2 - s^2 & j_2^2 j_3^2 s_2 s_3 - j_1^2 j_2^2 j_3^2 r_2 r_3 \\ (\mathbf{r}\times\mathbf{s})_3 & s_1 s_3 - j_1^2 r_1 r_3 & j_2^2 s_2 s_3 - j_1^2 j_2^2 r_2 r_3 & j_2^2 s_3^2 - j_1^2 j_2^2 j_3^2 r_3^2 - s^2 \end{pmatrix}.$$

The functions A,B,C,D in Eq. (5.23) are equal to

$$\begin{aligned} A &= z_1 \cos\sqrt{-z_2} - z_2 \cos\sqrt{-z_1}, \\ B &= z_1 \frac{\sin\sqrt{-z_1}}{\sqrt{-z_1}} - z_2 \frac{\sin\sqrt{-z_2}}{\sqrt{-z_2}}, \\ C &= \frac{\sin\sqrt{-z_1}}{\sqrt{-z_1}} - \frac{\sin\sqrt{-z_2}}{\sqrt{-z_2}}, \\ D &= \cos\sqrt{-z_1} - \cos\sqrt{-z_2}, \end{aligned} \tag{5.25}$$

where

$$z_{1,2} = -(r^2+s^2 \mp ((r^2+s^2)^2 - 4j_1^2 j_3^2(\mathbf{r},\mathbf{s}))^{1/2})/2. \tag{5.26}$$

We use the following denominations:

$$\begin{aligned} &r^2 = j_1^2 r_1^2 + j_1^2 j_2^2 r_2^2 + j_1^2 j_2^2 j_3^2 r_3^2, \quad s^2 = j_3^2 s_1^2 + j_2^2 j_3^2 s_2^2 + j_2^2 s_3^2, \\ &(\mathbf{r},\mathbf{s}) = r_1 s_1 + j_2^2 (r_2 s_2 + r_3 s_3), \quad (\mathbf{r}\times\mathbf{s})_1 = r_2 s_3 - j_3^2 r_3 s_2, \\ &(\mathbf{r}\times\mathbf{s})_2 = j_3^2 r_3 s_1 - r_1 s_3, \quad (\mathbf{r}\times\mathbf{s})_3 = r_1 s_2 - r_2 s_1. \end{aligned} \tag{5.27}$$

In the case of only contractions of the orthogonal group SO_4, i.e., $j_k = 1, \iota_k$, $k = 1,2,3$, the Jordan–Schwinger representation of the generators (5.20) is given by the expression $\hat{X}_k(\mathbf{j}) = \hat{\mathbf{a}}^+ X_k(\mathbf{j}) \hat{\mathbf{a}}$ and the kernel of the finite rotation operator is obtained by Eq. (4.8) with help of the matrix (5.23). We shall not write out this kernel.

VI. CONCLUDING REMARKS

In the previous sections we have defined the Cayley–Klein groups $SO_{n+1}(\mathbf{j})$ and have discussed their Jordan–Schwinger representations. Here we point out the connection of these representations with a properties of quantum systems. The general element $\hat{X}(\mathbf{r},\mathbf{j})$ of the algebra $so_{n+1}(\mathbf{j})$ in the Jordan–Schwinger representation is the linear function of the second quantized generators of $SO_{n+1}(\mathbf{j})$, therefore the replacement of the group parameters $\mathbf{r}$ by $(i/\hbar)t\mathbf{r}$, where t is the time variable, transform the finite rotation operator $\hat{U}_g = \exp(-\hat{X}(\mathbf{r},\mathbf{j}))$ into the evolution operator $\hat{U} = \exp(-(i/\hbar)t\hat{H}) = \exp(-(i/\hbar)t\hat{X}(\mathbf{r},\mathbf{j}))$ of the quantum system with the Hamiltonian $\hat{H} = \hat{X}(\mathbf{r},\mathbf{j})$. The last quantum systems we call the group quantum systems.[23] In the case of stationary systems (when the group parameters r do not depend on the time t) the kernel of the finite rotation operator is transformed into the matrix elements of the evolution operator (or Green's function) of corresponding quantum system as follows: $G(\alpha^*,\beta,\mathbf{r},\mathbf{j},t) = U(\alpha^*,\beta,(i/\hbar)t\mathbf{r},\mathbf{j})$. Thus, investigating the Jordan–Schwinger representations of the set of the groups $SO_{n+1}(\mathbf{j})$, we simultaneously investigate the set of the stationary quantum systems, which are corresponded to $SO_{n+1}(\mathbf{j})$.

[1] P. Jordan, Z. Phys. **94**, 531 (1935).
[2] L. C. Biedenharn and J. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, London, 1981).
[3] J. Schwinger, "On angular momentum," in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965).
[4] S. K. Kim, J. Math. Phys. **28**, 2540 (1987).
[5] V. V. Dodonov and V. I. Man'ko, Trudy FIAN (P. N. Lebedev Physical Institute Proceedings) **183**, 182 (1987) (in Russian).
[6] I. A. Malkin and V. I. Man'ko, *Dynamical Symmetries and Coherent States of Quantum Systems* (Nauka, Moscow, 1979) (in Russian).
[7] V. I. Man'ko and D. A. Trifonov, in *Group Theoretical Methods in Physics* (Harwood Academic, London, 1987), Vol. 3, p. 795.
[8] E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[9] N. A. Gromov, Teor. Mat. Fiz. **49**, 210 (1981) (in Russian).
[10] N. A. Gromov, Mat. Zametki **32**, 355 (1982) (in Russian).
[11] N. A. Gromov, Preprint Series "Scientific Reports," No. 95, Komi Branch Acad. Sci. USSR, Syktyvkar (1984) (in Russian).
[12] Reference 11, No. 126, (1985) (in Russian).
[13] N. A. Gromov, in *Group Theoretical Methods in Physics* (Nauka, Moscow, 1986), Vol. 2, p. 183 (in Russian).
[14] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[15] F. A. Beresin, *Second Quantization Method* (Nauka, Moscow, 1965) (in Russian).
[16] H. Bateman and A. Erde'lyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2.
[17] I. M. Jaglom, B. A. Rosenfled, and E. U. Jasinskaja, Usp. Mat. Nauk **19**(5), 51 (1964) (in Russian).
[18] R. I. Pimenov, Litovski Mat. Sbornik **5**, 457 (1965) (in Russian).
[19] X.-L.Quan, L.-J. Wang, and H.-S. Song, J. Math. Phys. **29**, 1780 (1988).
[20] G. A. Korn and T. M. Korn, *Mathematical Handbook* (McGraw-Hill, New York, 1961).
[21] N. A. Gromov and V. I. Man'ko, Preprint FIAN No. 155 (1988) (in Russian).
[22] V. I. Man'ko and N. A. Gromov, Preprint FIAN No. 196 (1988) (in Russian).
[23] N. A. Gromov and V. I. Man'ko, Preprint FIAN No. 245 (1988) (in Russian).

The Jordan–Schwinger representations of Cayley–Klein groups. II. The unitary groups

N. A. Gromov
Komi Scientific Centre, Academy of Sciences of the USSR, Syktyvkar, GSP, 167610, USSR

V. I. Man'ko
P. N. Lebedev Physical Institute, Academy of Sciences of the USSR, Moscow, 117333, USSR

(Received 12 May 1989; accepted for publication 25 October 1989)

The unitary Cayley–Klein groups are defined as the groups that are obtained by the contractions and analytical continuations of the special unitary groups. The Jordan–Schwinger representations of the groups under consideration are discussed based on the mixed sets of creation and annihilation operators of boson or fermion type. The matrix elements of finite group transformations are obtained in the bases of coherent states.

I. INTRODUCTION

The purpose of this series of papers is to present the Jordan–Schwinger representations[1–3] of the groups that are obtained from the classical one's by the contractions and analytical continuations of the group parameters. The Inonu–Wigner contractions[4] and analytical continuations of the groups are regarded on the base of the unified description of the groups.[5,6] The main feature of such an approach is the transformations of the well-known algebraic constructions (generators, commutators, etc.) of the classical groups into the algebraic constructions of the groups under consideration. The Jordan–Schwinger representations of the groups are built by both boson and fermion creation and annihilation operators. The matrix elements of the finite group transformations in the Glauber coherent state basis[7] are calculated with the help of the methods of the well-developed theory of the quantum systems that are quadratic in creation and annihilation operators Hamiltonians.[8–10]

The first paper in this series[11] (hereafter referred to as I and whose equations will be subsequently quoted by their number preceded by I) was devoted to a consideration of the Jordan–Schwinger representations of the orthogonal Cayley–Klein groups that were obtained from the special orthogonal groups by the contractions and analytical continuations of the group parameters. In this paper we discuss the case of the special unitary groups. In Sec. II we describe the unitary Cayley–Klein groups and construct their Jordan–Schwinger representations. In Sec. III we regard in detail $SU_2(j_1)$ and $SU_3(j_1,j_2)$ groups for which the calculations are made in explicit form.

II. THE SPECIAL UNITARY GROUPS IN COMPLEX CAYLEY–KLEIN SPACES

Let us define the map of the $(n+1)$-dimensional complex space C_{n+1} into the complex space $C_{n+1}(\mathbf{j})$ as follows:

$$\psi: \quad C_{n+1} \to C_{n+1}(\mathbf{j}),$$
$$\psi z_0 = z_0, \quad \psi z_k = z_k \prod_{m=1}^{k} j_m, \tag{2.1}$$

where $k = 1,2,...,n$; z_0, z_k are the complex Cartesian coordinates, $\mathbf{j} = (j_1, j_2, ..., j_n)$, and each of the parameters j_k may be equal to the real unit 1, the Clifford dual unit ι_k, or the imaginary unit i. The dual numbers are not often used and we briefly review their algebraic properties. Each of the dual units is not equal to zero $\iota_k \neq 0$; different dual units obey the commutative law of multiplication $\iota_k \iota_m = \iota_m \iota_k \neq 0$, $k \neq m$; the square of a dual unit is always equal to zero $\iota_k^2 = 0$. Division of a real or complex number by a dual unit is not defined, but division of a dual unit by itself is equal to a real unit $\iota_k/\iota_k = 1$. A function of a dual argument is defined by its Taylor expansion. The quadratic form $(\mathbf{z},\mathbf{z}) = \Sigma_{m=0}^{n} |z_m|^2$ of C_{n+1} transforms under the map (2.1) into the following quadratic form of the complex Cayley–Klein space $C_{n+1}(\mathbf{j})$ (Ref. 12):

$$(\mathbf{z},\mathbf{z}) = |z_0|^2 + \sum_{k=1}^{n} |z_k|^2 \prod_{m=1}^{k} j_m^2, \tag{2.2}$$

where $z_k = (x_k^2 + y_k^2)^{1/2}$ is the absolute value of the complex number $z_k = x_k + iy_k$ and $\mathbf{z} = (z_0, z_1, ..., z_n)$ is the complex vector.

The unitary Cayley–Klein group $SU_{n+1}(\mathbf{j})$ is defined as the group that keeps invariant the quadratic form (2.2) transformations in the space $C_{n+1}(\mathbf{j})$. The map (2.1) induces the transformation of the special unitary group SU_{n+1} into the group $SU_{n+1}(\mathbf{j})$. All $(n+1)^2 - 1$ generators of SU_{n+1} are Hermite matrices. The commutation relations for these Hermite generators are very complicated and usually the matrix generators $\tilde{Y}_{km}$, $k,m = 0,1,...,n$ of the general linear group $GL_{n+1}(R)$ are used.[13] The only nonzero matrix element of $\tilde{Y}_{km}$ is $(\tilde{Y}_{km})_{km} = 1$. The generators $\tilde{Y}$ of $GL_{n+1}(R)$ satisfy the commutation relations

$$[\tilde{Y}_{km}, \tilde{Y}_{pq}] = \delta_{mp} \tilde{Y}_{kq} - \delta_{kq} \tilde{Y}_{mp}. \tag{2.3}$$

The independent Hermite generators of SU_{n+1} are defined by the equations

$$\tilde{Q}_{\mu\nu} = (i/2)(\tilde{Y}_{\mu\nu} + \tilde{Y}_{\nu\mu}), \quad \tilde{L}_{\mu\nu} = (1/2)(\tilde{Y}_{\nu\mu} - \tilde{Y}_{\mu\nu}),$$
$$\tilde{P}_k = (i/2)(\tilde{Y}_{k-1,k-1} - \tilde{Y}_{kk}), \tag{2.4}$$

where $\mu = 0,1,...,n-1$, $\nu = \mu+1, \mu+2, ..., n$. The matrix

 0022-2488/90/051054-06$03.00

generators $\tilde{Y}$ are transformed under the map (2.1) as follows:

$$\tilde{Y}_{\mu\nu}(\mathbf{j}) = \Big(\prod_{m=\mu+1}^{\nu} j_m\Big)\tilde{Y}_{\mu\nu}(\to) = \Big(\prod_{m=\mu+1}^{\nu} j_m^2\Big)\tilde{Y}_{\mu\nu},$$
$$\tilde{Y}_{\nu\mu}(\mathbf{j}) = \Big(\prod_{m=\mu+1}^{\nu} j_m\Big)\tilde{Y}_{\nu\mu}(\to) = \tilde{Y}_{\nu\mu}, \quad \mu<\nu,$$
$$Y_{kk}(\mathbf{j}) = \tilde{Y}_{kk}, \tag{2.5}$$

where $\tilde{Y}_{\mu\nu}(\to)$ and $\tilde{Y}_{\nu\mu}(\to)$ denote the transformed generators $\tilde{Y}_{\mu\nu}, \tilde{Y}_{\nu\mu}$ with the following nonzero matrix elements:

$$(\tilde{Y}_{\mu\nu}(\to))_{\mu\nu} = \Big(\prod_{m=\mu+1}^{\nu} j_m\Big)(\tilde{Y}_{\mu\nu})_{\mu\nu} = \prod_{m=\mu+1}^{\nu} j_m,$$
$$(\tilde{Y}_{\nu\mu}(\to))_{\nu\mu} = \Big(\prod_{m=\mu+1}^{\nu} j_m^{-1}\Big)(\tilde{Y}_{\nu\mu})_{\nu\mu} = \prod_{m=\mu+1}^{\nu} j_m^{-1}, \tag{2.6}$$

The generators (2.5) satisfy the commutation relations

$$[Y_{km},Y_{pq}] = \prod_{l=l_1}^{l_2} j_l \prod_{l=l_3}^{l_4} j_l\Big(\delta_{mp}Y_{kq}\prod_{l=l_5}^{l_6} j_l^{-1} - \delta_{kq}Y_{mp}\prod_{l=l_7}^{l_8} j_l^{-1}\Big), \tag{2.7}$$

where $l_1 = 1+\min(k,m)$, $l_2 = \max(k,m)$, $l_3 = 1+\min(p,q)$, $l_4 = \max(p,q)$, $l_5 = 1+\min(k,q)$, $l_6 = \max(k,q)$, $l_7 = 1+\min(m,p)$, $l_8 = \max(m,p)$. The same laws of transformations as in Eq. (2.5) are held for the Hermite generators (2.4). Then we obtain the matrix generators of the unitary Cayley–Klein group $\mathrm{SU}_{n+1}(\mathbf{j})$ in the form

$$Q_{\mu\nu}(\mathbf{j}) = \Big(\prod_{m=\mu+1}^{\nu} j_m\Big)\tilde{Q}_{\mu\nu}(\to)$$
$$= (i/2)(Y_{\nu\mu}(\mathbf{j}) + Y_{\mu\nu}(\mathbf{j}))$$
$$= \frac{i}{2}\Big(\tilde{Y}_{\nu\mu} + \tilde{Y}_{\mu\nu}\prod_{m=\mu+1}^{\nu} j_m^2\Big),$$
$$L_{\mu\nu}(\mathbf{j}) = \Big(\prod_{m=\mu+1}^{\nu} j_m\Big)\tilde{L}_{\mu\nu}(\to)$$
$$= \tfrac{1}{2}(Y_{\nu\mu}(\mathbf{j}) - Y_{\mu\nu}(\mathbf{j}))$$
$$= \frac{1}{2}\Big(\tilde{Y}_{\nu\mu} - \tilde{Y}_{\mu\nu}\prod_{m=\mu+1}^{\nu} j_m^2\Big),$$
$$P_k(\mathbf{j}) = \tilde{P}_k = (i/2)(\tilde{Y}_{k-1,k-1} - \tilde{Y}_{kk}), \quad k = 1,2,\ldots,n. \tag{2.8}$$

The commutation relations of these generators may be derived with the help of Eq. (2.7), but they are very cumbersome and we do not write them here.

The finite group transformation

$$\Xi(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) = \exp Z(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) \tag{2.9}$$

of $\mathrm{SU}_{n+1}(\mathbf{j})$ correspond to the general element

$$Z(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) = \sum_{\lambda=1}^{n(n+1)/2}(r_\lambda Q_\lambda(\mathbf{j}) + s_\lambda L_\lambda(\mathbf{j})) + \sum_{k=1}^{n} w_k P_k(\mathbf{j}) \tag{2.10}$$

of the algebra $\mathrm{su}_{n+1}(\mathbf{j})$, where r_λ, s_λ, w_k are real group parameters and index λ is connected with the indices μ, ν by Eq. (I.3.8). The explicit form of the finite transformation $\Xi(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j})$ may be obtained for the groups $\mathrm{SU}_2(j_1)$ and $\mathrm{SU}_3(j_1,j_2)$.

The Jordan–Schwinger representations of $\mathrm{SU}_{n+1}(\mathbf{j})$ are provided by the operators

$$\hat{Q}_{\mu\nu}(\mathbf{j}) = \psi\hat{\mathbf{a}}^+ Q_{\mu\nu}(\mathbf{j})\psi\hat{\mathbf{a}}, \quad \hat{L}_{\mu\nu}(\mathbf{j}) = \psi\hat{\mathbf{a}}^+ L_{\mu\nu}(\mathbf{j})\psi\hat{\mathbf{a}},$$
$$P_k(\mathbf{j}) = \psi\hat{\mathbf{a}}^+ P_k(\mathbf{j})\psi\hat{\mathbf{a}}, \quad \mu<\nu, \quad k = 1,2,\ldots,n, \tag{2.11}$$

where the transformed sets $\psi\hat{\mathbf{a}}^+$, $\psi\hat{\mathbf{a}}$ of the creation and annihilation operators are given by Eqs. (I.4.1) and (I.4.2). Indeed, it is verified by direct calculations that the operators $\hat{Y}(\mathbf{j}) = \psi\hat{\mathbf{a}}^+ Y(\mathbf{j})\psi\hat{\mathbf{a}}$ satisfy the commutation relations (2.7) and we may conclude that the operators (2.11) satisfy the commutation relations of the group $\mathrm{SU}_{n+1}(\mathbf{j})$. The finite group transformation is represented by the operator

$$\hat{U}_g(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) = \exp(-\hat{Z}(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j})), \tag{2.12}$$

where the operator $\hat{Z}$ is given by Eq. (2.10) with the generators $Q_\lambda(\mathbf{j})$, $L_\lambda(\mathbf{j})$, $P_k(\mathbf{j})$ replaced by the operators $\hat{Q}_\lambda(\mathbf{j})$, $\hat{L}_\lambda(\mathbf{j})$, and $\hat{P}_k(\mathbf{j})$, respectively. The kernel of the operator $\hat{U}_g(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j})$ in a coherent state basis is obtained quite analogous to the case of orthogonal groups.[11] When the Cayley–Klein groups $\mathrm{SU}_{n+1}(\mathbf{j})$ are obtained from SU_{n+1} by only contractions ($j_k = 1, \iota_k$, $k = 1,2,\ldots,n$) this kernel is given by

$$U(\alpha^*,\beta,\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) = \exp(\alpha^*\Xi(-\mathbf{r},-\mathbf{s},-\mathbf{w},\mathbf{j})\beta), \tag{2.13}$$

[compare with Eq. (I.4.8)]. When the groups $\mathrm{SU}_{n+1}(\mathbf{j})$ are obtained from SU_{n+1} by both contractions and analytical continuations the kernel is given by the following equation:

$$U(\alpha^*,\beta,\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) = (\det\xi)^{-\varepsilon/2}\exp(\tfrac{1}{2}\alpha^*\xi^{-1}\eta\alpha^* - \alpha^*\xi^{-1}\beta + \tfrac{1}{2}\varepsilon\beta\eta_1\xi^{-1}\beta), \tag{2.14}$$

where the matrices ξ, η, η_1 are expressed through the matrix (2.9) by Eqs. (I.4.11).

III. EXAMPLES

We shall discuss in detail two sets of unitary Cayley–Klein groups $\mathrm{SU}_2(j_1)$ and $\mathrm{SU}_3(j_1,j_2)$ for which it is possible to obtain the explicit form of the finite group transformation matrix $\Xi(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j})$.

A. $\mathrm{SU}_2(j_1)$ groups

The map (2.1), namely $\psi z_0 = z_0$, $\psi z_1 = j_1 z_1$, $j_1 = 1, \iota_1, i$, gives the complex space $C_2(j_1)$ with the quadratic form $(\mathbf{z},\mathbf{z}) = |z_0|^2 + j_1^2|z_1|^2$. The transformations belonging to the group $\mathrm{SU}_2(j_1)$ keep this quadratic form invariant. The matrices (2.5) are as follows:

$$Y_{00} = \begin{pmatrix}1&0\\0&0\end{pmatrix}, \quad Y_{11} = \begin{pmatrix}0&0\\0&1\end{pmatrix}, \quad Y_{01} = \begin{pmatrix}0&j_1^2\\0&0\end{pmatrix},$$
$$Y_{10} = \begin{pmatrix}0&0\\1&0\end{pmatrix}, \tag{3.1}$$

and the commutation relations

$$[Y_{00},Y_{11}]=0,\quad [Y_{00},Y_{01}]=Y_{01},$$
$$[Y_{00},Y_{10}]=-Y_{10},$$
$$[Y_{11},Y_{01}]=-Y_{01},\quad [Y_{11},Y_{10}]=Y_{10},$$
$$[Y_{01},Y_{10}]=j_1^2 Y_{00} \tag{3.2}$$

are satisfied. The matrix generators of $SU_2(j_1)$ are given by Eq. (2.8) in the form

$$P_1=\frac{i}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad Q_{01}=\frac{i}{2}\begin{pmatrix}0 & j_1^2\\ 1 & 0\end{pmatrix},$$
$$L_{01}=\frac{1}{2}\begin{pmatrix}0 & -j_1^2\\ 1 & 0\end{pmatrix}, \tag{3.3}$$

and satisfy the following commutation relations

$$[P_1,Q_{01}]=L_{01},\quad [L_{01},P_1]=Q_{01},$$
$$[Q_{01},L_{01}]=j_1^2 P_1. \tag{3.4}$$

The generators (3.3) for $j_1=1$ are equal up to a coefficient to the Pauli matrices. The general element of algebra $su_2(j_1)$ in accordance with Eq. (2.10) is given by

$$Z(r_1,s_1,w_1,j_1)=r_1Q_{01}+s_1L_{01}+w_1P_1$$
$$=\frac{1}{2}\begin{pmatrix}iw_1 & -j_1^2(s_1-ir_1)\\ s_1+ir_1 & -iw_1\end{pmatrix}, \tag{3.5}$$

and for the finite group transformation matrix we have

$$\Xi(r_1,s_1,w_1,j_1)=E_2\cos\frac{v}{2}+Z\frac{2}{v}\sin\frac{v}{2}=\begin{pmatrix}\cos\frac{v}{2}+i\frac{w_1}{v}\sin\frac{v}{2} & -j_1^2(s_1-ir_1)\frac{1}{v}\sin\frac{v}{2}\\ (s_1+ir_1)\frac{1}{v}\sin\frac{v}{2} & \cos\frac{v}{2}-i\frac{w_1}{v}\sin\frac{v}{2}\end{pmatrix}, \tag{3.6}$$

where

$$v^2(j_1)=w_1^2+j_1^2(r_1^2+s_1^2). \tag{3.7}$$

When $j_1=1,\iota_1$ the following operators:

$$\hat{Q}_{01}(j_1)=\hat{\mathbf{a}}^+Q_{01}(j_1)\hat{\mathbf{a}}=(i/2)(\hat{a}_1^+\hat{a}_0+j_1^2\hat{a}_0^+\hat{a}_1),$$
$$\hat{L}_{01}(j_1)=\hat{\mathbf{a}}^+L_{01}(j_1)\hat{\mathbf{a}}=\tfrac{1}{2}(\hat{a}_1^+\hat{a}_0-j_1^2\hat{a}_0^+\hat{a}_1),$$
$$\hat{P}_1(j_1)=\hat{\mathbf{a}}^+P_1(j_1)\hat{\mathbf{a}}=(i/2)(\hat{a}_0^+\hat{a}_0-\hat{a}_1^+\hat{a}_1), \tag{3.8}$$

satisfy the commutation relations (3.4) and therefore provide the Jordan–Schwinger representation of $SU_2(j_1)$. The kernel of the finite group transformation operator is obtained by Eq. (2.13) with help of the matrix (3.6) and is given in the form

$$U(\boldsymbol{\alpha}^*,\boldsymbol{\beta},j_1)=\exp\left\{\alpha_0^*\beta_0\left(\cos\frac{v}{2}-i\frac{w_1}{v}\sin\frac{v}{2}\right)\right.$$
$$+\alpha_1^*\beta_1\left(\cos\frac{v}{2}+i\frac{w_1}{v}\sin\frac{v}{2}\right)$$
$$-\alpha_1^*\beta_0(s_1+ir_1)\frac{1}{v}\sin\frac{v}{2}$$
$$\left.+j_1^2\alpha_0^*\beta_1(s_1-ir_1)\frac{1}{v}\sin\frac{v}{2}\right\}. \tag{3.9}$$

For the contracted group $SU_2(\iota_1)$ we have from Eq. (3.7) $v=w_1$ and from Eq. (3.9)

$$U(\boldsymbol{\alpha}^*,\boldsymbol{\beta},\iota_1)=\exp\{\alpha_0^*\beta_0e^{-(i/2)w_1}+\alpha_1^*\beta_1e^{(i/2)w_1}$$
$$-\alpha_1^*\beta_0(s_1+ir_1)(1/w_1)\sin(w_1/2)\}. \tag{3.10}$$

When $j_1=i$ we introduce the new parameter $\tilde{j}_1$ as $j_1=i\tilde{j}_1,\tilde{j}=1,\iota_1$. The case $j_1=\iota_1$ corresponds to the contraction of the pseudounitary group $SU_2(i)=SU(1,1)$. Equations (I.4.1) give $\psi\hat{\mathbf{a}}=(\hat{a}_0,\hat{a}_1^+)$, $\psi\hat{\mathbf{a}}^+=(\hat{a}_0^+,-\varepsilon\hat{a}_1)$. Then from Eqs. (2.11) we obtain the operators

$$\hat{Q}_{01}(i\tilde{j}_1)=\psi\hat{\mathbf{a}}^+Q_{01}(i\tilde{j}_1)\psi\hat{\mathbf{a}}=-(i/2)(\varepsilon\hat{a}_1\hat{a}_0+j_1^2\hat{a}_0^+\hat{a}_1^+),$$
$$\hat{L}_{01}(i\tilde{j}_1)=\psi\hat{\mathbf{a}}^+L_{01}(i\tilde{j}_1)\psi\hat{\mathbf{a}}=-\tfrac{1}{2}(\varepsilon\hat{a}_1\hat{a}_0-j_1^2\hat{a}_0^+\hat{a}_1^+),$$
$$\hat{P}_1(i\tilde{j}_1)=\psi\hat{\mathbf{a}}^+P_1(i\tilde{j}_1)\psi\hat{\mathbf{a}}=(i/2)(\hat{a}_0^+\hat{a}_0+\varepsilon\hat{a}_1\hat{a}_1^+), \tag{3.11}$$

that provide the Jordan–Schwinger representation of $SU_2(i\tilde{j}_1)$. The diagonal matrices $\psi_1(i)$, $\psi_2(i)$ in Eq. (I.4.2) are in the form

$$\psi_1(i)=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\quad \psi_2(i)=\begin{pmatrix}0 & 0\\ 0 & -1\end{pmatrix}, \tag{3.12}$$

and using Eqs. (I.4.11) and (3.6) we obtain the matrices ξ,η,η_1

$$\xi=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\left(\cos\frac{\tilde{v}}{2}+i\frac{w_1}{\tilde{v}}\sin\frac{\tilde{v}}{2}\right),$$
$$\eta=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\tilde{j}_1^2(s_1-ir_1)\frac{1}{\tilde{v}}\sin\frac{\tilde{v}}{2},$$
$$\eta_1=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}(s_1+ir_1)\frac{1}{\tilde{v}}\sin\frac{\tilde{v}}{2}, \tag{3.13}$$
$$\tilde{v}^2=w_1^2-j_1^2(r_1^2+s_1^2).$$

After some calculations we find the matrices ξ^{-1}, $\xi^{-1}\eta$, $\eta_1\xi^{-1}$ and using Eq. (2.14) obtain the kernel of the finite group transformation operator of $SU_2(i\tilde{j}_1)$ in the form

$$U(\boldsymbol{\alpha}^*,\boldsymbol{\beta},i\tilde{j}_1)=\left(\cos\frac{\tilde{v}}{2}+i\frac{w_1}{\tilde{v}}\sin\frac{\tilde{v}}{2}\right)^{-\varepsilon}\exp\left\{\left(\cos\frac{\tilde{v}}{2}+i\frac{w_1}{\tilde{v}}\sin\frac{\tilde{v}}{2}\right)^{-1}\right.$$
$$\times\left[\alpha_0^*\beta_0+\alpha_1^*\beta_1-\frac{1}{2}(\alpha_0^*\alpha_1^*+\alpha_1^*\alpha_0^*)\tilde{j}_1^2(s_1-ir_1)\frac{1}{\tilde{v}}\sin\frac{\tilde{v}}{2}\right.$$
$$\left.\left.+\frac{1}{2}\varepsilon(\beta_0\beta_1+\beta_1\beta_0)(s_1+ir_1)\frac{1}{\tilde{v}}\sin\frac{\tilde{v}}{2}\right]\right\}. \tag{3.14}$$

For the contracted group $SU_2(i\iota_1)$ we have from Eq. (3.13) $v = w_1$ and from Eq. (3.14)

$$U(\alpha^*,\beta,i\iota_1) = e^{-(i/2)\varepsilon w_1}\exp\left\{e^{-(i/2)w_1}\left[\alpha_0^*\beta_0 + \alpha_1^*\beta_1 + \tfrac{1}{2}\varepsilon(\beta_0\beta_1+\beta_1\beta_0)(s_1+ir_1)\frac{1}{w_1}\sin\frac{w_1}{2}\right]\right\}. \tag{3.15}$$

Comparing Eq. (3.15) with Eq. (3.10) we conclude that they provide different Jordan–Schwinger representations of the same group $SU_2(\iota_1) = SU_2(i\iota_1)$. Notice that some recent works[13,14] were devoted to the connections SU(1,1) with SU(2) and to the problem of the evolution SU(2) and SU(1,1) coherent states regarding from a unified point of view.

B. $SU_3(j_1,j_2)$ groups

The $SU_3(j_1,j_2)$ group consists of all transformations of $C_{2+1}(j_1,j_2)$ keeping invariant the quadratic form $(\mathbf{z},\mathbf{z}) = |z_0|^2 + j_1^2|z_1|^2 + j_1^2 j_1^2|z_2|^2$. The matrix generators of the general linear group are given by Eq. (2.5) in the form

$$Y_{00} = \begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix},\quad Y_{11} = \begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix},\quad Y_{22} = \begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix},$$
$$Y_{10} = \begin{pmatrix}0&0&0\\1&0&0\\0&0&0\end{pmatrix},\quad Y_{01} = \begin{pmatrix}0&j_1^2&0\\0&0&0\\0&0&0\end{pmatrix},\quad Y_{20} = \begin{pmatrix}0&0&0\\0&0&0\\1&0&0\end{pmatrix}, \tag{3.16}$$
$$Y_{02} = \begin{pmatrix}0&0&j_1^2j_2^2\\0&0&0\\0&0&0\end{pmatrix},\quad Y_{21} = \begin{pmatrix}0&0&0\\0&0&0\\0&1&0\end{pmatrix},\quad Y_{12} = \begin{pmatrix}0&0&0\\0&0&j_2^2\\0&0&0\end{pmatrix}.$$

Then the matrix generators of $SU_3(j_1,j_2)$ are given by Eq. (2.8) as follows:

$$P_1 = \frac{i}{2}\begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix},\quad P_2 = \frac{i}{2}\begin{pmatrix}0&0&0\\0&1&0\\0&0&-1\end{pmatrix},$$
$$Q_1 = Q_{01} = \frac{i}{2}\begin{pmatrix}0&j_1^2&0\\1&0&0\\0&0&0\end{pmatrix},\quad L_1 = L_{01} = \frac{1}{2}\begin{pmatrix}0&-j_1^2&0\\1&0&0\\0&0&0\end{pmatrix},$$
$$Q_2 = Q_{02} = \frac{i}{2}\begin{pmatrix}0&0&j_1^2j_2^2\\0&0&0\\1&0&0\end{pmatrix},\quad L_2 = L_{02} = \frac{1}{2}\begin{pmatrix}0&0&-j_1^2j_2^2\\0&0&0\\1&0&0\end{pmatrix}, \tag{3.17}$$
$$Q_3 = Q_{12} = \frac{i}{2}\begin{pmatrix}0&0&0\\0&0&j_2^2\\0&1&0\end{pmatrix},\quad L_3 = L_{12} = \frac{1}{2}\begin{pmatrix}0&0&0\\0&0&-j_2^2\\0&1&0\end{pmatrix}.$$

They satisfy the commutation relations

$$[P_1,P_2]=0,\quad [P_1,Q_1]=L_1,\quad [P_1,L_1]=-Q_1,$$
$$[P_1,Q_2]=\tfrac{1}{2}L_2,\quad [P_1,L_2]=-\tfrac{1}{2}Q_2,\quad [P_1,Q_3]=-\tfrac{1}{2}L_3,$$
$$[P_1,L_3]=\tfrac{1}{2}Q_3,\quad [P_2,Q_1]=-\tfrac{1}{2}L_1,\quad [P_2,L_1]=\tfrac{1}{2}Q_1,$$
$$[P_2,Q_2]=\tfrac{1}{2}L_2,\quad [P_2,L_2]=-\tfrac{1}{2}Q_2,\quad [P_2,Q_3]=L_3,$$
$$[P_2,L_3]=-Q_3,\quad [Q_1,L_1]=j_1^2P_1,\quad [Q_3,L_3]=j_2^2P_2,$$
$$[Q_2,L_2]=j_1^2j_2^2(P_1+P_2),\quad [Q_1,L_2]=-\tfrac{1}{2}j_1^2Q_3,$$
$$[Q_2,L_3]=\tfrac{1}{2}j_2^2Q_1,\quad [L_1,Q_2]=\tfrac{1}{2}j_1^2Q_3,\quad [Q_3,L_1]=\tfrac{1}{2}Q_2,$$
$$[L_2,Q_3]=-\tfrac{1}{2}j_2^2Q_1,\quad [Q_1,L_3]=-\tfrac{1}{2}Q_2,\quad [Q_1,Q_3]=\tfrac{1}{2}L_2,$$
$$[Q_1,Q_2]=\tfrac{1}{2}j_1^2L_3,\quad [Q_2,Q_3]=\tfrac{1}{2}j_2^2,\quad [L_1,L_2]=\tfrac{1}{2}j_1^2L_3,$$
$$[L_1,L_3]=-\tfrac{1}{2}L_2,\quad [L_2,L_3]=\tfrac{1}{2}j_2^2L_1. \tag{3.18}$$

It is well known[15] that the structure of group (algebra) is changed under contraction. Let $j_1 = \iota_1$, then the simple classical algebra su_3 contracts into the algebra $su_3(\iota_1,j_2) = T\dot{+}u_2(j_2)$, where $T = \{Q_1,L_1,Q_2,L_2\}$ is the commutative ideal and the subalgebra $u_2(j_2) = \{P_1,P_2,Q_3,L_3\}$ is the Lie algebra of the unitary group in the complex Cayley–Klein space $C_2(j_2)$. From Eq. (3.18) for $j_1 = \iota_1$ we conclude that $[T,u_2(j_2)] \subset T$ as it must be for a semidirect sum. The structure of $SU_3(\iota_1,j_2)$ is the semidirect product $SU_3(\iota_1,j_2) = e^T\circledS U_2(j_2)$. Such groups are called inhomogeneous unitary groups.[16]

From Eq. (2.10) we have the general element of the algebra $su_3(\mathbf{j})$ in the form

$$Z(\mathbf{r},\mathbf{s},\mathbf{w},\mathbf{j}) = \sum_{k=1}^{3}(r_kQ_k + s_kL_k) + w_1P_1 + w_2P_2$$

$$= \frac{1}{2}\begin{pmatrix} iw_1 & -j_1^2\zeta_1^* & -j_1^2 j_2^2 \zeta_2^* \\ \zeta_1 & i(w_2-w_1) & -j_2^2\zeta_3^* \\ \zeta_2 & \zeta_3 & -iw_2 \end{pmatrix}, \quad (3.19)$$

where $\zeta_k = s_k + ir_k$, $k = 1,2,3$ and ζ_k^* is the complex conjugate. The finite group transformation matrix $\Xi(\boldsymbol{\zeta},\mathbf{w},\mathbf{j})$ is obtained from Eq. (3.19) by the exponential map (2.9). We shall find the matrix Ξ by the Cayley–Hamilton theorem.[17] The characteristic equation $\det(Z - \lambda E_3) = 0$ of the matrix Z is the following cubic equation:

$$\begin{aligned} &\lambda^3 + p\lambda + q = 0, \\ &p = w_1^2 - w_1 w_2 + w_2^2 + |\zeta|^2(\mathbf{j}), \\ &q = -iw_1 w_2(w_2 - w_1) + iw_2 j_1^2 |\zeta_1|^2 \\ &\quad - i(w_2 - w_1) j_1^2 j_2^2 |\zeta_2|^2 \\ &\quad - iw_1 j_2^2 |\zeta_3|^2 + 2ij_1^2 j_2^2 \operatorname{Im} \zeta_1 \zeta_2^* \zeta_3, \end{aligned} \quad (3.20)$$

where

$$|\zeta|^2(\mathbf{j}) = j_1^2|\zeta_1|^2 + j_1^2 j_2^2 |\zeta_2|^2 + j_2^2 |\zeta_3|^2. \quad (3.21)$$

The roots of Eq. (3.20) are as follows[15]:

$$\begin{aligned} \lambda_k &= \left(-\frac{q}{2} + \left(\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right)^{1/2}\right)^{1/3} \\ &\quad + \left(-\frac{q}{2} - \left(\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right)^{1/2}\right)^{1/3} \\ &= \lambda_k' + \lambda_k'', \end{aligned} \quad (3.22)$$

where $\lambda_k' + \lambda_k'' = -p/3$ and λ_k', $k = 1,2,3$ are the three distinct cube roots. Then by the Cayley–Hamilton theorem we obtain

$$\Xi(\boldsymbol{\zeta},\mathbf{w},\mathbf{j}) = A \cdot E_3 - B \cdot Z + C \cdot Z^2, \quad (3.23)$$

where

$$Z^2 = \frac{1}{4}\begin{pmatrix} -w_1^2 - j_1^2|\zeta_1|^2 - j_1^2 j_2^2 |\zeta_2|^2 & -j_1^2(iw_2\zeta_1^* + j_2^2 \zeta_2^* \zeta_3) & j_1^2 j_2^2 (i(w_2 - w_1)\zeta_2^* + \zeta_1^* \zeta_3^* \\ iw_2\zeta_1 - j_2^2 \zeta_2 \zeta_3^* & -(w_2 - w_1)^2 - j_1^2|\zeta_1|^2 - j_2^2|\zeta_3|^2 & j_2^2(iw_1\zeta_3^* - j_1^2 \zeta_1 \zeta_2^*) \\ -i(w_2 - w_1)\zeta_2 + \zeta_1\zeta_3 & -iw_1\zeta_3 - j_1^2 \zeta_1^* \zeta_2 & -w_2^2 - j_1^2 j_2^2 |\zeta_2|^2 - j_2^2|\zeta_3|^2 \end{pmatrix}, \quad (3.24)$$

and the functions A, B, C are expressed in the following way:

$$\begin{aligned} A &= D^{-1}[e^{\lambda_1}\lambda_2\lambda_3(\lambda_2 - \lambda_3) - e^{\lambda_2}\lambda_1\lambda_3(\lambda_1 - \lambda_3) \\ &\quad + e^{\lambda_3}\lambda_1\lambda_2(\lambda_1 - \lambda_2)], \\ B &= D^{-1}[e^{\lambda_1}(\lambda_2^2 - \lambda_3^2) - e^{\lambda_2}(\lambda_1^2 - \lambda_3^2) \\ &\quad + e^{\lambda_3}(\lambda_1^2 - \lambda_2^2)], \\ C &= D^{-1}[e^{\lambda_1}(\lambda_2 - \lambda_3) - e^{\lambda_2}(\lambda_1 - \lambda_3) + e^{\lambda_3}(\lambda_1 - \lambda_2)], \\ D &= (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3). \end{aligned} \quad (3.25)$$

From $\operatorname{Tr} Z = 0$ we conclude that $\det \Xi = 1$ and $\Xi^{-1}(\boldsymbol{\zeta},\mathbf{w},\mathbf{j}) = \Xi(-\boldsymbol{\zeta},-\mathbf{w},\mathbf{j})$.

We shall only discuss contractions of the special unitary group SU_3, i.e., $j_1 = 1, \iota_1$, $j_2 = 1, \iota_2$. Then $\psi\hat{\mathbf{a}}^+ = \hat{\mathbf{a}}^+$, $\psi\hat{\mathbf{a}} = \hat{\mathbf{a}}$ in Eq. (2.11) and the Jordan–Schwinger representation of the generators (3.17) is given by the operators

$$\begin{aligned} \hat{P}_1(\mathbf{j}) &= (i/2)(\hat{a}_0^+ \hat{a}_0 - \hat{a}_1^+ \hat{a}_1), \\ \hat{P}_2(\mathbf{j}) &= (i/2)(\hat{a}_1^+ \hat{a}_1 - \hat{a}_2^+ \hat{a}_2), \\ \hat{Q}_1(\mathbf{j}) &= (i/2)(\hat{a}_1^+ \hat{a}_0 + j_1^2 \hat{a}_0^+ \hat{a}_1), \\ \hat{L}_1(\mathbf{j}) &= \tfrac{1}{2}(\hat{a}_1^+ \hat{a}_0 - j_1^2 \hat{a}_0^+ \hat{a}_1), \\ \hat{Q}_2(\mathbf{j}) &= (i/2)(\hat{a}_2^+ \hat{a}_0 + j_1^2 j_2^2 \hat{a}_0^+ \hat{a}_2), \\ \hat{L}_2(\mathbf{j}) &= \tfrac{1}{2}(\hat{a}_2^+ \hat{a}_0 - j_1^2 j_2^2 \hat{a}_0^+ \hat{a}_2), \\ \hat{Q}_3(\mathbf{j}) &= (i/2)(\hat{a}_2^+ \hat{a}_1 + j_2^2 \hat{a}_0^+ \hat{a}_2), \\ \hat{L}_3(\mathbf{j}) &= \tfrac{1}{2}(\hat{a}_2^+ \hat{a}_1 - j_2^2 \hat{a}_1^+ \hat{a}_2). \end{aligned} \quad (3.26)$$

These operators satisfy the commutation relations (3.18). The general element of $su_3(\mathbf{j})$ is represented by the following operator:

$$\begin{aligned} \hat{Z}(\boldsymbol{\zeta},\mathbf{w},\mathbf{j}) &= \sum_{k=1}^{3} (r_k \hat{Q}_k(\mathbf{j}) + s_k \hat{L}_k(\mathbf{j})) + w_1 \hat{P}_1(\mathbf{j}) + w_2 \hat{P}_2(\mathbf{j}) \\ &= \frac{1}{2}\Big\{ iw_1 \hat{a}_0^+ \hat{a}_0 + i(w_2 - w_1)\hat{a}_1^+ \hat{a}_1 - iw_2 \hat{a}_2^+ \hat{a}_2 \\ &\quad + \sum_{k=1}^{3} \zeta_k \hat{a}_k^+ \hat{a}_0 - j_1^2 \zeta_1^* \hat{a}_0^+ \hat{a}_1 - j_1^2 j_2^2 \zeta_2^* \hat{a}_0^+ \hat{a}_2 \\ &\quad - j_2^2 \zeta_3^* \hat{a}_1^+ \hat{a}_2 \Big\}. \end{aligned} \quad (3.27)$$

The kernel of the finite group transformation operator (2.12) of $SU_3(\mathbf{j})$ in a coherent state basis is given by Eq. (2.13), using Eqs. (3.23)–(3.25). We shall not write this kernel. We write out only the kernel of operator of the group $SU_3(\iota_1,\iota_2)$ that is obtained from SU_3 by two-dimensional contraction. This kernel is as follows:

$$\begin{aligned} U(\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\zeta},\mathbf{w},\iota_1,\iota_2) = \exp\Big\{ &A' \sum_{k=1}^{3} \alpha_k^* \beta_k + (i/2)B' w_1(\alpha_1^* \beta_1 - \alpha_0^* \beta_0) + (i/2)B' w_2(\alpha_2^* \beta_2 - \alpha_1^* \beta_1) - \tfrac{1}{4} C' w_1^2 (\alpha_0^* \beta_0 + \alpha_1^* \beta_1) \\ &- \tfrac{1}{4} C' w_2^2 (\alpha_1^* \beta_1 + \alpha_2^* \beta_2) + \tfrac{1}{2} C' w_1 w_2 \alpha_1^* \beta_1 + \tfrac{1}{2}((i/2)w_2 C' - B')\zeta_1 \alpha_1^* \beta_0 \\ &- \tfrac{1}{2}[(i/2)(w_2 - w_1)C' + B']\zeta_2 \alpha_2^* \beta_0 - \tfrac{1}{2}((i/2)w_1 C' + B')\zeta_3 \alpha_2^* \beta_1 + \tfrac{1}{4} C' \zeta_1 \zeta_3 \alpha_2^* \beta_0 \Big\}, \end{aligned} \quad (3.28)$$

where the functions A', B', C' are given by Eqs. (3.25) and $\lambda_1,\lambda_2,\lambda_3$ are the roots of Eq. (3.20) with the following coefficients: $p = w_1^2 - w_1w_2 + w_2^2$, $q = -iw_1w_2(w_2 - w_1)$.

IV. CONCLUDING REMARKS

On the basis of ideas of the previous paper[11] we have regarded the unitary Cayley–Klein groups $SU_{n+1}(\mathbf{j})$ as the groups of motion (except for translations) of the complex Cayley–Klein spaces $C_{n+1}(\mathbf{j})$. The groups $SU_{n+1}(\mathbf{j})$ have been obtained from the classical group SU_{n+1} by contractions and analytical continuations of the group parameters. It has been shown that all these groups are described in the unified way by introducing n parameters $\mathbf{j} = (j_1, j_2, \ldots j_n)$ each of which were equal to the real, dual, or imaginary units. We have built the Jordan–Schwinger representation of the group under consideration. The only contractions of the Jordan–Schwinger representations permit of the unified description. In the case of analytical continuations each of the Jordan–Schwinger representation is built in a particular way.

[1]P. Jordan, Z. Phys. **94**, 531 (1935).
[2]L. C. Biedenharn and J. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, London, 1981).
[3]J. Schwinger, "On angular momentum," in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965), p. 229.
[4]E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[5]N. A. Gromov, Teor. Mat. Fiz. **49**, 210 (1981) (in Russian).
[6]N. A. Gromov, in *Group Theoretical Methods in Physics* (Nauka, Moscow, 1986), Vol. 2, p. 183 (in Russian).
[7]R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[8]I. A. Malkin and V. I. Man'ko, *Dynamical Symmetries and Coherent States of Quantum Systems* (Nauka, Moscow, 1979) (in Russian).
[9]V. V. Dodonov and V. I. Man'ko, Trudy FIAN (P. N. Lebedev Physical Institute Proceedings) **183**, 182 (1987) (in Russian).
[10]V. I. Man'ko and D. A. Trifonov, *Group Theoretical Methods in Physics* (Harwood Academic, London, 1987), Vol. 3, p. 795.
[11]N. A. Gromov and V. I. Man'ko, J. Math. Phys. **31**, 1047 (1990).
[12]N. A. Gromov, Preprint Series "Scientific Reports," No. 95, Komi Branch of the Acad. Sci. USSR, Syktyvkar (1984) (in Russian).
[13]G. Dattoli, M. Richetta, and A. Torre, J. Math. Phys. **29**, 2586 (1988).
[14]E. de Prunele, J. Math. Phys. **29**, 2523 (1988).
[15]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN-Polish Scientific Publishers, Warsaw, 1977).
[16]M. Perroud, J. Math. Phys. **24**, 1381 (1983).
[17]G. A. Korn and T. M. Korn, *Mathematical Handbook* (McGraw-Hill, New York, 1961).

The Jordan–Schwinger representations of Cayley–Klein groups. III. The symplectic groups

N. A. Gromov
Komi Scientific Centre, Academy of Sciences of the USSR, Syktyvkar, GSP, 167610, USSR
V. I. Man'ko
P. N. Lebedev Physical Institute, Academy of Sciences of the USSR, Moscow, 117333, USSR

(Received 12 May 1989; accepted for publication 25 October 1989)

The symplectic Cayley–Klein groups are defined as the groups that are obtained by the contractions and analytical continuations of the classical symplectic groups. The Jordan–Schwinger representations of the groups under consideration are discussed based on the mixed sets of creation and annihilation operators of boson or fermion type. The matrix elements of finite group transformations are obtained in the bases of coherent states.

I. INTRODUCTION

The previous papers[1,2] of this series (hereafter referred to as I or II and whose equations will be subsequently quoted by their number preceded by I or II) was devoted to a consideration of the Jordan–Schwinger representations of the orthogonal and unitary Cayley–Klein groups. In this concluding paper we discuss the case of the symplectic Cayley–Klein groups. The symplectic groups and their representations are used in different branches of physics.[3,4] The unitary representations of the symplectic Sp (n,R) and pseudosymplectic Sp (p,q) groups have been regarded in Refs. 5–7. The oscillator representation for the orthogonal and symplectic groups was discussed by Lohe and Hurst.[8] The present paper is organized as follows. In Sec. II we describe the symplectic Cayley–Klein groups as the groups which are obtained from the classical symplectic one's by the contractions and analytical continuations of the group parameters. We construct their Jordan–Schwinger representations and calculate the matrix elements of the finite group transformations in coherent state bases. In Sec. III we regard in detail two groups of low dimension Sp_1 and Sp_2 (j). The main statements of the work are briefly discussed in the conclusion.

II. THE SYMPLECTIC CAYLEY–KLEIN GROUPS

First we briefly review the necessary information about the symplectic group. As is well known[3], the symplectic group Sp_n includes all transformations of $2n$-dimensional space $R_n \times R_n$ under which the following bilinear form is invariant:

$$\mathrm{Sp}_n: \quad R_n \times R_n \to R_n \times R_n,$$
$$[\mathbf{x},\mathbf{y}] = \sum_{k=1}^{n} (x_k y_{-k} - x_{-k} y_k), \tag{2.1}$$

where $\{x_k, x_{-k}\}$ is the Cartesian coordinates in $R_n \times R_n$. In the space of infinitely differentiable functions $f: R_n \times R_n \to R$ the group Sp_n acts as $g: f(\mathbf{x},\mathbf{y}) \to f(g\mathbf{x}, g\mathbf{y})$ and their generators are in the form

$$\tilde{X}_{\alpha\beta}(\mathbf{x}) = x_\alpha \partial_\beta - \epsilon_\alpha \epsilon_\beta x_{-\beta} \partial_{-\alpha}, \tag{2.2}$$

where $\alpha,\beta = \pm 1, \pm 2, \ldots, \pm n$, $\partial_\alpha = \partial/\partial x_\alpha$, $\epsilon_\alpha = \operatorname{sign}\alpha$, i.e., $\epsilon_\alpha = 1$ for $\alpha > 0$, $\epsilon_\alpha = -1$ for $\alpha < 0$ and $\epsilon_\alpha = 0$ for $\alpha = 0$. The generators (2.2) are not independent. They satisfy the symmetry condition

$$\tilde{X}_{\alpha\beta} = -\epsilon_\alpha \epsilon_\beta \tilde{X}_{-\beta,-\alpha}. \tag{2.3}$$

Then, as independent generators we choose the following $n(2n+1)$ generators:

$$\begin{aligned}
&\tilde{X}_{\mu\mu}(\mathbf{x}) = x_\mu \partial_\mu - x_{-\mu}\partial_{-\mu}, \quad \mu = 1,2,\ldots,n \\
&\tilde{X}_{\mu,-\mu}(\mathbf{x}) = 2x_\mu \partial_{-\mu}, \quad \mu = \pm 1, \pm 2, \ldots, \pm n, \\
&\tilde{X}_{\nu\mu}(\mathbf{x}) = x_\nu \partial_\mu - \epsilon_\mu x_{-\mu}\partial_{-\nu}, \\
&\tilde{X}_{\mu\nu}(\mathbf{x}) = x_\mu \partial_\nu - \epsilon_\mu x_{-\nu}\partial_{-\mu}, \quad \nu = 2,3,\ldots,n, |\mu| < \nu.
\end{aligned} \tag{2.4}$$

The generators (2.2) of Sp_n satisfy the commutation relations

$$\begin{aligned}
[\tilde{X}_{\alpha\beta}, \tilde{X}_{\alpha'\beta'}] = {} & \delta_{\alpha'\beta}\tilde{X}_{\alpha\beta'} - \delta_{\alpha\beta'}\tilde{X}_{\alpha'\beta} + \epsilon_\alpha \epsilon_\beta \delta_{-\beta',\beta}\tilde{X}_{\alpha',-\alpha} \\
& + \epsilon_\beta \epsilon_{\alpha'} \delta_{\alpha',-\alpha}\tilde{X}_{-\beta,\beta'}.
\end{aligned} \tag{2.5}$$

We may write the generators (2.2), (2.4) in the matrix form using the relation

$$\tilde{X}_{\alpha\beta}(\mathbf{x}) = \partial \tilde{X}_{\alpha\beta} \mathbf{x}, \tag{2.6}$$

where $\partial = (\partial_1, \partial_2, \ldots, \partial_n, \partial_{-1}, \partial_{-2}, \ldots, \partial_{-n})$ is the row matrix, $\mathbf{x} = (x_1, x_2, \ldots, x_n, x_{-1}, x_{-2}, \ldots, x_{-n})^T$ is the column matrix, and the product in Eq. (2.6) is the ordinary matrix product. Then the generators $\tilde{X}_{\alpha\beta}$ are a $2n$-dimensional matrix. The independent generators (2.4) are characterized by the nonzero matrix elements as follows:

$$\begin{aligned}
&(\tilde{X}_{\mu\mu})_{\mu\mu} = 1, \quad (\tilde{X}_{\mu\mu})_{-\mu,-\mu} = -1, \quad \mu = 1,2,\ldots,n, \\
&(\tilde{X}_{\mu,-\mu})_{-\mu,\mu} = 2, \quad \mu = \pm 1, \pm 2, \ldots, \pm n, \\
&(\tilde{X}_{\nu\mu})_{\mu\nu} = 1, \quad (\tilde{X}_{\nu\mu})_{-\nu,-\mu} = -\epsilon_\mu, \\
&(\tilde{X}_{\mu\nu})_{\nu\mu} = 1, \quad (\tilde{X}_{\mu\nu})_{-\mu,-\nu} = -\epsilon_\mu, \\
&\qquad\qquad \nu = 2,3,\ldots,n, |\mu| < \nu.
\end{aligned} \tag{2.7}$$

Following Refs. 1 and 9, let us regard the map

$$\psi: \quad R_n \to R_n(\mathbf{j}),$$
$$\psi x_k = x_k \prod_{m=2}^{k} j_m, \quad k = 1,2,\ldots,n, \tag{2.8}$$

where $\mathbf{j} = (j_2, j_3, \ldots, j_n)$ and each of the parameters j_m may be equal to the real unit 1 or to the Clifford dual unit or to the imaginary unit i. Here $\iota_k \neq 0$, $\iota_m \neq 0$, $\iota_k \iota_m = \iota_m \iota_k \neq 0$ for $k \neq m$, but $\iota_k^2 = \iota_m^2 = 0$. Division by a dual unit is not defined, but division of a dual unit by itself is equal to the real

 0022-2488/90/051060-05$03.00

unit $\iota_k/\iota_k = 1$. We agree throughout the paper that $\Pi_{m=a}^{b} j_m = 1$ for $a > b$. Some of the applications of the dual numbers in geometry can be found in Refs. 10 and 11.

The symplectic Cayley–Klein group $\mathrm{Sp}_n(j)$ is defined[12,13] as the transformation group of $2n$-dimensional space $R_n(\mathbf{j}) \times R_n(\mathbf{j})$ leaving invariant the following bilinear form:

$$\mathrm{Sp}_n(\mathbf{j}): R_n(\mathbf{j}) \times R_n(\mathbf{j}) \to R_n(\mathbf{j}) \times R_n(\mathbf{j}),$$

$$[\mathbf{x},\mathbf{y}] = \sum_{k=1}^{n} \left(\prod_{m=2}^{k} j_m^2 \right) (x_k y_{-k} - x_{-k} y_k). \tag{2.9}$$

In accordance with our approach we obtain the generators of $\mathrm{Sp}_n(\mathbf{j})$ by the transformations of the generators of Sp_n. From the definition of the generator

$$X(\mathbf{x}) = \sum_{k=-n}^{n} \frac{x_k}{\partial a}\bigg|_{a=0} \partial_k,$$

where $\mathbf{x}' = g(a)\mathbf{x}$, $g(0) = 1$, $g(a) \in \mathrm{Sp}_n(\mathbf{j})$, $\mathbf{x} \in R_n(\mathbf{j}) \times R_n(\mathbf{j})$, using the map (2.8), we have the transformation law in the form

$$X_{\alpha\beta}(\mathbf{x}) = \left(\prod_{m=1+\min\{|\alpha|,|\beta|\}}^{\max\{|\alpha|,|\beta|\}} j_m \right) \tilde{X}_{\alpha\beta}(\psi\mathbf{x}). \tag{2.10}$$

Here $\psi\mathbf{x}$ is given by Eq. (2.8), where the upper limit of the production is equal to $|k|$ for negative k. The generators

$$\tilde{X}_{\alpha\beta}(\psi\mathbf{x}) = \left(\prod_{m=1+\min\{|\alpha|,|\beta|\}}^{\max\{|\alpha|,|\beta|\}} j_m^{\operatorname{sign}(|\alpha|-|\beta|)} \right) x_\alpha \partial_\beta - \left(\prod_{m=1+\min\{|\alpha|,|\beta|\}}^{\max\{|\alpha|,|\beta|\}} j_m^{-\operatorname{sign}(|\alpha|-|\beta|)} \right) \epsilon_\alpha \epsilon_\beta x_{-\beta} \partial_{-\alpha} \tag{2.11}$$

are the Inonu–Wigner[14] singular transformed generators (when some parameters j_k are equal to the dual units). Using Eqs. (2.10) and (2.11) we obtain the generators $X_{\alpha\beta}(\mathbf{x})$ of $\mathrm{Sp}_n(\mathbf{j})$ in the form

$$X_{\alpha\beta}(\mathbf{x}) = \left(\prod_{m=p}^{q} j_m^{1+\operatorname{sign}(|\alpha|-|\beta|)} \right) x_\alpha \partial_\beta - \left(\prod_{m=p}^{q} j_m^{1-\operatorname{sign}(|\alpha|-|\beta|)} \right) \epsilon_\alpha \epsilon_\beta x_{-\beta} \partial_{-\alpha}, \tag{2.12}$$

where $p = 1 + \min\{|\alpha|,|\beta|\}$, $q = \max\{|\alpha|,|\beta|\}$. They also satisfy the symmetry condition (2.3). The independent generators of $\mathrm{Sp}_n(\mathbf{j})$ are obtained from Eq. (2.4) by transformations (2.10) or directly from Eq. (2.12) and are given as follows:

$$\begin{aligned}
&X_{\mu\mu}(\mathbf{x}) = x_\mu \partial_\mu - x_{-\mu}\partial_{-\mu}, \quad \mu = 1,2,\ldots,n,\\
&X_{\mu,-\mu}(\mathbf{x}) = 2x_\mu \partial_{-\mu}, \quad \mu = \pm 1, \pm 2, \ldots, \pm n,\\
&X_{\nu\mu}(\mathbf{x}) = \left(\prod_{m=1+|\mu|}^{\nu} j_m^2 \right) x_\nu \partial_\mu - \epsilon_\mu x_{-\mu} \partial_{-\nu},\\
&X_{\mu\nu}(\mathbf{x}) = x_\mu \partial_\nu - \left(\prod_{m=1+|\mu|}^{\nu} j_m^2 \right) \epsilon_\mu x_{-\nu} \partial_{-\mu},\\
&\qquad |\mu| < \nu, \quad \nu = 2,3,\ldots,n.
\end{aligned} \tag{2.13}$$

From Eq. (2.10) we have

$$\tilde{X}_{\alpha\beta}(\psi\mathbf{x}) = \left(\prod_{m=p}^{q} j_m^{-1} \right) X_{\alpha\beta}(\mathbf{x}).$$

Substituting in Eq. (2.5) the generators $\tilde{X}_{\alpha\beta}$ for their expressions by $X_{\alpha\beta}$ we immediately find the commutation relations of $\mathrm{Sp}_n(\mathbf{j})$

$$\begin{aligned}
[X_{\alpha\beta}, X_{\alpha'\beta'}] = &\left(\prod_{m=p_1}^{q_1} j_m \right) \left(\prod_{m=p_1'}^{q_1'} j_m \right) \left\{ \left(\prod_{m=p_2}^{q_2} j_m^{-1} \right) \delta_{\alpha'\beta} X_{\alpha\beta'} \right.\\
&- \left(\prod_{m=p_2'}^{q_2'} j_m^{-1} \right) \delta_{\alpha\beta'} X_{\alpha'\beta}\\
&+ \left(\prod_{m=p_3}^{q_3} j_m^{-1} \right) \epsilon_\alpha \epsilon_\beta \delta_{-\beta',\beta} X_{\alpha',-\alpha}\\
&\left. + \left(\prod_{m=p_3'}^{q_3'} j_m^{-1} \right) \epsilon_\beta \epsilon_{\alpha'} \delta_{\alpha',-\alpha} X_{-\beta,\beta'} \right\},
\end{aligned} \tag{2.14}$$

where

$$\begin{aligned}
&p_1 = 1 + \min\{|\alpha|,|\beta|\}, \quad q_1 = \max\{|\alpha|,|\beta|\},\\
&p_1' = 1 + \min\{|\alpha'|,|\beta'|\}, \quad q_1' = \max\{|\alpha'|,|\beta'|\},\\
&p_2 = 1 + \min\{|\alpha|,|\beta'|\}, \quad q_2 = \max\{|\alpha|,|\beta'|\},\\
&p_2' = 1 + \min\{|\alpha'|,|\beta|\}, \quad q_2' = \max\{|\alpha'|,|\beta|\},\\
&p_3 = 1 + \min\{|\alpha|,|\alpha'|\}, \quad q_3 = \max\{|\alpha|,|\alpha'|\},\\
&p_3' = 1 + \min\{|\beta|,|\beta'|\}, \quad q_3' = \max\{|\beta|,|\beta'|\}.
\end{aligned}$$

At last the independent matrix generators of $\mathrm{Sp}_n(\mathbf{j})$ are obtained from Eq. (2.13) using Eq. (2.6). Their nonzero matrix elements are as follows:

$$\begin{aligned}
&(X_{\mu\mu})_{\mu\mu} = 1, \quad (X_{\mu\mu})_{-\mu,-\mu} = -1, \quad \mu = 1,2,\ldots,n,\\
&(X_{\mu,-\mu})_{-\mu,\mu} = 2, \quad \mu = \pm 1, \pm 2, \ldots, \pm n,\\
&(X_{\nu\mu})_{\mu\nu} = \prod_{m=1+|\mu|}^{\nu} j_m^2, \quad (X_{\nu\mu})_{-\nu,-\mu} = -\epsilon_\mu,\\
&(X_{\mu\nu})_{\nu\mu} = 1, \quad (X_{\mu\nu})_{-\mu,-\nu} = -\epsilon_\mu \prod_{m=1+|\mu|}^{\nu} j_m^2,\\
&\qquad |\mu| < \nu, \quad \nu = 2,3,\ldots,n.
\end{aligned} \tag{2.15}$$

The general element $Z(\mathbf{r},\mathbf{j}) = \Sigma r_{\alpha\beta} X_{\alpha\beta}(\mathbf{j})$ of the algebra $\mathrm{sp}_n(\mathbf{j})$, where the sum is doing over all independent generators, is mapped by exponent into the finite group transformation $\Xi(\mathbf{r},\mathbf{j}) = \exp Z(\mathbf{r},\mathbf{j})$.

The Jordan–Schwinger representation of $\mathrm{Sp}_n(\mathbf{j})$ is provided by the operators [cf. with Eq. (I.4.3.)]

$$\hat{X}_{\alpha\beta}(\mathbf{j}) = \psi\hat{\mathbf{a}}^+ X_{\alpha\beta}(\mathbf{j}) \psi\hat{\mathbf{a}}, \tag{2.16}$$

where

$$\begin{aligned}
&\psi\hat{\mathbf{a}}^+ = \left(\hat{a}_k^+ \prod_{m=2}^{k} j_m^{-1}, \quad \hat{a}_{-k}^+ \prod_{m=2}^{|k|} j_m^{-1} \right), \quad k = 1,2,\ldots,n,\\
&\psi\hat{\mathbf{a}} = \left(\hat{a}_k \prod_{m=2}^{k} j_m, \quad \hat{a}_{-k} \prod_{m=2}^{|k|} j_m \right).
\end{aligned} \tag{2.17}$$

Here ψ is identical, when $j_m = 1, \iota_m$, and for the imaginary values of parameters $\mathbf{j}$ we use the well known properties of the annihilation and creation operators: $i\hat{a}_{\pm k} = \hat{a}_{\pm k}^+$, $i\hat{a}_{\pm k}^+ = \epsilon\hat{a}_{\pm k}$. Then Eq. (2.17) may be written in the form

$$\begin{pmatrix} \psi\hat{\mathbf{a}} \\ \psi\hat{\mathbf{a}}^+ \end{pmatrix} = \begin{pmatrix} \psi_1(\mathbf{j}) & -\psi_2(\mathbf{j}) \\ \epsilon\psi_2(\mathbf{j}) & \psi_1(\mathbf{j}) \end{pmatrix} \begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^+ \end{pmatrix} = \Psi^{-1}(\mathbf{j}) \begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^+ \end{pmatrix}, \tag{2.18}$$

where $\psi_1(\mathbf{j})$, $\psi_2(\mathbf{j})$ are $2n$-dimensional diagonal matrices with the following nonzero matrix elements: $(\psi_1)_{\pm 1,\pm 1} = 1$, $(\psi_1)_{\pm k,\pm k} = \pm 1$, if $\Pi_{m=2}^{|k|} j_m = \pm\gamma$ and

γ is a positive real or dual number, $(\psi_1)_{\pm k,\pm k}=0$ otherwise; $(\psi_2)_{\pm 1,\pm 1}=0$, $(\psi_2)_{\pm k,\pm k}=0$, if $(\psi_1)_{\pm k,\pm k}=\pm 1$ and $(\psi_2)_{\pm k,\pm k}=\mp 1$, if $\Pi_{m=2}^{|k|} j_m=\pm i\gamma$. The $4n$-dimensional matrix $\Psi(\mathbf{j})$ has the property $\Psi(\mathbf{j})=(\Psi^{-1}(\mathbf{j}))^T$.

The finite group transformation is represented by the operator

$$\hat{U}_g(\mathbf{r},\mathbf{j})=\exp\{-\hat{Z}(\mathbf{r},\mathbf{j})\}=\exp\Big\{-\sum r_{\alpha\beta}\hat{X}_{\alpha\beta}(\mathbf{j})\Big\}.$$

The kernel of the operator $\hat{U}_g(\mathbf{r},\mathbf{j})$ in a coherent state basis is obtained quite analogous to the case of orthogonal or unitary groups.[1,2] When the Cayley–Klein groups $\mathrm{Sp}_n(\mathbf{j})$ are obtained from Sp_n only by contractions ($j_k=1,\iota_k$, $k=2,3,\ldots,n$) the kernel is given by

$$U(\boldsymbol{\alpha}^*,\boldsymbol{\beta},\mathbf{r},\mathbf{j})=\exp\{\boldsymbol{\alpha}^*\Xi(-\mathbf{r},\mathbf{j})\boldsymbol{\beta}\},\tag{2.19}$$

[see Eq. (I.4.8)], where $\boldsymbol{\alpha}^*=(\alpha_1^*,\ldots,\alpha_n^*,\alpha_{-1}^*,\ldots,\alpha_{-n}^*)$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_n,\beta_{-1},\ldots,\beta_{-n})$. When the groups $\mathrm{Sp}_n(\mathbf{j})$ are obtained from Sp_n by both contractions and analytical continuations the kernel is given by Eq. (I.2.7), where the matrices ξ,η,η_1 are expressed through the matrix $\Xi(\mathbf{r},\mathbf{j})$ by Eqs. (I.4.11) with $\psi_1(\mathbf{j})$ and $\psi_2(\mathbf{j})$ as in Eq. (2.18).

III. EXAMPLES

We shall discuss in detail the groups Sp_1 and $\mathrm{Sp}_2(j)$. Only for these groups we are able to obtain the explicit form of the finite group transformation matrix $\Xi(\mathbf{r},\mathbf{j})$.

A. Sp_1 group

The simplest symplectic group Sp_1 is the transformation group of the two dimensional space $R_1\times R_1$ which leaves invariant the bilinear form $[\mathbf{x},\mathbf{y}]=x_1y_{-1}-x_{-1}y_1$. The three independent generators of Sp_1 are given by Eq. (2.7) in the form

$$X_{11}=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad X_{1,-1}=\begin{pmatrix}0&0\\2&0\end{pmatrix},\quad X_{-1,1}=\begin{pmatrix}0&2\\0&0\end{pmatrix}\tag{3.1}$$

and satisfy the commutation relations

$$[X_{11},X_{1,-1}]=-2X_{1,-1},\quad [X_{11},X_{-1,1}]=2X_{-1,1},$$
$$[X_{1,-1},X_{-1,1}]=-4X_{11}.\tag{3.2}$$

The general element of the algebra sp_1 is given by the matrix

$$X(r_1,\mathbf{s})=r_1X_{11}+s_1X_{1,-1}+s_2X_{-1,1}=\begin{pmatrix}r_1&2s_2\\2s_1&-r_1\end{pmatrix},\tag{3.3}$$

and the finite transformation matrix Ξ is

$$\Xi(r_1,\mathbf{s})=\exp X(r_1,\mathbf{s})=E\cdot\mathrm{ch}\,p+X(\mathrm{sh}\,p)/p$$
$$=\begin{pmatrix}\mathrm{ch}\,p+r_1\dfrac{\mathrm{sh}\,p}{p}&2s_2\dfrac{\mathrm{sh}\,p}{p}\\[2ex]2s_1\dfrac{\mathrm{sh}\,p}{p}&\mathrm{ch}\,p-r_1\dfrac{\mathrm{sh}\,p}{p}\end{pmatrix},\tag{3.4}$$

where $p^2=r_1^2+4s_1s_2$.

Following (2.16), the Jordan–Schwinger representation of Sp_1 is provided by the operators

$$\hat{X}_{11}=\hat{a}_1^+\hat{a}_1-\hat{a}_{-1}^+\hat{a}_{-1},\quad \hat{X}_{1,-1}=2\hat{a}_{-1}^+\hat{a}_1,$$
$$\hat{X}_{-1,1}=2\hat{a}_1^+\hat{a}_{-1}.\tag{3.5}$$

The kernel of the operator $\hat{U}_g(r_1,\mathbf{s})=\exp\{-\hat{X}(r_1,\mathbf{s})\}$ in a coherent state basis is given by Eqs. (3.4), (2.19) or explicitly by

$$U(\boldsymbol{\alpha}^*,\boldsymbol{\beta},r_1,\mathbf{s})=\exp\{(\alpha_1^*\beta_1+\alpha_{-1}^*\beta_{-1})\mathrm{ch}\,p$$
$$-(\mathrm{sh}\,p)/p[r_1(\alpha_1^*\beta_1-\alpha_{-1}^*\beta_{-1})$$
$$+2s_1\alpha_{-1}^*\beta_1+2s_2\alpha_1^*\beta_{-1}]\}.\tag{3.6}$$

B. $\mathrm{Sp}_2(j)$ groups

The group $\mathrm{Sp}_2(j)$ (we omit the index 2 from the parameter j_2) consists of all transformations of $R_2(j)\times R_2(j)$ leaving invariant the bilinear form $[\mathbf{x},\mathbf{y}]=x_1y_{-1}-x_{-1}y_1+j^2(x_2y_{-2}-x_{-2}y_2)$. The ten independent matrix generators of $\mathrm{Sp}_2(j)$ are given by Eq. (2.15) as follows:

$$X_{11}=\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&-1&0\\0&0&0&0\end{pmatrix},\quad X_{1,-1}=\begin{pmatrix}0&0&0&0\\0&0&0&0\\2&0&0&0\\0&0&0&0\end{pmatrix},$$
$$X_{-1,1}=\begin{pmatrix}0&0&2&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad X_{22}=\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&-1\end{pmatrix},$$
$$X_{2,-2}=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&2&0&0\end{pmatrix},\quad X_{-2,2}=\begin{pmatrix}0&0&0&0\\0&0&0&2\\0&0&0&0\\0&0&0&0\end{pmatrix},$$
$$X_{12}(j)=\begin{pmatrix}0&0&0&0\\1&0&0&0\\0&0&0&-j^2\\0&0&0&0\end{pmatrix},$$
$$X_{21}(j)=\begin{pmatrix}0&j^2&0&0\\0&0&0&0\\0&0&0&0\\0&0&-1&0\end{pmatrix},$$
$$X_{2,-1}(j)=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&j^2&0&0\\1&0&0&0\end{pmatrix},\quad X_{-1,2}=\begin{pmatrix}0&0&0&j^2\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{pmatrix}.\tag{3.7}$$

Let us observe that each of the sets of generators $A_1=\{X_{11},X_{1,-1},X_{-1,1}\}$ and $A_2=\{X_{22},X_{2,-2},X_{-2,2}\}$ form the subalgebra of $\mathrm{sp}_2(j)$ isomorphic to the algebra sp_1 with the commutators (3.2) and $[A_1,A_2]=0$. Then $A=A_1\oplus A_2$ is the subalgebra of $\mathrm{sp}_2(j)$. The other nonzero commutation relations of $\mathrm{sp}_2(j)$ are obtained from Eq. (2.14) in the form

$$[X_{11},X_{12}]=-X_{12},\quad [X_{11},X_{21}]=X_{21},$$
$$[X_{11},X_{2,-1}]=-X_{2,-1},\quad [X_{11},X_{-1,2}]=X_{-1,2},$$
$$[X_{1,-1},X_{21}]=2X_{2,-1},\quad [X_{1,-1},X_{-1,2}]=-2X_{12},$$
$$[X_{-1,1},X_{12}]=-2X_{-1,2},\quad [X_{-1,1},X_{2,-1}]=2X_{21}.\tag{3.8a}$$

$$[X_{22},X_{12}]=X_{12},\quad [X_{22},X_{21}]=-X_{21},$$

$$\begin{aligned}
&[X_{22},X_{2,-1}]=-X_{2,-1},\quad [X_{22},X_{-1,2}]=X_{-1,2},\\
&[X_{2,-2},X_{12}]=2X_{2,-1},\quad [X_{2,-2},X_{-1,2}]=-2X_{21},\\
&[X_{-2,2},X_{21}]=-2X_{-1,2},\quad [X_{-2,2},X_{2,-1}]=2X_{12}.
\end{aligned}\tag{3.8b}$$

$$\begin{aligned}
&[X_{12},X_{21}]=j^2(X_{22}-X_{11}),\quad [X_{12},X_{2,-1}]=-j^2X_{1,-1},\\
&[X_{12},X_{-1,2}]=j^2X_{-2,2},\quad [X_{21},X_{2,-1}]=-j^2X_{2,-2},\\
&[X_{21},X_{-1,2}]=j^2X_{-1,1},[X_{2,-1},X_{-1,2}]=-j^2(X_{11}+X_{22}).
\end{aligned}\tag{3.8c}$$

For $j=\iota$ we conclude from Eq. (3.8c) that the set $T=\{X_{12},X_{21},X_{2,-1},X_{-1,2}\}$ is the commutative ideal of the algebra $\mathrm{sp}_2(\iota)$, since it follows from Eqs. (3.8a), (3.8b) that $[T,A_1]\subset T$ and $[T,A_2]\subset T$, i.e., $[T,A]\subset T$. Then the structure of the algebra $\mathrm{sp}_2(\iota)$ is the semidirect sum $\mathrm{sp}_2(\iota)=T\dot{+}A=T\dot{+}(A_1\oplus A_2)$ and for the contracted group $\mathrm{Sp}_2(\iota)$ we obtain the structure of the semidirect product $\mathrm{Sp}_2(\iota)=e^T\circledS(\mathrm{Sp}_1\times\mathrm{Sp}_1)$. It must be emphasized that unlike the case of unitary groups,[2] the contracted symplectic groups are not the inhomogeneous groups in the sense of Refs. 15 and 16 since the last one's have the structure $R_{2n}\circledS\mathrm{Sp}_n$.

The general element of the algebra $\mathrm{sp}_2(j)$ may be written in the form

$$\begin{aligned}
X(j)&=r_1X_{11}+r_2X_{22}+s_1X_{1,-1}+s_2X_{2,-2}+w_1X_{-1,1}\\
&\quad+w_2X_{-2,2}+u_1X_{12}+u_2X_{-1,2}+v_1X_{21}+v_2X_{2,-1}\\
&=\begin{pmatrix} r_1 & j^2v_1 & 2w_1 & j^2u_2\\ u_1 & r_2 & u_2 & 2w_2\\ 2s_1 & j^2v_2 & -r_1 & -j^2u_1\\ v_2 & 2s_2 & -v_1 & -r_2\end{pmatrix}.
\end{aligned}\tag{3.9}$$

We shall obtain the finite group transformation matrix $\Xi(j)$ by the Cayley–Hamilton[17] theorem. The characteristic equation $\det(X(j)-\lambda\cdot E)=0$ of the matrix $X(j)$ is the following biquadratic equation:

$$\lambda^4-p\lambda^2+q=0,$$

$$p=r^2+4(\mathbf{s},\mathbf{w})+2j^2(\mathbf{u},\mathbf{v}),$$

$$\begin{aligned}
q=&(r_1^2+4s_1w_1)(r_2^2+4s_2w_2)+j^4(\mathbf{u},\mathbf{v})^2+2j^2(r_1r_2u_2v_2\\
&-r_1r_2u_1v_1-2r_1u_1u_2s_2-2r_1v_1v_2w_2-2r_2u_2v_1s_1\\
&-2r_2u_1v_2w_1+2v_1^2s_1w_2+2u_1^2w_1s_2\\
&-2v_2^2w_1w_2-2u_2^2s_1s_2).
\end{aligned}\tag{3.10}$$

Let $\lambda^2=z$, then $z_{1,2}=\frac{1}{2}(p\pm\sqrt{p^2-4q})$ and the roots of Eq. (3.10) are as follows: $\lambda_{1,2}=\pm\sqrt{z_1}$, $\lambda_{3,4}=\pm\sqrt{z_2}$. The matrix $\Xi(j)=\exp X(j)$ is given by the Cayley–Hamilton theorem in the form

$$\begin{aligned}
\Xi(j)=&(p^2-4q)^{-1/2}\{E\cdot(z_1\mathrm{ch}\sqrt{z_2}-z_2\mathrm{ch}\sqrt{z_1})\\
&+X(j)\cdot(z_1(z_2)^{-1/2}\mathrm{sh}\sqrt{z_2}-z_2(z_1)^{-1/2}\mathrm{sh}\sqrt{z_1})\\
&+X^2(j)\cdot(\mathrm{ch}\sqrt{z_1}-\mathrm{ch}\sqrt{z_2})+X^3(j)\cdot((z_1)^{-1/2}\\
&\times\mathrm{sh}\sqrt{z_1}-(z_2)^{-1/2}\mathrm{sh}\sqrt{z_2})\},
\end{aligned}\tag{3.11}$$

where the matrices $X^2(j)$ and $X^3(j)$ are characterized by the matrix elements

$$\begin{aligned}
&(X^2(j))_{kk}=(X^2(j))_{-k,-k}\\
&\qquad=r_k^2+4s_kw_k+j^2(\mathbf{u},\mathbf{v}),\quad k=1,2,\\
&(X^2(j))_{21}=u_1(r_1+r_2)+2(s_1u_2+v_2w_2),\\
&(X^2(j))_{-1,-2}=j^2(X^2(j))_{21},\\
&(X^2(j))_{-2,-1}=v_1(r_1+r_2)+2(u_2s_2+v_2w_1),\\
&(X^2(j))_{12}=j^2(X^2(j))_{-2,-1},\\
&(X^2(j))_{-2,1}=v_2(r_1-r_2)+2(u_1s_2-v_1s_1),\\
&(X^2(j))_{-1,2}=-j^2(X^2(j))_{-2,1},\\
&(X^2(j))_{2,-1}=u_2(r_2-r_1)+2(u_1w_1-v_1w_2),\\
&(X^2(j))_{1,-2}=-j^2(X^2(j))_{2,-1},\\
&(X^2(j))_{1,-1}=(X^2(j))_{-1,1}=(X^2(j))_{2,-2}\\
&\qquad=(X^2(j))_{-2,2}=0.
\end{aligned}\tag{3.12}$$

$$\begin{aligned}
&(X^3(j))_{11}=-(X^3(j))_{-1,-1}\\
&\quad=r_1[r_1^2+4s_1w_1\\
&\qquad+2j^2(\mathbf{u},\mathbf{v})]+j^2[r_2(u_1v_1-u_2v_2)\\
&\qquad+2(u_1u_2s_2+v_1v_2w_2)],\\
&(X^3(j))_{22}=-(X^3(j))_{-2,-2}=r_2[r_2^2+4s_2w_2\\
&\qquad+2j^2(\mathbf{u},\mathbf{v})]+j^2[r_1(u_1v_1-u_2v_2)\\
&\qquad+2(u_1v_2w_1+u_2v_1s_1)],\\
&(X^3(j))_{21}=u_1[r_1^2+r_1r_2+r_2^2+4(\mathbf{s},\mathbf{w})+j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2(r_1v_2w_2+r_2u_2s_1-2v_1s_1w_2),\\
&(X^3(j))_{-2,1}=v_2[r_1^2-r_1r_2+r_2^2+4(\mathbf{s},\mathbf{w})+j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2(r_1u_1s_2+r_2v_1s_1+2u_2s_1s_2),\\
&(X^3(j))_{2,-1}=u_2[r_1^2-r_1r_2+r_2^2+4(\mathbf{s},\mathbf{w})+j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2(r_1v_1w_2+r_2u_1w_1+2v_2w_1w_2),\\
&(X^3(j))_{-2,-1}=-v_1[r_1^2+r_1r_2+r_2^2+4(\mathbf{s},\mathbf{w})+j^2(\mathbf{u},\mathbf{v})]\\
&\qquad-2(r_1u_2s_2+r_2v_2w_1-2u_1s_2w_1),\\
&(X^3(j))_{-1,1}=2s_1[r_1^2+4s_1w_1+2j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2j^2(r_2u_1v_2+v_2^2w_2-u_1^2s_2),\\
&(X^3(j))_{-2,2}=2s_2[r_2^2+4s_2w_2+2j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2j^2(r_1v_1v_2+v_2^2w_1-v_1^2s_1),\\
&(X^3(j))_{1,-1}=2w_1[r_1^2+4s_1w_1+2j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2j^2(r_2u_2v_1+u_2^2s_2-v_1^2w_2),\\
&(X^3(j))_{2,-2}=2w_2[r_2^2+4s_2w_2+2j^2(\mathbf{u},\mathbf{v})]\\
&\qquad+2j^2(r_1u_1u_2+u_2^2s_1-u_1^2w_1),\\
&(X^3(j))_{-1,-2}=-j^2(X^3(j))_{21},\\
&(X^3(j))_{12}=-j^2(X^3(j))_{-2,-1},\\
&(X^3(j))_{1,-2}=j^2(X^3(j))_{2,-1},\\
&(X^3(j))_{-1,2}=j^2(X^3(j))_{-2,1}.
\end{aligned}\tag{3.13}$$

According to Eq. (2.19) the matrix $\Xi(j)$ is only needed to obtain the kernel of the finite transformation operator of

$Sp_2(j), j = 1,\iota$ in the Jordan–Schwinger representation. We shall not write out this kernel.

IV. CONCLUSION

In this series of papers we have defined the orthogonal, unitary, and symplectic Cayley–Klein groups as the groups which are obtained from the classical one's by the Inonu–Wigner contractions and analytical continuations. It has been shown that all these groups are described in the unified way by the introducing n parameters $\mathbf{j} = (j_1 j_2, \ldots j_n)$ each of which were equal to the real, dual, or imaginary units. It must be emphasized that all Cayley–Klein groups in each series depend on the same number of independent group parameters as the corresponding classical groups.

Using the well-developed theory of quantum systems with the quadratic in creation and annihilation boson or fermion operator Hamiltonians, we have built the Jordan–Schwinger representations of the groups under consideration. The matrix elements of the Jordan–Schwinger representation of the finite group transformation operator in a Glauber coherent state basis are obtained with help of the finite transformation matrix Ξ. In the case of contractions these matrix elements are completely defined by the matrix Ξ and in the case of both contractions and analytical continuations we have introduced the map $\Psi(\mathbf{j})$ that transform the matrix Ξ into the matrix ξ,η,η_1. The last case includes a nonlinear operation of obtaining the inverse matrix ξ^{-1} and therefore for the matrix elements we have the more complicated equations, as in the first case. For the boson representations the matrix elements under consideration are the generating functions for the matrix elements in discrete Fock bases. The last matrix elements are expressed in terms of Hermite polynomials of several variables with zero arguments.

The Jordan–Schwinger representations of groups are closely connected with the properties of stationary quantum systems whose Hamiltonians are quadratic in creation and annihilation operators. The replacement of a group parameters $\mathbf{r}$ by $(i/\hbar)\, t\mathbf{r}$, where t is a time variable, transforms the matrix elements of the finite group operator into the Greens function of corresponding quantum systems. Thus, the unified description of the Jordan–Schwinger representations of the Cayley–Klein groups gives us opportunity to investigate the sets of the stationary quantum systems.

[1]N. A. Gromov and V. I. Man'ko, J. Math. Phys. **31**, 1047 (1990).
[2]N. A. Gromov and V. I. Man'ko, J. Math. Phys. **31**, 1054 (1990).
[3]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN—Polish Scientific, Warszawa, 1977).
[4]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison-Wesley, London, 1981).
[5]P. Pajas and R. Raczka, J. Math. Phys. **9**, 1188 (1968).
[6]P. Pajas, J. Math. Phys. **10**, 1777 (1969).
[7]P. I. Holod and A. U. Klimyk, Rep. Math. Phys. **21**, 127 (1985).
[8]M. A. Lohe and C. A. Hurst, J. Math. Phys. **12**, 1882 (1971).
[9]R. I. Pimenov, Litovski Mat. Sbornik **5**, 457 (1965) (in Russian).
[10]I. M. Jaglom, *Complex Numbers* (Fizmatgiz, Moscow, 1963) (in Russian).
[11]B. A. Rosenfeld, *Noneuclidean Spaces* (Nauka, Moscow, 1969) (in Russian).
[12]N. A. Gromov, Preprint Series "Scientific Reports," No. 126, Komi Branch of the Acad. Sci. USSR, Syktyvkar (1984) (in Russian).
[13]N. A. Gromov, in *Group Theoretical Methods in Physics* (Nauka, Moscow, 1986), Vol. 2, p. 183 (in Russian).
[14]E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[15]M. Perroud, J. Math. Phys. **24**, 1381 (1983).
[16]M. Chaichian, A. P. Demichev, and N. F. Nelipa, Commun. Math. Phys. **90**, 353 (1983).
[17]G. A. Korn and T. M. Korn, *Mathematical Handbook* (McGraw-Hill, New York, 1961).

Application of the eigenfunction method to the icosahedral group

Fa Liu and Jia-Lun Ping
Department of Physics, Nanjing University, Nanjing, People's Republic of China

Jin-Quan Chen
Center of Theoretical Physics, China Center of Advanced Science and Technology (World Lab.), Beijing, People's Republic of China[a] *and Department of Physics, Nanjing University, Nanjing, People's Republic of China*

(Received 20 January 1989; accepted for publication 20 December 1989)

The group table for the icosahedral group **I** is constructed by using the isomorphism between the group **I** and a subgroup of the permutation group S_{12}. The single-valued irreducible representations and Clebsch–Gordan (CG) coefficients of **I** are calculated by a computer code based on the eigenfunction method. The irreducible matrix elements for all the 60 group elements are given explicitly in the form of $\sqrt{m/n}\,[\exp(i\phi)]^p[2\cos\phi]^q[2\cos 2\phi]^r$, where m, n, p, q, and r are integers and $\phi = 2\pi/5$. The Clebsch–Gordan coefficients of **I** are all real under a new phase convention for time reverse states and tabulated in the form of $\sqrt{m/n}$.

I. INTRODUCTION

The icosahedral group **I** is the most complicated point group and has been the subject of many studies.[1–6] The discovery of the quasicrystal, or the icosahedral crystals,[7] has revived the interest in the group. Early works[2–4] are mainly concerned with the construction of the $\mathbf{SO}_3{\downarrow}\mathbf{I}$ subduced basis, namely the linear combinations of the spherical harmonics adapted to the symmetry of the group **I**. Although the primitive character of the group has been known for a long time, its irreducible matrices are not readily available except for three generators of the group.[1,4] Using the $\mathbf{O}_3{\downarrow}\mathbf{I}$ subduction coefficients given by McLellian,[4] Goulding[5] calculated the $3jm$-symbol of the group **I**, and later Pooler[6] calculated both the $3jm$- and $6j$-symbols of **I**. All the above studies are based on the fact that the group **I** is a subgroup of the rotation group SO_3 and use the subduction to construct the irreducible representations and $3jm$-symbols of the group **I** from their counterparts of the group SO_3.

Conscious of the fact that there is no universal and simple method for finding characters and irreps of a finite group, Chen[8] developed a new approach to group representation theory which, in turn, gives rise to a new method, the eigenfunction method (EFM) for calculating characters, irreps, CG coefficients, isoscalar factors, etc. The EFM has been successfully applied to point groups, permutation groups, unitary groups, and space groups (for an extensive review the reader is referred to the monograph[8] and Ref. 9). Recently, a versatile space group program package based on the EFM has been written by two of us[10] that can be used to compute *ab initia* the single- and double-valued irreps (projective irreps) of the 32 point groups (little cogroup of the 230 space groups), as well as the point group or space group CG coefficients. The program is written in FORTRAN-77 and implemented on the IBM-PC. The only input is the name of the space group or the point group and the wave vector to be considered [for point groups one only needs to set the wave vector to be (0,0,0)]. It is quite interesting to use the same program with minor modifications to calculate all the irreps and CG coefficients of the most complicated point group, the icosahedral group **I**. In this paper, the irreducible representations and CG coefficients of the group **I** are constructed solely from the group table of **I** without invoking any knowledge of the group SO_3.

II. RETROSPECT ON THE EFM

The essence of the EFM is best illustrated[11] in the three-dimensional rotation group SO_3. According to the terminology in Refs. 8 and 9, the Casimir operator $\mathbf{J}^2$ of SO_3 is called the first kind of complete set of commuting operators (CSCO-I) of SO_3, which is a CSCO in the class parameter space. The eigenfunction of $\mathbf{J}^2$ in the class parameter space is proportional to the complex conjugate of the primitive character. The operator set $(\mathbf{J}^2, J_z)$ is called the second kind of CSCO (CSCO-II) of SO_3, whose eigenfunction $|jm\rangle$ gives the $SO_3 \supset SO_2$ irreducible basis. The operator set $(\mathbf{J}^2, J_z, \overline{J}_z)$ is called the third kind of CSCO (CSCO-III), which is a CSCO in the group parameter space, where $\overline{J}_z$ is the Casimir operator of the subgroup $\overline{SO}_2$ of the intrinsic group $\overline{SO}_3$, which is commuting and anti-isomorphic with the rotation group SO_3 and describes the rotation of a system (such as a deformed nucleus) around its intrinsic (body fixed) axes.[11] Physically, $\overline{J}_z$ is the third component of the angular momentum in the intrinsic coordinate system,[11] usually denoted by J_3. The eigenfunction of the CSCO-III is the complex conjugate of the irreducible matrix of SO_3, $D^j_{mk}(\alpha,\beta,\gamma)^*$.

It was shown that the above approach can be extended to any compact group.[8,9] For a finite group **G**, the Casimir operator $\mathbf{J}^2$ is replaced by the CSCO-I of the finite group **G**, denoted by C, which is a linear combination of a few class operators of **G** and is the analogy of the Casimir operator in Lie groups. The eigenvectors of C in the class space are proportional to the complex conjugate of the character vectors. The characters are obtained by further using the normalization condition for the character stemmed from the orthogonal theorems of the characters. Suppose $\mathbf{G}(s)$ is a canonical subgroup chain of **G**, and $C(s)$ is an appropriate linear com-

[a] Permanent address.

 0022-2488/90/051065-11$03.00

bination of all the CSCO-I's of the subgroups contained in $\mathbf{G}(s)$. Then, $(C,C(s))$ is the CSCO-II of $\mathbf{G}$ and its eigenvectors give the irreducible basis of $\mathbf{G}$. Similarly suppose that $\overline{\mathbf{G}}(s)$ is the corresponding subgroup chain of the intrinsic group $\overline{\mathbf{G}}$, which is commuting and anti-isomorphic with the group $\mathbf{G}$. Note that $\overline{\mathbf{G}}(s)$ has the corresponding operator set $\overline{C}(s)$. Then, $(C,C(s),\overline{C}(s))$ is the CSCO-III of $\mathbf{G}$, whose eigenvectors (after proper normalization and using the standard phase convention[8,9]) in the group space give the complex conjugate of the irreducible matrix vector $\{D^{\nu}_{ab}(R_1), D^{\nu}_{ab}(R_2),...,D^{\nu}_{ab}(R_{|G|})\}$, where ν, a, and b are the eigenvalues of C, $C(s)$, and $\overline{C}(s)$, respectively, while $R_1, R_2,..., R_{|G|}$ are the elements of the group $\mathbf{G}$ of order $|\mathbf{G}|$. In practice it is very convenient to linearly combine the operators C, $C(s)$, and $\overline{C}(s)$ into a single operator K which is a CSCO in the group space and can be served as the CSCO-III of $\mathbf{G}$. Having done this, we only need to solve the eigenequation of a single operator K to find the irreducible matrices of $\mathbf{G}$.

The prerequisite condition for the application of the eigenfunction method is that the group table is known. In the following section we describe a way to construct the group table of the group $\mathbf{I}$.

III. GROUP TABLE

A regular icosahedron(icosahedron for short) has 12 vertices, 20 faces (which consist of identical regular triangles), and 30 edges. The rotation axes of the group $\mathbf{I}$ consists of 6 fivefold axes (joining the two opposite vertices), A_{5j} $(j=1,2,...,6)$, 10 threefold axes (joining the centers of two opposite faces), A_{3j} $(j=1,2,...,10)$, and 15 twofold axes (joining the midpoints of two opposite edges), A_{2j} $(j=1,2,...,15)$. The vertices of the icosahedron, the centers of the triangles, and the midpoints of the edges are indexed as in Figs. 1 and 2, and the rotation axes are listed in Table I by listing the indices of the vertices, centers, or midpoints through which they pass.

The rotation operators of the group $\mathbf{I}$ are denoted by

$$C^{m}_{5,j} \quad (j=1,2,...,6; \quad m=1,2,3,4),$$
$$C^{m}_{3,j} \quad (j=1,2,...,10; \quad m=1,2), \qquad (1)$$
$$C_{2,j} \quad (j=1,2,...,15).$$

Together with the identity, they form the group of $\mathbf{I}$ with order 60. The group elements are denoted by R_i, $i=1,2,...,60$.

For constructing the group table, it is convenient to use the permutations of the 12 vertices fo the icosahedron under the rotations, which form a subgroup of the permutation group S_{12}, to replace the rotation operations. (The isomorphism between the group $\mathbf{I}$ and the subgroup of S_{12} is shown in the first table in the Appendix.) With this as input, the group table of the group $\mathbf{I}$ is generated by the computer by using the multiplication rule of the permutation group and is shown in the second table in the Appendix.

IV. THE CSCO-I AND CHARACTERS

From the group table, the program[10] will find the class operators and the class multiplication tables, the CSCO-I and the primitive characters. For a detailed description of the program, the reader is referred to Ref. 10.

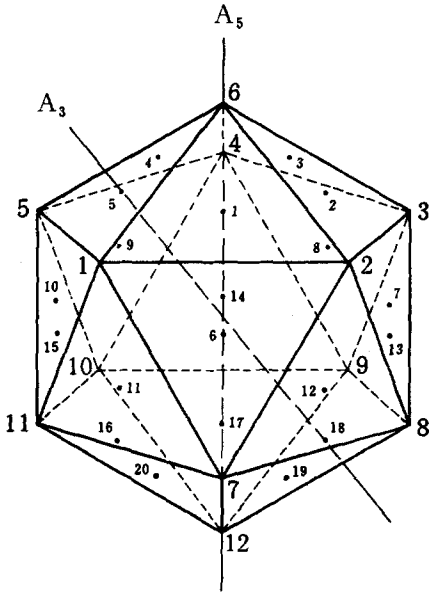


FIG. 1. The large size integers label the 12 vertices and the small size integers the centers of the 20 faces.

The five class operators of the group $\mathbf{I}$ are found as follows:

$$C_1 = E = R_1, \quad C_2 = \sum_{j=1}^{6} (C^{1}_{5,j} + C^{4}_{5,j}) = \sum_{i=2}^{13} R_i, \qquad (2a)$$

$$C_3 = \sum_{j=1}^{6} (C^{2}_{5,j} + C^{3}_{5,j}) = \sum_{i=14}^{25} R_i, \qquad (2b)$$

$$C_4 = \sum_{j=1}^{10} (C^{1}_{3,j} + C^{2}_{3,j}) = \sum_{i=26}^{45} R_i,$$

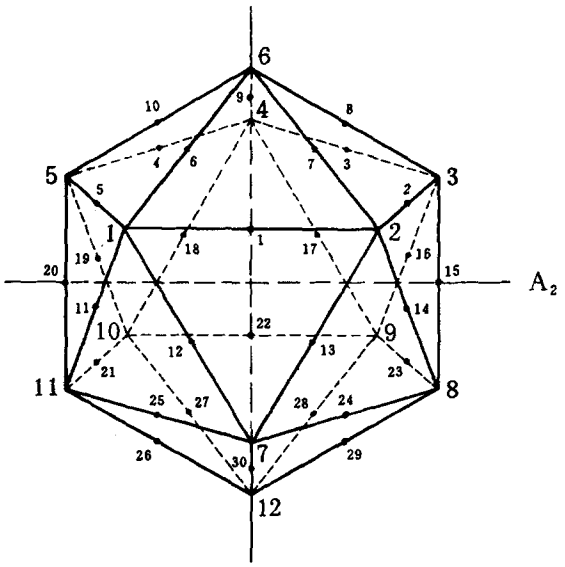


FIG. 2. The large size integers label the 12 vertices and the small size integers the midpoints of the 30 faces.

TABLE I. The rotation axes of the group I. Here $m \to n$ denotes an axis going from the point m to the point n, its positive direction being toward the point n.

Fivefold axes		Threefold axes		Twofold axes	
$A_{5,1}$	$9 \to 1$	$A_{3,1}$	$17 \to 1$	$A_{2,1}$	$22 \to 1$
$A_{5,2}$	$10 \to 2$	$A_{3,2}$	$16 \to 2$	$A_{2,2}$	$21 \to 2$
$A_{5,3}$	$11 \to 3$	$A_{3,3}$	$20 \to 3$	$A_{2,3}$	$25 \to 3$
$A_{5,4}$	$7 \to 4$	$A_{3,4}$	$19 \to 4$	$A_{2,4}$	$24 \to 4$
$A_{5,5}$	$8 \to 5$	$A_{3,5}$	$18 \to 5$	$A_{2,5}$	$23 \to 5$
$A_{5,6}$	$12 \to 6$	$A_{3,6}$	$14 \to 6$	$A_{2,6}$	$28 \to 6$
		$A_{3,7}$	$15 \to 7$	$A_{2,7}$	$27 \to 7$
		$A_{3,8}$	$11 \to 8$	$A_{2,8}$	$26 \to 8$
		$A_{3,9}$	$12 \to 9$	$A_{2,9}$	$30 \to 9$
		$A_{3,10}$	$13 \to 10$	$A_{2,10}$	$29 \to 10$
				$A_{2,11}$	$16 \to 11$
				$A_{2,12}$	$17 \to 12$
				$A_{2,13}$	$18 \to 13$
				$A_{2,14}$	$19 \to 14$
				$A_{2,15}$	$20 \to 15$

$$C_5 = \sum_{j=1}^{15} C_{2,j} = \sum_{i=46}^{60} R_i. \tag{2c}$$

The CSCO-I of the group **I** is found as $C = C_2$. In the class space spanned by $C_1,\ldots,C_5$, the representation matrix of C_2 is found to be

$$D(C_2) = \begin{pmatrix} 0 & 12 & 0 & 0 & 0 \\ 1 & 5 & 1 & 5 & 0 \\ 0 & 1 & 1 & 5 & 5 \\ 0 & 3 & 3 & 3 & 3 \\ 0 & 0 & 4 & 4 & 4 \end{pmatrix}. \tag{3}$$

By diagonalizing $D(C_2)$, we obtain five distinct eigenvalues that can be served as the irrep label and five character vectors that are well known and not listed here. The correspondence between the eigenvalues of C_2 and the Mülliken notation is as follows:

The eigenvalues of C_2:

$$12 \quad 8\cos(\pi/5) \quad 8\cos(3\pi/5) \quad -3 \quad 0.$$

The Mülliken notation:

$$A \quad T_1 \quad T_2 \quad G \quad H.$$

V. THE CSCO-III AND IRREDUCIBLE MATRICES

The subgroup or subgroup chain $G(s)$ used for classifying the irreducible matrices can be specified either by the user according to one's need or by the computer. We choose the cyclic group $\mathbb{C}_5 = (E, C_{5,1}^1, C_{5,1}^2, C_{5,1}^3, C_{5,1}^4) = (R_1, R_2, R_{14}, R_{15}, R_3)$ as the subgroup $\mathbf{G}(s)$. Once $\mathbf{G}(s)$ is specified, the program[10] will find the CSCO-I of $\mathbf{G}(s)$. In our case it is trivial, since any element of $\mathbb{C}_5$ except the identity can be chosen as its CSCO-I. The operator $C_{5,1}^1 = R_2$ has been chosen as $C(s)$ and its rotation axis is chosen as the z axis,

$$C(s) = R_2 = R_z(2\pi/5) = \exp(-2\pi i J_z/5). \tag{4a}$$

The corresponding intrinsic operator is $\overline{C}(s) = \overline{C}_{5,1}^1 = \overline{R}_2$. The program will find a single operator as the CSCO-III K of the group **I** with the result

$$K = 7C + C(s) + 9\overline{C}(s). \tag{4b}$$

From (4) and (2a), and the group table, the program will find the representation matrix of the CSCO-III K in the group space, and its 60 eigenvectors corresponding to the 60 distinct eigenvalues. With proper normalization and taking complex conjugate, they yield all the irreducible matrix elements of the group **I**. The program contains a subroutine to check that the calculated matrices do form a representation of the group. The rows and columns of the matrices are indexed according to the eigenvalue $\exp(-2\pi\mu i/5)$ of $C(s)$, where μ is the eigenvalue of J_z modulo 5. The values of the integer μ along with its index a for different irreps are listed below:

irreps	A	T_1	T_2	G	H
μ:	(0)	(1,0,−1)	(2,0,−2)	(2,1,−1,−2)	(2,1,0,−1,−2)
a:	(1)	(1,2,3)	(1,2,3,)	(1,2,3,4)	(1,2,3,4,5).

(5)

The original output of the matrix elements are complex decimals. To convert the complex decimals to the exact values, we use the following procedure. From the character table and the matrices for the generators of the group **I** given in Ref. 4, we use the ansatz that the entries are of the following form:

$$\sqrt{m/n}\, z^p P^q Q^r, \tag{6a}$$

where m, n, p, q, and r are integers, and

$$z = \exp\left(i\frac{2\pi}{5}\right), \quad P = \frac{\sqrt{5}-1}{2} = 2\cos\frac{2\pi}{5},$$
$$Q = -\frac{\sqrt{5}+1}{2} = 2\cos\frac{4\pi}{5}. \tag{6b}$$

Among the 60×60 entries, besides 0 and 1, there are only 30 distinct values, 10 being real and 20 complex, denoted by capital and small letters, respectively. With the help of a computer, all of the 30 decimal values are converted into the form of (6) and are listed in Table II. The irreducible matrix elements of the group **I** are given in Table III.

VI. THE CSCO-II AND CG COEFFICIENTS

The eigenfunction method for the CG coefficients is discussed in Refs. 8–10. Here we only give some key points. Let $|v_i a_i\rangle$, $i = 1,2$, be the two irreducible bases. By using the CG coefficients they can be linearly combined into another irreducible basis,

TABLE II. The symbols used in Table III.

A	B	D	E	F	G	H	L	M	N
$\frac{1}{\sqrt{5}}$	$-AQ$	$-\sqrt{2}A$	AP	$-A^2$	B^2	A^2P^2	$\sqrt{6}A^2$	$-2A^2P$	$-2A^2Q$
a	b	d	e	f	g	h	j	k	m
z^{-1}	$-Bz^2$	z^{-2}	$-Ez$	Dz^2	Dz^{-1}	Ez^2	Bz^{-1}	$-Az^{-1}$	Az^2
n	p	q	r	s	t	u	v	w	y
Gz^{-1}	Hz^2	Nz^{-1}	Nz^{-2}	Mz^2	Mz	Lz^{-1}	Lz^{-2}	Hz^{-1}	Gz^2

$$|\nu\tau\alpha\rangle = \sum_{a_1a_2} \langle \nu_1 a_1, \nu_2 a_2 | \nu\tau a \rangle | \nu_1 a_1 \rangle \nu_2 a_2 \rangle, \tag{7}$$

where the first factor in the right-hand side is the CG coefficient and τ is a multiplicity label, $\tau = 1,2,\ldots,(\nu_1\nu_2\nu)$. According to Sec. II, $|\nu\tau a\rangle$ is necessarily an eigenfunction of the CSCO-II, $(C,C(s))$, with the eigenvalues ν and a. Therefore, the CG coefficients result from a diagonalization of the representative matrix of the CSCO-II in the product basis,

$$\sum_{a_1a_2} \left[\left\langle \nu_1 a_1', \nu_2 a_2' \middle| \begin{matrix} C \\ C(s) \end{matrix} \middle| \nu_1 a_1, \nu_2 a_2 \right\rangle - \binom{\nu}{a} \delta_{a_1' a_1} \delta_{a_2' a_2} \right] \langle \nu_1 a_1, \nu_2 a_2 | \nu\tau a \rangle = 0, \tag{8}$$

where a can be understood either as the eigenvalue $\exp(-2\pi\mu i/5)$ or its index ordered in Eq. (5). Solving the eigenequation (8) we can get the eigenvalue (ν,a) and its

TABLE III. The single-valued irreducible representation of the icosahedral group.

NU	a	b	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	(1,1)		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	(1,1)		1	a	a*	b	b*	b*	b	b*	b	b	b*	b	b*	d	d*	e	e*	e*	e	e*	e	e	e*	e	e*	-j*	-j	-h	-h*	-h*
	(2,1)		0	0	0	f	D	f*	D	g	g*	g	f	f*	g*	0	0	-g	-f*	-g*	-f	-f	-f*	-g*	-D	-D	-g	f	g*	-D	-f*	-g*
	(3,1)		0	0	0	h	h	h*	h*	E	E	-e	-e	-e*	-e*	0	0	-b	-b	-b*	-b*	B	B	j*	j*	j	j	h*	h*	-b*	-b*	j
	(1,2)		0	0	0	D	f*	D	f	g	g*	f*	g*	g	f	0	0	-f	-g*	-f*	-g	-f	-f*	-D	-g	-g*	-D	g	f*	-f	-D	-f
	(2,2)		1	1	1	A	A	A	A	A	A	A	A	A	A	1	1	-A	-A	-A	-A	-A	-A	-A	-A	-A	A	A	-A	-A	-A	-A
	(3,2)		0	0	0	D	f	D	f*	g*	g	f	g	g*	f*	0	0	-f*	-g	-f	-g*	-f*	-f	-D	-g*	-g	-D	g*	f	-f*	-D	-f*
	(1,3)		0	0	0	h*	h*	h	h	E	E	-e*	-e*	-e	-e	0	0	-b*	-b*	-b	-b	B	B	j	j*	j*	j*	h	h	-b	-b	j*
	(2,3)		0	0	0	f*	D	f	D	g*	g	g*	f*	f	g	0	0	-g*	-f	-g	-f*	-f*	-f	-g	-D	-D	-g*	f*	g	-D	-f	-g
	(3,3)		1	a*	a	b*	b	b	b*	b	b*	b*	b	b*	b	d*	d	e*	e	e	e*	e	e*	e*	e	e*	e	-j	-j*	-h*	-h	-h
3	(1,1)		1	d	d*	e*	e	e	e*	e	e*	e*	e	e*	e	a*	a	b	b*	b*	b	b*	b	b	b*	b	b*	-h	-h*	-j	-j*	-j*
	(2,1)		0	0	0	-g	-D	-g*	-D	-f*	-f	-f*	-g	-g*	-f	0	0	f*	g*	f	g	g	g*	f	D	D	f*	-g	-f	D	g*	f
	(3,1)		0	0	0	j	j	j*	j*	B	B	-b	-b	-b*	-b*	0	0	-e*	-e*	-e	-e	E	E	h	h	h*	h*	j*	j*	-e	-e	h*
	(1,2)		0	0	0	-D	-g*	-D	-g	-f*	-f	-g*	-f	-f*	-g	0	0	g	f	g*	f*	g	g*	D	f*	f	D	-f*	-g*	g	D	g
	(2,2)		1	1	1	-A	-A	-A	-A	-A	-A	-A	-A	-A	-A	1	1	A	A	A	A	A	A	A	A	A	A	-A	-A	A	A	A
	(3,2)		0	0	0	-D	-g	-D	-g*	-f	-f*	-g	-f*	-f	-g*	0	0	g*	f*	g	f	g*	g	D	f	f*	D	-f	-g	g*	D	g*
	(1,3)		0	0	0	j*	j*	j	j	B	B	-b*	-b*	-b	-b	0	0	-e	-e	-e*	-e*	E	E	h*	h*	h	h	j	j	-e*	-e*	h
	(2,3)		0	0	0	-g*	-D	-g	-D	-f	-f*	-f	-g*	-g	-f*	0	0	f	g	f*	g*	g*	g	f*	D	D	f	-g*	-f*	D	g	f*
	(3,3)		1	d*	d	e	e*	e*	e	e*	e	e	e*	e	e*	a	a*	b*	b	b	b*	b	b*	b*	b	b*	b	-h*	-h	-j*	-j	-j
4	(1,1)		1	d	d*	k	k*	k*	k	k*	k	k	k*	k	k*	a*	a	m	m*	m*	m	m*	m	m	m*	m	m*	-m	-m*	-k	-k*	-k*
	(2,1)		0	0	0	e*	-h*	e	-h	-h	-h*	e	-E	-E	e*	0	0	-B	b	-B	b*	-j*	-j	b	-j	-j*	b*	-h*	-E	b	-j*	-j
	(3,1)		0	0	0	j	-b	j*	-b*	j	j*	B	-b*	-b	B	0	0	-e	E	-e*	E	h	h*	h	-e	-e*	h*	B	j	h*	-e	E
	(4,1)		0	0	0	k	k	k*	k*	-A	-A	-m	-m	-m*	-m*	0	0	-k	-k	-k*	-k*	A	A	m	m	m*	m*	k*	k*	-k*	-k*	m*
	(1,2)		0	0	0	-h	e	-h*	e*	-h	-h*	-E	e*	e	-E	0	0	b*	-B	b	-B	-j*	-j	-j*	b*	-h*	b	-j	-E	-h	-j	b*	-B
	(2,2)		1	a	a*	m	m*	m*	m	m*	m	m	m*	m	m*	d	d*	k*	k	k	k*	k	k*	k*	k	k*	k	-k*	-k	-m	-m*	-m*
	(3,2)		0	0	0	m	m	m*	m*	A	A	-k*	-k*	-k	-k	0	0	-m	-m	-m*	-m*	-A	-A	k*	k*	k	k	m*	m*	-m*	-m*	k
	(4,2)		0	0	0	-b	j	-b*	j*	j*	j	-b*	B	B	-b	0	0	E	-e	E	-e*	h*	h	-e	h	h*	-e*	j	B	-e	h*	h
	(1,3)		0	0	0	-b*	j*	-b	j	j	j*	-b	B	B	-b*	0	0	E	-e*	E	-e	h	h*	-e*	h*	h	-e	j*	B	-e*	h	h*
	(2,3)		0	0	0	m*	m*	m	m	A	A	-k	-k	-k*	-k*	0	0	-m*	-m*	-m	-m	-A	-A	k	k	k*	k*	m	m	-m	-m	k*
	(3,3)		1	a*	a	m*	m	m	m*	m	m*	m*	m	m*	m	d*	d	k	k*	k*	k	k*	k	k	k*	k	k*	-k	-k*	-m*	-m	-m
	(4,3)		0	0	0	-h*	e*	-h	e	-h*	-h	-E	e	e*	-E	0	0	b	-B	b*	-B	-j	-j*	-j	b	b*	-j*	-E	-h*	-j*	b	-B
	(1,4)		0	0	0	k*	k*	k	k	-A	-A	-m*	-m*	-m	-m	0	0	-k*	-k*	-k	-k	A	A	m*	m*	m	m	k	k	-k	-k	m
	(2,4)		0	0	0	j*	-b*	j	-b	j*	j	B	-b	-b*	B	0	0	-e*	E	-e	E	h*	h	h*	-e*	-e	h	B	j*	h	-e*	E
	(3,4)		0	0	0	e	-h	e*	-h*	-h*	-h	e*	-E	-E	e	0	0	-B	b*	-B	b	-j	-j*	b*	-j*	-j	b	-h	-E	b*	-j	-j*
	(4,4)		1	d*	d	k*	k	k	k*	k	k*	k*	k	k*	k	a	a*	m*	m	m	m*	m	m*	m*	m	m*	m	-m*	-m	-k*	-k	-k
5	(1,1)		1	d	d*	n	n*	n*	n	n*	n	n	n*	n	n*	a*	a	p	p*	p*	p	p*	p	p	p*	p	p*	y	y*	w	w*	w*
	(2,1)		0	0	0	q	r	q*	r*	r*	r	q*	N	N	q	0	0	M	s	M	s*	t	t*	s	t*	t	s*	r	N	s	t	t*
	(3,1)		0	0	0	u	L	u*	L	v	v*	v	u	u*	v*	0	0	v	u*	v*	u	u	u*	v*	L	L	v	u	v*	L	u*	v*
	(4,1)		0	0	0	t*	s	t	s*	t*	t	M	s*	s	M	0	0	q*	N	q	N	r*	r	r*	q*	q	r	M	t*	r	q*	N
	(5,1)		0	0	0	w	w	w*	w*	H	H	p	p	p*	p*	0	0	n	n	n*	n*	G	G	y	y	y*	y*	w*	w*	n*	n*	y*
	(1,2)		0	0	0	r*	q*	r	q	r*	r	N	q	q*	N	0	0	s*	M	s	M	t	t*	t	s*	s	t*	N	r*	t*	s*	M
	(2,2)		1	a	a*	p	p*	p*	p	p*	p	p	p*	p	p*	d	d*	n*	n	n	n*	n	n*	n*	n	n*	n	w*	w	y	y*	y*
	(3,2)		0	0	0	-v*	-L	-v	-L	-u	-u*	-u	-v*	-v	-u*	0	0	-u	-v	-u*	-v*	-v*	-v	-u*	-L	-L	-u	-v*	-u*	-L	-v	-u*
	(4,2)		0	0	0	y	y	y*	y*	G	G	n*	n*	n	n	0	0	p	p	p*	p*	H	H	w*	w*	w	w	y*	y*	p*	p*	w
	(5,2)		0	0	0	s	t*	s*	t	t	t*	s*	M	M	s	0	0	N	q*	N	q	r	r*	q*	r*	r	q	t*	M	q*	r	r*
	(1,3)		0	0	0	L	u*	L	u	v	v*	u*	v*	v	u	0	0	u	v*	u*	v	u	u*	L	v	v*	L	v	u*	u	L	u
	(2,3)		0	0	0	-L	-v	-L	-v*	-u	-u*	-v	-u*	-u	-v*	0	0	-v*	-u*	-v	-u	-v*	-v	-L	-u	-u*	-L	-u	-v	-v*	-L	-v*
	(3,3)		1	1	1	F	F	F	F	F	F	F	F	F	F	1	1	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
	(4,3)		0	0	0	-L	-v*	-L	-v	-u*	-u	-v*	-u	-u*	-v	0	0	-v	-u	-v*	-u*	-v	-v*	-L	-u*	-u	-L	-u*	-v*	-v	-L	-v
	(5,3)		0	0	0	L	u	L	u*	v*	v	u	v	v*	u*	0	0	u*	v	u	v*	u*	u	L	v*	v	L	v*	u	u*	L	u*
	(1,4)		0	0	0	s*	t	s	t*	t*	t	s	M	M	s*	0	0	N	q	N	q*	r*	r	q	r	r*	q*	t	M	q	r*	r
	(2,4)		0	0	0	y*	y*	y	y	G	G	n	n	n*	n*	0	0	p*	p*	p	p	H	H	w	w	w*	w*	y	y	p	p	w*
	(3,4)		0	0	0	-v	-L	-v*	-L	-u*	-u	-u*	-v	-v*	-u	0	0	-u*	-v*	-u	-v	-v	-v*	-u	-L	-L	-u*	-v	-u	-L	-v*	-u
	(4,4)		1	a*	a	p*	p	p	p*	p	p*	p*	p	p*	p	d*	d	n	n*	n*	n	n*	n	n	n*	n	n*	w	w*	y*	y	y
	(5,4)		0	0	0	r	q	r*	q*	r	r*	N	q*	q	N	0	0	s	M	s*	M	t*	t	t*	s	s*	t	N	r	t	s	M
	(1,5)		0	0	0	w*	w*	w	w	H	H	p*	p*	p	p	0	0	n*	n*	n	n	G	G	y*	y*	y	y	w	w	n	n	y
	(2,5)		0	0	0	t	s*	t*	s	t	t*	M	s	s*	M	0	0	q	N	q*	N	r	r*	r	q	q*	r*	M	t	r*	q	N
	(3,5)		0	0	0	u*	L	u	L	v*	v	v*	u*	u	v	0	0	v*	u	v	u*	u*	u	v	L	L	v*	u*	v	L	u	v
	(4,5)		0	0	0	q*	r*	q	r	r	r*	q	N	N	q*	0	0	M	s*	M	s	t*	t	s*	t	t*	s	r*	N	s*	t*	t
	(5,5)		1	d*	d	n*	n	n	n*	n	n*	n*	n	n*	n	a	a*	p*	p	p	p*	p	p*	p*	p	p*	p	y*	y	w*	w	w

TABLE III. (*Continued*)

NU	a	b	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	(1,1)		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	(1,1)		-h	-h	-h*	-j*	-j	-j*	-j	-h*	-h	-j	-j*	-h*	-h	-j*	-j	-B	0	-E	0	-B	-B	-E	0	0	-E	-B	-B	-E	0	-E
	(2,1)		-f*	-g*	-g	f*	f	g*	D	-D	-f	g	D	-f	-g	g	f*	g*	0	-f	0	f*	f	-g	0	0	-D	g	D	-f*	0	-g*
	(3,1)		j	B	B	E	E	-e	-e	-b	-b	-e*	-e*	j*	j*	h	h	h	-1	j	-d	-e	-e*	-b*	-a*	-d*	B	h*	E	j*	-a	-b
	(1,2)		-g	-g*	-g	f*	f	D	g	-f*	-D	D	g*	-g*	-f*	f	g*	g	0	-f*	0	f	f*	-g*	0	0	-D	g*	D	-f	0	-g
	(2,2)		-A	-A	-A	A	A	A	A	-A	-A	A	A	-A	-A	A	A	A	-1	-A	-1	A	A	-A	-1	-1	-A	A	A	-A	-1	-A
	(3,2)		-g*	-g	-g*	f	f*	D	g*	-f	-D	D	g	-g	-f	f*	g	g*	0	-f	0	f*	f	-g	0	0	-D	g	D	-f*	0	-g*
	(1,3)		j*	B	B	E	E	-e*	-e*	-b*	-b*	-e	-e	j	j	h*	h*	h*	-1	j*	-d*	-e*	-e	-b	-a	-d	B	h	E	j	-a*	-b*
	(2,3)		-f	-g	-g*	f	f*	g	D	-D	-f*	g*	D	-f*	-g*	g*	f	g	0	-f*	0	f	f*	-g*	0	0	-D	g*	D	-f	0	-g
	(3,3)		-h*	-h*	-h	-j	-j*	-j	-j*	-h	-h*	-j*	-j	-h	-h*	-j	-j*	-B	0	-E	0	-B	-B	-E	0	0	-E	-B	-B	-E	0	-E
3	(1,1)		-j	-j	-j*	-h	-h*	-h	-h*	-j*	-j	-h*	-h	-j*	-j	-h	-h*	-E	0	-B	0	-E	-E	-B	0	0	-B	-E	-E	-B	0	-B
	(2,1)		g*	f	f*	-g*	-g	-f	-D	D	g	-f*	-D	g	f*	-f*	-g*	-f	0	g	0	-g*	-g	f*	0	0	D	-f*	-D	g*	0	f
	(3,1)		h*	E	E	B	B	-b	-b	-e*	-e*	-b*	-b*	h	h	j	j	j	-1	h*	-a*	-b	-b*	-e	-d*	-a	E	j*	B	h	-d	-e*
	(1,2)		f*	f	f*	-g*	-g	-D	-f*	g*	D	-D	-f	f	g*	-g	-f	-f*	0	g*	0	-g	-g*	f	0	0	D	-f	-D	g	0	f*
	(2,2)		A	A	A	-A	-A	-A	-A	A	A	-A	-A	A	A	-A	-A	-A	-1	A	-1	-A	-A	A	-1	-1	A	-A	-A	A	-1	A
	(3,2)		f	f*	f	-g	-g*	-D	-f	g	D	-D	-f*	f*	g	-g*	-f*	-f	0	g	0	-g*	-g	f*	0	0	D	-f*	-D	g*	0	f
	(1,3)		h	E	E	B	B	-b*	-b*	-e	-e	-b	h*	h*	j*	j*	j*	j*	-1	h	-a	-b*	-b	-e*	-d	-a*	E	j	B	h*	-d*	-e
	(2,3)		g	f*	f	-g	-g*	-f*	-D	D	g*	-f	-D	g*	f	-f	-g	-f*	0	g*	0	-g	-g*	f	0	0	D	-f	-D	g	0	f*
	(3,3)		-j*	-j*	-j	-h*	-h	-h*	-h	-j	-j*	-h	-h*	-j	-j*	-h*	-h	-E	0	-B	0	-E	-E	-B	0	0	-B	-E	-E	-B	0	-B
4	(1,1)		-k	-k	-k*	-m	-m*	-m	-m*	-k*	-k	-m*	-m	-k*	-k	-m	-m*	-A	0	A	0	-A	-A	A	0	0	A	-A	-A	A	0	A
	(2,1)		-B	b*	b	e*	e	-h	e*	b*	-j	-h*	e	-B	-j*	-E	-h	e	0	b	0	-h*	-h	-j	0	0	-B	e*	-E	b*	0	-j*
	(3,1)		h	-e	-e*	-b*	-b	-b	j*	h	-e*	-b*	j	h*	E	j*	B	-b*	0	-e	0	j	j*	h	0	0	E	-b	B	-e*	0	h*
	(4,1)		m*	A	A	-A	-A	-m	-m	-k	-k	-m*	-m*	m	m	k	k	k	-1	m*	-a*	-m	-m*	-k*	-d*	-a	A	k*	-A	m	-d	-k
	(1,2)		-j*	b*	b	e*	e	-h*	-j*	b	e*	-h	-j	-B	-h*	-E	e*	0	b*	0	-h	-h*	-j*	0	0	-B	e	-E	b	0	-j
	(2,2)		-m	-m	-m*	-k*	-k	-k*	-k	-m*	-m	-k	-k*	-m*	-m	-k*	-k	A	0	-A	0	A	A	-A	0	0	-A	A	A	-A	0	-A
	(3,2)		k	-A	-A	A	A	-k*	-k*	-m	-m	-k	-k	k*	k*	m	m	m	-1	k	-d	-k*	-k	-m*	-a*	-d*	-A	m*	A	k*	-a	-m
	(4,2)		E	-e*	-e	-b	-b*	j*	-b	-e*	h	j	-b*	E	h*	B	j*	-b*	0	-e	0	j	j*	h	0	0	E	-b	B	-e*	0	h*
	(1,3)		E	-e	-e*	-b*	-b	j	-b*	-e	h*	j*	-b	E	h	B	j	-b	0	-e*	0	j*	j	h*	0	0	E	-b*	B	-e	0	h
	(2,3)		k*	-A	-A	A	A	-k	-k	-m*	-m*	-k*	-k*	k	k	m*	m*	m*	-1	k*	-d*	-k	-k*	-m	-a	-d	-A	m	A	k	-a*	-m*
	(3,3)		-m*	-m*	-m	-k	-k*	-k	-k*	-m	-m*	-k*	-k	-m	-m*	-k	-k*	A	0	-A	0	A	A	-A	0	0	-A	A	A	-A	0	-A
	(4,3)		-j	b	b*	e	e*	e*	-h	-j	b*	e	-h*	-j*	-B	-h	-E	e	0	b	0	-h*	-h	-j	0	0	-B	e*	-E	b*	0	-j*
	(1,4)		m	A	A	-A	-A	-m*	-m*	-k*	-k*	-m	-m	m*	m*	k*	k*	k*	-1	m	-a	-m*	-m	-k	-d	-a*	A	k	-A	m*	-d*	-k*
	(2,4)		h*	-e*	-e	-b	-b*	-b*	j	h*	-e	-b	j*	h	E	j	B	-b	0	-e*	0	j*	j	h*	0	0	E	-b*	B	-e	0	h
	(3,4)		-B	b	b*	e	e*	-h*	e	b	-j*	-h	e*	-B	-j	-E	-h*	e*	0	b*	0	-h	-h*	-j*	0	0	-B	e	-E	b	0	-j
	(4,4)		-k*	-k*	-k	-m*	-m	-m*	-m	-k	-k*	-m	-m*	-k	-k*	-m*	-m	-A	0	A	0	-A	-A	A	0	0	A	-A	-A	A	0	A
5	(1,1)		w	w	w*	y	y*	y	y*	w*	w	y*	y	w*	w	y	y*	G	0	H	0	G	G	H	0	0	H	G	G	H	0	H
	(2,1)		M	s*	s	q	q*	r*	q	s*	t*	r	q*	M	t	N	r*	q*	0	s	0	r	r*	t*	0	0	M	q	N	s*	0	t
	(3,1)		u*	v*	v	u*	u	v*	L	L	u	v	L	u	v	v	u*	v*	0	u	0	u*	u	v	0	0	L	v	L	u*	0	v*
	(4,1)		r*	q*	q	s*	s	s	t	r*	q	s*	t*	r	N	t	M	s*	0	q*	0	t*	t	r*	0	0	N	s	M	q	0	r
	(5,1)		y*	G	G	H	H	p	p	n	n	p*	p*	y	y	w	w	w	1	y*	a*	p	p*	n*	d*	a	G	w*	H	y	d	n
	(1,2)		t	s*	s	q	q*	q*	r	t	s	q	r*	t*	M	r	N	q	0	s*	0	r*	r	t	0	0	M	q*	N	s	0	t*
	(2,2)		y	y	y*	w*	w	w*	w	y*	y	w	w*	y*	y	w*	w	H	0	G	0	H	H	G	0	0	G	H	H	G	0	G
	(3,2)		-v	-u*	-u	-v	-v*	-u*	-L	-L	-v*	-u	-L	-v*	-u	-u	-v	-u*	0	-v*	0	-v	-v*	-u	0	0	-L	-u	-L	-v	0	-u*
	(4,2)		w	H	H	G	G	n*	n*	p	p	n	n	w*	w*	y	y	y	1	w	d	n*	n	p*	a*	d*	H	y*	G	w*	a	p
	(5,2)		N	q	q*	s	s*	t	s	q	r*	t*	s*	N	r	M	t	s*	0	q*	0	t*	t	r*	0	0	N	s	M	q	0	r
	(1,3)		v	v*	v	u*	u	L	v	u*	L	L	v*	v*	u*	u	v*	v	0	u*	0	u	u*	v*	0	0	L	v*	L	u	0	v
	(2,3)		-u	-u*	-u	-v	-v*	-L	-u	-v	-L	-L	-u*	-u*	-v	-v*	-u*	-u	0	-v	0	-v*	-v	-u*	0	0	-L	-u*	-L	-v*	0	-u
	(3,3)		F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	1	F	1	F	F	F	1	1	F	F	F	F	1	F
	(4,3)		-u*	-u	-u*	-v*	-v	-L	-u*	-v*	-L	-L	-u	-u	-v*	-v	-u	-u*	0	-v*	0	-v	-v*	-u	0	0	-L	-u	-L	-v	0	-u*
	(5,3)		v*	v	v*	u	u*	L	v*	u	L	L	v	v	u	u*	v	v*	0	u	0	u*	u	v	0	0	L	v	L	u*	0	v*
	(1,4)		N	q*	q	s*	s	t*	s*	q*	r	t	s	N	r*	M	t*	s	0	q	0	t	t*	r	0	0	N	s*	M	q*	0	r*
	(2,4)		w*	H	H	G	G	n	n	p*	p*	n*	n*	w	w	y*	y*	y*	1	w*	d*	n	n*	p	a	d	H	y	G	w	a*	p*
	(3,4)		-v*	-u	-u*	-v*	-v	-u	-L	-L	-v	-u*	-L	-v	-u*	-u*	-v*	-u	0	-v	0	-v*	-v	-u*	0	0	-L	-u*	-L	-v*	0	-u
	(4,4)		y*	y*	y	w	w*	w	w*	y	y*	w*	w	y	y*	w	w*	H	0	G	0	H	H	G	0	0	G	H	H	G	0	G
	(5,4)		t*	s	s*	q*	q	q	r*	t*	s*	q*	r	t	M	r*	N	q*	0	s	0	r	r*	t*	0	0	M	q	N	s*	0	t
	(1,5)		y	G	G	H	H	p*	p*	n*	n*	p	p	y*	y*	w*	w*	w*	1	y	a	p*	p	n	d	a*	G	w	H	y*	d*	n*
	(2,5)		r	q	q*	s	s*	s*	t*	r	q*	s	t	r*	N	t*	M	s	0	q	0	t	t*	r	0	0	N	s*	M	q*	0	r*
	(3,5)		u	v	v*	u	u*	v	L	L	u*	v*	L	u*	v*	v*	u	v	0	u*	0	u	u*	v*	0	0	L	v*	L	u	0	v
	(4,5)		M	s	s*	q*	q	r	q*	s	t	r*	q	M	t*	N	r	q	0	s*	0	r*	r	t	0	0	M	q*	N	s	0	t*
	(5,5)		w*	w*	w	y*	y	y*	y	w	w*	y	y*	w	w*	y*	y	G	0	H	0	G	G	H	0	0	H	G	G	H	0	H

degeneracy d, which is just the coefficients $(\nu_1\nu_2\nu)$ in the CG series. For the eigenvalue (ν,a) we can obtain the $(\nu_1\nu_2\nu)$ orthogonal eigenvectors,

$$\{\langle \nu_1 a_1, \nu_2 a_2 | \nu\tau a\rangle\}, \quad \tau = 1,2,\ldots,(\nu_1\nu_2\nu). \tag{9}$$

To ensure that the CG coefficients with the same ν and τ but different a have the correct relative phase, we use the same procedure as used in Sec. III of Ref. 10 for the space group CG coefficients.

We can choose a single operator M as the CSCO-II of the group **I**. Once the CSCO-III, K, of a group **G** is found, the CSCO-II of **G** is readily obtained by deleting the $\overline{C}(s)$ term in K. Then the set of eigenequations (8) can be replaced by a single eigenequation of the operator M. From (4) it is known that the CSCO-II of the group **I** is

$$M = 7C + C(s). \tag{10}$$

The CG coefficients of the group **I** are calculated by the subroutine CG in the space group program package[10] and the results are given in Table IV. The meaning of the table heading is as follows: *NU*, the index of the irrep; a, the component index of the irrep; MUL, the multiplicity label; xx, the double index a_1a_2.

Notice that the CG series and CG coefficients are obtained simultaneously by the EFM. Collecting the first lines in each subtable, we have the CG series for the group **I** listed in Table V.

As we see from Tables III and IV that although the irreducible matrices of the group **I** are rather complicated (i.e., the matrices are not spare ones and many matrix elements are not square roots of simple fractions), the CG coefficients are rather simple with many zeroes and the nonzero elements being square roots of simple fractions.

From Table IV, we see that the CG coefficients are all real. Besides the unitarity, they have the following properties.

(1) Due to Eq. (4a), $C(s)$ can be replaced by the operator $\mathbf{J}_z$ with the eigenvalue μ modulo 5. Similar to the condition $m_1 + m_2 = m$ for the SO_3 CG coefficients $\langle j_1 m_1 j_2 m_2 | jm\rangle$, for the group **I** we have

$$\langle j_1\mu_1 j_2\mu_2 | j\mu\rangle = 0, \quad \text{unless} \quad \mu_1 + \mu_2 = \mu \pmod 5. \tag{11}$$

TABLE IV. The CG coefficients of the icosahedral group.

THE CG SERIES:

NU1(2)* NU2(2) = + 1* NU(1) + 1* NU(2) + 1* NU(5)

IRREP NU1*NU2 = 2 * 2

CG COEFFICIENTS

NU	a	MUL	11	12	13	21	22	23	31	32	33
1	1	1	0	0	A	0	A	0	A	0	0
2	1	1	0	B	0	-B	0	0	0	0	0
	2	1	0	0	-B	0	0	0	B	0	0
	3	1	0	0	0	0	0	B	0	-B	0
5	1	1	1	0	0	0	0	0	0	0	0
	2	1	0	B	0	B	0	0	0	0	0
	3	1	0	0	D	0	E	0	D	0	0
	4	1	0	0	0	0	0	B	0	B	0
	5	1	0	0	0	0	0	0	0	0	1

LIST OF SYMBOLS USED
A=SQRT(1/3) B=SQRT(1/2) D=-SQRT(1/6) E=SQRT(2/3)

THE CG SERIES:

NU1(2)* NU2(3) = + 1* NU(4) + 1* NU(5)

IRREP NU1*NU2 = 2* 3

CG COEFFICIENTS

NU	a	MUL	11	12	13	21	22	23	31	32	33
4	1	1	0	0	0	A	0	0	0	0	B
	2	1	0	A	0	0	0	0	B	0	0
	3	1	0	0	-B	0	0	0	0	-A	0
	4	1	-B	0	0	0	-A	0	0	0	0
5	1	1	0	0	0	B	0	0	0	0	-A
	2	1	0	-B	0	0	0	0	A	0	0
	3	1	0	0	0	0	1	0	0	0	0
	4	1	0	0	A	0	0	0	0	-B	0
	5	1	-A	0	0	0	0	B	0	0	0

LIST OF SYMBOLS USED
A=SQRT(2/3) B=SQRT(1/3)

THE CG SERIES:

NU1(2)* NU2(4) = + 1* NU(3) + 1* NU(4) + 1* NU(5)

IRREP NU1*NU2 = 2* 4

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	21	22	23	24	31	32	33	34
3	1	1	0	A	0	0	B	0	0	0	0	0	0	-A
	2	1	0	0	-B	0	0	0	0	0	B	0	0	0
	3	1	A	0	0	0	0	0	0	-B	0	0	-A	0
4	1	1	0	D	0	0	E	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	-E	0	0	D	0	0	0
	3	1	0	0	0	-D	0	0	E	0	0	0	0	0
	4	1	0	0	0	0	0	0	0	-E	0	0	-D	0
5	1	1	0	F	0	0	G	0	0	0	0	0	0	H
	2	1	0	0	0	0	0	-D	0	0	-E	0	0	0
	3	1	0	0	-B	0	0	0	0	0	0	-B	0	0
	4	1	0	0	0	-E	0	0	-D	0	0	0	0	0
	5	1	H	0	0	0	0	0	0	G	0	0	F	0

LIST OF SYMBOLS USED
A=SQRT(1/4) B=SQRT(1/2) D=SQRT(2/3) E=-SQRT(1/3)
F=SQRT(1/12) G=SQRT(1/6) H=SQRT(3/4)

THE CG SERIES:

NU1(2)* NU2(5) = + 1* NU(2) + 1* NU(3) + 1* NU(4) + 1* NU(5)

IRREP NU1*NU2 = 2* 5

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	15	21	22	23	24	25	31	32	33	34	35
2	1	1	0	0	A	0	0	0	B	0	0	0	D	0	0	0	0
	2	1	0	0	0	B	0	0	0	E	0	0	0	B	0	0	0
	3	1	0	0	0	0	D	0	0	0	B	0	0	0	A	0	0
3	1	1	0	-E	0	0	0	F	0	0	0	0	0	0	0	0	E
	2	1	0	0	0	-F	0	0	0	-D	0	0	0	-F	0	0	0
	3	1	E	0	0	0	0	0	0	0	0	F	0	0	0	-E	0
4	1	1	0	G	0	0	0	H	0	0	0	0	0	0	0	0	-D
	2	1	0	0	E	0	0	0	J	0	0	0	K	0	0	0	0
	3	1	0	0	0	0	K	0	0	0	J	0	0	0	E	0	0
	4	1	-D	0	0	0	0	0	0	0	0	H	0	0	0	G	0
5	1	1	0	L	0	0	0	M	0	0	0	0	0	0	0	0	0
	2	1	0	0	N	0	0	0	P	0	0	0	L	0	0	0	0
	3	1	0	0	0	-N	0	0	0	0	0	0	0	N	0	0	0
	4	1	0	0	0	0	-L	0	0	0	-P	0	0	0	-N	0	0
	5	1	0	0	0	0	0	0	0	0	0	-M	0	0	0	-L	0

LIST OF SYMBOLS USED
A=SQRT(1/10) B=-SQRT(3/10) D=-SQRT(3/5) E=-SQRT(2/5)
F=SQRT(1/5) G=SQRT(4/15) H=SQRT(2/15) J=-SQRT(8/15) K=SQRT(1/15)
L=SQRT(1/3) M=-SQRT(2/3) N=SQRT(1/2) P=-SQRT(1/6)

THE CG SERIES:

NU1(3)* NU2(3) = + 1* NU(1) + 1* NU(3) + 1* NU(5)

IRREP NU1*NU2 = 3 * 3

CG COEFFICIENTS

NU	a	MUL	11	12	13	21	22	23	31	32	33
1	1	1	0	0	A	0	A	0	A	0	0
3	1	1	0	B	0	-B	0	0	0	0	0
	2	1	0	0	-B	0	0	0	B	0	0
	3	1	0	0	0	0	B	0	-B	0	0
5	1	1	0	B	0	B	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	1
	3	1	0	0	D	0	E	0	D	0	0
	4	1	1	0	0	0	0	0	0	0	0
	5	1	0	0	0	0	B	0	B	0	0

LIST OF SYMBOLS USED
A=SQRT(1/3) B=SQRT(1/2) D=SQRT(1/6) E=-SQRT(2/3)

THE CG SERIES:

NU1(3)* NU2(4) = + 1* NU(2) + 1* NU(4) + 1* NU(5)

IRREP NU1*NU2 = 3* 4

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	21	22	23	24	31	32	33	34
2	1	1	0	0	A	0	0	B	0	0	0	0	0	A
	2	1	0	0	0	-B	0	0	0	0	B	0	0	0
	3	1	-A	0	0	0	0	0	-B	0	0	-A	0	0
4	1	1	0	0	0	0	D	0	0	0	0	0	E	0
	2	1	0	0	0	0	0	D	0	0	0	0	0	-E
	3	1	E	0	0	0	0	0	-D	0	0	0	0	0
	4	1	0	-E	0	0	0	0	0	-D	0	0	0	0
5	1	1	0	0	0	0	-E	0	0	0	0	0	D	0
	2	1	0	0	F	0	0	G	0	0	0	0	0	H
	3	1	0	0	0	B	0	0	0	0	B	0	0	0
	4	1	H	0	0	0	0	0	G	0	0	F	0	0
	5	1	0	D	0	0	0	0	0	-E	0	0	0	0

LIST OF SYMBOLS USED
A=SQRT(1/4) B=-SQRT(1/2) D=SQRT(1/3) E=-SQRT(2/3)
F=SQRT(3/4) G=SQRT(1/6) H=-SQRT(1/12)

THE CG SERIES:

NU1(3)* NU2(5) = + 1* NU(2) + 1* NU(3) + 1* NU(4) + 1* NU(5)

IRREP NU1*NU2 = 3* 5

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	15	21	22	23	24	25	31	32	33	34	35
2	1	1	0	0	0	A	0	0	B	0	0	0	0	0	0	0	-A
	2	1	0	0	0	0	-B	0	0	D	0	0	-B	0	0	0	0
	3	1	-A	0	0	0	0	0	0	0	B	0	0	A	0	0	0
3	1	1	0	0	0	E	0	0	F	0	0	0	0	0	0	0	D
	2	1	0	0	0	0	F	0	0	-A	0	0	F	0	0	0	0
	3	1	0	D	0	0	0	0	0	0	0	F	0	0	E	0	0
4	1	1	0	0	A	0	0	G	0	0	0	0	0	0	0	H	0
	2	1	0	0	0	-D	0	0	J	0	0	0	0	0	0	0	K
	3	1	K	0	0	0	0	0	0	0	J	0	0	-D	0	0	0
	4	1	0	H	0	0	0	0	0	0	0	G	0	0	A	0	0
5	1	1	0	0	L	0	0	M	0	0	0	0	0	0	0	N	0
	2	1	0	0	0	0	0	0	P	0	0	0	0	0	0	0	-N
	3	1	0	0	0	0	-L	0	0	0	0	0	L	0	0	0	0
	4	1	N	0	0	0	0	0	0	0	-P	0	0	0	0	0	0
	5	1	0	-N	0	0	0	0	0	0	0	-M	0	0	-L	0	0

LIST OF SYMBOLS USED
A=SQRT(2/5) B=-SQRT(1/5) D=SQRT(3/5) E=SQRT(1/10) F=SQRT(3/10)
G=-SQRT(8/15) H=SQRT(1/15) J=-SQRT(2/15) K=-SQRT(4/15)
L=SQRT(1/2) M=SQRT(1/6) N=-SQRT(1/3) P=-SQRT(2/3)

THE CG SERIES:

NU1(4)* NU2(4) = + 1* NU(1) + 1* NU(2) + 1* NU(3)
+ 1* NU(4) + 1* NU(5)

IRREP NU1*NU2 = 4* 4

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
1	1	1	0	0	0	A	0	0	A	0	0	A	0	0	A	0	0	0
2	1	1	0	0	B	0	0	0	0	0	-B	0	0	0	0	0	0	0
	2	1	0	0	0	-A	0	0	A	0	0	-A	0	0	A	0	0	0
	3	1	0	0	0	0	0	0	0	B	0	0	0	0	0	-B	0	0
3	1	1	0	0	0	0	0	0	0	0	0	0	B	0	0	-B	0	0
	2	1	0	0	0	-A	0	0	-A	0	0	A	0	0	A	0	0	0
	3	1	0	B	0	0	-B	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	D	0	0	0	0	-D	0	0	-D	0	0
	2	1	0	0	-D	0	0	0	0	0	-D	0	0	0	0	0	0	-D
	3	1	D	0	0	0	0	0	0	D	0	0	0	0	D	0	0	0
	4	1	0	D	0	0	D	0	0	0	0	-D	0	0	0	0	0	0
5	1	1	0	0	0	0	0	E	0	0	0	0	F	0	0	F	0	0
	2	1	0	0	-F	0	0	0	0	0	-F	0	0	0	0	0	0	E
	3	1	0	0	0	-A	0	0	A	0	0	A	0	0	-A	0	0	0
	4	1	0	0	-F	0	0	0	0	0	-F	0	0	0	0	-F	0	0
	5	1	0	F	0	0	F	0	0	0	0	0	E	0	0	0	0	0

LIST OF SYMBOLS USED
A=SQRT(1/4) B=SQRT(1/2) D=SQRT(1/3) E=SQRT(2/3) F=SQRT(1/6)

TABLE IV. (*Continued*)

THE CG SERIES:

NU1 (4) * NU2(5) = = + 1* NU(2) + 1* NU(3) + 1* NU(4) + 2* NU(5)

IRREP NU1* NU2 = 4* 5

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	15	21	22	23	24	25	31	32	33	34	35	41	42	43	44	45
2	1	1	0	0	0	A	0	0	0	B	0	0	D	0	0	0	0	0	0	0	0	E
	2	1	0	0	0	0	F	0	0	0	G	0	0	G	0	0	0	F	0	0	0	0
	3	1	E	0	0	0	0	0	0	0	0	D	0	0	B	0	0	0	A	0	0	0
3	1	1	0	0	-B	0	0	0	-E	0	0	0	0	0	0	0	-A	0	0	0	0	D
	2	1	0	0	0	0	G	0	0	0	-F	0	0	-F	0	0	0	G	0	0	0	0
	3	1	0	D	0	0	0	-A	0	0	0	0	0	0	0	-E	0	0	0	-B	0	0
4	1	1	0	0	A	0	0	0	H	0	0	0	0	0	0	0	-H	0	0	0	J	0
	2	1	0	0	0	H	0	0	0	-A	0	0	J	0	0	0	0	0	0	0	0	-H
	3	1	-H	0	0	0	0	0	0	0	0	J	0	0	-A	0	0	0	H	0	0	0
	4	1	0	J	0	0	0	-H	0	0	0	0	0	0	0	H	0	0	0	A	0	0
5	1	1	0	0	K	0	0	0	L	0	0	0	0	0	0	0	0	0	0	0	L	0
	2	1	0	0	0	0	0	0	0	-K	0	0	L	0	0	0	0	0	0	0	0	-L
	3	1	0	0	0	0	-K	0	0	0	K	0	0	-K	0	0	0	K	0	0	0	0
	4	1	L	0	0	0	0	0	0	0	0	-L	0	0	K	0	0	0	0	0	0	0
	5	1	0	-L	0	0	0	0	0	0	0	0	0	0	0	-L	0	0	0	-K	0	0
5	1	2	0	0	K	0	0	0	-M	0	0	0	0	0	0	0	N	0	0	0	-M	0
	2	2	0	0	0	N	0	0	0	K	0	0	M	0	0	0	0	0	0	0	0	-M
	3	2	0	0	0	0	K	0	0	0	K	0	0	-K	0	0	0	-K	0	0	0	0
	4	2	M	0	0	0	0	0	0	0	0	-M	0	0	-K	0	0	0	-N	0	0	0
	5	2	0	M	0	0	0	-N	0	0	0	0	0	0	0	M	0	0	0	-K	0	0

LIST OF SYMBOLS USED

A=SQRT(1/5) B=-SQRT(3/10) D=SQRT(1/20) E=SQRT(9/20) F=SQRT(1/10)

G=-SQRT(2/5) H=SQRT(2/15) J=-SQRT(8/15) K=SQRT(1/4) L=SQRT(3/8)

M=SQRT(1/24) N=SQRT(2/3)

THE CG SERIES:

NU1(5)* NU2(5) == + 1* NU(1) + 1* NU(2) + 1* NU(3)
+ 2* NU(4) + 2* NU(5)

IRREP NU1*NU2 = 5* 5

CG COEFFICIENTS

NU	a	MUL	11	12	13	14	15	21	22	23	24	25	31	32	33	34	35	41	42	43	44	45	51	52	53	54	55
1	1	1	0	0	0	0	A	0	0	0	A	0	0	0	A	0	0	0	A	0	0	0	A	0	0	0	0
2	1	1	0	0	0	A	0	0	0	B	0	0	0	-B	0	0	0	-A	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	D	0	0	0	E	0	0	0	0	0	0	0	-E	0	0	0	-D	0	0	0	0
	3	1	0	0	0	0	0	0	0	0	0	A	0	0	0	B	0	0	0	-B	0	0	0	-A	0	0	0
3	1	1	0	0	B	0	0	0	0	0	0	0	-B	0	0	0	0	0	0	0	0	-A	0	0	0	A	0
	2	1	0	0	0	0	-E	0	0	0	D	0	0	0	0	0	0	0	-D	0	0	0	E	0	0	0	0
	3	1	0	-A	0	0	0	A	0	0	0	0	0	0	0	0	B	0	0	0	0	0	0	0	-B	0	0
4	1	1	0	0	-A	0	0	0	0	0	0	0	A	0	0	0	0	0	0	0	0	-B	0	0	0	B	0
	2	1	0	0	0	-B	0	0	0	A	0	0	0	-A	0	0	0	B	0	0	0	0	0	0	0	0	0
	3	1	0	0	0	0	0	0	0	0	0	-B	0	0	0	A	0	0	0	-A	0	0	0	B	0	0	0
	4	1	0	-B	0	0	0	B	0	0	0	0	0	0	0	0	-A	0	0	0	0	0	0	0	A	0	0
4	1	2	0	0	A	0	0	0	F	0	0	0	A	0	0	0	0	0	0	0	0	G	0	0	0	G	0
	2	2	0	0	0	-G	0	0	0	A	0	0	0	A	0	0	0	-G	0	0	0	0	0	0	0	0	-F
	3	2	F	0	0	0	0	0	0	0	0	G	0	0	0	-A	0	0	0	-A	0	0	0	G	0	0	0
	4	2	0	-G	0	0	0	-G	0	0	0	0	0	0	0	0	-A	0	0	0	-F	0	0	0	-A	0	0
5	1	1	0	0	H	0	0	0	J	0	0	0	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	J	0	0	0	K	0	0	0	K	0	0	0	J	0	0	0	0	0	0	0	0	0
	3	1	0	0	0	0	H	0	0	0	K	0	0	0	-H	0	0	0	K	0	0	0	H	0	0	0	0
	4	1	0	0	0	0	0	0	0	0	0	J	0	0	0	K	0	0	0	K	0	0	0	J	0	0	0
	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H	0	0	0	J	0	0	0	H	0	0
5	1	2	0	0	L	0	0	0	M	0	0	0	L	0	0	0	0	0	0	0	0	N	0	0	0	N	0
	2	2	0	0	0	M	0	0	0	P	0	0	0	P	0	0	0	M	0	0	0	0	0	0	0	0	N
	3	2	0	0	0	0	L	0	0	0	P	0	0	0	Q	0	0	0	P	0	0	0	L	0	0	0	0
	4	2	N	0	0	0	0	0	0	0	0	M	0	0	0	P	0	0	0	P	0	0	0	M	0	0	0
	5	2	0	N	0	0	0	N	0	0	0	0	0	0	0	0	L	0	0	0	M	0	0	0	L	0	0

LIST OF SYMBOLS USED

A=SQRT(1/5) B=SQRT(3/10) D=-SQRT(2/5) E=-SQRT(1/10) F=SQRT(8/15)
G=SQRT(1/30) H=SQRT(2/7) J=-SQRT(3/7) K=-SQRT(1/14) L=SQRT(1/70)
M=SQRT(4/105) N=-SQRT(7/15) P=-SQRT(8/35) Q=SQRT(18/35)

TABLE V. The CG series of the icosahedral group.

	A	T_1	T_2	G	H
A	A	T_1	T_2	G	H
T_1		$A+T_1+H$	$G+H$	T_2+G+H	T_1+T_2+G+H
T_2			$A+T_2+H$	T_1+G+H	T_1+T_2+G+H
G				$A+T_1+T_2+G+H$	T_1+T_2+G+2H
H					$A+T_1+T_2+2G+2H$

(2) Similar to the symmetries of the $3jm$-symbols of the group I (Ref. 6), we have

$$\begin{aligned}&\langle \nu_1\mu_1,\nu_2\mu_2|\nu_3\tau\mu_3\rangle \\ &\quad=(-)^{\nu_1+\nu_2+\nu_3}(-)^{q(\nu_1\nu_2\nu_3\tau)}\langle \nu_2\mu_2,\nu_1\mu_1|\nu_3\tau\mu_3\rangle \\ &\quad=([\nu_3]/[\nu_1])^{1/2}(-)^{\nu_1+\nu_2+\nu_3}(-)^{q(\nu_1\nu_2\nu_3\tau)} \\ &\qquad\times\langle \nu_3-\mu_3,\nu_2\mu_2|\nu_1\tau-\mu_1\rangle, \end{aligned} \tag{12a}$$

where $[\nu]$ is the dimension of the irrep ν, the phase $(-)^{\nu}$ is defined by

$$(-)^{\nu}=\begin{cases}-1, \\ 1,\end{cases}\quad\text{for }\nu=\begin{cases}T_1,\ T_2, \\ A,\ G,\ H,\end{cases} \tag{12b}$$

and $(-)^{q(\nu_1\nu_2\nu_3\tau)}$ equals -1 if $(\nu_1\nu_2\nu_3\tau)$ is $(HHG\,1)$ or any permutation of $(HHG\,1)$.

(3) From Table IV we se that

$$\begin{aligned}&\langle \nu_1-\mu_1,\nu_2-\mu_2|\nu_3\tau-\mu_3\rangle \\ &\quad=\theta_{\nu_1+\nu_2+\nu_3}\langle \nu_1\mu_1,\nu_2\mu_2|\nu_3\tau\mu_3\rangle, \end{aligned} \tag{13a}$$

where $\theta_{\nu_1+\nu_2+\nu_3}=-1$, if $(\nu_1\nu_2\nu_3)$ is one of the following triples:

$$\begin{aligned}&(T_iT_iT_i),\ (T_1T_2G),\ (HHT_i),\ (HHG), \\ &(GGT_i),\ (GGG),\quad\text{for } i=1,2,\end{aligned} \tag{13b}$$

and $\theta_{\nu_1+\nu_2+\nu_3}=1$, otherwise. This leads us to introduce the following phase convention for the time reverse state:

$$\mathrm{T}|\mu\nu\rangle=\theta_{\nu}|\nu-\mu\rangle, \tag{14}$$

$$\theta_{\nu}=\begin{cases}-1, \\ 1\end{cases}\quad\text{for }\nu=\begin{cases}T_1,\ T_2,\ G, \\ A,\ H.\end{cases} \tag{13a'}$$

Then the phase $\theta_{\nu_1+\nu_2+\nu_3}$ in Eq. (13a) can be expressed as

$$\theta_{\nu_1+\nu_2+\nu_3}=\theta_{\nu_1}\theta_{\nu_2}\theta_{\nu_3}. \tag{13b'}$$

V. DISCUSSIONS

A. The irreducible representations

Speiser[1] and McLellian[4] have given the irreducible matrices for the three generators A, B, and C of the group I, where A is a $2\pi/5$ rotation about one of the fivefold axes that was chosen as the z axis, B is a π rotation about a twofold axis which was chosen as the x axis, and C represents the similar π rotation whose axis is perpendicular to that of B. In our notation, they are

$$\begin{aligned}&A=R_2=C_{5,1}^{1}=R_z(2\pi/5), \\ &B=R_{47}=C_{2,2}=R_x(\pi), \\ &C=R_{57}=C_{2,12}.\end{aligned} \tag{15}$$

The irrep labels A, T_1, T_2, G, and H are named Γ_1, Γ_2, Γ_3, Γ_4, and Γ_5 in Ref. 4, respectively. The symbols z, P, and Q in (6) are designated ϵ, α, and β in Ref. 4, respectively. Comparing our Table II with Table I and Eq. (21) in Ref. 4, we see that the matrices of the three generators for all five irreps obtained by the EFM are exactly the same as obtained by McLellian.

The irreps given here differ from that of Speiser[1] by a similarity transformation. For example, for the irrep T_1, if we make the following basis transformation $\mathbf{e}_1\to\mathbf{e}_1$, $\mathbf{e}_2\to i\mathbf{e}_2$, $\mathbf{e}_3\to\mathbf{e}_3$, then the matrices for R_2 and R_{47} remain unchanged:

$$\begin{aligned}D^{T_1}(R_2)&=\begin{vmatrix}z^{-1}&0&0\\0&1&0\\0&0&z\end{vmatrix}, \\ D^{T_1}(R_{47})&=\begin{vmatrix}0&0&-1\\0&-1&0\\-1&0&0\end{vmatrix},\end{aligned} \tag{16}$$

while the matrix of R_{57} undergoes the following transformation:

$$D^{T_1}(R_{57})=\frac{1}{\sqrt5}\begin{vmatrix}Q&-\sqrt2&P\\-\sqrt2 i&1&-\sqrt2\\P&-\sqrt2 i&Q\end{vmatrix}\to\frac{1}{\sqrt5}\begin{vmatrix}Q&-\sqrt2 i&P\\\sqrt2 i&1&\sqrt2 i\\P&-\sqrt2 i&Q\end{vmatrix}. \tag{17}$$

The transformed matrices are identical with Speiser's result.

B. The CG coefficients

The $3jm$-symbols of the group I defined by Golding[5] and Pooler[6] are related to the CG coefficients by

$$\langle \nu_1\mu_1,\nu_2\mu_2|\nu_3\tau\mu_3\rangle=[\nu_3]^{1/2}\begin{pmatrix}\nu_1&\nu_2&\nu_3\\ \mu_1&\mu_2&-\mu_3\end{pmatrix}^{\tau}. \tag{18}$$

Comparing our results with theirs it is seen that both are identical (including the multiplicity separation) up to absolute phases. The CG coefficients here are all real, while theirs are imaginary for the triples in (13b). The discrepancy comes from the different conventions for the phase of time reverse states. Instead of (14), they impose

$$\mathrm{T}|\mu\nu\rangle=|\nu-\mu\rangle, \tag{19}$$

which, in turn, leads to the following symmetry for the CG coefficient of I,

$$\langle \nu_1-\mu_1,\nu_2-\mu_2|\nu_3\tau-\mu_3\rangle=\langle \nu_1\mu_1,\nu_2\mu_2|\nu_3\tau\nu_3\rangle^{*}. \tag{20}$$

Since the property (13a) is independent of absolute phase choices, the imposition of the symmetry (20) will force the CG coefficients for the triples in (13b) to be imaginary.

C. The $SO_3\downarrow I$ subduction coefficients

The $SO_3\supset I\supset C_5$ irreducible basis $|j\beta\nu\mu\rangle$ can be expressed in terms of the $SO_3\supset SO_2$ basis $|jm\rangle$,

$$|j\beta\nu\mu\rangle=\sum_{m=\mu(\mathrm{mod}\,5)}C(jm,\beta\nu\mu)|jm\rangle, \tag{21}$$

where β is the label for distinguishing the multiple occurrence of the irrep ν of I in the irrep j of SO_3 and $C(jm,\beta\nu\mu)$ is called the subduction coefficient. (For j up to seven, β is

redundant.[4]) Under time reverse the basis $|jm\rangle$ transforms as

$$\mathbb{T}|jm\rangle = (-)^{j+m}|j-m\rangle. \tag{22}$$

From (14) and (22) we have

$$C(jm,\beta\nu\mu) = (-)^{j+m}\theta_\nu C(j-m,\beta\nu-\mu)^*. \tag{23}$$

McLellian calculated the subduction coefficients for j up to eight by using the projection operator method. Since the phase convention (19) is used in Refs. 4 and 6, their subduction coefficients satisfy the following relation:

$$C(jm,\beta\nu\mu) = (-)^{j+m}C(j-m,\beta\nu-\mu)^*. \tag{24}$$

Therefore, their subduction coefficients differ from ours by a factor of i for the irreps T_1, T_2, and G. Starting from the subduction coefficients for $j=1$,

$$|1T_1\pm1\rangle = i|1\pm1\rangle, \quad |1T_1,0\rangle = |10\rangle, \tag{25}$$

by using the following formula and the CG coefficients of the group **I** we can obtain both the subduction coefficients and $SO_3\supset\mathbf{I}$ isoscalar factor $\langle j_1\beta_1\nu_1, j_2\beta_2\nu_2\|j\beta\nu\tau\rangle$, recursively:

$$\sum_\beta \langle j_1\beta_1\nu_1\, j_2\beta_2\nu_2\|j\beta\nu\tau\rangle^* C(jm,\beta\nu\mu) = \sum_{\mu_1\mu_2 m_1 m_2} \langle \nu_1\mu_1,\nu_2\mu_2|\nu\tau\mu\rangle \times \langle j_1m_1,j_2m_2|jm\rangle \prod_{i=1}^{2} C(j_im_i,\beta_i\nu_i\mu_i). \tag{26}$$

Under our phase convention, the $SO_3\supset\mathbf{I}$ isoscalar factors remain to be real. For example, the subduction coefficients for $j=1-4$ and the isoscalar factors are listed below:

$$|1T_1,\pm1\rangle = i|1\pm1\rangle, \quad |1T_1,0\rangle = |10\rangle; \quad |2H,\pm2\rangle = -|2\pm2\rangle, \quad |2H,\pm1\rangle = i|2\pm1\rangle, \quad |2H,0\rangle = |20\rangle;$$
$$|3T_2,\pm2\rangle = -\sqrt{3/5}|3\pm2\rangle + i\sqrt{2/5}|3\mp3\rangle, \quad |3T_2,0\rangle = |30\rangle; \quad |3G,\pm2\rangle = -\sqrt{2/5}|3\pm2\rangle - i\sqrt{3/5}|3\mp3\rangle,$$
$$|3G,\pm1\rangle = -i|3\pm1\rangle; |4G,\pm2\rangle = \pm\sqrt{14/15}|4\pm2\rangle \pm\sqrt{1/15}i|4\mp3\rangle, \quad |4G,\pm1\rangle = \mp\sqrt{7/15}i|4\pm1\rangle \pm\sqrt{8/15}|4\mp4\rangle;$$
$$|4H,\pm2\rangle = -\sqrt{1/15}|4\pm2\rangle + \sqrt{14/15}i|4\pm3\rangle, \quad |4H,\pm1\rangle - \sqrt{8/15}i|4\pm1\rangle - \sqrt{7/15}|4\mp4\rangle, \quad |4H,0\rangle = |40\rangle; \tag{27}$$

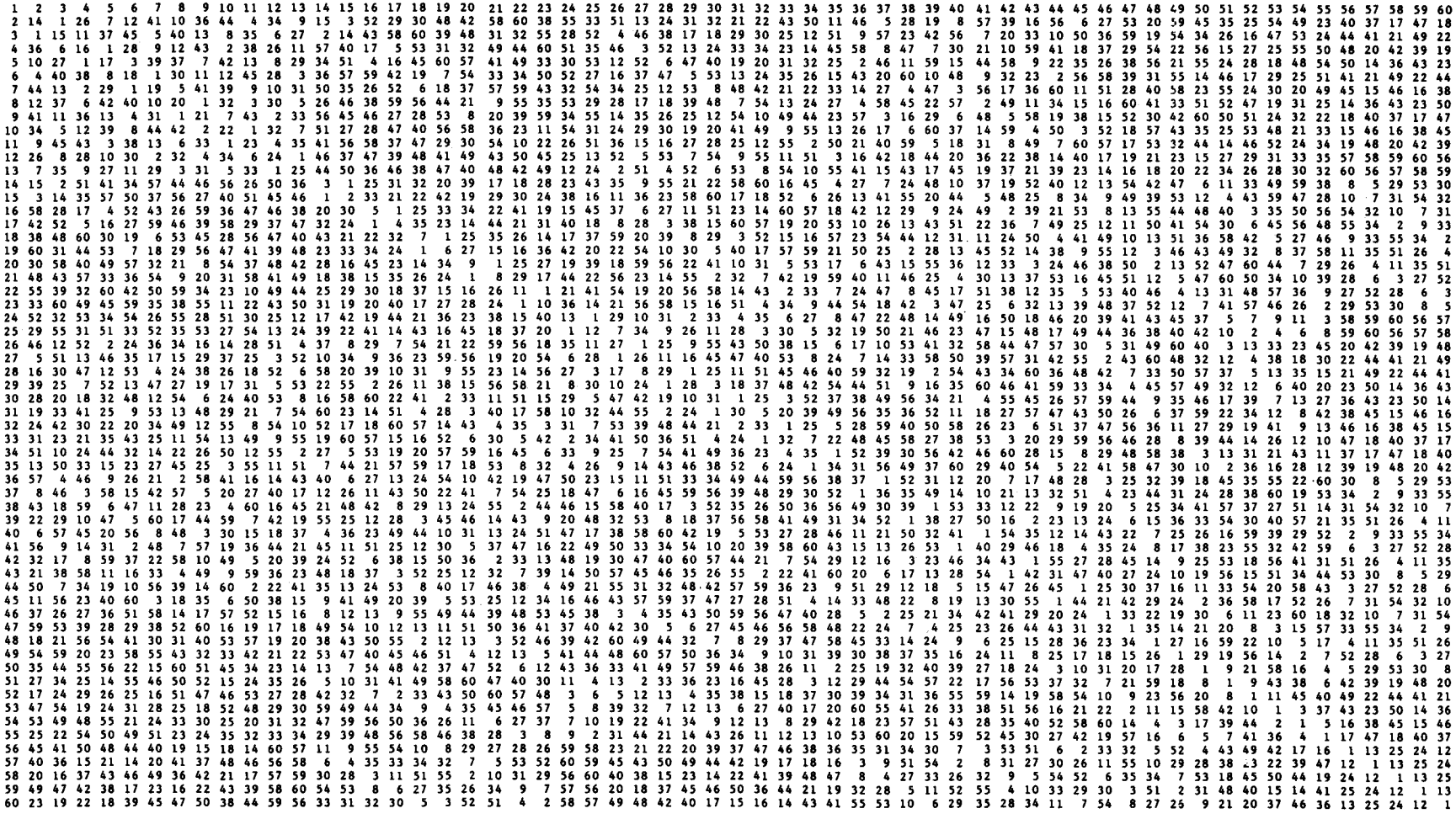
$$\langle 3T_2,1T_1\|4G\rangle = -1/2, \quad \langle 3T_2,1T_1\|3G\rangle = -\sqrt{3}/2, \quad \langle 3T_2,1T_2\|4H\rangle = \sqrt{4/7}, \quad \langle 3T_2,1T_1\|2H\rangle = -\sqrt{3/7}. \tag{28}$$

The phases for the subduction coefficients $C(jm,\nu\mu)$ in (27) are more elegant than those of Refs. 4 and 6 in the sense that the coefficients $C(jm=\text{even},\nu\mu)$ are all real, while $C(jm=\text{odd},\nu\mu)$ are all imaginary.

In summary, with the isomorphism shown in the first table in the Appendix, the program found the CSCO-I, CSCO-II, and CSCO-III of the icosahedral group **I** as well as their eigenvectors that give the primitive characters, the CG coefficients, and all the irreducible matrix elements of the group **I**. It once again shows the power of the eigenfunction method. With the new phase convention for the time reversal state, the CG coefficients of the group **I** can be made to be real even though the bases are still complex as shown in (27). The introducing of the extra phase factor $\theta_{\nu_1+\nu_2+\nu_3}$ in the symmetry (13a) is the only price one has to pay for real CG coefficients.

APPENDIX: THE ISOMORPHISM BETWEEN I AND A SUBGROUP OF S_{12} AND THE GROUP TABLE OF I

R_1	E	E	R_1	E	E
$R_2=R_3^{-1}$	$C_{5.1}^1$	(265,11,7)(34,10,12,8)	$R_{40}=R_{41}^{-1}$	$C_{3.8}^1$	(17,11)(2,12,5)(394)(68,10)
$R_4=R_5^{-1}$	$C_{5.2}^1$	(17836)(45,11,12,9)	$R_{42}=R_{43}^{-1}$	$C_{3.9}^1$	(13,12)(287)(4,10,5)(69,11)
$R_6=R_7^{-1}$	$C_{5.3}^1$	(17,12,10,5)(28946)	$R_{44}=R_{45}^{-1}$	$C_{3.10}^1$	(15,11)(24,12)(398)(6,10,7)
$R_8=R_9^{-1}$	$C_{5.4}^1$	(128,12,11)(39,10,56)	R_{46}	$C_{2.1}$	(12)(3,11)(4,12)(58)(67)(9,10)
$R_{10}=R_{11}^{-1}$	$C_{5.5}^1$	(164,10,11)(239,12,7)	R_{47}	$C_{2.2}$	(19)(23)(47)(5,12)(68)(10,11)
$R_{12}=R_{13}^{-1}$	$C_{5.6}^1$	(12345)(789,10,11)	R_{48}	$C_{2.3}$	(1,12)(2,10)(34)(58)(69)(7,11)
$R_{14}=R_{15}^{-1}$	$C_{5.1}^2$	(2576,11)(3,10,84,12)	R_{49}	$C_{2.4}$	(19)(2,12)(3,11)(45)(6,10)(78)
$R_{16}=R_{17}^{-1}$	$C_{5.2}^2$	(18673)(4,11,95,12)	R_{50}	$C_{2.5}$	(15)(2,10)(3,12)(47)(6,11)(89)
$R_{18}=R_{19}^{-1}$	$C_{5.3}^2$	(1,12,57,10)(29684)	R_{51}	$C_{2.6}$	(16)(25)(3,11)(47)(8,10)(9,12)
$R_{20}=R_{21}^{-1}$	$C_{5.4}^2$	(18,11,2,12)(3,10,695)	R_{52}	$C_{2.7}$	(13)(26)(47)(58)(9,11)(10,12)
$R_{22}=R_{23}^{-1}$	$C_{5.5}^2$	(14,11,6,10)(2973,12)	R_{53}	$C_{2.8}$	(19)(24)(36)(58)(7,10)(11,12)
$R_{24}=R_{25}^{-1}$	$C_{5.6}^2$	(13524)(79,11,8,10)	R_{54}	$C_{2.9}$	(19)(2,10)(35)(46)(7,12)(8,11)
$R_{26}=R_{27}^{-1}$	$C_{3.1}^1$	(126)(357)(4,11,8)(9,10,12)	R_{55}	$C_{2.10}$	(14)(2,10)(3,11)(56)(70)(8,12)
$R_{28}=R_{29}^{-1}$	$C_{3.2}^1$	(184)(236)(579)(10,11,12)	R_{56}	$C_{2.11}$	(1,11)(2,10)(39)(48)(57)(6,12)
$R_{30}=R_{31}^{-1}$	$C_{3.3}^1$	(18,10)(295)(346)(7,12,11)	R_{57}	$C_{2.12}$	(17)(2,11)(3,10)(49)(58)(6,12)
$R_{32}=R_{33}^{-1}$	$C_{3.4}^1$	(13,10)(29,11)(456)(78,12)	R_{58}	$C_{2.13}$	(18)(27)(3,11)(4,10)(59)(6,12)
$R_{34}=R_{35}^{-1}$	$C_{3.5}^1$	(165)(24,11)(3,10,7)(89,12)	R_{59}	$C_{2.14}$	(19)(28)(37)(4,11)(5,10)(6,12)
$R_{36}=R_{37}^{-1}$	$C_{3.6}^1$	(172)(35,12)(4,10,9)(6,11,8)	R_{60}	$C_{2.15}$	(1,10)(29)(38)(47)(5,11)(6,12)
$R_{38}=R_{39}^{-1}$	$C_{3.7}^1$	(1,12,4)(283)(5,11,10)(679)			



[1]A. Speiser, *Die Theorie der Gruppen von Endlicher Ordnung* (Dover, New York, 1937).
[2]O. Laporte, Z. Naturforsch. **3a**, 447 (1948).
[3]N. V. Cohan, Proc. Cambridge Philos. Soc. Math. Phys. Sci. **54**, 28 (1958).
[4]A. G. McLellian, J. Chem. Phys. **34**, 1350 (1961).
[5]R. M. Golding, Molecular Phys. **26**, 661 (1973).
[6]D. R. Pooler, J. Math. Phys. Gen. **13**, 1197 (1988).
[7]D. Levine and P. Steinhardt, Phys. Rev. Lett. **53**, 2477 (1984).
[8]J. Q. Chen, *Group Representation Theory for Physicists* (World Scientific, Singapore, 1989); *A New Approach to Group Representation Theory* (Science and Technology, Shanghai, 1984).
[9]J. Q. Chen, M. J. Gao, and G. Q. Ma, Rev. Mod. Phys. **57**, 211 (1985).
[10]J. L. Ping, Q. R. Zheng, B. Q. Chen, and J. Q. Chen, Comp. Phys. Commun. (1989).
[11]J. Q. Chen, F. Wang, and M. J. Gao, J. Phys. A **16**, 1347 (1983).

Decomposition of the enveloping algebra of $\mathfrak{sl}_3$

Daniel E. Flath
Department of Mathematics, University of South Alabama, Mobile, Alabama 36688

(Received 15 August 1989; accepted for publication 20 December 1989)

The adjoint representation of $\mathfrak{sl}_3$ on its universal enveloping algebra $\mathcal{U}$ is explicitly decomposed. The result is expressed as a presentation by generators and relations of the commutant in $\mathcal{U}$ of the raising operators in $\mathfrak{sl}_3$. Application is made to the analysis of the representations End(W) for finite-dimensional simple W.

I. INTRODUCTION

Applications of representation theory often rest on rather detailed knowledge of the decomposition of tensor product representations. A special but important case is the decomposition of the representations $\mathrm{End}_{\mathbb{C}}(W)$ for irreducible W. When W is a finite-dimensional irreducible representation of a simple Lie algebra $\mathfrak{g}$, $\mathrm{End}_{\mathbb{C}}(W)$ is a quotient of the universal enveloping algebra $\mathcal{U}$ of $\mathfrak{g}$, and so a good first step is the explicit decomposition of the adjoint representation of $\mathfrak{g}$ on $\mathcal{U}$.

For $\mathfrak{g}=\mathfrak{sl}_2$, this program is very easily carried out, leading to the following elegant, classical statement. Let (τ,W) be a finite-dimensional irreducible representation of $\mathfrak{sl}_2$. The vectors of highest weight in $\mathrm{End}_{\mathbb{C}}(W)$ [that is, commuting with $\tau(E_{12})$] are precisely the nonzero vectors of the form $\tau(E_{12})^n$, $n=0,1,2,\ldots$.

In this paper we consider the next case, that of $\mathfrak{sl}_3$. Our main result, a completely explicit decomposition of the adjoint representation of $\mathfrak{sl}_3$ on its universal enveloping algebra, is formulated in Sec. II and proved in Sec. III. As applications, we deduce some corollaries in Sec. IV, including the analog of the $\mathfrak{sl}_2$ statement mentioned above. Corollaries 3 and 4 are special cases of a very general theorem of Kostant.[1]

For a related approach to the representations $\mathrm{Hom}_{\mathbb{C}}(V,W)$ of $\mathfrak{sl}_3$, see Ref. 2. The simplicity of the relations in Corollary 6.7 of Ref. 2 may be contrasted with the complexity of the relations (8)–(12) in the present paper. This is perhaps an indication that for some problems the universal enveloping algebra is not the most appropriate or convenient tool available.

II. THE DECOMPOSITION OF THE ENVELOPING ALGEBRA

Let $\mathfrak{g}=\mathfrak{sl}_3$ denote the Lie algebra of 3×3 traceless complex matrices. Let $\mathcal{U}$ be the universal enveloping algebra of $\mathfrak{g}$. We view $\mathcal{U}$ as the space of a $\mathfrak{g}$ representation ρ via the adjoint action:

$$\rho(x)a=[x,a], \quad \text{for } x\in\mathfrak{g}, \quad a\in\mathcal{U}.$$

In order to decompose the representation ρ it is enough to determine the space $\mathcal{B}$ of highest weight vectors in $\mathcal{U}$. Because $\mathcal{B}$ is the set of $a\in\mathcal{U}$ such that $\rho(E_{12})a=\rho(E_{23})a=0$, the space $\mathcal{B}$ is a subalgebra of $\mathcal{U}$, the commutant of $\{E_{12},E_{23}\}$. (By E_{ij} we mean the 3×3 matrix whose only nonzero entry is the ijth, which equals 1.)

We introduce a notation H_1,H_2 for two elements of $\mathfrak{g}$:

$$H_1=E_{11}-E_{22}, \quad H_2=E_{22}-E_{33}. \tag{1}$$

An element $A\in\mathcal{U}$ is of *weight* (p,q) if $[H_1,A]=pA$ and $[H_2,A]=qA$. In this paper all weights (p,q) are *integral* in the sense that $p,q\in\mathbb{Z}$.

We next list six elements (2)–(7) of $\mathcal{U}$ which lie in $\mathcal{B}$:

$$B_1=E_{13}, \tag{2}$$

$$B_2=3E_{12}E_{23}+(H_1-H_2)E_{13}, \tag{3}$$

$$I_2=\tfrac{1}{3}(H_1^2+H_1H_2+H_2^2)+H_1+H_2+E_{21}E_{12}+E_{31}E_{13}+E_{32}E_{23}, \tag{4}$$

$$\begin{aligned}I_3=&\tfrac{2}{27}H_1^3+\tfrac{1}{9}H_1^2H_2-\tfrac{1}{9}H_1H_2^2-\tfrac{2}{27}H_2^3+E_{32}E_{21}E_{13}\\&+E_{31}E_{12}E_{23}+\tfrac{1}{3}E_{31}(H_1-H_2)E_{13}+\tfrac{1}{3}E_{21}\\&\times(H_1+2H_2)E_{12}-\tfrac{1}{3}E_{32}(2H_1+H_2)E_{23}+E_{21}E_{12}\\&-E_{31}E_{13}-E_{32}E_{23}+\tfrac{1}{3}(H_1^2-H_2^2)+\tfrac{1}{3}(H_1-H_2),\end{aligned} \tag{5}$$

$$C_1=E_{32}E_{13}^2-E_{23}E_{12}^2+H_2E_{12}E_{13}, \tag{6}$$

$$C_2=-E_{21}E_{13}^2+E_{12}E_{23}^2+H_1E_{23}E_{13}. \tag{7}$$

The elements B_1 and B_2 are of weight (1,1), I_2 and I_3 are $\mathfrak{g}$ invariants of weight (0,0), C_1 is of weight (3,0), and C_2 is of weight (0,3).

We give 16 relations (8)–(12) in $\mathcal{B}$ satisfied by the above six elements:

$$[B_2,C_1]=3B_1C_1, \tag{8}$$

$$[B_2,C_2]=-3B_1C_2, \tag{9}$$

$$[C_1,C_2]=-B_1^3I_2-B_1^2B_2+\tfrac{1}{3}B_1B_2^2. \tag{10}$$

The 12 remaining brackets between the six elements (2)–(7) are zero, (11)

$$C_1C_2=-B_1^3(I_2+I_3)+\tfrac{1}{3}B_1^2B_2(I_2-2)+\tfrac{1}{3}B_1B_2^2-\tfrac{1}{27}B_2^3. \tag{12}$$

Theorem: (i) The commutant of $\{E_{12},E_{23}\}$ in $\mathcal{U}$, $\mathcal{B}$, is generated by the six elements (2)–(7). (ii) The ideal of relations in $\mathcal{B}$ among the elements (2)–(7) is generated by the 16 relations (8)–(12).

III. THE PROOF

We will begin the proof by establishing the sufficiency of the 16 relations.

Lemma 1: The four commuting elements B_1,B_2,I_2,I_3 are algebraically independent.

Proof: We will denote by $\bar{A}$ the element of the symmetric algebra $\mathrm{Sym}(\mathfrak{g})$ on $\mathfrak{g}$ that is the image of $A\in\mathcal{U}$ under the

 0022-2488/90/051076-02$03.00

canonical linear isomorphism from $\mathcal{U}$ to $\mathrm{Sym}(\mathcal{g})$ whose inverse is symmetrization [from the Poincaré–Birkhoff–Witt (PBW) theorem].

One actually shows that the four elements $\overline{B}_1,\overline{B}_2,\overline{I}_2,\overline{I}_3$ are algebraically independent.

By symmetrizing, one verifies the following which is in fact how I_2 and I_3 were found:

$$\overline{B}_1 = E_{13}, \tag{13}$$

$$\overline{B}_2 = 3E_{12}E_{23} + (H_1 - H_2)E_{13} + \tfrac{3}{2}E_{13}, \tag{14}$$

$$\overline{I}_2 = \tfrac{1}{3}(H_1^2 + H_1H_2 + H_2^2) + E_{21}E_{12} + E_{31}E_{13} + E_{32}E_{23}, \tag{15}$$

$$\overline{I}_3 = \tfrac{2}{27}H_1^3 + \tfrac{1}{9}H_1^2H_2 - \tfrac{1}{9}H_1H_2^2 - \tfrac{2}{27}H_2^3 + E_{32}E_{21}E_{13} + E_{31}E_{12}E_{23} + \tfrac{1}{3}E_{31}(H_1 - H_2)E_{13} + \tfrac{1}{3}E_{21}(H_1 + 2H_2)E_{12} - \tfrac{1}{3}E_{32}(2H_1 + H_2)E_{23}. \tag{16}$$

It is now an easy exercise to show that $\overline{B}_1,\overline{B}_2,\overline{I}_2,\overline{I}_3$ are algebraically independent even after specializing H_1, H_2, and E_{21} to zero. ■

Let $\mathcal{C}$ be the algebra generated by elements (2)–(7). We want to show of course that $\mathcal{C} = \mathcal{B}$.

By using the given relations one sees that C is spanned as a vector space by elements of the form:

$$B_1^{\,a}B_2^{\,b}I_2^{\,c}I_3^{\,d}C_1^{\,e}C_2^{\,f}, \quad \text{with } ef = 0. \tag{17}$$

Lemma 2: The elements (17) are linearly independent.

Proof: The element (17) has weight $(a + b + 3e, a + b + 3f)$ and so a dependence must express itself among elements with *fixed* e and f. But since $\mathcal{U}$ has no zero divisors, Lemma 1 makes such a dependence impossible. ■

The proof that $\mathcal{C} = \mathcal{B}$, or equivalently, that the $\mathcal{g}$ subrepresentation of $\mathcal{U}$ generated by $\mathcal{B}$ is all of $\mathcal{U}$, is based on dimension considerations.

We recall the canonical filtration on $\mathcal{U}$. The space $\mathcal{U}_n$ is the subspace of $\mathcal{U}$ spanned by elements that are products of at most n elements of $\mathcal{g}$. By the PBW theorem,

$$\dim\left(\frac{\mathcal{U}_n}{\mathcal{U}_{n-1}}\right) = \dim \mathrm{Sym}^n(\mathcal{g}) = \binom{n+7}{7}. \tag{18}$$

Let $\dim(a,b,c,d,e,f)$ be the dimension of the irreducible $\mathcal{g}$ representation generated by the element (17). By the Weyl dimension formula, for an element of weight (p,q), $\dim = \frac{1}{2}(p+1)(q+1)(p+q+2)$.

The proof of the theorem is now completed straightforwardly by induction, showing that $\Sigma \dim = \binom{n+7}{7}$, where the sum is over a,b,c,d,e,f such that $a + 2(b + c) + 3(d + e + f) = n$ and $ef = 0$. One must consider each of the six residue classes of n mod six separately. ■

IV. CONSEQUENCES

Let X be the set of elements of the form (17) with $c = d = 0$.

Corollary 1: The algebra B is a free $\mathbb{C}[I_2,I_3]$ module for which X is a basis.

The elements I_2 and I_3 act as scalar multiplications on each irreducible $\mathcal{g}$ module. We deduce the following corollary:

Corollary 2: Let (τ, W) be a finite-dimensional irreducible $\mathcal{sl}_3$ representation. The set $\{\tau(x) | x \in X\}$ spans the space of highest weight vectors in the $\mathcal{sl}_3$ representation $\mathrm{End}_{\mathbb{C}}(W)$.

Define $\mathcal{B}(p,q)$ to be the subspace of elements of $\mathcal{B}$ of $\mathcal{sl}_3$ weight (p,q).

Let $V(p,q)$ be a finite-dimensional irreducible $\mathcal{sl}_3$ representation of highest weight (p,q). It is defined only when (p,q) is dominant, that is, when $p,q \geqslant 0$.

The elements (17) are all of dominant weight. For $\mathcal{B}(p,q)$ to be nonzero, we must have $p \equiv q \pmod 3$, which is the condition that (0,0) be a weight of $V(p,q)$. In that case, the number of elements of X of weight (p,q) is easily seen to be $\min\{p+1, q+1\}$, which by Ref. 2 equals the multiplicity of the weight (0,0) in $V(p,q)$.

Corollary 3: (i) $\mathcal{B}(0,0) = \mathbb{C}[I_2,I_3]$. (ii) $\mathcal{B}(p,q)$ is a free $\mathbb{C}[I_2,I_3]$ module of rank equal to the multiplicity of the weight (0,0) in the $\mathcal{sl}_3$ module $V(p,q)$.

Corollary 4: Let W be a finite-dimensional, irreducible $\mathcal{sl}_3$ representation. Then the multiplicity of $V(p,q)$ in $\mathrm{End}_{\mathbb{C}}(W)$ is bounded (independently of W) by the multiplicity of (0,0) in $V(p,q)$.

It is in fact known[2] that if $p + q - 1 \leqslant \min\{r,s\}$, then the multiplicity of $V(p,q)$ in $\mathrm{End}_{\mathbb{C}}(V(r,s))$ exactly equals the multiplicity of (0,0) in $V(p,q)$.

Corollary 5: If $p + q - 1 \leqslant \min\{r,s\}$, then the set $\{\tau(x) | x \in X, \text{ weight } x = (p,q)\}$ is a basis for the space of highest weight vectors of weight (p,q) in $\mathrm{End}_{\mathbb{C}}(V(r,s))$.

We will not comment here on the linear dependences that arise among the $\tau(x)$ when (r,s) does not satisfy the hypothesis of Corollary 5.

[1] B. Kostant, Am. J. Math. **85**, 327 (1963).
[2] D. E. Flath and L. C. Biedenharn, Can. J. Math. **37**, 710 (1985).

Highest weight irreducible unitarizable representations of Lie algebras of infinite matrices. The algebra A_∞

Tchavdar D. Palev[a)]
Arnold Sommerfeld Institute for Mathematical Physics, Technical University of Clausthal, D-3392 Clausthal-Zellerfeld, West Germany

(Received 2 May 1989; accepted for publication 29 November 1989)

In this paper all unitarizable irreducible highest weight representations of the infinite-dimensional Lie algebra A_∞, which is a completion and central extension of the general linear Lie algebra gl_∞, are considered. Within each representation space a basis is introduced and expressions for the transformations of the basis under the action of the Chevalley generators are written.

I. INTRODUCTION

Since it was defined in 1981,[1,2] the algebra A_∞ and, in particular, its highest weight irreducible unitarizable representations, turned out to be a useful tool in several areas of theoretical physics. In this respect we mention the results of the Kyoto school on the theory of integrable field equations and solitons (see Refs. 1, 3–6) and the interpretation of the Dirac positron theory in terms of the infinite wedge space representations of A_∞.[7,8] In a more indirect way the algebra A_∞ has applications in the theory of two-dimensional statistical systems, quantum string theory, etc. (see the review article[9] and the references therein): This is due to the circumstance that A_∞ contains as subalgebras the infinite-dimensional Heisenberg algebra, the Virasoro algebra, and other Kac–Moody algebras (for a systematic exposition of the relations of A_∞ to these algebras and their applications see Ref. 8).

The criterion for selecting highest weight representations usually comes from natural physical requirements. In the known interesting examples, the algebras under consideration contain observables whose spectrums should be bounded from below for some physical reasons: In two-dimensional field theories this is the energy operator, in string theory this is the mass spectrum operator, etc.[9] In most cases the representation space plays the role of a state space; therefore, it has to be accomplished with a scalar product, so that the observables are self-adjoint operators. This leads to the requirement that the representation be unitarizable.

In order to be more concrete let us give an example. Take a free spinor field locked (for simplicity) in a finite volume. The field is characterized by the operators $f^\nu_{\eta A}$, corresponding to the creation ($\nu = +$) or annihilation ($\nu = -$) of a particle with a charge $\eta = \pm 1$ and all other (discrete in this case) indices $A \in \mathbb{N}$ (see the notation at the end of this section) and which satisfy the Fermi anticommutation relations

$$\{f^\nu_i f^\epsilon_j\} = \tfrac{1}{4}(\nu - \epsilon)^2 \delta_{ij}, \quad \nu,\epsilon = \pm \text{ or } \pm 1, \quad i,j \in \mathbb{Z}\setminus 0 (\mathbb{Z}\setminus 0 \equiv \pm 1, \pm 2, \ldots). \tag{1.1}$$

The state space is the usual Fock space Φ_∞ (1), characterized with a vacuum $|0\rangle$:

$$f^-_i |0\rangle = 0, \quad \forall i \in \mathbb{Z}\setminus 0. \tag{1.2}$$

It is known[10] that any n pairs of operators $f^\pm_{i_1}, \ldots, f^\pm_{i_n}$ generate the simple Lie algebra $\mathrm{so}(2n+1) \equiv B_n$ of the rotation group. If $i \in \mathbb{Z}\setminus 0$, then $f^\pm_i$ generate an infinite-dimensional Lie algebra B_∞:

$$B_\infty = \text{lin.env.}\{f^\nu_i f^\nu_j f^\epsilon_j - f^\epsilon_j f^\nu_i \equiv [f^\nu_i f^\epsilon_j] \,|\, i,j \in \mathbb{Z}\setminus 0, \nu,\epsilon = \pm\}. \tag{1.3}$$

As an ordered basis in the Cartan subalgebra H of B_∞ one can take all generators[11]

$$h_i = \tfrac{1}{2}[f^-_i f^+_i], \quad i \in \mathbb{Z}. \tag{1.4}$$

Then

$$N^+ = \text{lin.env.}\{[f^-_i f^\pm_j] f^-_k \,|\, i,j,k \in \mathbb{Z}\setminus 0, \quad i<j\} \tag{1.5}$$

is a Borel subalgebra in B_∞ and Eqs. (1.1) and (1.2) yield

$$h_i|0\rangle = \tfrac{1}{2}|0\rangle, \quad N^+|0\rangle = 0, \quad i \in \mathbb{Z}\setminus 0. \tag{1.6}$$

Hence, the vacuum is a highest weight vector in Φ_∞ (1), with a weight $(\ldots, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, \ldots)$. Moreover, the usual scalar product in Φ_∞ (1) is contravariant with respect to the antilinear anti-involution ω: $B_\infty \to B_\infty$ defined by $\omega(f^+_i) = f^-_i$ and, therefore, the representation of the infinite-dimensional simple Lie algebra B_∞, realized in the ordinary Fermi–Fock space, is a unitarizable highest weight irreducible representation. The relation of the Fermi field to B_∞ has far-reaching consequences: It leads to generalized statistics, namely the para-Fermi[12] statistics, with state spaces, that are irreducible B_∞ modules. Also, in this case, imposing requirements that the vacuum be unique and nondegenerate and the Hermiticity condition $\omega(f^+_i) = f^-_i$ to hold, one comes to the conclusion[13] that the Fock spaces $\Phi(p)$, $p \in \mathbb{N}$ have to carry highest weight unitarizable representations of B_∞ corresponding to one of the highest weights $(\ldots, p/2, \ldots, p/2, p/2, p/2, \ldots, p/2, \ldots)$, $p \in \mathbb{N}$. For physical reasons (for instance, to have the four-momentum well defined in the Fock space) one is forced to enlarge B_∞ to its completion $\overline{B}_\infty$. However, since $\overline{B}_\infty$ has no representations in $\Phi(p)$ (see the beginning of Sec. III B for an idea why this happens) one has to go further to the central extension of $\overline{B}_\infty$ (which in the Fermi case is equivalent to the normal ordering postulate).

Our motivation for studying the highest weight unitarizable representations of A_∞ originates from an attempt to introduce new quantum statistics. Our interest is based on the observation that the statistics of the spinor fields admits a logically self-consistent generalization such that the creation

[a)] On leave of absence from the Institute for Nuclear Research and Nuclear Energy, blvd. Lenin 72, 1184 Sofia, Bulgaria.

 0022-2488/90/051078-07$03.00

and annihilation operators of the field generate the algebra gl_∞ (instead of B_∞).[14] Exactly as in the example considered above, one shows that the gl_∞ representations, which are physically admissible, should be necessarily highest weight unitarizable representations and that rigorous results along this line, free of divergences, can be obtained on the grounds of the completion and central extension A_∞ of gl_∞ [see Sec. III B].[15]

The highest weight unitarizable irreducible modules of A_∞ have been classified by Kac and Peterson.[7] Any highest weight irreducible module $V([M])$ is characterized with its signature [= the coordinates of the highest weight]

$$[M] \equiv [\ldots,M_p,\ldots,M_{-1},M_0,M_1,\ldots,M_q,\ldots] \equiv \{M_i | M_i \in \mathbb{C},\ i \in \mathbb{Z}\}. \tag{1.7}$$

The module $V([M])$ carries a unitarizable representation of A_∞ if and only if (see the notation at the end of this section)

$$M_i - M_j \in \mathbb{Z}_+,\ \forall i<j \in \mathbb{Z}; \tag{1.8a}$$

there exists $p<q \in \mathbb{Z}$, such that $M_{p-k} = M_p$,

$$M_{q+k} = M_q,\quad \forall k \in \mathbb{N}; \tag{1.8b}$$

all coordinates M_i are real numbers $M_i \in \mathbb{R}$. (1.8c)

Thus far explicit expressions for the transformation of the basis under the action of the generators of A_∞ are available for the fundamental representations,[2] i.e., those corresponding to weights (1.7) such that $M_i = 1$ for $i \leqslant m$ and $M_i = 0$ for $i > m$, $m \in \mathbb{Z}$. In the present paper we introduce a basis and write expressions for its transformations for every module $V([M])$ whose signature $[M]$ satisfies conditions (1.8a)–(1.8c).

Recently,[16] we have studied the class of highest weight irreducible modules of gl_∞ with signatures obeying condition (1.8a). Here we shall use essentially those results. Therefore, in Sec. II we give a résumé of the necessary results from Ref. 16. In Sec. III A (as a convenient intermediate step) we construct all unitarizable irreducible highest weight representations of the Lie algebra a_∞. The latter is a central extension of gl_∞. In Sec. III B we construct all unitarizable irreducible highest weight modules of A_∞.

Throughout the paper we use the following abbreviations and notation:

LA, LA's—Lie algebra, Lie algebras,

lin. env. $\{X\}$—the linear envelope of X,

$\mathbb{C}$—the complex numbers,

$\mathbb{R}$—the real numbers,

$\mathbb{Z}$—all integers,

$\mathbb{Z}_+$—all non-negative integers,

$\mathbb{N}$—all positive integers,

$$[M] = \{M_i | i \in \mathbb{Z},\ M_i \in \mathbb{C}\}, \tag{1.9}$$

$$L_{ij} = M_{ij} - i, \tag{1.10}$$

$$(x,y)_0 = x,y,\quad (x,y)_1 = y,x, \tag{1.11}$$

If $p \leqslant q \in \mathbb{Z}$, then

$$[p,q] = \{k \,|\, p \leqslant k \leqslant q,\ k \in \mathbb{Z}\}, \tag{1.12}$$

$$\theta(x) = \begin{cases} 1, & \text{for } x \geqslant 0, \\ 0, & \text{for } x < 0, \end{cases} \tag{1.13}$$

and

$$S(i,j) = \begin{cases} 1, & \text{for } i \leqslant j, \\ -1, & \text{for } i > j. \end{cases} \tag{1.14}$$

II. HIGHEST WEIGHT IRREDUCIBLE gl_∞ MODULES

The infinite-dimensional general Lie algebra gl_∞ can be defined as the set of all squared infinite matrices such that each matrix has only a finite number of nonzero entries, i.e.,

$$gl_\infty = \{(A_{ij}) | i,j \in \mathbb{Z}, \text{ all but a finite number of } A_{ij} \in \mathbb{C} \text{ are zero}\}. \tag{2.1}$$

The Lie bracket on gl_∞ is the ordinary matrix commutator. A convenient basis in gl_∞ is given with the set of all Weyl matrices E_{ij}, $i,j \in \mathbb{Z}$, where E_{ij} is a matrix with 1 on the ith row, the jth column, and zero elsewhere. The commutation relations on gl_∞ read as

$$[E_{ij},E_{kl}] = \delta_{jk}E_{il} - \delta_{li}E_{kj},\quad i,j,k,l \in \mathbb{Z}. \tag{2.2}$$

In Ref. 16, we have constructed a class of highest weight irreducible gl_∞ modules $\mathfrak{F}$ corresponding to the "Borel" subalgebra

$$N_+ = \text{lin. env. } \{E_{ij} | i<j \in \mathbb{Z}\}. \tag{2.3}$$

By definition this means[8] that each gl_∞ module $V \in \mathfrak{F}$ contains a unique (up to a multiplicative constant) vector x_Λ, the highest weight vector, with the properties

$$E_{ij}x_\Lambda = 0,\quad \forall i<j \in \mathbb{Z}, \tag{2.4}$$

$$E_{ii}x_\Lambda = M_i x_\Lambda,\quad \forall i \in \mathbb{Z}. \tag{2.5}$$

Any two modules V_1 and V_2 from $\mathfrak{F}$ and corresponding to different signatures $[M^1] \neq [M^2]$ carry nonequivalent representations. The signature $[M]$ of $V \in \mathfrak{F}$ uniquely labels the module. Therefore, we set $V \equiv V([M])$. The possibility that there might exist a nonequivalent module to the $V([M])$ irreducible highest weight gl_∞ module with the same signature is not excluded. However, if such a module exists, it cannot be from the class $\mathfrak{F}$.

We now proceed to describe more precisely the modules from $\mathfrak{F}$.

Definition 1: Let

$$[M] \equiv [\ldots,M_p,\ldots,M_{-1},M_0,M_1,\ldots,M_q,\ldots] \equiv \{M_i | M_i \in \mathbb{C},\quad i \in \mathbb{Z}\} \tag{2.6}$$

be a sequence of complex numbers such that

$$M_i - M_j \in \mathbb{Z}_+,\quad \forall i<j \in \mathbb{Z}. \tag{2.7}$$

A pattern (M) consisting of all complex numbers

$$M_{i,2k+\theta-1},\quad \forall k \in \mathbb{N},\quad \theta = 0,1,\quad i \in [1-\theta-k,k-1], \tag{2.8}$$

which satisfy the following conditions:

there exists $N([(M)] \in \mathbb{N}$ such that

$$M_{i,2k+\theta-1} = M_i,\quad \forall 2k+\theta-1 \geqslant N([(M)],$$
$$\theta = 0,1,\quad i \in [1-\theta-k,k-1], \tag{2.9}$$
$$M_{i+\theta-1,2k+\theta} - M_{i,2k+\theta-1} \in \mathbb{Z}_+,$$

$$M_{i,2k+\theta-1} - M_{i+\theta,2k+\theta} \in \mathbb{Z}_+, \quad \forall k \in \mathbb{N},$$
$$\theta = 0,1, \quad i \in [1-\theta-k, k-1], \tag{2.10}$$

is called a C pattern (corresponding to $[M]$).

The entries $M_{i,2k+\theta-1}$ are said to be coordinates of the C pattern: It is convenient to order them as indicated in the pattern below, writing as a first row the sequence $[M]$:

$$(M) = \begin{bmatrix} \ldots, & M_{1-\theta-k}, & \ldots, & M_{-1}, & M_0, & M_1, & \ldots, & M_{k+\theta-1}, & \ldots \\ & \vdots & & \vdots & \vdots & \vdots & & \vdots & \\ \ldots, & M_{1-\theta-k,2k+\theta-1}, & \ldots, & M_{-1,2k+\theta-1}, & M_{0,2k+\theta-1}, & M_{1,2k+\theta-1}, & \ldots, & M_{k+\theta-1,2k+\theta-1} & \\ & & & \vdots & \vdots & \vdots & & & \\ & & & M_{-1,3}, & M_{03}, & M_{13} & & & \\ & & & M_{-1,2}, & M_{02} & & & & \\ & & & & M_{01} & & & & \end{bmatrix}, \tag{2.11}$$

where $k \in \mathbb{N}$, $\theta = 0,1$.

Proposition 1: To each sequence (2.6) there corresponds an irreducible highest weight gl_∞ module $V([M])$ with a signature $[M]$. The basis $\Gamma([M])$ in $V([M])$ consists of all C patterns corresponding to $[M]$.

The transformations of the C basis $\Gamma([M])$ of $V([M]) \in \mathfrak{F}$ under the action of the generators $E_{k,k}$, $\forall k \in \mathbb{Z}$ and $E_{-k,k-\nu}$, $E_{k-\nu,-k}$, $\forall \nu = 0,1$, $k \in \mathbb{N}$ were given in Ref. 16 [see Ref. 16, Eqs. (3.43) and (3.44)]. Using these generators and the commutation relations (2.2) one derives the transformations of the basis under the action of the Chevalley generators $E_{k,k-1}$, $E_{k-1,k}$, $k \in \mathbb{Z}$. To this end denote by $(M)_{\pm[j,p]}$ and $(M)^{\pm[l,q]}_{\pm[j,p]}$, $p \neq q$ the C patterns obtained from the C pattern (M) by the replacements

$$M_{lq} \to M_{lq} \pm 1, \quad M_{jp} \to M_{jp} \pm 1 \tag{2.12}$$

and let

$$S(j,l;\nu) = \begin{cases} (-1)^\nu, & \text{for} \quad j = l, \\ 1, & \text{for} \quad j < l, \\ -1, & \text{for} \quad j > l. \end{cases} \tag{2.13}$$

Then one has [see (1.11)]

$$E_{(-1,0)_\mu}(M) = |(L_{-1,2} - L_{0,1} - \mu)(L_{0,2} - L_{0,1} - \mu)|^{1/2} (M)_{-(-1)^\mu[0,1]}, \tag{2.14}$$

$$E_{(-1)^\nu k-\mu,(-1)^\nu k+\mu-1}(M) = -\sum_{j=1-k-\nu}^{k-1} \sum_{l=-k}^{k+\nu-1} S(j,l;\nu)$$
$$\times \left| \frac{\Pi_{i\neq l=-k}^{k+\nu-1}(L_{i,2k+\nu} - L_{j,2k+\nu-1} - (-1)^\nu\mu)\Pi_{i=1-k}^{k+\nu-2}(L_{i,2k+\nu-2} - L_{j,2k+\nu-1} - (-1)^\nu\mu)}{\Pi_{i\neq j=1-k-\nu}^{k-1}(L_{i,2k+\nu-1} - L_{j,2k+\nu-1})(L_{i,2k+\nu-1} - L_{j,2k+\nu-1} + (-1)^{\mu+\nu})} \right|^{1/2}$$
$$\times \left| \frac{\Pi_{i=-k-\nu}^{k}(L_{i,2k+\nu+1} - L_{l,2k+\nu} + (-1)^\nu(1-\mu))\Pi_{i\neq j=1-k-\nu}^{k-1}(L_{i,2k+\nu-1} - L_{l,2k+n} + (-1)^\nu(1-\mu))}{\Pi_{i\neq l=-k}^{k+\nu-1}(L_{i,2k+\nu} - L_{l,2k+\nu})(L_{i,2k+\nu} - L_{l,2k+\nu} + (-1)^{\mu+\nu})} \right|^{1/2}$$
$$\times (M)^{-(-1)^{\mu+\nu}[l,2k]}_{-(-1)^{\mu+\nu}[j,2k-1]}, \quad \text{where } k \in \mathbb{N}, \ \mu,\nu = 0,1, \tag{2.15}$$

$$E_{kk}(M) = \left[\sum_{i=-|k|}^{|k|+\theta(k)-1} M_{i,2|k|+\theta(k)} - \sum_{i=-|k|+1-\theta(k)}^{|k|-1} M_{i,2|k|+\theta(k)-1} \right](M), \tag{2.16}$$

where $k \in \mathbb{Z}$ and $M_{00} = M_{-1,0} = 0$. The gl_∞ highest weight vector $(M)_0$ is the one from $\Gamma([M])$ for which

$$M_{i,2k+\theta-1} = M_i, \quad \forall k \in \mathbb{N},$$
$$\theta = 0,1, \quad i \in [1-\theta-k, k-1]. \tag{2.17}$$

By $\mathfrak{F}$ we denote all gl_∞ modules $V([M])$ described by Proposition 1. The elements of $\mathfrak{F}$ are labeled by all possible sequences $[M]$ that satisfy condition (2.7).

Definition 2: We say that the signature

$$[M] \equiv \{M_i | i \in \mathbb{Z}; \ M_p - M_q \in \mathbb{Z}_+, \quad \forall p < q \in \mathbb{Z}\} \tag{2.18}$$

of $V([M])$ is real (resp., integer), if $M_0 \in \mathbb{R}$ (resp., $M_0 \in \mathbb{Z}$).

By $\mathfrak{F}_R$ (resp., $\mathfrak{F}_Z$) we denote all gl_∞ modules from $\mathfrak{F}$ corresponding to a real (resp., integer) signature. Then

$$\mathfrak{F}_Z \subset \mathfrak{F}_R \subset \mathfrak{F}. \tag{2.19}$$

Let $\omega: A \to A$ be an antilinear anti-involution in the Lie algebra A, i.e.,

$$\omega(a+b) = \omega(a) + \omega(b), \quad \forall a, b \in A, \tag{2.20}$$
$$\omega(\lambda a) = \lambda^* \omega(a), \quad \lambda^* \text{ is the complex conjugate to } \lambda \in \mathbb{C}, \tag{2.21}$$
$$\omega([a,b]) = [\omega(b), \omega(a)], \quad \forall a, b \in A, \tag{2.22}$$

and let (,) be a scalar product in the A module V. The representation of A in V is said to be unitarizable[8] if (we write the same letter a for $a\in A$ and for its representative as an operator in V)

$$(ax,y)=(x,\omega(a)y), \quad \forall a\in A,\ x,y\in V, \tag{2.23}$$

i.e., if the scalar product is contravariant with respect to ω (here we accept the more appropriate terminology from Ref. 17; in Ref. 16 we have called the unitarizable representations unitary representations). One natural [also called " compact" in case of $\mathrm{gl}(n)$] antilinear anti-involution on gl_∞ is given with Eqs. (2.20), (2.21), and

$$\omega(E_{ij})=E_{ji}, \quad \forall i,j\in\mathbb{Z}. \tag{2.24}$$

In Ref. 16 we determined all unitarizable irreducible highest weight gl_∞ modules corresponding to the compact anti-involution (2.24). It turns out that any such unitarizable module is one from the class $\mathfrak{F}$. More precisely (see Ref. 16), we have the following proposition.

Proposition 2: The irreducible highest weight gl_∞ module $V([M])$ carries a unitarizable representation of gl_∞ if and only if $V([M])\in\mathfrak{F}_R$. The real form u_∞ of gl_∞, which is a linear span (over $\mathbf{R}$) of all generators $i(E_{pq}+E_{qp})$, $(E_{pq}-E_{qp})$, $p,q\in\mathbb{Z}$ is integrable to a (unitary) representation of the corresponding group U_∞ if and only if $V([M])\in\mathfrak{F}_Z$.

III. UNITARIZABLE REPRESENTATIONS OF A_∞

The Lie algebra A_∞ is a completion and central extension of gl_∞. The completion $\overline{A}_\infty$ of gl_∞ (see Ref. 8) consists of all square infinite matrices with the property that each matrix has a finite number of nonzero diagonals, i.e.,

$$\overline{A}_\infty=\{(A_{ij})\,|\,i,j\in\mathbb{Z},\ A_{ij}=0 \text{ for all } |i-j|\geqslant 0\}. \tag{3.1}$$

The Lie bracket on $\overline{A}_\infty$ is also the usual matrix commutator. Clearly, gl_∞ is a subalgebra of $\overline{A}_\infty$. Each element $a\in\overline{A}_\infty$ can be represented as an infinite "linear combination" of Weyl generators:

$$a=\sum_{k=p}^{q}\sum_{i=-\infty}^{\infty}\lambda_{ik}E_{i,i+k}, \quad \lambda_{ik}\in\mathbb{C},\ i,j\in\mathbb{Z},\ p\leqslant q\in\mathbb{Z}. \tag{3.2}$$

The central extension A_∞ of $\overline{A}_\infty$ is a direct sum of $\overline{A}_\infty$, considered as a linear space, with a one-dimensional space $\mathbb{C}c$ spanned on c:

$$A_\infty=\overline{A}_\infty\oplus\mathbb{C}c. \tag{3.3}$$

In order to avoid possible confusion, we denote the vectors $E_{ij}\in\mathrm{gl}_\infty$, considered as elements from A_∞, by e_{ij}, $i,j\in\mathbb{Z}$. Then A_∞, considered as a linear space, is a subspace of the direct space sum

$$\sum_{i,j\in\mathbb{Z}}\oplus\,\mathbb{C}e_{ij}\oplus\mathbb{C}c \tag{3.4}$$

consisting of all sums (for simplicity we skip $\oplus$)

$$a=\sum_{k=p}^{q}\sum_{i=-\infty}^{\infty}\lambda_{ik}e_{i,i+k}+\lambda c, \quad \lambda_{ik},\ \lambda\in\mathbb{C},$$
$$i,k\in\mathbb{Z},\ p\leqslant q\in\mathbb{Z}. \tag{3.5}$$

The Lie bracket [,] between any two elements of A_∞ is an (infinite) linear extension of the relations

$$[c,e_{ij}]=0, \quad \forall i,j\in\mathbb{Z}, \tag{3.6}$$

$$[e_{ij},e_{kl}]=\delta_{jk}e_{il}-\delta_{li}e_{kj}+\delta_{il}\delta_{jk}\times[\theta(-i)-\theta(-j)]c, \quad \forall i,j,k,l\in\mathbb{Z}. \tag{3.7}$$

In this article we study all unitarizable highest weight representations of A_∞. More precisely, we construct all irreducible A_∞ modules V with the following properties.

(i) Each module V contains a highest weight vector x_Λ, i.e., the vector x_Λ is annihilated by (the representatives of) the generators e_{ij}, $\forall i<j\in\mathbb{Z}$,

$$e_{ij}x_\Lambda=0, \quad \forall i<j\in\mathbb{Z}, \tag{3.8}$$

and is an eigenvector of all possible elements of the form

$$\sum_{i=-\infty}^{\infty}\lambda_i e_{ii}+\lambda c, \quad \forall\lambda,\lambda_i\in\mathbb{C}, \quad i\in\mathbb{Z}. \tag{3.9}$$

(ii) Each module V is unitarizable with respect to the antilinear anti-involution ω defined from (2.20), (2.21), and

$$\omega(c)=c, \quad \omega(e_{ij})=e_{ji}, \quad \forall i,j\in\mathbb{Z}. \tag{3.10}$$

A. The Lie algebra a_∞

By a_∞ we denote the subalgebra of A_∞, which is a linear span of all e_{ij}, $i,j\in\mathbb{Z}$, and c:

$$a_\infty=\text{lin.env.}\{e_{ij},c\,|\,i,j\in\mathbb{Z}\}, \tag{3.11}$$

i.e., it consists of all linear combinations

$$\sum_{k=p}^{q}\sum_{i=r}^{s}\lambda_{ik}e_{i,i+k}+\lambda c, \quad \lambda_{ik},\lambda\in\mathbb{C},$$
$$i,k\in\mathbb{Z}, \quad p\leqslant q\in\mathbb{Z}, \quad r\leqslant s\in\mathbb{Z}. \tag{3.12}$$

This algebra is a central extension of gl_∞. As a convenient intermediate step we extend in this section each gl_∞ module $V([M])\in\mathfrak{F}$ to an a_∞ (highest weight irreducible) module.

Proposition 3: The central extension a_∞ of gl_∞ is isomorphic ($\simeq$) to a (Lie algebra) direct sum $\hat{\oplus}$ of gl_∞ and $\mathbb{C}c$:

$$a_\infty\simeq\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c. \tag{3.13}$$

Proof: Let E_{ij}, $i,j\in\mathbb{Z}$ be the Weyl generators of gl_∞ in (3.13). Then the linear map f of $\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c$ on a_∞, which is a linear extension of the relations

$$f(c)=c, \quad f(E_{ij})=e_{ij}+\delta_{ij}[\theta(-i)+\alpha]c, \quad \forall i,j\in\mathbb{Z}, \tag{3.14}$$

is an isomorphism of $\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c$ on a_∞ for every $\alpha\in C$. ■

We shall identify a_∞ with $\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c$, $a_\infty=\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c$. Then the a_∞ generators $e_{ij}\in\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c$ read as

$$e_{ij}=E_{ij}-\delta_{ij}[\theta(-i)+\alpha]c, \quad \forall i,j\in\mathbb{Z}. \tag{3.15}$$

Let π be an irreducible representation of gl_∞ in the linear space $V([M])\in\mathfrak{F}$. Setting $\pi(c)=K\in\mathbb{C}$, we enlarge $V([M])$ to an irreducible $\mathrm{gl}_\infty\,\hat{\oplus}\,\mathbb{C}c=a_\infty$ module. The expressions for the a_∞ generators e_{ij} as operators in V read as

$$\pi(e_{ij})=\pi(E_{ij})-\delta_{ij}[\theta(-i)+\alpha]K. \tag{3.16}$$

Setting $K=\xi_0-\xi_1$ and $\alpha K=\xi_1$, we write (3.16) as

$$\pi(e_{ij})=\pi(E_{ij})-\delta_{ij}[(\xi_0-\xi_1)\theta(-i)+\xi_1], \tag{3.17}$$

$$\pi(c)=\xi_0-\xi_1, \tag{3.18}$$

where ξ_0 and ξ_1 are arbitrary, but fixed complex constants.

On the contrary, if V is an irreducible a_∞ module, then $\pi(c)$ is a constant, say $\pi(c)=\xi_0-\xi_1$, and the relations

$$\pi(E_{ij})=\pi(e_{ij})+\delta_{ij}[(\xi_0-\xi_1)\theta(-i)+\xi_1] \quad (3.19)$$

turn V into an irreducible gl_∞ module. Thus, we have the following corollary.

Corollary 1: Every irreducible a_∞ module V can be obtained as an extension of an irreducible gl_∞ module V. The restriction of every irreducible a_∞ module V to gl_∞ is irreducible.

Let V be an irreducible highest weight gl_∞ module with a highest weight vector x_Λ. Then [see (3.19)]

$$\pi(E_{ij})x_\Lambda=\pi(e_{ij})x_\Lambda=0, \quad \forall i<j\in\mathbb{Z}. \quad (3.20)$$

Corollary 2: Every irreducible highest weight a_∞ module V can be obtained as an extension of a highest weight irreducible gl_∞ module V. The restriction of every irreducible highest weight a_∞ module V to gl_∞ is an irreducible highest weight V module.

The extension of a gl_∞ module $V([M])\in\mathfrak{F}$ to a highest weight irreducible a_∞ module by means of relations (3.17) and (3.18) will be denoted as $V([M];\xi_0,\xi_1)$ and its basis—as $\Gamma([M];\xi_0,\xi_1)$. We denote by $\mathfrak{F}^0$ all such possible extensions, i.e.,

$$\mathfrak{F}^0=\{V([M];\xi_0,\xi_1)\,|\,V([M])\in\mathfrak{F},\xi_0,\xi_1\in\mathbb{C}\}, \quad (3.21)$$

and by $\mathfrak{F}^0_R$ (resp. $\mathfrak{F}^0_Z$), the subset of $\mathfrak{F}^0$ consisting of $V([M];\xi_0,\xi_1)$ with real (resp., integer) signatures $[M]$ (see Definition 2) and real (resp., integer) ξ_0,ξ_1.

Let V be an arbitrary unitarizable [with respect to ω as defined by (3.10)] highest weight irreducible a_∞ module. Setting in (3.17) ξ_0,ξ_1 as real numbers, we turn V into a unitarizable gl_∞ module [i.e., (2.24) holds]. Hence (Proposition 2), V is a module from $\mathfrak{F}_R$, $V=V([M])\in\mathfrak{F}_R$ and its extension to an a_∞ module V [by means of Eqs. (3.17) and (3.18)] with the same real ξ_0,ξ_1 gives $V=V([M];\xi_0,\xi_1)\in\mathfrak{F}^0_R$. We summarize.

Corollary 3: Every unitarizable irreducible highest weight a_∞ module V is an extension of a gl_∞ module $V([M])\in\mathfrak{F}_R$. The a_∞ generators are defined as operators in $V([M])$ with Eqs. (3.17) and (3.18), where $\xi_0,\xi_1\in\mathbb{R}$. Hence, $V=V([M];\xi_0,\xi_1)\in\mathfrak{F}^0_R$.

From now on we simplify the notation and write E_{ij}, e_{ij} instead of $\pi(E_{ij})$, $\pi(e_{ij})$. Therefore, Eqs. (3.17) and (3.18) read as

$$e_{ij}=E_{ij}-\delta_{ij}[(\xi_0-\xi_1)\theta(-i)+\xi_1], \quad c=\xi_0-\xi_1. \quad (3.22)$$

The transformation of the basis $\Gamma([M];\xi_0,\xi_1)$ in each $V([M];\xi_0,\xi_1)\in\mathfrak{F}^0$ is easily found from Eq. (3.22) and the transformation relations (2.14)–(2.16). Since $e_{ij}=E_{ij}$ for $i\neq j$, one has simply to replace E by e in the lhs of Eqs. (2.14) and (2.15). The action of e_{kk}, $k\in\mathbb{Z}$ on any C vector $(M)\in V([M];\xi_0,\xi_1)$ reads as

$$\begin{aligned}e_{kk}(M)=&\Bigg[\sum_{i=-|k|}^{|k|+\theta(k)-1}M_{i,2|k|+\theta(k)}\\&-\sum_{i=-|k|+1-\theta(k)}^{|k|-1}M_{i,2|k|+\theta(k)-1}\\&+(\xi_1-\xi_0)\theta(-i)-\xi_1\Bigg](M). \quad (3.23)\end{aligned}$$

B. Unitarizable representations of A_∞

Observe first that (apart from the trivial case $M_i=0$, $\forall i\in\mathbb{Z}$) the gl_∞ modules $V([M])\in\mathfrak{F}$ cannot be extended by linearity to $\bar{A}_\infty$ modules. Indeed, let $(M)_0$ be the $V([M])$ highest weight vector [see (2.17)]. For $(M)_0$ Eq. (2.16) reduces to

$$E_{ii}(M)_0=M_i(M)_0, \quad i\in\mathbb{Z}. \quad (3.24)$$

Therefore, the action of

$$I_1=\sum_{i\in\mathbb{Z}}E_{ii}\in\bar{A}_\infty \quad (3.25)$$

on $(M)_0$ gives an undefined, divergent expression:

$$I_1(M)_0=\Bigg[\sum_{i=-\infty}^{\infty}M_i\Bigg](M)_0, \quad (3.26)$$

i.e., the Lie algebra $\bar{A}_\infty$ has no nontrivial highest weight representations [with respect to the Borel subalgebra (2.3)]. This is the reason why (also in all physical applications) one has to go from the completion $\bar{A}_\infty$ of gl_∞ to its central extension A_∞ (for the same reason—see Sec. I—one has to enlarge $\bar{B}_\infty$ to its central extension).

We now proceed to investigate which a_∞ modules $V([M];\xi_0,\xi_1)$ can be turned into A_∞ modules.

Definition 3: We say that the a_∞ module $V([M];\xi_0,\xi_1)\in\mathfrak{F}^0$ [resp., the gl_∞ module $V([M])\in\mathfrak{F}$] is of a finite signature or, more precisely, of (p,q) signature and write $[M]=[M]_{p,q}$ if there exist integers $p<q$ such that

$$M_{p-k}=M_p>M_{p+1}, \quad M_{q-1}>M_q=M_{q+k}, \quad \forall k\in\mathbb{N}. \quad (3.27)$$

Let $(M)_0$ be the highest weight vector in the a_∞ module $V([M];\xi_0,\xi_1)$. From (3.23) we find that

$$e_{ii}(M)_0=[M_i+(\xi_1-\xi_0)\theta(-i)-\xi_1](M)_0, \quad (3.28)$$

Therefore, the action of $I=\Sigma_{i=-\infty}^{\infty}e_{ii}\in A_\infty$ on $(M)_0$ gives

$$I(M)_0=\Bigg[\sum_{i=-\infty}^{0}(M_i-\xi_0)+\sum_{i=1}^{\infty}(M_i-\xi_1)\Bigg](M)_0. \quad (3.29)$$

The expression in the square brackets of (3.29) is not divergent only in the finite-signature modules $V([M]_{p,q};M_p,M_q)$. Thus we conclude with the following corollary.

Corollary 4: The a_∞ module $V([M];\xi_0,\xi_1)$ cannot be extended to an A_∞ module if it is not of a finite signature. The finite-signature module $V([M]_{p,q};\xi_0,\xi_1)$ cannot be extended to an A_∞ module if $\xi_0\neq M_p$ and $\xi_1\neq M_q$.

It remains to investigate all finite-signature a_∞ modules:

$$\begin{aligned}\mathfrak{F}^{\mathrm{FS}}=\{&V([M]_{p,q};M_p,M_q)\,|\,p\leqslant q\in\mathbb{Z},\\&M_p,M_q\in\mathbb{C}, \quad M_p-M_q\in\mathbb{Z}_+\}\subset\mathfrak{F}^0. \quad (3.30)\end{aligned}$$

We now proceed to prove that each a_∞ module from $\mathfrak{F}^{\mathrm{FS}}$ can be extended by linearity to an A_∞ module. To this end it suffices to show that the action of each k-diagonal element $k\in\mathbb{Z}$ [see (3.5)],

$$a_k=\sum_{i=-\infty}^{\infty}\lambda_i e_{i,i+k}\in A_\infty, \quad \lambda_i\in\mathbb{C} \quad (3.31)$$

is well defined as an operator in every a_∞ module from $\mathfrak{F}^{\mathrm{FS}}$.

Let $(M)\in V([M];\xi_0,\xi_1)$ [resp., $(M)\in V([M])$] be an arbitrary C-basis vector. If $N\equiv N[(M)]\in\mathbb{N}$ is an integer, for which (2.9) holds, we write

$$(M)\equiv(M)_N. \tag{3.32}$$

Proposition 4: Let $(M)_N\in V([M]_{p,q};M_p,M_q)$. Then

$$\begin{aligned} &e_{kk}(M)_N=0, \quad \forall k\leqslant\min[p,-(N+1)/2],\\ &e_{kk}(M)_N=0, \quad \forall k>\max[q,(N-1)/2]. \end{aligned} \tag{3.33}$$

Proof: From (3.27) we have that $M_k=M_p$, $\forall k\leqslant p$; $M_k=M_q$, $\forall k\geqslant q$. If $k>\max[q,(N-1)/2]$, then $k\geqslant 0$, $2|k|+\theta(k)-1\geqslant N$ and according to (2.9),

$$M_{i,2|k|+\theta(k)}=M_i, \quad M_{i,2|k|+\theta(k)-1}=M_i. \tag{3.34}$$

Then (3.23) yields

$$e_{kk}(M)_N=\left[\sum_{i=-k}^{k}M_i-\sum_{i=-k}^{k-1}M_i-M_q\right](M)_N$$
$$=(M_k-M_q)(M)_N=0.$$

The proof of (3.33), if $k\leqslant\min[p,-(N+1)/2]$, is similar. ∎

From Proposition 4 we draw an immediate conclusion.

Corollary 5: Every element

$$a_0=\sum_{i=-\infty}^{\infty}\lambda_i e_{ii}\in A_\infty, \quad \lambda_i\in\mathbb{C} \tag{3.35}$$

from the principal diagonal of A_∞ is defined as a linear operator in any finite-signature a_∞ module $V([M]_{p,q};M_p,M_q)\in\mathfrak{F}^{FS}$.

The action of a_0 on $(M)_N$ reads as

$$a_0(M)_N=\left\{\sum_{i>\min(p,-1/2(N+1))}^{i\leqslant\max(q,1/2(N-1))}\lambda_i\left[\sum_{k=-|i|}^{|i|+\theta(i)-1}M_{k,2|i|+\theta(i)}-\sum_{k=-|i|+1-\theta(i)}^{|i|-1}M_{k,2|i|+\theta(i)-1}+(M_q-M_p)\theta(-i)-M_q\right]\right\}(M)_N. \tag{3.36}$$

We recall that [see (2.16)] $M_{00}=M_{-1,0}=0$.

Proposition 5: Let $(M)\equiv(M)_N\in V([M]_{p,q};\xi_0,\xi_1)\in\mathfrak{F}^0$ and

$$Q[(M)_N]=\max[-p,q,(1/2)(N+1)]. \tag{3.37}$$

Then

$$e_{k,k-1}(M)=e_{-k,-k-1}(M)=e_{k-1,k}(M)=e_{-k-1,-k}(M)=0, \quad \forall k>Q[(M)_N]. \tag{3.38}$$

Proof: If $k>Q[(M)_N]$, $M_{i,2k+\eta-2}=M_i$, $\forall\eta=0,1$ and in (2.15) we can replace

$$L_{i,2k+\eta-2},\ L_{i,2k+\eta-1},\ L_{i,2k+\eta},\ L_{i,2k+\eta+1} \text{ by } L_i\equiv M_i-i. \tag{3.39}$$

Consider the action of $e_{k,k-1}=E_{k,k-1}$ on $(M)_N\equiv(M)$. From (3.39) and (2.15) we obtain

$$E_{k,k-1}(M)=-\sum_{l=-k}^{k-1}\sum_{j=1-k}^{k-1}S(j,l)\left|\frac{\Pi_{i\neq l=-k}^{k-1}(L_i-L_j)\Pi_{i=1-k}^{k-2}(L_i-L_j)}{\Pi_{i\neq j=1-k}^{k-1}(L_i-L_j+1)(L_i-L_j)}\right|^{1/2}$$
$$\times\left|\frac{\Pi_{i=-k}^{k}(L_i-L_l+1)\Pi_{i\neq j=1-k}^{k-1}(L_i-L_l+1)}{\Pi_{i\neq l=-k}^{k-1}(L_i-L_l+1)(L_i-L_l)}\right|^{1/2}(M)_{-[l,2k]}^{-[l,2k-1]}. \tag{3.40}$$

The multiple $\Pi_{i=1-k}^{k-2}(L_i-L_j)$ in the numerator of (3.40) is different from zero only if $j=k-1$. Thus (3.40) reduces to

$$E_{k,k-1}(M)=-\sum_{l=-k}^{k-1}S(k-1,l)\left|\frac{\Pi_{i\neq l=-k}^{k-1}(L_i-L_{k-1})}{\Pi_{i=1-k}^{k-2}(L_i-L_{k-1}+1)}\right|^{1/2}$$
$$\times\left|\frac{\Pi_{i=-k}^{k}(L_i-L_l+1)\Pi_{i=1-k}^{k-2}(L_i-L_l+1)}{\Pi_{i\neq l=-k}^{k-1}(L_i-L_l+1)(L_i-L_l)}\right|^{1/2}(M)_{-[l,2k]}^{-[l,2k]}. \tag{3.41}$$

Since $\Pi_{i\neq l=-k}^{k-1}(L_i-L_{k-1})\neq 0$ only if $i\neq k-1$, we obtain [see (3.39)]

$$E_{k,k-1}(M)=|L_k-L_{k-1}+1|^{1/2}(M)_{-[k-1,2k]}^{-[k-1,2k-1]}=0. \tag{3.42}$$

The proof for the other cases in (3.38) is similar. ∎

Proposition 6: Let $(M)\equiv(M)_N\in V([M]_{p,q};M_p,M_q)$ and $Q[(M)_N]=\max[-p,q,1/2(N+1)]$. Then

$$\begin{aligned} &e_{ij}(M)_N=0, \quad \forall i,j>Q[(M)_N],\\ &e_{ij}(M)_N=0, \quad \forall i,j<-Q[(M)_N]. \end{aligned} \tag{3.43}$$

The proof is a consequence of (3.33), (3.38), and the identities

$$e_{ij}=[[[\ldots[[e_{i,i+1},e_{i+1,i+2}],e_{i+2,i+3}],\ldots],e_{j-2,j-1}],e_{j-1,j}], \quad i<j,$$
$$e_{ij}=[[[\ldots[[e_{i,i-1},e_{i-1,i-2}],e_{i-2,i-3}],\ldots],e_{j+2,j+1}],e_{j+1,j}], \quad i>j.$$

Proposition 7: Each finite-signature a_∞ module $V([M]_{p,q};M_p,M_q)$ can be extended by linearity to an A_∞ module.

Proof: Choose an arbitrary C-basis vector $(M)_N\in V([M]_{p,q};M_p,M_q)$ and consider the action of a k-diagonal element a_k [see (3.31)] on $(M)_N$:

$$a_k(M)_N = \sum_{i=-\infty}^{\infty} \lambda_i e_{i,i+k}(M)_N. \tag{3.44}$$

Proposition 6 tells us that $e_{i,i+k}(M)_N$ can be different from zero only for those $i\in\mathbb{Z}$ for which $\min(i,i+k)\leqslant Q[(M)_N]$ and at the same time $\max(i,i+k)\geqslant -Q[(M)_N]$. Setting $\min(i,i+k)=i+k\theta(-k)$ and $\max(i,i+k)=i+k\theta(k)$, we conclude that $e_{i,i+k}(M)_N$ may differ from zero only for those $i\in\mathbb{Z}$ for which

$$-Q[(M)_N]-k\theta(k)\leqslant i\leqslant Q[(M)_N]-k\theta(-k). \tag{3.45}$$

Then (3.44) reduces to

$$a_k(M)_N = \sum_{i=-Q[(M)_N]-k\theta(k)}^{Q[(M)_N]-k\theta(-k)} \lambda_i e_{i,i+k}(M)_N. \tag{3.46}$$

Every $e_{i,i+k}(M)_N$ is a (finite) linear combination of C-basis vectors and, therefore, the same holds for $a_k(M)_N$. Finally, since any $a\in A_\infty$ is a finite sum of k-diagonal elements and λc, i.e.,

$$a=\sum_{k=p}^{q} a_k+\lambda c,\quad p\leqslant q\in\mathbb{Z},\quad \lambda\in\mathbb{C}, \tag{3.47}$$

$a(M)_N$ is well-defined vector from $V([M]_{p,q};M_p,M_q)$. ■

From (3.46) and (3.47) we conclude that for any C-basis vector (M) and every $a\in A_\infty$ there exists $\hat{a}\in a_\infty$, so that $a(M)=\hat{a}(M)$. More generally, for every (M) and every polynomial P of elements from A_∞ [considered as operators in $V([M]_{p,q};M_p,M_q)$] there exists a polynomial $\hat{P}$ of a_∞ generators, so that

$$P(M)=\hat{P}(M), \tag{3.48}$$

where $\hat{P}$ may depend on (M).

Since the antilinear anti-involution ω on A_∞ and a_∞ is one and the same, every unitarizable A_∞ module has to be unitarizable with respect to a_∞. Therefore (see Corollary 3), $V([M]_{p,q};M_p,M_q)$ is unitarizable if and only if $M_p,M_q\in R$ (it suffices to assume $M_0\in\mathbb{R}$). Denote by $\mathfrak{F}_R^{FS}$ [see (3.30)] all such modules

$$\mathfrak{F}_R^{FS}=\{V([M]_{p,q};M_p,M_q)\,|\,p\leqslant q\in\mathbb{Z},\quad M_p,M_q\in\mathbb{R},\quad M_p-M_q\in\mathbb{Z}_+\}\subset\mathfrak{F}^{FS}\subset\mathfrak{F}^0. \tag{3.49}$$

Proposition 8: The highest weight irreducible A_∞ module V is unitarizable if and only if $V\in\mathfrak{F}_R^{FS}$.

Proof: Let x_Λ be the highest weight vector in V (with respect to $gl_\infty, a_\infty, A_\infty$). Denote by $U[A_\infty]$ and $U[a_\infty]$ all polynomials of the elements of A_∞ and a_∞ (considered as operators in V), respectively. Then

$$V=U[A_\infty]x_\Lambda. \tag{3.50}$$

The subspace

$$V_0=U[a_\infty]x_\Lambda\subseteq V \tag{3.51}$$

is an unitarizable highest weight irreducible a_∞ module (V_0 cannot be indecomposible since the representation of a_∞ in it is unitarizable). Hence (Corollary 3), $V_0=V([M];\xi_0,\xi_1)\in\mathfrak{F}_R^0$ and $x_\Lambda\in V([M];\xi_0,\xi_1)$. The action of A_∞ is defined on V and, in particular, on $x_\Lambda\equiv(M)_0\in V_0$. Then (3.29), Corollary 4, and Proposition 7 yield that $V_0=V([M]_{p,q};M_p,M_q)\in\mathfrak{F}_R^{FS}$. Now applying (3.48) and (3.51), we have

$$V=U[A_\infty]x_\Lambda\subseteq U[a_\infty]x_\Lambda = V_0=V([M]_{p,q};M_p,M_q)\subseteq V. \tag{3.52}$$

Therefore, $V=V([M]_{p,q};M_p,M_q)\in\mathfrak{F}_R^{FS}$. ■

The class of all A_∞ modules, which are in addition integrable, consists of all $V([M]_{p,q};M_p,M_q)\in\mathfrak{F}_R^{FS}$, for which $M_0\in\mathbb{Z}$. This is evident, for instance, from the observation that the subalgebra $gl(2)\subset A_\infty$ has to be integrable in $V([M]_{p,q};M_p,M_q)$ and in particular, on the subspace, which $gl(2)$ generate from the highest weight vector $(M)_0$.

ACKNOWLEDGMENTS

I am thankful to Professor H. D. Doebner for the kind hospitality in the Arnold Sommerfeld Institute for Mathematical Physics, where the present paper has been completed. Some of the results have been obtained during my visit to the Department of Mathematics at the University of Queensland. I am thankful to Dr. A. J. Bracken, Dr. M. D. Gould, and Dr. R. B. Zhang for the fruitful discussions.

[1] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, J. Phys. Soc. Jpn. **50**, 3806 (1981).
[2] V. G. Kac and D. H. Peterson, Proc. Natl. Acad. Sci. USA **78**, 3308 (1981).
[3] E. Date, M. Kashiwara, and T. Miwa, Proc. Jpn. Acad. **57**, 387 (1981).
[4] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, Physica **4D**, 343 (1982).
[5] M. Jimbo and T. Miwa, Publ. Res. Inst. Math. Sci. **19**, 943 (1983).
[6] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, in *Proceedings of the RIMS Symposium*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).
[7] V. G. Kac and D. H. Peterson, in Séminaire de Mathématique Supérieure 102, *Systèmes dynamiques non linéaires* (Les Presses de l'Université de Montreal, Montreal, 1986).
[8] V. G. Kac and A. K. Raina, *Advanced Series in Mathematics, Vol. 2* (World Scientific, Singapore, 1987).
[9] P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 303 (1986).
[10] C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[11] T. Palev, Ann. Inst. Henri Poincaré **XXIII**, 49 (1975); Rep. Math. Phys. **14**, 317 (1978).
[12] H. S. Green, Phys. Rev. **90**, 270 (1953).
[13] O. W. Greenberg and A. M. L. Messiah, Phys. Rev. B **138**, 1155 (1965).
[14] T. D. Palev, preprint JINR, Dubna, E17-10550 (1977); Rep. Math. Phys. **18**, 117, 129 (1980); A-superquantization, Communication JINR, Dubna, E2-11942 (1978); J. Math. Phys. **21**, 2560 (1980).
[15] T. D. Palev, in *Proceedings of the Workshop on Group Theoretical Methods and Physical Applications*, ASI TU Clausthal, 1988, edited by H. D. Doebner and J. D. Hennig (World Scientific, Singapore).
[16] T. D. Palev, J. Math. Phys. **31**, 579 (1990).
[17] H. P. Jakobson and V. G. Kac, J. Funct. Anal. **82**, 69 (1989).

A simple derivation of the quantum Clebsch–Gordan coefficients for $SU(2)_q$

H. Ruegg
Département de Physique Théorique, Université de Genève, CH-1211 Genève 4, Switzerland

(Received 7 September 1989; accepted for publication 10 January 1990)

To $SU(2)_q$, the quantum deformation of SU(2), the van der Waerden method for calculating the Clebsch–Gordan (CG) coefficients is genaralized. The polynomial basis for irreducible representations of $SU(2)_q$, the relevant polynomial invariants, and the deduction of the q analog of the CG coefficients are given.

I. INTRODUCTION

Quantum deformations of Lie groups and Lie algebras or "quantum groups"[1] are new mathematical objects developed in the theory of quantum integrable systems, where the Yang–Baxter equation plays a crucial role.[2] The physical applications[3] comprise factorizable S matrices; lattice statistical models at the critical temperature; string theory in particle physics; and, more generally, conformal quantum field theories in two dimensions.[4,5] There are also deep relations to the theory of knot invariants, Jones polynomials,[6] and Hecke algebras.[4,5]

One of the simplest examples[7] of a quantum group is $SU(2)_q$, the quantum deformation of SU(2). In particular, the q analogs of the Racah $6j$ symbols are closely related[4,5,8] to the crossing and braiding properties of conformal blocks, due to the bootstrap conditions.[9] The $3j$ and $6j$ symbols of $SU(2)_q$ have been explicitly given,[8,10] but it seems desirable to offer an alternative derivation of the general formula for Clebsch–Gordan (CG) coefficients that sheds light on different aspects of quantum groups.

The method we propose in this paper is inspired by the derivation of SU(2) CG coefficients due to van der Waerden.[11] The method uses a polynomial basis[5] for the representations of $SU(2)_q$. This requires the introduction of "q numbers" and a "q derivative" D.[12] With these devices, the relevant expressions are very similar for SU(2) and $SU(2)_q$, where the irreducible representations are characterized by the same positive half-integer or integer j and the states are again characterized by $-j \leqslant m \leqslant j$. However, the polynomial invariants used by van der Waerden become more complicated as a result of the non-Leibniz properties of the q derivative and the action of the $SU(2)_q$ generators on tensor products.

The paper is organized as follows. In Sec. II we recall the van der Waerden procedure for SU(2). In Sec. III we give the polynomial basis for $SU(2)_q$ and the expressions of the generators as q derivatives. From this we derive the commutation relations and show that one has to modify the action on the tensor product. In Sec. IV we discuss the $SU(2)_q$ invariants and deduce the CG coefficients.

II. THE CG COEFFICIENTS OF SU(2)

Van der Waerden's method[11] uses as basis vectors for irreducible representations of SU(2) the monomials in two variables s and t:

$$T^j_m \equiv s^{j+m}t^{j-m}/\{(j+m)!(j-m)!\}^{1/2} \equiv |jm\rangle. \quad (2.1)$$

These vectors form an orthonormal system with respect to the scalar product

$$(f,g) = f(\partial_s,\partial_t)g(s,t)|_{s=t=0}, \quad (2.2)$$

where $\partial_s = \partial/\partial_s$. The generators of the Lie algebra SU(2) are the differential operators

$$J_+ = s\partial_t, \quad J_- = t\partial_s, \quad J_3 = \tfrac{1}{2}(s\partial_s - t\partial_t) \equiv \tfrac{1}{2}H. \quad (2.3)$$

One can also form the equivalent conjugate representation with the basis

$$\bar{T}^j_m \equiv \bar{s}^{j+m}\bar{t}^{j-m}/\{(j+m)!(j-m)!\}^{1/2}, \quad (2.4)$$

where the generators are given by

$$J_+ = -\bar{t}\partial_{\bar{s}}, \quad J_- = -\bar{s}\partial_{\bar{t}}, \quad H = -\bar{s}\partial_{\bar{s}} + \bar{t}\partial_{\bar{t}}. \quad (2.5)$$

Here $\bar{T}^j_m$ is contravariant to T^j_m because we can form the invariant

$$(s\bar{s} + t\bar{t})^{2j} = \sum_{m=-j}^{j} \frac{(2j)!(s\bar{s})^{j+m}(t\bar{t})^{j-m}}{(j+m)!(j-m)!} = (2j)!\sum_m T^j_m \bar{T}^j_m. \quad (2.6)$$

We can now form the invariant in six variables:

$$I = (s_1t_2 - t_1s_2)^n(s_1\bar{s}_3 + t_1\bar{t}_3)^{n_1}(s_2\bar{s}_3 + t_2\bar{t}_3)^{n_2}, \quad n = j_1 + j_2 - j, \quad n_1 = -j_1 + j_2 + j, \quad n_2 = j_1 - j_2 + j, \quad (2.7)$$

which can also be written as

$$I = \sum_{mm_1m_2} \langle j_1m_1j_2m_2|jm\rangle T^{j_1}_{m_1} T^{j_2}_{m_2} \bar{T}^j_m, \quad m = m_1 + m_2. \quad (2.8)$$

Equating equal powers one easily obtains the CG coefficients $\langle j_1m_1j_2m_2|jm\rangle$ up to a normalization factor depending on j, j_1, and j_2.

III. POLYNOMIAL BASIS FOR $SU(2)_q$

We restrict q to the real number

$$0 < q < 1. \quad (3.1)$$

One defines the "q number" $[\lambda]$ as

$$[\lambda] = (q^{\lambda/2} - q^{-\lambda/2})/(q^{1/2} - q^{-1/2}), \quad [0] = 0. \quad (3.2)$$

For $q \to 1$, one finds $[\lambda] = \lambda$. The "q derivative" is defined by[12]

 0022-2488/90/051085-03$03.00

$$D_x f(x) = [f(q^{1/2}x) - f(q^{-1/2}x)]/(q^{1/2}x - q^{-1/2}x). \tag{3.3}$$

For $q \to 1$, (3.3) goes over into the usual derivative.

It follows easily that

$$D_x x^n = [n]x^{n-1}. \tag{3.4}$$

One finds a "q-Leibniz" rule since

$$\begin{aligned}(q^{1/2}x - q^{-1/2}x)D_x(f(x)g(x)) &= f(q^{1/2}x)g(q^{1/2}x) - f(q^{-1/2}x)g(q^{-1/2}x)\\ &= (f(q^{1/2}x) - f(q^{-1/2}x))g(q^{1/2}x)\\ &\quad + f(q^{-1/2}x)(g(q^{1/2}x) - g(q^{-1/2}x)).\end{aligned}$$

Hence

$$\begin{aligned}D_x(f(x)g(x)) &= (D_x f(x))g(q^{1/2}x) + f(q^{-1/2}x)D_x g(x) \qquad (3.5)\\ &= f(q^{1/2}x)D_x g(x) + (D_x f(x))g(q^{-1/2}x).\end{aligned}$$

We can now use these concepts to write a polynomial basis for the space of irreducible representations of $SU(2)_q$ that are again labeled by $j = 0, \frac{1}{2}, 1, \ldots$ and the states labeled by $-j \leqslant m \leqslant j$:

$$|jm\rangle_q \equiv s^{j+m}t^{j-m}/\{[j+m]![j-m]!\}^{1/2} \equiv Q^j_m, \tag{3.6}$$

where the only difference with the corresponding expression (2.1) for $SU(2)$ is the replacement of $(j+m)$ and $(j-m)$ by their q numbers (3.2) and

$$[n]! = [n][n-1]\cdots[1], \quad [0]! = 1, \quad [-n]! = \infty. \tag{3.7}$$

However, the scalar product is now, in view of (3.4), given by

$$(f,g) = f(D_s, D_t)g(s,t)|_{s=t=0}, \tag{3.8}$$

so that the vectors (3.6) are again orthonormal.

To obtain the generators $SU(2)_q$, one simply replaces in $J_\pm$ the usual derivative by the q derivative, while H remains unchanged:

$$X_+ = sD_t, \quad X_- = tD_s, \quad H = s\partial_s - t\partial_t. \tag{3.9}$$

Using (3.4) one easily calculates

$$\begin{aligned}X_\pm |jm\rangle_q &= [j \mp m]^{1/2}[j \pm m + 1]^{1/2}|jm \pm 1\rangle_q, \qquad (3.10)\\ H|jm\rangle_q &= 2m|jm\rangle_q.\end{aligned}$$

Notice that in the fundamental representation $j = \frac{1}{2}$, $X_\pm$ are given by the same matrices as $J_\pm$ for $SU(2)$. The commutation relations[7] of $SU(2)_q$ are

$$[H, X_\pm] = \pm 2X_\pm, \tag{3.11a}$$

$$[X_+, X_-] = [H] \equiv (q^{H/2} - q^{-H/2})/(q^{1/2} - q^{-1/2}). \tag{3.11b}$$

While (3.11a) follows easily from (3.5) and (3.6), we verify (3.11b) using the q-Leibniz rule:

$$\begin{aligned}[X_+X_-]g(s,t) &= (sD_t tD_s - tD_s sD_t)g(s,t)\\ &= sD_s g(s, q^{-1/2}t) + q^{1/2}stD_tD_s g(s,t)\\ &\quad - tD_t g(q^{-1/2}s, t) - q^{1/2}tsD_sD_t g(s,t)\\ &= sD_s g(s, q^{-1/2}t) - tD_t g(q^{-1/2}s, t).\end{aligned}$$

Writing $g(s,t) = s^\alpha t^\beta$ we obtain, using (3.4) and (3.2),

$$\begin{aligned}[X_+, X_-]s^\alpha t^\beta &= ([\alpha]q^{-\beta/2} - [\beta]q^{-\alpha/2})s^\alpha t^\beta\\ &= [\alpha - \beta]s^\alpha t^\beta\end{aligned}$$

or

$$\begin{aligned}[X_+, X_-]s^{j+m}t^{j-m} &= [2m]s^{j+m}t^{j-m}\\ &= [H]s^{j+m}t^{j-m},\end{aligned}$$

which is (3.11b). One obtains the same result applying $[X_+, X_-]$ directly on $s^\alpha t^\beta$ and using the identity

$$[\alpha][\beta + 1] - [\alpha + 1][\beta] = [\alpha - \beta].$$

The commutation relations (3.11) are defined when the generators act on a vector space V spanned, for example, by the basis (3.6). It is now necessary to define the action of the generators on a tensor product $V \otimes V$. Here H is an ordinary differential operator and we try the usual definition

$$H(f \otimes g) = Hf \otimes g + f \otimes Hg, \quad f \otimes g \in V \otimes V. \tag{3.12}$$

We write, when H acts on $V \otimes V$,

$$\Delta(H) = H \otimes 1 + 1 \otimes H. \tag{3.13}$$

From (3.13) it follows that

$$\Delta(q^{\alpha H}) = q^{\alpha H} \otimes q^{\alpha H}. \tag{3.14}$$

Here $\Delta(X_\pm)$ should be defined in such a way that Δ is a homomorphism of the algebra $SU(2)_q$ into $SU(2)_q \otimes SU(2)_q$. In particular, we require that

$$\begin{aligned}&[\Delta(X_+), \Delta(X_-)]\\ &\quad = [\Delta(q^{H/2}) - \Delta(q^{-H/2})]/(q^{1/2} - q^{-1/2}). \qquad (3.15)\end{aligned}$$

One finds, using (3.9) and (3.6), that a definition similar to (3.13) is not compatible with (3.15), but instead one has to define[1]

$$\Delta(X_\pm) = X_\pm \otimes q^{H/4} + q^{-H/4} \otimes X_\pm \tag{3.16}$$

and, when acting on $V \otimes V \otimes V$,

$$\begin{aligned}\Delta^{(3)}(X_\pm) &= X_\pm \otimes q^{H/4} \otimes q^{H/4} + q^{-H/4} \otimes X_\pm \otimes q^{H/4}\\ &\quad + q^{-H/4} \otimes q^{-H/4} \otimes X_\pm. \qquad (3.17)\end{aligned}$$

The map Δ is called a *comultiplication* and is an essential ingredient of a Hopf algebra.[1]

The polynomial basis of $V \otimes V$ will be given in terms of suitably normalized monomials in four variables:

$$s_1^{j_1+m_1}t_1^{j_1-m_1}s_2^{j_2+m_2}t_2^{j_2-m_2}, \tag{3.18}$$

on which the generators act, in agreement with (3.13) and (3.16), as

$$\begin{aligned}H &= s_1\partial_{s_1} - t_1\partial_{t_1} + s_2\partial_{s_2} - t_2\partial_{s_2} \equiv H_1 + H_2,\\ X_+ &= s_1 D_{t_1} q^{H_2/4} + q^{-H_1/4}s_2 D_{t_2}, \qquad (3.19)\\ X_- &= t_1 D_{s_1} q^{H_2/4} + q^{-H_1/4}t_2 D_{s_2}.\end{aligned}$$

A short calculation shows that (3.19) satisfies (3.15).

IV. $SU(2)_q$ INVARIANTS AND CG COEFFICIENTS

As a result of the modified actions (3.16) and (3.19) of $X_\pm$ on tensor products and as a result of the q-Leibniz rule (3.5) for the q derivative D, the $SU(2)_q$ invariants differ from those of $SU(2)$. With two covariant doublets $(s_1 t_1)$ and $(s_2 t_2)$ we can form the invariant

$$I_2 = q^{1/4} s_1 t_2 - q^{-1/4} t_1 s_2 \tag{4.1}$$

and if one of the doublets is the contravariant $(\bar{t}_3 \bar{s}_3)$,

$$I_2' = q^{1/4} s_1 \bar{s}_3 + q^{-1/4} t_1 \bar{t}_3. \tag{4.2}$$

In (4.2) the generators of $SU(2)_q$ are, instead of (3.19),

$$\begin{aligned} H &= s_1 \partial_{s_1} - t_1 \partial_{t_1} - \bar{s}_3 \partial_{\bar{s}_3} + \bar{t}_3 \partial_{\bar{t}_3} \equiv H_1 + H_3, \\ X_+ &= s_1 D_{t_1} q^{H_3/4} - q^{-H_1/4} \bar{t}_3 D_{\bar{s}_3}, \\ X_- &= t_1 D_{s_1} q^{H_3/4} - q^{-H_1/4} \bar{s}_3 D_{\bar{t}_3}. \end{aligned} \tag{4.3}$$

However, the nth powers of I_2 or I_2' are no longer invariant because of (3.4). This can be cured if one defines a q-binomial formula

$$(ax + by)_q^n \equiv \sum_{\nu=0}^{n} \binom{n}{\nu}_q (ax)^{n-\nu} (by)^{\nu}, \tag{4.4}$$

$$\binom{n}{\nu}_q \equiv \frac{[n]!}{[\nu]![n-\nu]!}. \tag{4.5}$$

Then the result of a q derivation is

$$D_x (ax + by)_q^n = [n]! (ax + by)_q^{n-1} D_x (ax + by) \tag{4.6}$$

and now $(I_2)_q^n$ and $(I_2')_q^n$ are invariant.

Next, in agreement with (4.3), we define a conjugate tensor

$$\bar{Q}^j_m \equiv \bar{s}^{j+m} \bar{t}^{j-m} / \{[j+m]![j-m]!\}^{1/2} \equiv {}_q\langle jm|. \tag{4.7}$$

Taking into account (4.2) and (4.4), we have to modify the SU(2) expression (2.6) in the following way in order to obtain an $SU(2)_q$ invariant:

$$\begin{aligned} &(q^{1/4} s_1 \bar{s}_3 + q^{-1/4} t_1 \bar{t}_3)_q^{2j} \\ &\quad = \sum_m \binom{2j}{j+m}_q (q^{1/4} s_1 \bar{s}_3)^{j+m} (q^{-1/4} t_1 \bar{t}_3)^{j-m} \\ &\quad = [2j]! \sum_m q^{m/2} Q^j_m (s_1 t_1) \bar{Q}^j_m (\bar{s}_3 \bar{t}_3). \end{aligned} \tag{4.8}$$

One verifies explictly that (4.8) is invariant under (4.3).

We now have the necessary elements in order to write the q analog of the van der Waerden invariants (2.7) and (2.8), consisting in three factors of six variables. There are two difficulties. (i) The variables (s_1, t_1), (s_2, t_2), and $(\bar{s}_3, \bar{t}_3)$ live in three different Hilbert spaces, so one has to apply the comultiplication $\Delta^{(3)}$ [(3.17)] (ii) Each variable appears in two factors; hence one must use the q-Leibniz rule (3.5). As a consequence, a product of invariant factor is not invariant. The invariant polynomial[13] turns out to be

$$\begin{aligned} I_3 &= (q^a s_1 t_2 - q^{-a} t_1 s_2)_q^n (q^{-a_2} s_1 \bar{s}_3 + q^{a_2} t_1 \bar{t}_3)_q^{n_2} \\ &\quad \times (q^{a_1} s_2 \bar{s}_3 + q^{-a_1} t_2 \bar{t}_3)_q^{n_1}, \end{aligned} \tag{4.9}$$

$$a = \tfrac{1}{4}(j+1), \quad n = j_1 + j_2 - j,$$

$$a_2 = \tfrac{1}{4}(j_2 - 1), \quad n_2 = j_1 - j_2 + j, \tag{4.10}$$

$$a_1 = \tfrac{1}{4}(j_1 + 1), \quad n_1 = -j_1 + j_2 + j.$$

Taking into account (4.8), I_3 can be written, up to a normalization factor $f(j_1 j_2 j)$, as

$$\begin{aligned} I_3 &= \sum_{m, m_1, m_2} (j_1 m_1 j_2 m_2 | jm)_q Q^{j_1}_{m_1} Q^{j_2}_{m_2} \bar{Q}^j_m q^{m/2} \\ &\quad m = m_1 + m_2, \quad |j_1 - j_2| \leqslant j \leqslant j_1 + j_2. \end{aligned} \tag{4.11}$$

From (4.11) one easily obtains the q-CG coefficients, which are very similar to the SU(2) CG coefficients[14] except for q factors and replacement of numbers by q numbers. We obtain, using (3.6), (4.4), and (4.7),

$$\begin{aligned} \langle j_1 m_1 j_2 m_2 | jm \rangle_q &= f(j_1 j_2 j) q^{1/4nN + 1/2(j_1 m_2 - j_2 m_1)} \\ &\quad \times \left\{ \prod_{i=1}^{3} [j_i + m_i]! [j_i - m_i]! \right\}^{1/2} \\ &\quad \times \sum_{\nu \geq 0} (-)^{\nu} q^{-1/2\nu N} D^{-1}, \end{aligned} \tag{4.12}$$

$$\begin{aligned} N &= j_1 + j_2 + j + 1, \quad j_3 = j, \quad m_3 = m, \\ D &= [\nu]! [j_1 + j_2 - j - \nu]! [j_1 - m_1 - \nu]! \\ &\quad \times [j_2 + m_2 - \nu]! [j - j_2 + m_1 + \nu]! \\ &\quad \times [j - j_1 - m_2 + \nu]!. \end{aligned} \tag{4.13}$$

This agrees with Ref. 8 and 10 if the normalization factor is taken to be

$$f(j_1 j_2 j) = \{[2j+1][n_1]![n_2]![n]!([N]!)^{-1}\}^{1/2}. \tag{4.14}$$

ACKNOWLEDGMENTS

I would like to thank Cesar Gomez for his lectures on quantum groups and Philippe Zaugg for useful comments. I also thank Professor H. D. Doebner for his invitation to the 8th International Workshop on Mathematical Physics—Quantum Groups and Geometry, Clausthal, July 1989, where the main results of this paper were presented.

This research was partially supported by the Swiss National Science Foundation.

[1] V. Drinfeld, Dokl. Akad. Nauk USSR **283**, 1060 (1985); M. Jimbo, Lett. Math. Phys. **10**, 63 (1985) and **11**, 247 (1986); E. Sklyanin, Usp. Mat. Nauk, **214**, N2 (1985); V. Drinfeld, Proceedings of the International Congress on Mathematicians, Berkeley, edited by A. M. Gleason (AMS), Providence, 1987), p. 798; S. L. Woronowicz, Publ. RIMS Kyoto Univ. **23**, 117 (1987); Yu. I. Manin, *Quantum Groups and Non-commutative Geometry* (CRM, Montreal, 1988).

[2] L. Faddeev and L. Takhtajan, Russ. Math. Surveys **34**, N5, 11 (1979).

[3] For a review and references see H. J. de Vega, Int. J. Mod. Phys. A **4**, 2371 (1989).

[4] A. Kohno, Ann. l'Inst. Fourier **37**, fasc. 4 (1987); G. Moore and N. Seiberg Commun. Math. Phys. **123**, 177 (1989); G. Moore and N. Reshetikhin, Nucl. Phys. B **328**, 557 (1989) and references therein.

[5] L. Alvarez-Gaumé, C. Gomez, and G. Sierra, Nucl. Phys. B **319**, 155 (1989); L. Alvarez-Gaumé, C. Gomez, and G. Sierra, Phys. Lett. **220**, 142 (1989); L. Alvarez-Gaumé, C. Gomez, and G. Sierra, preprint CERN-TH5369/89 UGVA-DPT-3/605/89 (1989).

[6] J. Fröhlich, in *Nonperturbative Quantum Field Theory*, edited by G. 't Hooft A. Jaffe, G. Mack, P. K. Mitter, and R. Stora (Plenum, New York, 1988); N. Reshetikhin, LOMI preprints E-4-87, E-17-87 (1988).

[7] P. P. Kulish and N. Reshetikhin, J. Sov. Math. **23**, 2435 (1983) [Russian original in: "Zapiski Nauch" Semin LOMI **101**, 101 (1981)].

[8] A. N. Kirillov and N. Reshetikhin, LOMI preprint E-9-88 (1988).

[9] A. Belavin, A. M. Polyakov, and A. Zamolodchikov, Nucl. Phys. B **241**, 33 (1984).

[10] B.-Y. Hou, B.-Y. Hou, and Z.-Q. Ma, preprints BIHEP-TH-89-7-NWU-IMP-89-11, BIHEP-TH-89-8 NWU-IMP-89-12 (1989); L. C. Biedenharn, Invited paper at the 1989 Clausthal Summer Workshop on Mathematical Physics (Quantum Groups) to be published in *Proceedings* (Springer-Verlag, Berlin, 1990).

[11] B. L. van der Waerden, *Die Gruppentheoretische Methode in der Quantenmechanik* (Springer, Berlin 1932) [English translation, *Group Theory and Quantum Mechanics* (Springer, Verlag, Berlin 1974)].

[12] F. N. Jackson, Q. J. Pure Appl. Math. **41**, 193 (1910).

[13] An expression similar, but not identical to (4.9), using a different basis, has been obtained by A. Ch. Ganchev and V. B. Petkova, preprint IC/89/158 (revised) (1989).

[14] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1957), p. 45, Eq. (3.6.11).

Infinite-dimensional algebras and a trigonometric basis for the classical Lie algebras

D. B. Fairlie and P. Fletcher
Department of Mathematical Sciences, University of Durham, Durham DH1 3LE, England
C. K. Zachos
High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois 60439

(Received 15 August 1989; accepted for publication 18 October 1989)

This paper explores features of the infinite-dimensional algebras that have been previously introduced. In particular, it is shown that the classical simple Lie algebras (A_N,B_N,C_N,D_N) may be expressed in an "egalitarian" basis with trigonometric structure constants. The transformation to the standard Cartan–Weyl basis, and the particularly transparent $N\to\infty$ limit that this formulation allows is provided.

I. INTRODUCTION

This article is a further contribution to our investigations of a class of infinite-dimensional Lie algebras whose structure constants are simple trigonometric functions.[1] The generators of the basic algebra are indexed by two-vectors $\mathbf{m}=(m_1,m_2)$ which, in the simplest case, are taken to lie on an integral square lattice, although in general the components need not be integral or even real. We also consider cases where they lie on a triangular lattice. The basic algebra is, in a convenient normalization,

$$[K_{\mathbf{m}+\mathbf{b}},K_{\mathbf{n}+\mathbf{b}}]=(1/k)\sin k(\mathbf{m}\times\mathbf{n})K_{\mathbf{m}+\mathbf{n}+\mathbf{b}}+\mathbf{a}\cdot\mathbf{m}\,\delta_{\mathbf{m}+\mathbf{n},0}, \tag{1.1}$$

where $\mathbf{m}\times\mathbf{n}=m_1n_2-m_2n_1$ and $\mathbf{a},\mathbf{b}$ are arbitrary two-vectors. If $\mathbf{b}$ is a lattice vector then $K_{\mathbf{m}}$ can be redefined to eliminate it. In what follows, we shall assume that this is the case and this has been done.

This algebra admits a superextension, but only if $\mathbf{a}=\mathbf{0}$, i.e., there is no bosonic c-number central extension, by adjoining the relations

$$\{F_{\mathbf{m}},F_{\mathbf{n}}\}=\cos k(\mathbf{m}\times\mathbf{n})K_{\mathbf{m}+\mathbf{n}},\qquad [K_{\mathbf{m}},F_{\mathbf{n}}]=(1/k)\sin k(\mathbf{m}\times\mathbf{n})F_{\mathbf{m}+\mathbf{n}}. \tag{1.2}$$

There are special cases of (1.1) that are centerless with $\mathbf{m}$ on a two-dimensional integral lattice and have $k=2\pi/N$ for some integer N. As this imposes a modulo-N arithmetic on the structure constants, the generators can be considered to be indexed by a toroidal integral lattice, in the sense that the generators $K_{\mathbf{m}+N\mathbf{a}}$ are identified with $K_{\mathbf{m}}$ for all integral two-vectors $\mathbf{a}$. The generators $\mathscr{K}_{\mathbf{m}}$ resulting from this identification, in a convenient normalization, satisfy the finite algebra

$$[\mathscr{K}_{\mathbf{m}},\mathscr{K}_{\mathbf{n}}]=\sin(2\pi/N)(\mathbf{m}\times\mathbf{n})\mathscr{K}_{\mathbf{m}+\mathbf{n}}. \tag{1.3}$$

There is a particular realization of the superalgebra, in which $F_{\mathbf{m}}$ and $K_{\mathbf{m}}$ are identified, going back to Weyl[2] and his correspondence rule, given by

$$K_{\mathbf{m}}=(1/2ik)e^{i(2km_1P+m_2X)}=(1/ik)F_{\mathbf{m}}, \tag{1.4}$$

where (X,P) are canonically conjugate quantum variables with $[X,P]=i$. Using the familiar Baker–Campbell–Hausdorf expansion, the product is

$$K_{\mathbf{m}}K_{\mathbf{n}}=(1/2ik)e^{ik(m_1n_2-m_2n_1)}K_{\mathbf{m}+\mathbf{n}}, \tag{1.5}$$

and therefore

$$K_{\mathbf{m}}K_{\mathbf{n}}=(1/2ik)e^{2ik(m_1n_2-m_2n_1)}K_{\mathbf{n}}K_{\mathbf{m}}. \tag{1.6}$$

For the finite algebras (1.3) the relationship (1.6), evocative of quantum groups, is familiar from the work of 't Hooft and Belavin.[3] Also see Refs. 4. Naturally, it satisfies (1.1) and (1.2).

In fact, when N is odd, (1.3) is just the algebra of $SU(N)\times U(1)$. Although the fact that there exists a basis for $SU(N)$ in which the structure constants are simple trigonometric functions has been recorded several times, it still elicits surprise, and there are as yet only a few articles[5] developing the theory of semi-simple Lie algebras from this maximal grading "egalitarian" point of view, in which all the generators appear on the same footing.

In this paper, we explore several features of the algebra (1.1) and the finite algebras (1.3), as well as their $N\to\infty$ limit,

$$[L_{\mathbf{m}},L_{\mathbf{n}}]=(\mathbf{m}\times\mathbf{n})L_{\mathbf{m}+\mathbf{n}}+\mathbf{a}\cdot\mathbf{m}\,\delta_{\mathbf{m}+\mathbf{n},0}. \tag{1.7}$$

This constitutes the algebra of infinitesimal area-preserving diffeomorphisms of the torus, $SDiff(T^2)$,[6,7] which we have identified with that of $SU(\infty)$.[1]

(i) We find the Casimir invariants of (1.1), and its related limit algebra (1.7).

(ii) We explain how both algebras may be realized as algebras of differential operators on surface coordinates and show how they act as algebras of derivations.

(iii) We transform the finite algebras to their Cartan–Weyl bases and demonstrate that for N even the algebra (1.3) is that of $U(N/2)^4$.

(iv) We identify several subalgebras of (1.3) corresponding to $SO(N)$ and $USp(N)$ and express them in a similar neat form.

(v) We also consider algebras whose generators are indexed by a triangular lattice.

(vi) Finally, we utilize the surface coordinate formalism to express gauge theories for $SU(\infty)$, $SO(\infty)$, and $USp(\infty)$ and identify the Schild string action present within the Yang–Mills action.

II. CASIMIR INVARIANTS

The construction of Casimir invariants is modeled upon that for the finite algebras (1.3) discussed by Patera and

 0022-2488/90/051088-07$03.00

Zassenhaus.[5] The quadratic Casimir is

$$\sum_{\mathbf{m}} K_{\mathbf{m}} K_{-\mathbf{m}}. \tag{2.1}$$

There are, in general, two Casimir invariants of each degree above the quadratic. They are the real and imaginary parts of

$$\sum_{\mathbf{m},\mathbf{n}} e^{ik\mathbf{m}\times\mathbf{n}} K_{\mathbf{m}} K_{\mathbf{n}} K_{-\mathbf{m}-\mathbf{n}},\cdots,$$

$$\sum_{\mathbf{m}_1,\mathbf{m}_2,\ldots,\mathbf{m}_r} \left(\prod_{\alpha<\beta} e^{ik(\mathbf{m}_\alpha\times\mathbf{m}_\beta)}\right) K_{\mathbf{m}_1} K_{\mathbf{m}_2}\cdots K_{\mathbf{m}_r} K_{-\mathbf{m}_1-\mathbf{m}_2-\cdots-\mathbf{m}_r}. \tag{2.2}$$

Taking the imaginary part, a generic coefficient will be of the form

$$\sin(k(\mathbf{m}\times\mathbf{n}+\mathbf{m}\times\mathbf{p}+\mathbf{n}\times\mathbf{p}+\cdots)).$$

By use of the addition formula for sines, this will always be reducible to terms with a typical $\sin k(\mathbf{m}\times\mathbf{n})$ factor. Whenever the remaining factor in such a term is symmetric in $\mathbf{m}$ and $\mathbf{n}$, after use of the commutation relations to make $K_{\mathbf{m}}$ and $K_{\mathbf{n}}$ adjacent, it is easy to see that this contribution to the Casimir may be reduced to one of one degree lower. For example, in the case of the cubic,

$$\sum_{\mathbf{m},\mathbf{n}} \sin(k(\mathbf{m}\times\mathbf{n})) K_{\mathbf{m}} K_{\mathbf{n}} K_{-\mathbf{m}-\mathbf{n}}$$

$$= \sum_{\mathbf{m},\mathbf{n}} \sin^2(k(\mathbf{m}\times\mathbf{n})) K_{\mathbf{m}+\mathbf{n}} K_{-\mathbf{m}-\mathbf{n}}. \tag{2.3}$$

Re-summing over $\mathbf{m}+\mathbf{n}$ and $\mathbf{m}-\mathbf{n}$ we see that the right-hand side diverges, without an infinite renormalization of $K_{\mathbf{m}}$. Such a renormalization, however, would make the cosine-like contributions vanish.

The Casimirs of (1.7) follow by a $k\to 0$ limiting procedure. Again there are apparently two for each degree: one of which can be reduced in degree as above, again with a divergent result.

III. DERIVATIONS

The algebra (1.7) is known to be, in a particular basis optimal for the torus, that of the generic area-preserving (symplectic) reparametrizations of a two-surface. Taking x and p to be local (commuting) coordinates for the surface, and f and g to be differentiable functions of them, a basis-independent realization for the generators of the centerless algebra is [6,7]

$$L_f = \frac{\partial f}{\partial x}\frac{\partial}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial}{\partial x} \Rightarrow \tag{3.1}$$

$$[L_f, L_g] = L_{\{f,g\}}, \quad [L_f, g] = \{f,g\}, \tag{3.2}$$

where

$$\{f,g\} = \frac{\partial f}{\partial x}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial x} \tag{3.3}$$

is the Poisson bracket of classical phase space. The generator L_f transforms (x,p) to $(x-\partial f/\partial p, p+\partial f/\partial x)$. Infinitesimally, this is a canonical transformation generated by f, which preserves the phase space area element $dx\,dp$. This may be regarded as the flow generated by an arbitrary Hamiltonian f. For a small patch of two-surface, the functions $f(x,p)$ may be expanded in any coordinate basis. If the surface is a torus, the preferred basis is $\exp(inx+imp)$; if it is a sphere, spherical harmonics; if it is a plane, powers; and so on. Nevertheless, any coordinate basis will do for the infinitesimal transformations effected by the algebra in a local patch.

Another realization and application of this algebra appears in the work of Case and Monge,[8] which investigates the algebra of conserved currents of the Kadomtsev–Petviashvili equation. Their algebra is seen to be contained in (1.1) with $k\to 0$, after a change of variables.

We found a basis-independent differential operator realization of K_f, corresponding to (3.1):

$$K_f = \frac{1}{2ik} f\left(x+ik\frac{\partial}{\partial p}, p-ik\frac{\partial}{\partial x}\right). \tag{3.4}$$

In the torus basis, this becomes

$$K_{(m_1,m_2)} = \frac{i}{2k}\exp\left(im_1 x + km_2\frac{\partial}{\partial x} + im_2 p - km_1\frac{\partial}{\partial p}\right)$$

$$= \frac{i}{2k}\exp(im_1 x + im_2 p)$$

$$\times\exp\left(km_2\frac{\partial}{\partial x} - km_1\frac{\partial}{\partial p}\right), \tag{3.5}$$

somewhat analogous to the one-variable realization found by Hoppe.[7] Note the triviality in this realization of the Casimir operators, as the indices of each of their terms sum to zero.

Just as the algebra (1.7) may be thought of as the Fourier transform of the Poisson bracket algebra, (3.2), the algebra with general k is the Fourier transform of the "sine bracket" algebra

$$[K_f, K_g] = K_{\sin\{f,g\}}, \tag{3.6}$$

where the analog of the Poisson bracket in this case is the sine, or Moyal, bracket $\sin\{f,g\}$. This is the extension of the Poisson brackets $\{f,g\}$ to statistical distributions on phase space, introduced by Weyl[2] and Moyal.[9] It is generalized convolution that reduces to the Poisson bracket as $\hbar$, replaced by $2k$ in our context, is taken to zero:

$$\sin\{f,g\} = \frac{-1}{4\pi^2 k^3}\int dp'\,dp''\,dx'\,dx\; f(x',p')$$

$$\times g(x'',p'')\sin(1/k)(p(x'-x'')$$

$$+ x(p''-p') + p'x'' - p''x'). \tag{3.7}$$

The argument of the sine above is

$$\frac{1}{k}\det\begin{pmatrix} 1 & p & x \\ 1 & p' & x' \\ 1 & p'' & x'' \end{pmatrix} = \frac{1}{k}\int p\cdot dq, \tag{3.8}$$

i.e., $2/k$ times the area of the equilateral phase-space triangle with vertices at (x,p), (x',p'), and (x'',p''). The antisymmetry of f with g is evident in the determinant. The sine brackets satisfy the Jacobi identities,[10] just as their Fourier components (1.1) do, and thus determine a Lie algebra. These brackets help reformulate quantum mechanics in terms of Wigner's phase-space distribution.[11]

The Poisson bracket $\{f,h\}$ acts as a derivation for both

the Poisson bracket and the ordinary product fg, i.e.,

$$\delta(fg) = (\delta f)\, g + g(\delta f), \tag{3.9}$$

where $\delta(f) = \{f,h\}$ for a given h. Similarly, the sine bracket acts as a derivation for the sine bracket and also for the cosine bracket as "product."

$$\delta \cos(f, g) = \cos(\delta f, g) + \cos(f, \delta g), \tag{3.10}$$

where $\delta(f) = \sin\{f,h\}$. The cosine bracket $\cos(f, g)$ is the counterpart of (3.7) for the cosine. These relations are a simple consequence of the graded Jacobi identities for the respective algebras. [In the special case where $h = p$, (3.9) is just the Leibniz rule for the differentiation of a product.] The sine bracket relation with $h = e^{ap}$ generates finite central differences.

IV. CARTAN–WEYL BASIS

We have already shown[1] that for N odd the algebra (1.3) describes $U(N)$, and that for N even it contains a $U(N/2)$ subalgebra. By finding the combinations of the $\mathscr{K}$'s which form the Cartan–Weyl basis, we shall demonstrate that the full algebra in the N even case is $U(N/2)$[4].

Patera and Zassenhaus[5] and Pope and Stelle[12] have transformed the algebra between the trigonometrical basis and the standard $GL(N)$ basis. We, instead, exhibit the connection of the generators $\mathscr{K}_m$ to the Cartan–Weyl basis h_i and e_α, where the h_i are the members of the Cartan subalgebra H, and α is in the root space Σ.

Here we shall carry out the transformation of the finite algebras (1.3) to the Cartan–Weyl basis, first for N odd, showing that it is $SU(N)\times U(1)$. The generator $\mathscr{K}_{0,0}$ factors out of the algebra, as it commutes with the other N^2-1 and cannot result as a commutator of any of them. This is the $U(1)$ part of the algebra.

The Cartan–Weyl basis for $SU(N)$ has the following commutation relations, in the usual notation:

$$[e_\alpha, e_\beta] = \begin{cases} N_{\alpha\beta} e_{\alpha+\beta}, & \text{if } \alpha+\beta \in \Sigma, \\ (e_\alpha, e_{-\alpha}) h_\alpha, & \text{if } \alpha+\beta = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.1}$$

$$[h_i, e_\alpha] = \alpha(h) e_\alpha, \tag{4.2}$$

$$[h_i, h_j] = 0. \tag{4.3}$$

In the case of N odd, the combinations of $\mathscr{K}$'s that give this basis are

$$E^p_q = \sum_{j=0}^{N-1} \omega^{(2j-q)p} \mathscr{K}_{j,q-j}, \quad \omega^N = 1, \tag{4.4}$$

where $q = 0$ and $p = 1,\ldots,\frac{1}{2}(N-1)$ for the Cartan subalgebra, and $q = 1,\ldots,N-1$ and $p = 0,\ldots,N-1$, for the remaining generators. This may be shown by checking the commutation relations as follows:

$$\begin{aligned}
[E^{p_1}_{q_1}, E^{p_2}_{q_2}] &= \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} \omega^{(2j-q_1)p_1 + (2k-q_2)p_2} \\
&\quad \times [\mathscr{K}_{j,q_1-j}, \mathscr{K}_{k,q_2-k}] \\
&= \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} \omega^{(2j-q_1)p_1 + (2k-q_2)p_2} \\
&\quad \times \sin\frac{2\pi}{N}(jq_2 - kq_1) \mathscr{K}_{j+k, q_1+q_2-j-k} \\
&= \frac{1}{2i}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1} \omega^{(2j-q_1)p_1 + (2k-q_2)p_2} \\
&\quad \times (\omega^{jq_2 - kq_1} - \omega^{-jq_2 + kq_1}) \mathscr{K}_{j+k, q_1+q_2-j-k}.
\end{aligned} \tag{4.5}$$

Putting $s = j + k$ and $q = q_1 + q_2$ and then resumming, using the invariance modulo N:

$$\begin{aligned}
[E^{p_1}_{q_1}, E^{p_2}_{q_2}] &= \frac{1}{2i}\left(\sum_{s=0}^{N-1}\sum_{k=0}^{s} + \sum_{s=N}^{2N-2}\sum_{k=s-N+1}^{N-1}\right) \\
&\quad \times \omega^{(2s-2k-q_1)p_1 + (2k-q_2)p_2} \\
&\quad \times (\omega^{sq_2 - kq} - \omega^{-sq_2 + kq}) \mathscr{K}_{s,q-s} \\
&= \frac{1}{2i}\sum_{s=0}^{N-1} \omega^{(2s-q_1)p_1 - q_2p_2} \sum_{k=0}^{N-1} (\omega^{2k(p_2-p_1) - kq + sq_2} \\
&\quad - \omega^{2k(p_2-p_1) + kg - sq_2}) \mathscr{K}_{s,q-s}.
\end{aligned} \tag{4.6}$$

Using $\sum_{k=0}^{N-1} \omega^{ak} = N\delta_{a,0}$, and then comparing with our expression for E^p_q, we obtain

$$\begin{aligned}
[E^{p_1}_{q_1}, E^{p_2}_{q_2}] &= \frac{1}{2i}\sum_{s=0}^{N-1} \omega^{(2s-q_1)p_1 - q_2p_2} N(\omega^{sq_2}\delta_{2(p_2-p_1)-q,0} \\
&\quad - \omega^{-sq_2}\delta_{2(p_2-p_1)+q,0}) \mathscr{K}_{s,q-s} \\
&= (N/2i)(E^{p_1 + q_2/2}_{q_1+q_2}\delta_{2(p_2-p_1)-q,0} \\
&\quad \times - E^{p_1 - q_2/2}_{q_1+q_2}\delta_{2(p_2-p_1)+q,0}),
\end{aligned} \tag{4.7}$$

where the $q_2/2$ may be defined as an integer (as N may be added to the power of ω to ensure that this power is even), so the halving of the index is defined by:

$$\frac{q}{2} \begin{cases} q/2, & q \text{ even}, \\ (q+N)/2, & q \text{ odd}. \end{cases} \tag{4.8}$$

And so

$$[E^{p_1}_{q_1}, E^{p_2}_{q_2}] = \begin{cases} \pm (N/2i) E^{p_1 \pm q_2/2}_{q_1+q_2}, & \text{if } 2(p_2-p_1) = \pm(q_1+q_2), \\ (N/2i)(E^{p_1+q_2/2}_0 - E^{p_1-q_2/2}_0), & \text{if } q_1+q_2 = 0 = p_2-p_1, \\ 0, & \text{otherwise,} \end{cases} \tag{4.9}$$

which corresponds to (4.1).

$$[E^{p_1}_0, E^{p_2}_q] = \begin{cases} \pm (N/2i) E^{p_2}_q, & \text{if } 2(p_2-p_1) = \pm q, \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

showing the basis is diagonal (4.2), and

$$[E^{p_1}_0, E^{p_2}_0] = 0, \tag{4.11}$$

as required.

For the case N even, the situation is more complicated. This time four of the $\mathscr{K}$'s disconnect into U(1)'s, $\mathscr{K}_{0,0}$, $\mathscr{K}_{N/2,0}$, $\mathscr{K}_{0,N/2}$, and $\mathscr{K}_{N/2,N/2}$. This leaves N^2-4 generators, which we shall show span four commuting SU($N/2$)'s. There are slight differences between the cases (a) $N\equiv 0 \bmod 4$ and (b) $N\equiv 2 \bmod 4$, but the principle of construction is the same. As before, the Cartan subalgebra is spanned by the elements whose indices sum to 0 mod $N/2$. Note that there are $2N-4$ such operators, after excluding the four U(1)'s.

For (a), $N\equiv 0 \bmod 4$, the generators in the Cartan–Weyl basis are

$$E_q^{s,p}=\sum_{a=0}^{1}\sum_{j=0}^{N-1}\omega^{pj}(-1)^{s(j+a)+a(j+1)}\mathscr{K}_{j,q-j+aN/2}, \tag{4.12}$$

and for (b), $N\equiv 2 \bmod 4$,

$$E_q^{s,p}=\sum_{a=0}^{1}\sum_{j=0}^{N-1}\omega^{pj}(-1)^{s(j+1)+(a+1)(j+1+p)}\times\mathscr{K}_{j,q-j+aN/2}, \tag{4.13}$$

where the q labels the sum of the indices, $q=0,\ldots,N/2-1$, and s,p take the values $s=0,1, p=0,\ldots,N-1$. Then for case (a) the elements $E_q^{s,p}$ for $s=0,1$; $q+p$ even, odd span the four commuting SU($N/2$)'s. For case (b), the splitting is into those with $s+q$ even, odd and p even, odd.

That the above combinations are the generators of SU$(N/2)^4$ in the Cartan–Weyl basis may be shown by checking the commutation relations in a similar fashion to the N odd case above. Commuting two of the E's gives an expression which may be resummed so that the coefficients of the $\mathscr{K}$'s can be read off. The resummed expression for (a), with $a=a_1+a_2$, $j=j_1+j_2$, $q=q_1+q_2$, and $p_-=p_1-p_2$ is

$$\begin{aligned}[E_{q_1}^{s_1,p_1},E_{q_2}^{s_2,p_2}]=&\sum_{a=0}^{1}\sum_{j=0}^{N-1}\sum_{a_1=0}^{1}\sum_{j_1=0}^{N-1}\omega^{j_1p_-+p_2j}\\&\times\sin\frac{2\pi}{N}(qj_1-jq_1)\\&\times(-1)^{(s_1+s_2)(a_1+j_1)+s_2(a+j)+a(j+1)}\mathscr{K}_{j,q-j+(N/2)a}.\end{aligned} \tag{4.14}$$

The separation into $s=0,1$ is evident as the only dependence of the coefficient of the $\mathscr{K}$ on a_1 is of the form $(-1)^{(s_1+s_2)a_1}$, so if $s_1\neq s_2$ then the two terms in that sum exactly cancel. When $s_1=s_2$, the coefficient indexed by a,j becomes

$$\begin{aligned}&(-1)^{s_1(a+j)+a(j+1)}\omega^{p_2j}\frac{1}{i}\\&\quad\times\sum_{j_1=0}^{N-1}\omega^{j_1p_-}(\omega^{j_1q-jq_1}-\omega^{-j_1q+jq_1})\\&=(-1)^{s_1(a+j)+a(j+1)}\omega^{p_2j}(N/i)\\&\quad\times(\omega^{-jq_1}\delta_{p_-+q,0}-\omega^{jq_1}\delta_{p_--q,0}).\end{aligned}$$

The δ's are both zero if p_1+q_1 and p_2+q_2 have different parity, showing the overall split into four commuting subspaces. This coefficient may be compared that in (4.12), and the commutator rewritten as

$$\begin{aligned}[E_{q_1}^{s_1,p_1},E_{q_2}^{s_2,p_2}]=(N/i)\delta_{s_1,s_2}&(E_{q_1+q_2}^{s_1,p_2-q_1}\delta_{p_1-p_2+q_1+q_2,0}\\&-E_{q_1+q_2}^{s_1,p_2+q_1}\delta_{p_1-p_2-(q_1+q_2),0}).\end{aligned} \tag{4.15}$$

Similarly for (b),

$$\begin{aligned}[E_{q_1}^{s_1,p_1},E_{q_2}^{s_2,p_2}]=&\sum_{a=0}^{1}\sum_{j=0}^{N-1}\sum_{a_1=0}^{1}\sum_{j_1=0}^{N-1}\omega^{j_1p_-+p_2j}(-1)^{(s_1+s_2)(j_1+1)+a_1p_-+a(j+p_2+1)+j(s_2+1)+p_-}\\&\times\sin 2\pi/N(qj_1-jq_1)\mathscr{K}_{j,q-j+(N/2)a}.\end{aligned} \tag{4.16}$$

This time, the coefficient of a_1 in the exponent is p_-, so the space splits into p even, odd. The coefficient of the $\mathscr{K}$ here is

$$\begin{aligned}&(-1)^{s_1+s_2+(s_2+a+1)j+a(p_1+1)}\omega^{p_2j}\frac{1}{i}\sum_{j_1=0}^{N-1}(-1)^{j_1(s_1+s_2)}\omega^{j_1p_-}(\omega^{j_1q-jq_1}-\omega^{-j_1q+jq_1})\\&=(-1)^{s_1+s_2+(s_2+a+1)j+a(p_1+1)}\omega^{p_2j}(N/i)(\omega^{-jq_1}\delta_{p_-+q,(s_1+s_2)(N/2)}-\omega^{jq_1}\delta_{p_--q,(s_1+s_2)(N/2)})\\&=N/i(-1)^{(s_1+q_1+p_1+1)+(j+1)(s_2+q_1)+(a+1)(j+1+p_1)}\begin{pmatrix}\omega^{(p_2-q_1(1+N/2))j}\delta_{p_-+q,(s_1+s_2)(N/2)}\\-\omega^{(p_2+q_1(1+N/2))j}\delta_{p_--q,(s_1+s_2)(N/2)}\end{pmatrix}.\end{aligned}$$

Both δ functions are zero if s_1+q_1 and s_2+q_2 have different parity. In performing the above manipulations the fact that $\omega^{N/2}=-1$ has been used. Thus, for $p_1\equiv p_2 \bmod 2$,

$$\begin{aligned}[E_{q_1}^{s_1,p_1},E_{q_2}^{s_2,p_2}]=(N/i)(-1)^{s_1+q_1+p_1+1}&(E_q^{s,p_2-q_1(1+N/2)}\\&\times\delta_{p_-+q,(s_1+s_2)N/2}-E_q^{s,p_2+q_1(1+N/2)}\delta_{p_--q,(s_1+s_2)N/2}),\end{aligned} \tag{4.17}$$

where $q \equiv q_1 + q_2 \bmod N/2$, and $s \equiv q + q_1 + q_2 \bmod 2$, so that $s + q \equiv s_1 + q_1 \equiv s_2 + q_2 \bmod 2$.

Saveliev and Vershik[13] have discussed infinite algebras of the type SDiff(T^2) and their generalizations directly in the Cartan–Weyl basis. In our notation, the Cartan–Weyl basis for the infinite algebras L's in (1.7) is

$$E^p_q = \sum_j e^{ipj} L_{1/2(q+j),(1/2)(q-j)}. \tag{4.18}$$

This result may be obtained by checking the commutation relation

$$[E^{p_1}_{q_1}, E^{p_2}_{q_2}] = \tfrac{1}{2}(q_1+q_2)\delta'(p_1-p_2) E^{(1/2)(p_1+p_2)}_{q_1+q_2} - \tfrac{1}{2}(q_1-q_2)\delta(p_1-p_2)\frac{\partial}{\partial p_1} E^{(1/2)(p_1+p_2)}_{q_1+q_2}. \tag{4.19}$$

In Saveliev and Vershik's notation,

$$X_q(f) = \int_{-\infty}^{\infty} f(p) E^p_q \, dp. \tag{4.20}$$

Thus, multiplying equation (4.19) by $f(p_1)g(p_2)$ and performing this integral transform yields precisely their equation for the SU(∞) commutator

$$[X_{q_1}(f), X_{q_2}(g)] = X_{q_1+q_2}(q_2 f'(p_1+p_2) g(p_1+p_2) - q_1 f(p_1+p_2) g'(p_1+p_2)). \tag{4.21}$$

V. SUBALGEBRAS

In this section, we exhibit all the (nonexceptional) classical Lie algebras in a 'trigonometrical' basis analogous to that of SU(N). Since these can fit as subalgebras in SU(N), we can extract them from it, and hence SO(∞) and USp(∞) out of the limit SU(∞). To simplify the analysis, we introduce matrices J which satisfy (1.3), but with structure constant π/N, instead of $2\pi/N$. These matrices provide a basis for one of the four copies of $U(N)^4$, and hence for SU(N) equally well for N even or odd.

Consider matrices g,h (see Ref. 1),

$$g = \sqrt{\omega}\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \omega & 0 & \cdots & 0 \\ 0 & 0 & \omega^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \omega^{N-1} \end{pmatrix},$$

$$h = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad g^N = h^N = -\mathbb{1}, \tag{5.1}$$

with $\omega = e^{2\pi i/N}$, $\sqrt{\omega} = e^{\pi i/N}$. They obey the identity

$$hg = \omega gh. \tag{5.2}$$

Then there is a complete set of unitary $N \times N$ matrices

$$J_{(m_1,m_2)} = \omega^{m_1 m_2/2} g^{m_1} h^{m_2}, \tag{5.3}$$

which satisfy

$$\mathrm{Tr}\, J_{(m_1,m_2)} = 0, \quad \text{except for } m_1 = m_2 = 0 \bmod N, \tag{5.4}$$

and span the algebra of SU(N). Like the Pauli matrices, they close under multiplication to just one such, (a finite group), by virtue of (5.2):

$$J_{\mathbf{m}} J_{\mathbf{n}} = \omega^{\mathbf{n}\times\mathbf{m}/2} J_{\mathbf{m}+\mathbf{n}}. \tag{5.5}$$

They therefore satisfy the algebra

$$[J_{\mathbf{m}}, J_{\mathbf{n}}] = -2i \sin((\pi/N)\mathbf{m}\times\mathbf{n}) J_{\mathbf{m}+\mathbf{n}}. \tag{5.6}$$

It might appear that the fundamental period is $2N$ instead of N. However, note that, by virtue of the symmetry

$$J_{\mathbf{m}+\mathbf{N}(r,s)} = (-1)^{(m_1+1)s+(m_2+1)r} J_{\mathbf{m}}, \tag{5.7}$$

only indices in the fundamental cell $N \times N$ need be considered.

The subalgebras may be written as combinations of J's which close on themselves. Those of most interest are,

$$J_{m_1,m_2} - (-)^a J_{m_1,-m_2} \begin{cases} a = 0, & & \mathrm{SO}(N) \\ a = m_1, & N \text{ even}, & \mathrm{USp}(N) \\ a = m_2, & N \text{ even}, & \mathrm{SO}(N) \\ a = m_1 + m_2, & N = 4M, & \mathrm{USp}(N) \\ a = m_1 + m_2, & N = 4M+2, & \mathrm{SO}(N) \end{cases},$$

$$J_{m_1,m_2} - (-)^a J_{m_2,m_1} \begin{cases} a = 0, & & \mathrm{SO}(N), \\ a = m_1 + m_2, & N \text{ even}, & \mathrm{SO}(N). \end{cases}$$

As an example, consider the second case, with $a = 0$, N odd. We denote

$$J_{[m_1,m_2]} = J_{m_1,m_2} - J_{m_2,m_1}. \tag{5.8}$$

The number of generators of these algebras is $(1/2)N(N-1)$. The commutation relations are

$$[J_{[m_1,m_2]}, J_{[n_1,n_2]}] = -2i \begin{pmatrix} \sin(\pi/N)(m_1 n_2 - m_2 n_1) J_{[m_1+n_1,m_2+n_2]} \\ -\sin(\pi/N)(m_1 n_1 - m_2 n_2) J_{[m_1+n_2,m_2+n_1]} \end{pmatrix}. \tag{5.9}$$

These algebras are presently shown to be SO(N).

It is convenient to label the generators by $q = m_1 + m_2$, the sum of the indices. Those with $q = 0 \bmod N$ all mutually commute, and this is taken as the Cartan subalgebra. Forming the Cartan–Weyl basis amounts to simultaneously diagonalizing the matrices which are the elements of the Cartan subalgebra in the adjoint representation, i.e., the matrices of structure constants on commutation of h with the e_α's,

$$[h, e_\alpha] = \sum_\beta M_{\alpha\beta} e_\beta. \tag{5.10}$$

These matrices are block diagonal, with a block for each $q = 1,\ldots,N-1$, of size $r = \frac{1}{2}(N-1)$. The blocks are all of the form, independent of q, of

$$M_r = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{5.11}$$

Thus the combinations of the generators with a given q proportional to the eigenvectors of M_r are diagonal.

The characteristic polynomial, P_r, of M_r is given (up to sign) by the recurrence relation

$$P_r = \lambda P_{r-1} - P_{r-2}, \tag{5.12}$$

with $P_0 = 1$ and $P_1 = \lambda - 1$. This may be solved by writing $\lambda = 2\cos\phi$, then

$$\begin{aligned} P_r &= \cos r\phi - [(1-\cos\phi/\sin\phi]\sin r\phi \\ &= (1/\sin\phi)(\sin(r+1)\phi - \sin r\phi) \\ &= (2/\sin\phi)\cos(r+\tfrac{1}{2})\phi \sin(\phi/2). \end{aligned} \tag{5.13}$$

This vanishes when $\phi = \phi^k = (2k-1)\pi/N$, $k = 1,\ldots, N$.

We define

$$P_j^k = (\sin\phi^k)^{-1}(\sin(j+1)\phi^k - \sin j\phi^k), \tag{5.14}$$

so that $P_r^k = P_r = 0$. Then the eigenvector of M_r corresponding to the eigenvalue $\lambda^k = 2\cos\phi^k$ is

$$(P_0^k, P_1^k, \ldots, P_{r-1}^k).$$

Now it is clear that the combinations of generators that diagonalize the basis are

$$S_q^k = \sum_{j=0}^{r-1} P_j^k J_{[1/2(q+2j-1),1/2(q-2j+1)]}. \tag{5.15}$$

The Cartan elements are $H^\alpha = J_{[\alpha,-\alpha]}$.

Working out the commutation relations in this basis,

$$\begin{aligned} [H^\alpha, S_q^k] &= \left[J_{[\alpha,N-\alpha]}, \sum_{j=0}^{r-1} P_j^k J_{[(1/2)(q+2j-1)(1/2)(q-2j+1)]}\right] \\ &= \sum_{j=0}^{r-1} (\sin\phi^k)^{-1}(\sin(j+1)\phi^k - \sin j\phi^k)[J_{[\alpha,-\alpha]}, J_{[(1/2)(q+2j-1),(1/2)(q-2j+1)]}] \\ &= \frac{-2i\sin(\pi/N)\alpha q}{\sin\phi^k} \sum_{j=0}^{r-1} (\sin(j+1)\phi^k - \sin j\phi^k) \\ &\quad \times (J_{[(1/2)(q+2j-1)+\alpha,(1/2)(q-2j+1)-\alpha]} - J_{[(1/2)(q-2j+1)+\alpha(1/2)(q+2j-1)-\alpha]}). \end{aligned}$$

Now consider the coefficient of $J_{[(1/2)(q+2l-1),(1/2)(q-2l+1)]}$. This is

$$\begin{aligned} &\frac{-2i\sin(\pi/N)\alpha q}{\sin\phi^k}(\sin(l-\alpha)\phi^k - \sin(l-\alpha-1)\phi^k \\ &\quad + \sin(l+\alpha)\phi^k - \sin(l+a-1)\phi^k) \\ &= \frac{-4i\sin(\pi/N)\alpha q\cos\alpha\phi^k}{\sin\phi^k}(\sin l\phi^k - \sin(l-1)\phi^k), \end{aligned}$$

therefore,

$$\begin{aligned} [H^\alpha, S_q^k] &= \frac{-4i\sin(\pi/N)\alpha q\cos\alpha\phi^k}{\sin\phi^k} \\ &\quad \times \sum_{l=0}^{r-1} P_l^k J_{[1/2(q+2l-1),1/2(q-2l+1)]} \\ &= \frac{-4i\sin(\pi/N)\alpha q\cos\alpha\phi^k}{\sin\phi^k} S_q^k, \end{aligned}$$

showing that this is the diagonal basis.

In a recent paper, a similar identification of some subalgebras has been made by Pope and Romans.[14] They introduce a basis for SO(N) and another for USp(N) which are extensions of the 't Hooft[3] basis to make the identification. Furthermore, they identify the infinite limits of these subalgebras with the group of diffeomorphisms acting on two dimensional manifolds with different topology from the torus; in the one case a Klein bottle, in the other a projective plane.

VI. TRIANGULAR LATTICES

As was mentioned in the Introduction, it is possible to realize similar infinite algebras on other lattices. In this section we report results for triangular lattices.

It is convenient to choose a system of barycentric coordinates, and index K by three integers m_1, m_2, m_3, where $m_1 + m_2 + m_3 = 0$. (Barycentric coordinates measure the perpendicular distances of any point from the edges of the fundamental reference triangle, as in the Dalitz plot.)

For this case, the relations are

$$[K_{\mathbf{m}}, K_{\mathbf{n}}] = \sin(k\mathbf{u}\cdot(\mathbf{m}\times\mathbf{n}))K_{\mathbf{m}+\mathbf{n}}, \tag{6.1}$$

where $k = \pi/N$, and $\mathbf{u}$ is the vector (1,1,1). As before, we find finite algebras by identifying generators at lattice points equivalent modulo N in each index. When $N \equiv 0 \bmod 3$ the generators whose indices are congruent modulo N all disconnect into U(1)'s. This leaves a hexagonal lattice, and the algebras obtained are $U(N/3)$[6]. When $N \not\equiv 0 \bmod 3$, the fundamental lattice vectors of points reduced mod N contains only one disconnected member, (0,0,0), and the remaining $N^2 - 1$ points are associated with generators which close on SU(N). This situation is parallel with that for the square lattice.

VII. LARGE N LIMITS, AND SU(∞) YANG-MILLS

The two-index SU(N) basis considered here has a particularly simple large N limit. As N increases, the fundamental $N \times N$ cell covers the entire index lattice; the operators $\mathscr{K}$ are supplanted by the K's and, in turn, since $k \to 0$, by the operators L of (1.7).

More directly, it is immediately evident by inspection that, as $N \to \infty$, the SU(N) algebra (5.6) goes over to the centerless algebra (1.7) of SDiff(T^2) through the identification:

$$(iN/2\pi)J_{\mathbf{m}} \to L_{\mathbf{m}}. \tag{7.1}$$

An identification of this type was first noted by Hoppe[7] in the context of membrane physics: He connected the infinite N limit of the SU(N) algebra in a special basis to that of SDiff(S^2), i.e., the infinitesimal symplectic diffeomor-

phisms in the sphere basis. A discussion of the group topology of SU(N), or SDiff(T^2) vs SDiff(S^2), or other two-dimensional manifolds for that matter[14] goes beyond the scope of this type of local analysis. In view of the SO(N) subalgebras described in (5.8) we may also simply identify the SO(∞) subalgebra with the Poisson bracket subalgebra whose shift potentials f are odd under interchange of x with p, corresponding to Hamiltonians which evolve even functions to even ones, and odd to odd ones. Likewise, USp(∞) is generated by shift potentials of the form $\exp(im_1x)\sin(m_2p - m_1\pi/2)$, i.e., toroidal phase-space Hamiltonians odd under $p \to -p$, $x \to x + \pi$.

Floratos, Iliopoulos, and Tiktopoulos[15] utilized Hoppe's identification to take the limit of SU(N) gauge theory. Their results are immediately reproduced without ambiguity, again by inspection, on the basis of the orthogonality condition dictated by (5.4) and (5.5);

$$\mathrm{Tr}\, J_{\mathbf{m}}J_{\mathbf{n}} = N\delta_{\mathbf{m}+\mathbf{n},0} \to \mathrm{Tr}\, L_{\mathbf{m}}L_{\mathbf{n}} = -[N^3/(2\pi)^2]\delta_{\mathbf{m}+\mathbf{n},0}. \tag{7.2}$$

As a result, for a gauge field A_μ in an SU(N) matrix normalization with trace 1, the analog of

$$A_\mu \equiv (1/\sqrt{N})A_\mu^{\mathbf{m}}J_{\mathbf{m}} \to [2\pi/(iN^{3/2})]A_\mu^{\mathbf{m}}L_{\mathbf{m}} = \tilde{A}_\mu^{\mathbf{m}}L_{\mathbf{m}}, \tag{7.3}$$

where summation over repeated $\mathbf{m}$'s is implied, and the gauge field $\tilde{A}_\mu^{\mathbf{m}}$ is as defined above. As $N \to \infty$, the indices $\mathbf{m}$ cover the entire integer lattice, and hence we may define

$$a_\mu^{(x,p)} \equiv -\sum_{\mathbf{m}} \tilde{A}_\mu^{\mathbf{m}} e^{i(m_1x + m_2p)}. \tag{7.4}$$

By Eq. (3.2),

$$[A_\mu, A_\nu] \to [L_{a_\mu}, L_{a_\nu}] = L_{\{a_\mu, a_\nu\}}. \tag{7.5}$$

Hence, by virtue of the linearity of L in its arguments,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \to L_{f_{\mu\nu}}, \tag{7.6}$$

$$f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu + \{a_\mu, a_\nu\}. \tag{7.7}$$

The group trace defining the Yang–Mills Lagrangian density is thus

$$\begin{aligned}\mathrm{Tr}\, F_{\mu\nu}F_{\mu\nu} &\to -N^3/(2\pi)^2 \tilde{F}_{\mu\nu}^{\mathbf{m}}\tilde{F}_{\mu\nu}^{-\mathbf{m}} \\ &= -N^3/(16\pi^4)\int dx\, dp \\ &\quad \times \sum_{\substack{m_1,m_2 \\ n_1,n_2}} e^{ix(m_1+m_2)+ip(m_2+n_2)} \\ &\quad \times \tilde{F}_{\mu\nu}^{m_1,m_2}\tilde{F}_{\mu\nu}^{n_1,n_2}.\end{aligned} \tag{7.8}$$

Thus, in the limit of the gauge theory, the group indices are surface (torus) coordinates, and the fields are rescaled Fourier transforms of the original SU(N) fields; the group composition rule for them is given by the Poisson bracket, and the trace by surface integration. For SO(∞) and USp(∞) the a_μ's must have the above-mentioned symmetries.

Now note that an intriguing connection to strings emerges, for the first time *directly at the level of the action*: for gauge fields independent of x^μ (e.g., vacuum configurations), this Lagrangian density reduces to $\{a_\mu, a_\nu\}\{a_\mu, a_\nu\}$, the quadratic Schild–Eguchi action density for strings,[16] where the a_μ now serve as string variables, and the surface serves as the world-sheet. This means that the classical vacuum states of SU(∞) Yang–Mills are equivalent to the configurations of the classical string. Whether a superstring follows analogously from super-Yang–Mills is an interesting question.

The Lagrangian (7.8) with the sine bracket supplanting the Poisson bracket is also a gauge invariant theory, provided that the gauge transformation also involves the sine instead of the Poisson bracket,

$$\delta a_\mu = \partial_\mu \Lambda - \sin\{\Lambda, a_\mu\}, \tag{7.9}$$

and hence, by virtue of the Jacobi identity,

$$\delta f_{\mu\nu} = -\sin\{\Lambda, f_{\mu\nu}\}. \tag{7.10}$$

It then follows that

$$\delta \int dx\, dp\, f_{\mu\nu} f_{\mu\nu} = -2\int dx\, dp\, f_{\mu\nu} \sin\{\Lambda, f_{\mu\nu}\} = 0. \tag{7.11}$$

At the moment, however, it is not clear what system is described by the corresponding space-time-independent Lagrangian density $\sin\{a_\mu, a_\nu\}\sin\{a_\mu, a_\nu\}$.

ACKNOWLEDGMENTS

The work of P.F. is supported by a Durham University Research Studentship. The work of C.K.Z. is supported by the U. S. Department of Energy, Division of High Energy Physics, Contract W-31-109-ENG-38.

[1] D. B. Fairlie, P. Fletcher, and C. K. Zachos, Phys. Lett. B **218**, 203 (1989); D. B. Fairlie and C. K. Zachos, Phys. Lett. B **224**, 101 (1989); C. K. Zachos, 'Hamiltonian Flows, SU(∞), SO(∞), USp(∞) and Strings', ANL-HEP-CP-89-55 (1989).

[2] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1931).

[3] J. Schwinger, Proc. Natl. Acad. Sci. USA **46**, 570 (1960); G.'t Hooft, Nucl. Phys. B **138**, 1 (1978); A. Belavin, JETP Lett. **32**, 169 (1980); G.'t Hooft, Commun. Math. Phys. **81**, 267 (1981).

[4] V. Rittenberg and D. Wyler, Nucl. Phys. B **139**, 189 (1978); K. Fabricius and C. P. Korthals–Altes, Nucl. Phys. B **240**, 237 (1984); A. Balinskii and S. Novikov, Sov. Math. Dokl. **32**, 228 (1985); M. Lüscher and P. Weisz, Nucl. Phys. B **266**, 309 (1986); R. Balian and C. Itzykson, C. R. Acad. Sci. Paris **303**, 773 (1986); E. Floratos, Phys. Lett. B **228**, 335 (1989).

[5] J. Patera and H. Zassenhaus, J. Math. Phys. **29**, 665 (1988).

[6] V. Arnold, Ann. Inst. Fourier **XVI**(1), 319 (1966); V. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978). Appendix 2.K, p. 339; D. Fuks, Funct. Anal. Appl. **19**, 305 (1985); E. Floratos and J. Iliopoulos, Phys. Lett. B **201**, 237 (1988); I. Bakas, "The Large N Limit of Extended Conformal Symmetries," PP#89-164.

[7] J. Hoppe, Ph.D. thesis M.I.T., 1982; J. Hoppe, *Constraints Theory and Relativistic Dynamics*, Florence, 1986, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987) p. 267; J. Hoppe, Phys. Lett. B **215**, 706 (1988); J. Hoppe, J. Mod. Phys. A **4**, 8235 (1989).

[8] K. M. Case and A. Monge, J. Math. Phys. **30**, 1250 (1989).

[9] J. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).

[10] G. Baker, Phys. Rev. **109**, 2198 (1958); D. B. Fairlie, Proc. Cambridge Philos. Soc. **60**, 581 (1964).

[11] E. P. Wigner, Phys. Rev. **40**, 749 (1932).

[12] C. N. Pope and K. S. Stelle, Phys. Lett. B **226**, 257 (1989).

[13] M. V. Saveliev and A. M. Vershik, Commun. Math. Phys. **126**, 367 (1989).

[14] C. N. Pope and L. J. Romans, "Local area-preserving algebras for two-dimensional surfaces," USC-89/HEP.

[15] E. Floratos, J. Iliopoulos, and G. Tiktopoulos, Phys. Lett. B **217**, 285 (1989).

[16] A. Schild, Phys. Rev. D **16**, 1722 (1977); T. Eguchi, Phys. Rev. Lett. **44**, 126 (1980); J.-L. Gervais and A. Neveu, Nucl. Phys. B **192**, 463 (1981); Y. Hosotani, Phys. Rev. Lett. **55**, 1719 (1985).

Nonlinear equations invariant under the Poincaré, similitude, and conformal groups in two-dimensional space-time

G. Rideau
L.P.T.M., Université Paris 7, T.C. 3e Etage, 2, place Jussieu, 75251 Paris Cedex 05, France

P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, C. P. 6128A, Montréal, Québec H3C 3J7, Canada

(Received 13 July 1989; accepted for publication 15 November 1989)

All realizations of the Lie algebras p(1,1), sim(1,1), and conf(1,1) are classified under the action of the group of local diffeomorphisms of $\mathbb{R}^3$. The result is used to obtain all second-order scalar differential equations, invariant under the corresponding Poincaré, similitude, and conformal groups. The invariant equations are, in general, nonlinear, and the requirement of linearity turns out to be very restrictive. Group invariant solutions of some of the conformally invariant equations are obtained either by quadratures or by a linearizing transformation.

I. INTRODUCTION

A useful tool for obtaining exact analytic particular solutions of nonlinear partial differential equations is the method of symmetry reduction.[1-3] This method has recently been used to obtain solutions of various nonlinear multidimensional Klein–Gordon equations, invariant under Poincaré, similitude, or conformal groups.[4,5] The same method has yielded a large number of quite unexpected solutions of a quintic and cubic nonlinear Schrödinger equation in $3+1$ dimensions.[6,7]

In this paper, we turn the question around and ask, "What is the most general equation invariant under a given local Lie group?" More specifically, in this paper, we concentrate on second-order scalar equations, invariant under the Poincaré group P(1,1), the similitude group Sim(1,1), and the conformal group Conf(1,1) of $(1+1)$-dimensional Minkowski space. We request that these groups act as local point symmetries of the obtained equations and that their action on the space-time variables be the standard geometrical one. The action on the dependent variable $F(x,t)$ is allowed to be general, restricted only by the requirement that the corresponding Lie algebra of vector fields satisfy the correct commutation relations.

Our motivation is multifold. First, we wish to determine how restricting fundamental symmetry requirements are, i.e., how general can the field equations of a classical relativistic field theory be.

If we restrict ourselves to linear equations, then Poincaré invariance alone leads to the linear Klein–Gordon equation, or more generally, in $3+1$ dimensions to the Bargmann–Wigner equations.[8] The intimate relationship between the theory of irreducible unitary linear representations of the Poincaré group and the solutions of relativistically invariant linear equations is well known.[8-10]

A sizable literature on nonlinear group representations exists,[11-14] in particular on representations of the Poincaré group. This theory should be relevant for linearizing, or otherwise solving, nonlinear invariant equations. The second part of our motivation is to derive physically interesting nonlinear equations to which this nonlinear group representation theory could be applied.

Finally, all closed continuous subgroups of Poincaré, similitude, and conformal groups of low-dimensional Minkowski space are known.[15,16] They can hence be applied to obtain solutions of any equations invariant under these groups.

This present paper is organized as follows. In Sec. II we obtain all realizations of the Lie algebras p(1,1), sim(1,1), and conf(1,1) by vector fields in three variables (x,t,F). We classify these realizations up to arbitrary diffeomorphisms $(x,t,F)\to(x',t',F')$. We also discuss the prolongations of these vector fields, up to second order, since we are, at this stage, interested in first- and second-order equations. In Sec. III, we obtain the differential invariants of the prolonged group action and hence also the invariant equations. We restrict ourselves to the case of fiber preserving transformations. Section IV is devoted to the relation between some of the obtained invariant equations and the theory of nonlinear group representations. In Sec. V, we apply the symmetry reduction method for getting solutions of some of the conformally invariant equations obtained in Sec. III.

Throughout this paper, when invoking conformal invariance, we restrict ourselves to the finite-dimensional Lie group conf(1,1) $\sim$ O(2,2) of $(1+1)$-dimensional Minkowski space. "Invariance" is always meant in the "strong" sense, i.e., if a second-order equation $\Delta=0$ is invariant under some one-parameter group $\exp \lambda X$, then this equation is annihilated by the second-order prolongation $\mathrm{pr}^{(2)}X$ of X everywhere, not only on the solution set of Δ.

II. REALIZATION OF THE LIE ALGEBRAS

A. The Lie algebras p(1,1), sim(1,1), and conf(1,1) realized by vector fields

The first step in the derivation of invariant equations is to realize the Lie algebra of the assumed symmetry group in terms of vector fields on the space $X\otimes U$ of independent and dependent variables. In our case, X is the two-dimensional

 0022-2488/90/051095-11$03.00

Minkowski space with coordinates x,t and U is the space of real scalar functions $F(x,t)$.

The vector fields will all have the form

$$V=\xi(x,t,F)\partial_x+\tau(x,t,F)\partial_t+\phi(x,t,F)\partial_F. \tag{2.1}$$

We shall denote the generators of translations, Lorentz boosts, dilations, and proper conformal transformations P_μ, K, D, and C_μ $(\mu=0,1)$, respectively. Their commutation relations are the usual ones, namely,

$$[P_0,K]=P_1,\quad [P_1,K]=P_0,\quad [P_\mu,D]=P_\mu,$$
$$[C_0,K]=C_1,\quad [C_1,K]=C_0,\quad [C_\mu,D]=-C_\mu,$$
$$[P_\mu,C_\nu]=2(g_{\mu\nu}D-\varepsilon_{\mu\nu}K),$$
$$g_{00}=-g_{11}=1,\quad \varepsilon_{01}=-\varepsilon_{10}=1,\quad \mu=0,1, \tag{2.2}$$

where all remaining commutators and all unlisted components of the metric tensor $g_{\mu\nu}$ and antisymmetric tensor $\varepsilon_{\mu\nu}$ are zero.

We shall classify realizations of the conformal Lie algebra conf(1,1) and its similitude sim(1,1) and Poincaré p(1,1) subalgebras up to diffeomorphisms, i.e., up to arbitrary smooth invertible changes of variables:

$$x'=f(x,t,F),\quad t'=g(x,t,F),\quad F'=h(x,t,F). \tag{2.3}$$

We start by realizing the translations P_μ. According to Lie's theorem on the straightenning out of vector fields (see, e.g., Ref. 1), we can always transform, say, P_0 into $P_0=\partial_t$. Using a further transformation leaving P_0 unchanged, we can transform P_1 into $P_1=\partial_x$. From the commutation relations (2.2), we find that K must have the form $K=(x+\eta_0(F))\partial_t+(t+\eta_1(F))\partial_x+A(F)\partial_F$. Performing a transformation of the form $x'=x+\alpha(F)$, $t'=t+\beta(F)$, $F'=\gamma(F)$ with appropriate choice of α, β, and γ, we find that two classes of realizations of p(1,1) exist. They can be represented by

$$L_1:\quad \{P_0=\partial_t,\quad P_1=\partial_x,\quad K=x\partial_t+t\partial_x\}, \tag{2.4}$$
$$L_2:\quad \{P_0=\partial_t,\quad P_1=\partial_x,\quad K=x\partial_t+t\partial_x+F\partial_F\}. \tag{2.5}$$

Each of these realizations can be extended to realizations of the similitude algebra sim(1,1) by adding a dilation operator D. Its form is dictated by the commutation relations (2.2) and it can be simplified by transformations (2.3), leaving the already standardized algebras, L_1 or L_2, invariant. The realization L_1 leads to two possibilities, represented by

$$S_1:\quad \{\{L_1\},\quad D=x\partial_x+t\partial_t\}, \tag{2.6}$$
$$S_2:\quad \{\{L_1\},\quad D=x\partial_x+t\partial_t+F\partial_F\}. \tag{2.7}$$

The realization L_2 leads to a family of realizations:

$$S_3^{\lambda,a,b}:\quad \{\{L_2\},\ D=(x+aF-b/F)\partial_x+(t+aF+b/F)\partial_t+\lambda F\partial_F\},$$
$$\lambda\in\mathbb{R},\ (a,b)=(0,0),(0,1),(1,0),(1,1)\ \text{or}\ (1,-1). \tag{2.8}$$

If we wish the corresponding group transformations to be fiber preserving, i.e., such that the new space-time variables depend only on the old ones (and not on F), we must take $(a,b)=(0,0)$ in (2.8).

The obtained realizations of sim(1,1) can be further extended to conf(1,1) by adding the proper conformal transformations C_0 and C_1. Once the commutation relations (2.2) are satisfied, further transformations respecting the form of the sim(1,1) realizations can be performed. The calculations are rather lengthy, though entirely straightforward. They were partly performed using a MACSYMA routine. We present the results only.

Starting from the realization S_1, we obtain a single realization of conf(1,1). The corresponding transformations are fiber preserving:

$$K_1:\quad \{\{S_1\},\ C_0=(x^2+t^2)\partial_t+2xt\partial_x,\quad C_1=-(x^2+t^2)\partial_x-2xt\partial_t\}. \tag{2.9}$$

Starting from S_2, we obtain three inequivalent realizations:

$$K_2^a:\quad \{\{S_2\},\ C_0=(x^2+t^2+aF^2)\partial_t+2xt\partial_x+2tF\partial_F,\quad C_1=-(x^2+t^2-aF^2)\partial_x-2xt\partial_t-2xF\partial_F\},\quad a=0,+1,-1. \tag{2.10}$$

Fiber preserving transformations are obtained for $a=0$ only.

The realization $S_3^{\lambda,a,b}$ leads to three types of realizations of conf(1,1).

For $\lambda\neq\pm1$, we obtain

$$K_3^{\lambda ab}=\{\{S_3^{\lambda ab}\},\quad C_1=-[K,C_0],$$
$$C_0=[x^2+t^2+2t(aF+bF^{-1})-2\lambda a^2(\lambda-1)^{-2}F^2+2\lambda b^2(\lambda+1)^{-2}F^{-2}-4ab(\lambda^2-1)^{-1}]\partial_t$$
$$+[2xt+2t(aF-bF^{-1})-2\lambda a^2(\lambda-1)^{-2}F^2-2\lambda b^2(\lambda+1)^{-2}F^{-2}+4ab(\lambda^2-1)^{-1}]\partial_x$$
$$+[2F(x+\lambda t)-2(\lambda+1)(\lambda-1)^{-1}aF^2+2(\lambda-1)(\lambda+1)^{-1}b]\partial_F\}, \tag{2.11}$$

with (λ,a,b) as in (2.8). The corresponding transformations are fiber preserving if and only if $a=b=0$. In this case $\lambda\in\mathbb{R}$ without restriction.

For $\lambda=+1$, we obtain

$$K_4^{b,c,e}=\{\{S_3^{10b}\},\quad C_1=-[K,C_0],$$
$$C_0=[x^2+t^2+2btF^{-1}+cF^2+\tfrac{1}{2}b^2F^{-2}+\tfrac{1}{2}be]\partial_t$$
$$+[2xt-2btF^{-1}+cF^2-\tfrac{1}{2}b^2F^{-2}-\tfrac{1}{2}be]\partial_x+[2(x+t)F+eF^2]\partial_F\},$$
$$(b,c,e)=(0,\kappa,0),(0,\kappa,1),(1,\kappa,e),\quad \kappa=0,+1,-1,\quad e\in\mathbb{R}. \tag{2.12}$$

Transformations (2.12) are fiber preserving iff $b = c = 0$, in which case we can put $e = 1$.

For $\lambda = -1$, we obtain

$$K_5^{a,c,e} = \{\{S_3^{-1,a,0}\},\quad C_1 = -[K,C_0],$$
$$C_0 = [x^2 + t^2 + 2atF + \tfrac{1}{2}a^2F^2 + cF^{-2} - \tfrac{1}{4}ae]\partial_t + [2xt + 2atF + \tfrac{1}{2}a^2F^2 - cF^{-2} + \tfrac{1}{4}ae]\partial_x + [2(x-t)F - \tfrac{1}{2}e]\partial_F\},$$
$$(a,c,e) = (0,\kappa,0),(0,\kappa,1),(1,\kappa,e),\quad \kappa = 0,+1,-1,\quad e\in\mathbf{R}. \tag{2.13}$$

Transformations (2.13) are fiber preserving iff $a = c = 0$, in which case we can take $e = 1$.

The distinct fiber preserving realizations of conf(1,1) are K_1, K_2^0, $K_3^{\lambda 00}$, K_4^{001}, and K_5^{001}.

B. Prolongation of the vector fields

In order to construct invariant second-order partial differential equations, we need to know how the considered group acts on first and second derivatives F_x, F_t, F_{xx}, F_{xt}, and F_{tt}, once we know its action on (x,t,F). Since our entire approach is infinitesimal, it is sufficient for us to know the prolonged infinitesimal group action. In other words, we need the second prolongations of all the vector fields obtained above.

The general formula for the k th prolongation of a vector field is given, e.g., by Olver.[1] We are interested in the second prolongation of the vector field V given in (2.1) and we have

$$\mathrm{pr}^2 V = V + \phi^x\,\partial_{F_x} + \phi^t\,\partial_{F_t} + \phi^{xx}\,\partial_{F_{xx}} + \phi^{xt}\,\partial F_{xt} + \phi^{tt}\,\partial_{F_{tt}}; \tag{2.14a}$$

$$\phi^x = D_x(\phi - \xi F_x - \tau F_t) + \xi F_{xx} + \tau F_{xt},$$
$$\phi^t = D_t(\phi - \xi F_x - \tau F_t) + \xi F_{tx} + \tau F_{tt}; \tag{2.14b}$$

$$\phi^{xx} = D_x\phi^x - (D_x\xi)F_{xx} - (D_x\tau)F_{xt},$$
$$\phi^{xt} = D_t\phi^x - (D_t\xi)F_{xx} - (D_t\tau)F_{xt}, \tag{2.14c}$$
$$\phi^{tt} = D_t\phi^t - (D_t\xi)F_{tx} - (D_t\tau)F_{tt};$$

where D_x and D_t are total derivatives.

Using Eqs. (2.14), we calculate the second prolongations of all the vector fields P_μ, C_μ, K, and D for each realization obtained above. We obviously have

$$\mathrm{pr}^{(2)} P_\mu = P_\mu,\quad \mu = 0,1. \tag{2.15}$$

The other prolongations are more complicated and we do not reproduce them here. In each case, we shall need $\mathrm{pr}^{(2)} K$, $\mathrm{pr}^{(2)} D$, and $\mathrm{pr}^{(2)} C_0$. The expression for $\mathrm{pr}^{(2)} C_1$ is not needed, since we have $\mathrm{pr}^{(2)} C_1 = -[\mathrm{pr}^{(2)} K, \mathrm{pr}^{(2)} C_0]$ and hence any invariant of $\exp C_0$ and $\exp K$ is automatically invariant under $\exp C_1$.

III. DIFFERENTIAL INVARIANTS AND INVARIANT EQUATIONS

A. General comments

The procedure used to obtain the invariant equations is standard. Thus let X_a, $a = 1,\ldots,N$, be a basis for the Lie algebra L of the symmetry group G, acting on the space $X \otimes U$. In our case, $X \otimes U$ is the space $\{x,t,F\}$ and all X_a have the form (2.1). The invariant equation will have the form

$$\Delta(x,t,F,F_x,F_t,F_{xx},F_{xt},F_{tt}) = 0, \tag{3.1}$$

where the function Δ satisfies

$$\mathrm{pr}^{(2)} X_a \cdot \Delta = 0,\quad \forall\, a. \tag{3.2}$$

Thus all we have to do is to find the characteristics for the set of equations (3.2), in which all the arguments in (3.1) are viewed as independent variables. The characteristics will provide us with a set of elementary invariants $I_k(x,t,F,F_\mu,F_{\mu\nu})$ $(\mu,\nu = x,t)$, and the invariant equation is

$$H(I_1,\ldots,I_N) = 0. \tag{3.3}$$

We shall look for p(1,1), sim(1,1), and conf(1,1) invariants for each realization of the corresponding Lie algebras found in Sec. II. The number of variables in (3.1) and (3.2) is 8. The algebras p(1,1) and sim(1,1) are solvable; the generic orbits of the corresponding prolonged group action are three and four dimensional, respectively. Hence we shall obtain five and four functionally independent invariants, respectively. The group $\mathrm{conf}(1,1) \sim \mathrm{O}(2,2) \sim \mathrm{O}(2,1) \times \mathrm{O}(2,1)$ is semisimple. The genric orbits of the prolonged group action are six dimensional. We shall hence obtain two invariants.

In view of (2.15) all invariants will be independent of x and t.

B. Poincaré invariant equations

1. Realization L_1

Consider first the more "natural" usual realization L_1 of p(1,1), given in (2.4). The second prolongation of the boost K is

$$\mathrm{pr}^{(2)} K = x\,\partial_t + t\,\partial_x - F_t\,\partial_{F_x} - F_x\,\partial F_t - 2F_{xt}(\partial_{F_{xx}} + \partial_{F_{tt}}) - (F_{tt} + F_{xx})\partial_{F_{xt}}. \tag{3.4}$$

Translational invariance [see (2.15)] implies that Δ in (3.1) does not depend on x and t. Solving (3.2) for $\mathrm{pr}^{(2)} K$ given in (3.4), we obtain five elementary invariants, namely,

$$I_1 = F,\quad I_2 = F_t^2 - F_x^2,\quad I_3 = F_{tt} - F_{xx},$$
$$I_4 = (F_t - F_x)^2(F_{xx} + 2F_{xt} + F_{tt}), \tag{3.5}$$
$$I_5 = (F_t + F_x)^2(F_{xx} - 2F_{xt} + F_{tt}).$$

The most general P(1,1) invariant equation is hence

$$H(I_1,I_2,I_3,I_4,I_5) = 0. \tag{3.6}$$

Notice that (3.6) includes the linear Klein–Gordon equation

$$F_{tt} - F_{xx} = mF,\quad m\in\mathbf{R}, \tag{3.7}$$

where the mass m may, or may not, be zero. Other linear equations, not involving any parameters, are obtained by setting $I_4 = 0$ or $I_5 = 0$, yielding independently

$$F_t - F_x = 0,\quad F_{xx} + 2F_{xt} + F_{tt} = 0,$$
$$F_t + F_x = 0,\quad F_{xx} - 2F_{xt} + F_{tt} = 0. \tag{3.8}$$

Notice that we also obtain a reasonably general first-order P(1,1) invariant equation, namely,

$$F_t^2 - F_x^2 = h(F), \tag{3.9}$$

where h is an arbitrary function of F. In particular, (3.9) includes the physically interesting eikonal, or Hamilton–Jacobi, equation

$$F_t^2 - F_x^2 = \mu F, \quad \mu \in \mathbb{R}. \tag{3.10}$$

2. Realization L_2

The realization L_2 is less standard than L_1. Indeed, in this realization a scalar function $F(x,t)$ transforms under Lorentz transformations with a certain weight, i.e.,

$$F'(x',t') = (\exp\lambda)F(x'\cosh\lambda - t'\sinh\lambda, \\ -x'\sinh\lambda + t'\cosh\lambda). \tag{3.11}$$

Skipping all details, we present the five elementary invariants for the realization L_2 of (2.5), namely,

$$I_1 = F_x + F_t, \quad I_2 = (F_t - F_x)F^{-2}, \\ I_3 = (F_{tt} - F_{xx})F^{-1}, \tag{3.12} \\ I_4 = (F_{xx} + 2F_{xt} + F_{tt})F, \quad I_5 = (F_{xx} - 2F_{xt} + F_{tt})F^{-3}.$$

The invariant equation (3.6) again includes the linear massive or massless Klein–Gordon equation (3.7), obtained by putting $I_3 = m$. As a matter of fact, several linear equations are obtained, namely,

$$I_1 = \mu, \quad I_2 = 0, \quad I_3 = m, \quad I_4 = 0, \quad I_5 = 0. \tag{3.13}$$

A general first-order invariant equation is

$$F_t - F_x = F^2 h(F_x + F_t). \tag{3.14}$$

An eikonal type equation with no free (mass) parameter is obtained by putting $I_1 \cdot I_2 = 0$.

C. Equations invariant under the similitude group sim(1,1)

We now have to deal with three different realizations of sim(1,1), given in (2.6), (2.7), and (2.8). The realizations S_1 and S_2 are the more standard ones, in that they incorporate the L_1 realization of p(1,1), rather than L_2. The realization $S_3^{\lambda ab}$ is fiber preserving only for $a = b = 0$ and we restrict ourselves to this case. In all three cases the additional dilational invariance reduces the number of functionally independent elementary invariants to 4. We skip all details and simply present the invariants.

1. Realization S_1

We have

$$J_1 = F, \quad J_2 = (F_t^2 - F_x^2)^{-1}(F_{tt} - F_{xx}), \\ J_3 = (F_t + F_x)^{-2}(F_{tt} + 2F_{xt} + F_{xx}), \tag{3.15} \\ J_4 = (F_t - F_x)^{-2}(F_{tt} - 2F_{xt} + F_{xx}).$$

Linear invariant equations are

$$J_2 = 0, \quad J_3^{\epsilon} = 0, \; J_4^{\epsilon} = 0, \quad \epsilon = \pm 1, \tag{3.16}$$

where $J_2 = 0$ is the massless Klein–Gordon equation. Similarly $(J_2)^{-1} = 0$ yields an eikonal equation with vanishing right-hand side.

Notice that in realization S_1 the group sim(1,1) acts only on the independent variables, not on the function F.

2. Realization S_2

Since S_2 also contains L_1, the elementary sim(1,1) invariants will be constructed out of $I_1,\ldots,I_5$ of (3.5). We have

$$J_1 = F_t^2 - F_x^2, \quad J_2 = F(F_{tt} - F_{xx}), \\ J_3 = F(F_t - F_x)^2(F_{xx} + 2F_{xt} + F_{tt}), \tag{3.17} \\ J_4 = F(F_t + F_x)^2(F_{xx} - 2F_{xt} + F_{tt}).$$

The massless Klein–Gordon equation is obtained by putting $J_2 = 0$; an eikonal type equation is $J_1 = 0$. Setting individually $J_i = 0$ for J_i as in (3.17) we get a variety of invariant linear equations,

$$F_t - F_x = 0, \quad F_t + F_x = 0, \quad F_{tt} - F_{xx} = 0, \\ F_{xx} + 2F_{xt} + F_{tt} = 0, \quad F_{xx} - 2F_{xt} + F_{tt} = 0. \tag{3.18}$$

To explain the pronounced difference between (3.15) and (3.17), remember that in realization S_2 the dilations act not only on (x,t) but also on the function F.

3. Realization $S_3^{\lambda\infty}$

Calculating $\mathrm{pr}^{(2)} D$ for D as in (2.8) with $a = b = 0$ and applying it to the p(1,1) invariants (3.12), we find it convenient to consider the cases $\lambda \neq 1$ and $\lambda = 1$ separately.

(a) $\lambda \neq 1$: The elementary sim(1,1) invariants in this case are

$$J_1 = (F_t^2 - F_x^2)(F_t + F_x)^{2/(\lambda-1)}F^{-2}, \\ J_2 = (F_{tt} - F_{xx})(F_t + F_x)^{2/(\lambda-1)}F^{-1}, \\ J_3 = (F_{xx} + 2F_{xt} + F_{tt})(F_t + F_x)^{-2}F, \tag{3.19} \\ J_4 = (F_{xx} - 2F_{xt} + F_{tt})(F_x + F_t)^{2(\lambda+1)/(\lambda-1)}F^{-3}.$$

Massless Klein–Gordon and eikonal type equations are obtained by putting $J_1 = 0$ and $J_2 = 0$, respectively.

(b) $\lambda = 1$: The elementary sim(1,1) invariants are

$$J_1 = F_t + F_x, \\ J_2 = (F_{tt} - F_{xx})(F_t - F_x)^{-1}F, \\ J_3 = (F_{xx} + 2F_{xt} + F_{tt})F, \tag{3.20} \\ J_4 = (F_{xx} - 2F_{xt} + F_{tt})(F_t - F_x)^{-2}F.$$

Massless Klein–Gordon and eikonal type equations are obtained as $J_2 = 0$ and $J_2^{-1}J_1 = 0$, respectively; $J_1 = \text{const}$ is a linear first-order invariant equation.

D. Conformally invariant equations

We have five realizations of conf(1,1) to deal with, namely, K_1, K_2^0, $K_3^{\lambda 00}$, K_4^{001}, and K_5^{001}.

1. Realization K_1

The realization of (2.9) acts on space-time only; hence F is one of the conformal invariants. We calculated $\mathrm{pr}^{(2)} C_0$ and let it act on the S_1 invariants (3.15) (since K_1 includes S_1). We have

$$\mathrm{pr}^{(2)} C_0 \cdot J_1 = \mathrm{pr}^{(2)} C_0 \cdot J_2 = 0,$$

$$\mathrm{pr}^{(2)}\, C_0 \cdot J_3 = -4(F_t + F_x)^{-1},$$
$$\mathrm{pr}^{(2)}\, C_0 \cdot J_4 = -4(F_t - F_x)^{-1}.$$

Since we have $C_1 = [C_0, K]$, $\mathrm{pr}^{(2)}\, C_1$ will annihilate any expression annihilated by K and C_0. The conclusion is that two conf(1,1) invariants exist, namely,

$$\Sigma_1 = F, \quad \Sigma_2 = (F_{tt} - F_{xx})(F_t^2 - F_x^2)^{-1}. \tag{3.21}$$

The general conf(1,1) invariant equation for the realization K_1 is hence

$$F_{tt} - F_{xx} = (F_t^2 - F_x^2)h(F), \tag{3.22}$$

where h is an arbitrary function of F. In particular, $h(F) = 0$ yields the massless Klein–Gordon equation. Another limit of (3.22) is the eikonal equation $F_t^2 - F_x^2 = 0$.

2. The realization K_2^0

This realization acts nontrivially on the dependent function F. Since K_2^0 includes S_2, we shall start out from the invariants (3.17) of S_2. The condition $\mathrm{pr}^{(2)}\, C_0 \cdot H(J_1, J_2, J_3, J_4) = 0$ yields

$$F_t\left(\frac{\partial H}{\partial J_1} + \frac{\partial H}{\partial J_2} + \frac{J_3}{J_1}\frac{\partial H}{\partial J_3} + \frac{J_4}{J_1}\frac{\partial H}{\partial J_4}\right) + F_x\left(\frac{J_3}{J_1}\frac{\partial H}{\partial J_3} - \frac{J_4}{J_1}\frac{\partial H}{\partial J_4}\right) = 0. \tag{3.23}$$

Since F_t and F_x cannot be expressed in terms of invariants, the coefficients of these two expressions must vanish independently. Solving the corresponding characteristic equations, we obtain two conf(1,1) invariants.

We choose them to be

$$\Sigma_1 = J_2 - J_1 = (F_{tt} - F_{xx})F - F_t^2 + F_x^2,$$
$$\Sigma_2 = J_3 J_4 J_1^{-2} = [(F_{tt} + F_{xx})^2 - 4F_{xt}^2]F^2, \tag{3.24}$$

and the corresponding invariant equation is

$$H(\Sigma_1, \Sigma_2) = 0, \tag{3.24'}$$

where H is an arbitrary function of its arguments. Notice that (3.25) does not include the Klein–Gordon equation (even for zero mass).

3. Realization $K_3^{\lambda\infty}$

We must consider the cases $\lambda \neq 1$ and $\lambda = 1$ separately, but the final result can be written in a unified manner. We start from the sim(1,1) invariants (3.19) or (3.20), respectively. Proceeding as in the previous case, we find that the equation $\mathrm{pr}^{(2)}\, C_0 \cdot H(J_1, J_2, J_3, J_4) = 0$ provides two independent equations involving the sim(1,1) invariants J_i. The final result consists of two conf(1,1) invariants which, for arbitrary values of λ, can be written as

$$\Sigma_1 = [\lambda(F_t + F_x)^2 - (\lambda + 1)F(F_{xx} + 2F_{xt} + F_{tt})] \times F^{-2(\lambda-1)}[F(F_{tt} - F_{xx}) - (F_t^2 - F_x^2)]^{\lambda-1},$$
$$\Sigma_2 = [(\lambda - 1)(F_{xx} - 2F_{xt} + F_{tt}) - (\lambda - 2)(F_t - F_x)^2] \times F^{2(\lambda-1)}[F(F_{tt} - F_{xx}) - (F_t^2 - F_x^2)]^{-(\lambda+1)}. \tag{3.25}$$

Clearly these two expression simplify considerably for $\lambda = 1$ and $\lambda = -1$. Linear equations are obtained from $\Sigma_1 = 0$ for $\lambda = 0$ or $\lambda = -1$ and from $\Sigma_2 = 0$ for $\lambda = 1$ or $\lambda = 2$.

4. Realization K_4^{001}

The two elementary conf(1,1) invariants are, in this case,

$$\Sigma_1 = [F(F_{tt} - F_{xx}) - 2(F_t^2 - F_x^2) - 4(F_t - F_x)] \times (F_t - F_x)^{-1}(F_t + F_x + 2)^{-1/2},$$
$$\Sigma_2 = [F(F_{tt} + 2F_{xt} + F_{xx}) - 2(F_x + F_t)(F_x + F_t + 2)] \times (F_t + F_x + 2)^{-3/2}. \tag{3.26}$$

Linear equations

$$F_t - F_x = 0, \quad F_t + F_x + 2 = 0$$

are obtained from $(\Sigma_1^{-1}) = 0$, $(\Sigma_2^{-1}) = 0$.

5. Realization K_5^{001}

The two elementary conf(1,1) invariants in this case are

$$\Sigma_1 = (F_{tt} - F_{xx} - 8F(F_t + F_x))^2(F_t + F_x)^{-2} \times (F_t - F_x - 4F^2)^{-1},$$
$$\Sigma_2 = [F_{xx} - 2F_{xt} + F_{tt} - 24F(F_t - F_x) + 64F^3] \times (F_t - F_x - 4F^2)^{-3/2}. \tag{3.27}$$

No Klein–Gordon type equation is obtained.

Concerning conformally invariant equations, we mention that we do not obtain, for any of the realizations, a nonlinear Klein–Gordon equation of the form

$$F_{tt} - F_{xx} = \lambda F^p. \tag{3.28}$$

In an $(n+1)$-dimensional Minkowski space the corresponding equation $\Box F = \lambda F^p$ is conformally invariant for $p = (n+3)/(n-1)$. This relation breaks down for $n = 1$, i.e., the case under consideration.

IV. INVARIANT EQUATIONS AND NONLINEAR REPRESENTATIONS

As mentioned in the Introduction, we can expect to extend to the nonlinear case (at least partially) the well-known relations between linear representations and linear invariant equations.

We do not intend to present here an accomplished theory which seems out of reach presently. Indeed the nonlinear representation theory is far from being as firmly found as the linear one. Consequently, we limit ourselves to few relevant examples concerning the p(1,1) group, where the considered nonlinear representations are given by formal series according to the general method in Ref. 11 (FPS theory). We think that these examples will be sufficient to display how much the nonlinear representations are powerful tools to solve and classify, at least partially, the invariant nonlinear equations.

A. Generalities

Let us explain briefly the main features of FPS theory, referring to the original articles for further details.[11–14]

Let G be a group, $U(g)$, $g \in G$, a linear representation of G in some linear topological space E. We denote by $E^{\hat{\otimes} n}$ the nth symmetrical projective tensor power of E. For $f \in E$, or $A \in \mathcal{L}(E)$, $f^{\hat{\otimes} n}$ and $A^{\hat{\otimes} n}$ will denote their nth projective tensor power, respectively: $f^{\hat{\otimes} n} \in E^{\hat{\otimes} n}$, $A^{\hat{\otimes} n} \in \mathcal{L}(E^{\hat{\otimes} n})$.

Then to $f\in E$, $g\in G$, we associate the mapping

$$f\to U(g)f+\sum_{n\geq 2} S_n(g)(U(g)f)^{\hat{\otimes} n}, \tag{4.1}$$

where $S_n(g)$ is a linear mapping of $E^{\hat{\otimes} n}$ into E and the expansion is understood in the sense of a formal series. The mapping (4.1) is a formal representation of G if the $S_n(g)$ verify the following recurrent set of equations:

$$\begin{aligned} S_n(gg') - S_n(g) - U(g)S_n(g')U(g^{-1})^{\hat{\otimes} n} &= \sum_{q=2}^{n-1} S_q(g)U(g)^{\hat{\otimes} q} \\ &\times \sum_{l_1+\cdots+l_q=n} S_{l_1}(g')\hat{\otimes}\cdots\hat{\otimes} S_{l_q}(g')\sigma_n U(g^{-1})^{\hat{\otimes} n}, \end{aligned} \tag{4.2}$$

where σ_n is the usual symmetrization operator. By convention, $S_1(g)=I$ and the right-hand side of (4.2) is zero when $n=2$.

Two formal representations are equivalent if there exists a formal invertible transformation:

$$f\to\sum_1^\infty t_n f^{\hat{\otimes} n}, \quad t_1 \text{ invertible}, \tag{4.3}$$

transforming one representation into the other. As usual, we shall identify two equivalent representations.

It is noteworthy that the nonlinear representation defined by (4.1) and (4.2) can be seen as a deformation of the linear representation $U(g)$. This means we assume implicitly the existence of some parameter by the cancellation of which the representation becomes linear.

B. Application to p(1,1)

In the following (a_+,a_-,θ) is any element of P(1,1), where θ is the parameter of the hyperbolic rotation and, if a_0, a_1 are the time and space components of the translation, then $a_\pm = a_0 \pm a_1$.

The composition law is written

$$\begin{aligned}(a_+,a_-,\theta)(a'_+,a'_-,\theta') &= (a_+ + a'_+\exp\theta, a_- + a'_-\exp(-\theta), \theta+\theta').\end{aligned} \tag{4.4}$$

In the space $\mathscr{D}(\mathbb{R})$ of C^∞ functions on $\mathbb{R}$ with compact support, we define the following representations:

$$U_\pm(a_+,a_-,\theta)f(k) = \exp(ika_\pm)f(k\exp(\pm\theta)), \quad f(k)\in\mathscr{D}(\mathbb{R}). \tag{4.5}$$

Let F be an auxiliary Fock space with annihilation and creation operators $A(k), A^*(k)$ with the usual commutation relations:

$$[A(k),A^*(l)] = \delta(k-l).$$

Then let us consider the mapping

$$f(k)\to\langle 0|A(k)\exp(a_\mp Z)FU_\pm(a_+,a_-,\theta)\exp f\rangle, \tag{4.6}$$

where $\exp f\rangle$ is the coherent vector

$$\exp\int dk f(k)A^*(k)|0\rangle,$$

$FU_\pm(a_+,a_-,\theta)$ is the extension to F of $U_\pm(a_+,a_-,\theta)$, and the operator Z is written

$$Z=\int dk\, A^*(k)Z(k), \tag{4.7}$$

with $Z(k)$ polynomial in the annihilation operators:

$$Z(k)=\sum_n\int Z_n(k;k_1,\ldots,k_n)\prod_1^n A(k_i)dk_i, \tag{4.8}$$

where the kernel $Z_n(k;k_1,\ldots,k_n)$ is in $\mathscr{D}'(\mathbb{R}^n)$ for each k. Let us denote $\exp(a_\mp Z)FU_\pm(a_+,a_-,\theta)$ by $V_\pm(a_+,a_-,\theta)$.

Proposition 1: The mapping (4.6) provides a nonlinear representation of p(1,1) iff $V_\pm(a_+,a_-,\theta)$ is a (linear) representation of p(1,1) in F.

The proof results from the following lemma.

Lemma: If Z is given by (4.7) and (4.8), we have the factorization property:

$$\langle 0|A(k_1)\cdots A(k_n)V_\pm(a_+,a_-,\theta)\exp f\rangle = \prod_{i=1}^n\langle 0|A(k_i)V_\pm(a_+,a_-,\theta)\exp f\rangle. \tag{4.9}$$

Proof: By recurrence after we have observed that

(i) $\exp(-a_\mp Z)A(k)\exp(a_\mp Z)$

contains only annihilators,

(ii) $$\exp\Big(-\int dk f(k)A^*(k)\Big)(FU_\pm(a_+,a_-,\theta))^{-1}\exp(-a_\mp Z)A(k)\exp(a_\mp Z)FU_\pm(a_+,a_-,\theta)\exp f\rangle = (\langle 0|A(k)\exp(a_\mp Z)FU_\pm(a_+,a_-,\theta)\exp f\rangle)|0\rangle.$$

Suppose now that $V_\pm(a_+,a_-,\theta)$ is a representation. We can write

$$\langle 0|A(k)V_\pm((a_+,a_-,\theta)(a'_+,a'_-,\theta'))\exp f\rangle = \sum_n\frac{1}{n!}\int\prod_1^n dk_i\langle 0|A(k)V_\pm(a_+,a_-,\theta)A^*(k_1)\cdots A^*(k_n)|0\rangle\langle 0|A(k_1)\cdots A(k_n)V_\pm(a'_+,a'_-,\theta')\exp f\rangle.$$

But, by (4.9), the right-hand side becomes, as expected,

$$\langle 0|A(k)V_\pm(a_+,a_-,\theta)\exp\int dk\, A^*(k) \times\langle 0|A(k)V_\pm(a'_+,a'_-,\theta')\exp f\rangle|0\rangle.$$

Conversely, if (4.6) is a representation, we must have

$$\langle 0|A(k)V_\pm((a_+,a_-,\theta)(a'_+,a'_-,\theta'))\exp f\rangle = \langle 0|A_kV_\pm(a_+,a_-,\theta)\exp\int dk\, A^*(k)$$

$$\times\langle 0|A_k V_\pm(a'_+,a'_-,\theta')\exp f\rangle|0\rangle.$$

By the lemma, the right-hand side becomes

$$\langle 0|A(k)V_\pm(a_+,a_-,\theta)V_\pm(a'_+,a'_-,\theta')\exp f\rangle.$$

Applying the lemma once more, we get

$$\langle \exp g\, V_\pm((a_+,a_-,\theta)(a'_+,a'_-,\theta'))\exp f\rangle$$
$$=\langle \exp g\, V_\pm(a_+,a_-,\theta)V_\pm(a'_+,a'_-,\theta')\exp f\rangle.$$

As the finite linear combinations of coherent vectors are dense in F, we obtain finally

$$V_\pm((a_+,a_-,\theta)(a'_+,a'_-,\theta'))$$
$$=V_\pm(a_+,a_-,\theta)V_\pm(a'_+,a'_-,\theta').$$

The statement on the inequivalence derives from the obvious nontriviality of $\exp(a_\mp Z)$ as a multiplicative one-cocycle; indeed, a trivial cocycle would depend on $a_\pm$.

Proposition 2: $V_\pm(a_+,a_-,\theta)$ is a representation when the distribution kernels are given by

$$Z_n(k;k_1,\ldots,k_n)=\delta(k-k_1-\cdots-k_n)\mathcal{Z}_n(k_1,\ldots,k_n), \tag{4.10}$$

where $\mathcal{Z}_n(k_1,\ldots,k_n)$ in $\mathcal{D}'(\mathbb{R}^n)$ has the homogeneity property

$$\mathcal{Z}_n(k_1,\ldots,k_n)=\exp(n\theta)\mathcal{Z}_n(k_1\exp\theta,\ldots,k_n\exp\theta). \tag{4.11}$$

Proof: $V_\pm(a_+,a_-,\theta)$ is a representation if

$$\exp(-\theta)Z=FU_\pm(a_+,a_-,\theta)Z(FU_\pm(a_+,a_-,\theta))^{-1}. \tag{4.12}$$

But, by the definition of $FU_\pm(a_+,a_-,\theta)$ [note that $FU_\pm(a_+,a_-,\theta)$ is not unitary in F],

$$FU_\pm(a_+,a_-,\theta)A^*(k)(FU_\pm(a_+,a_-,\theta))^{-1}$$
$$=\exp(ika_\pm)\exp(\mp\theta)A^*(k\exp(\mp\theta)), \tag{4.13}$$
$$FU_\pm(a_+,a_-,\theta)A(k)(FU_\pm(a_+,a_-,\theta))^{-1}$$
$$=\exp(-ika_\pm)A(k\exp(\mp\theta)). \tag{4.14}$$

Then (4.12) reads

$$(k-k_1-\cdots-k_n)Z_n(k;k_1,\ldots,k_n)=0,$$
$$Z_n(k;k_1,\ldots,k_n)$$
$$=\exp(n+1)\theta Z_n(k\exp\theta;k_1\exp\theta,\ldots,k_n\exp\theta).$$

This readily implies (4.10) and (4.11).

C. Wave functions and associated invariant equations

Henceforth, we assume (4.10) and (4.11) to be valid. Let us define mappings of $\mathcal{D}(\mathbb{R})$ into a subset of the space of C^∞ functions on $\mathbb{R}$:

$$f(k)\to F^{(n)}_\pm(x_+,x_-)=\int k^{-n}\,dk\,\langle 0|A(k)\exp(-x_\pm Z)$$
$$\times FU_\pm(-x_+,-x_-,0)\exp f\rangle$$
$$=\int k^{-n}\,dk\exp(-ikx_\pm)$$
$$\times\langle 0|A(k)\exp(-x_\mp Z)\exp f\rangle, \tag{4.15}$$

where we have used (4.12). The integration has to be performed in the sense of distribution theory, and n is some integer. We call $F^{(n)}_\pm(x_+,x_-)$ the wave function associated to the nonlinear representation fixed by Z. The group acts on the wave function by transport of structure. So, we have

$$F^{(n)}_\pm(x_+,x_-)\to\int k^{-n}\,dk\,\langle 0|A(k)V_\pm(-x_+,-x_-,0)$$
$$\times V_\pm(a_+,a_-,\theta)\exp f\rangle$$
$$=\exp(\pm(n-1)\theta)$$
$$\times F^{(n)}_\pm((x_+-a_+)\exp(-\theta),$$
$$(x_--a_-)\exp\theta).$$

In other words, the integer n fixes the kind of covariance ascribed to $F^{(n)}_\pm(x_+,x_-)$. In particular, $n=1$ is relevant to the L_1 realization and $n=2$ to the L_2 realization of Sec. III B.

First example: Let us take $n=2$. Then we have, according to the lemma,

$$\frac{\partial}{\partial x_-}F^{(2)}_+(x_+,x_-)=-\int k^{-2}\,dk\exp(-ikx_+)\sum_n\int\delta\Big(k-\sum_1^n k_i\Big)\mathcal{Z}_n(k_1,\ldots,k_n)\prod_{i=1}^n\langle 0|A(k_i)\exp(-x_-Z)\exp f\rangle dk_i$$
$$=-\sum_n\int(k_1+\cdots+k_n)^{-2}\mathcal{Z}_n(k_1,\ldots,k_n)\prod_{i=1}^n\exp(-ik_ix_+)\langle 0|A(k_i)\exp(-x_-Z)\exp f\rangle dk_i. \tag{4.16}$$

Let us choose

$$\mathcal{Z}_n(k_1,\ldots,k_n)$$
$$=-2\frac{\zeta_n}{n(n-1)}(k_1+\cdots+k_n)^n k_1^{-1}\cdots k_n^{-1}$$
$$\times\sum_{i<j}k_i^{-1}k_j^{-1},$$

which satisfies the homogeneity condition (4.11). Since we have

$$\frac{\partial}{\partial x_+}F^{(2)}(x_+,x_-)=-i\int k^{-1}\,dk\exp(-ikx_+)$$
$$\times\langle 0|A(k)\exp(-x_-Z)\exp f\rangle,$$

we finally get

$$\frac{\partial}{\partial x_-}F^{(2)}_+(x_+,x_-)=(F^{(2)}_+(x_+,x_-))^2\sum_{n\ge 2}(i)^n\zeta_n$$
$$\times\Big(\frac{\partial}{\partial x_+}F^{(2)}_+(x_+,x_-)\Big)^{n-2},$$

i.e., Eq. (3.14) of Sec. III B with

$$h(x)=\sum_{n\ge 2}(i)^n\zeta_n x^{n-2}.$$

Second example: Let us choose in the previous example

$$\mathcal{Z}_{2n+1}(k_1,\ldots,k_{2n+1})=0,$$

$$\mathcal{F}_{2n+2}(k_1,\ldots,k_{2n+2}) = \zeta_n \frac{n!(n+2)!}{(2n+2)!} \frac{(k_1+\cdots+k_{2n+2})^2}{k_1^2\cdots k_{2n+2}^2} \times \sum_{i_1<\cdots<i_n} k_{i_1}^2\cdots k_{i_n}^2, \quad n\geqslant 0.$$

The homogeneity condition (4.11) is satisfied. Since we have

$$\frac{\partial^2}{\partial x_+^2} F^{(2)}(x_+,x_-) = -\int dk \exp(-ikx_+) \times \langle 0|A(k)\exp(-x_-Z)\exp f\rangle,$$

we obtain the following equations:

$$\frac{\partial}{\partial x_-} F_+^{(2)}(x_+,x_-) = (F_+^{(2)}(x_+,x_-))^2 \times \sum_0^\infty \zeta_n \left(F_+^{(2)}(x_+,x_-) \times \frac{\partial^2 F_+^{(2)}}{\partial x_+^2}(x_+,x_-)\right)^n,$$

i.e., in terms of the invariants (3.12) the invariant equation

$$I_2 = h(I_4), \quad h(x) = \sum_0^\infty \zeta_n x^n.$$

More generally, if we take, in (4.16),

$$\mathcal{F}_n(k_1,\ldots,k_n) = (k_1+\cdots+k_n)^2 k_1^{-2}\cdots k_n^{-2} P_n(k_1,\ldots,k_n),$$

where $P_n(k_1,\ldots,k_n)$ is a symmetrical homogeneous polynomial of degree $(n-2)$, in which each k_i appears at most to the power 2, we can obtain from (4.16) any invariant equation of the form

$$I_2 = h(I_1,I_4),$$

where $h(x,y)$ is some analytic function.

Third example: Let us start now with $\mathcal{F}_n(k_1,\ldots,k_n) = 0$, except for $n=2$. Then we have

$$\frac{\partial^2}{\partial x_-^2} F_+^{(2)}(x_+,x_-) = \frac{2}{3}\int dk_1\,dk_2\,dk_3 (k_1+k_2+k_3)^{-2}(\mathcal{F}_2(k_1,k_2+k_3)\mathcal{F}_2(k_2,k_3) + \mathcal{F}_2(k_2,k_1+k_3)\mathcal{F}_2(k_1,k_3) + \mathcal{F}_2(k_3,k_1+k_2)\mathcal{F}_2(k_1,k_2))\prod_{i=1}^3 \exp(-ik_ix_+)\langle 0|A(k_i)\exp(-x_-Z)\exp f\rangle. \tag{4.17}$$

$$\mathcal{F}_2(k_1,k_2) = (\lambda/2)(k_1+k_2)^2 k_1^{-2}k_2^{-2},$$

the right-hand side of (4.17) reads

$$\lambda\int dk_1\,dk_2\,dk_3 \prod_{i=1}^3 k_i^{-2}\exp(-ik_ix_+) \times \langle 0|A(k_i)\exp(-x_-Z)\exp f\rangle = \lambda(F_+^{(2)}(x_+,x_-))^3.$$

So, we finally get the invariant equation

$$\frac{\partial^2}{\partial x_-^2} F_+^{(2)}(x_+,x_-) = \lambda(F_+^{(2)}(x_+,x_-))^3, \tag{4.18}$$

i.e., in terms of invariants (3.12),

$$I_5 = \lambda.$$

Fourth example: This last example is related to the L_1 realization of Sec. III B. In this case, we have readily

$$\frac{\partial}{\partial x_-}\frac{\partial}{\partial x_+} F_\pm^{(1)}(x_+,x_-) = i\sum_n \int \mathcal{F}_n(k_1,\ldots,k_n)\prod_{i=1}^n dk_i \exp(-ik_ix_\pm) \times \langle 0|A(k_i)\exp(-x_\mp Z)\exp f\rangle.$$

Let us take

$$\mathcal{F}_n(k_1,\ldots,k_n) = -i\zeta_n/k_1k_2\cdots k_n;$$

we get the well-known perturbed Klein–Gordon equation

$$\frac{\partial}{\partial x_-}\frac{\partial}{\partial x_+} F_\pm^{(1)}(x_+,x_-) = h(F_\pm^{(1)}(x_+,x_-)), \quad h(x) = \sum \zeta_n x^n. \tag{4.19}$$

Remark: Insofar as the wave function defined by (4.15) depends only on an arbitrary function, it does not provide the general solution of Eqs. (4.19) and (4.18). We can expect to reach this objective only when starting with the direct sum

$$U_+(a_+,a_-,\theta)\oplus U_-(a_+,a_-,\theta),$$

as $U(g)$ in the formal expansion (4.1) for (4.19) or with a linear deformation of it for (4.18). But then, we cannot get a compact form such as (4.6) and the above considerations would become much more lengthy and involved.

V. SYMMETRY REDUCTION FOR CONFORMALLY INVARIANT EQUATIONS

By construction, all equations of Sec. III are invariant under the Poincaré group p(1,1), the similitude group sim(1,1), or the conformal group conf(1,1). This invariance can be used to reduce the obtained equations to ordinary differential equations (ODE's). In many cases, the obtained ODE's can be solved directly.

In order to perform symmetry reduction in a systematic manner, we need to classify the subalgebras of the corresponding symmetry algebra into conjugacy classes, under

TABLE I. Representative one-dimensional subalgebras of conf(1,1) and the corresponding invariants ξ, satisfying (5.11). If the classifying group is the connected component $\mathrm{conf}_0(1,1)$, we have $\varepsilon_1 = \pm 1, \varepsilon_2 = \pm 1, a \neq 0, b \in \mathbf{R}$, and $c > 0$. If the classifying group is conf(1,1), including parity and time reversal, we have $\varepsilon_1 = \varepsilon_2 = 1$, $a > 0$, $b \geqslant 0$, and $c > 0$.

Number	Vector field	ξ (for $\varepsilon_1 = \varepsilon_2 = 1$)
$p_1(\varepsilon_1)$	$P_0 - \varepsilon_1 P_1$	$t + x$
p_2	P_0	x
p_3	P_1	t
p_4	K	$\ln(t+x) + \ln(t-x)$
s_5	D	$\ln(t+x) - \ln(t-x)$
$s_6(a)$	$D + aK$	$(a-1)\ln(t+x) + (a+1)\ln(t-x)$
$s_7(\varepsilon_1,\varepsilon_2)$	$D + \varepsilon_1 K + \varepsilon_2(P_0 - \varepsilon_1 P_1)$	$x - t + \ln(t+x)$
k_8	$P_1 - C_1$	$\arctan(x+t) + \arctan(t-x)$
$k_9(b)$	$P_0 + C_0 + b(P_1 - C_1)$	$(b-1)\arctan(t+x) + (b+1)\arctan(t-x)$
$k_{10}(c)$	$(1+c)(P_1 + C_0) + (1-c)(P_0 - C_1)$	$2c\arctan(t+x) + \ln\dfrac{t-x+1}{t-x-1}$
k_{11}	$P_0 + P_1 + C_0$	$\arctan(t+x)/\sqrt{2} - \sqrt{2}(t-x)^{-1}$
k_{12}	$P_0 + P_1 - C_1$	$\arctan(t+x)/\sqrt{2} - \sqrt{2}(t-x)^{-1}$

the action of the symmetry group. Here we shall concentrate on equations invariant under $\mathrm{conf}(1,1) \sim \mathrm{O}(2,2)$. The sublgebras of the Lie algebra conf(1,1) can be extracted from Ref. 16, or can be obtained directly, using the methods of Refs. 15 and 16. Alternatively, use can be made of the isomorphism $\mathrm{o}(2,2) = \mathrm{o}(2,1) \oplus \mathrm{o}(2,1)$ and the Goursat method for classifying subalgebras of direct sum algebras.[16,17] In any case, we only need to know the one-dimensional subalgebras.

A list of representatives of one-dimensional subalgebras of conf(1,1) is given in Table I. In column 1 we have $p_i \in \mathrm{p}(1,1)$, $s_i \in \mathrm{sim}(1,1)$, $s_i \notin \mathrm{p}(1,1)$, $k_i \in \mathrm{conf}(1,1)$, and $k_i \notin \mathrm{sim}(1,1)$.

Let us now consider the conformally invariant equation (3.22). Each of the subalgebras of Table I provides a different reduction to an ODE. Each of the ODE's can be solved by quadratures for an arbitrary function $h(F)$ in (3.22). This equation corresponds to the realization K_1 of conf(1,1). Since the action of conf(1,1) on the solution F in this realization is trivial, each one-dimensional subgroup will have two invariants of the form

$$\xi = \xi(x,t) \quad \text{and} \quad F. \tag{5.1}$$

Equation (3.22) is hence reduced to an ODE by putting

$$F(x,t) = f(\xi). \tag{5.2}$$

Substituting (5.2) into (3.22), we obtain the equation

$$\begin{aligned} &f_{\xi\xi} + (\Box\xi/(\nabla\xi)^2) f_\xi = h(f) f_\xi^2, \quad (\nabla\xi)^2 \neq 0, \\ &\Box\xi = \xi_{tt} - \xi_{xx}, \quad (\nabla\xi)^2 = \xi_t^2 - \xi_x^2. \end{aligned} \tag{5.3}$$

Equation (5.3) is an ODE if we have

$$(\Box\xi)/(\nabla\xi)^2 = \alpha(\xi), \tag{5.4}$$

a condition that is always satisfied if ξ is subgroup invariant.

Let us assume that (5.4) is satisfied and rewrite (5.3) as

$$f_{\xi\xi}/f_\xi + \alpha(\xi) = h(f) f_\xi, \quad f_\xi \neq 0. \tag{5.5}$$

In order to integrate Eq. (5.5) by quadratures, we define

$$G(f) = \int_{f_0}^{f} du \exp\left[-\int_{u_0}^{u} h(s)\,ds \right]. \tag{5.6}$$

In terms of $G(f)$, a well-defined transform of $h(f)$, the solution of Eq. (5.5) is

$$f(\xi) = G^{-1}(c_1 \gamma(\xi) + c_2), \tag{5.7}$$

where c_1 and c_2 are integration constants and we have put

$$\gamma(\xi) = \int_{\xi_0}^{\xi} dv \exp\left[-\int_{v_0}^{v} \alpha(s)\,ds \right]. \tag{5.8}$$

In order to obtain explicit solutions of Eq. (3.22) for a given function $h(F)$, all we need to do is to specify ξ for each subalgebra of conf(1,1) and then to calculate $\gamma(\xi)$ as in (5.8). We can actually simplify and unify further. If ξ satisfies (5.4), then so does any function $\eta(\xi)$, and we have

$$\alpha(\eta) = \eta_\xi^{-1}(\alpha(\xi) + \eta_{\xi\xi}\eta_\xi^{-1}). \tag{5.9}$$

Choosing $\eta(\xi)$ to satisfy $\eta_{\xi\xi} + \alpha(\xi)\eta(\xi) = 0$, we obtain $\alpha(\eta) = 0$, $\gamma(\xi) = \xi$.

The final result can be summed up as follows.

Theorem 1: Group invariant solutions of Eq. (3.22) have the form

$$f(\xi) = G^{-1}(c_1 \xi + c_2), \tag{5.10}$$

where $G(f)$ is defined in terms of $h(f)$ in (5.6), c_1 and c_2 are integration constants, and $\xi(x,t)$ is an invariant of one of the subgroups of conf(1,1), chosen so as to satisfy

$$\Box\xi = 0. \tag{5.11}$$

The variables ξ for each subgroup are given in the last column of Table I.

The results of Theorem 1 can be generalized to wider classes of solutions. Indeed, the PDE (3.22) reduces to an ODE for $F(x,t)$ satisfying (5.2) and ξ satisfying (5.11). Equation (5.11) has the general solution

$$\xi = \xi_1(x+t) + \xi_2(x-t). \tag{5.12}$$

If $\xi_1 = 0$ or $\xi_2 = 0$, we also have $(\nabla\xi)^2 = 0$, and

$$F(x,t) = f(x \pm t) \tag{5.13}$$

is a solution for any function $f(\xi)$. For ξ as in (5.12) and ξ_1 and ξ_2 arbitrary, we obtain a solution in the form (5.10). All group invariant solutions of Theorem 1 (and Table I) are special cases of the solution involving ξ as in (5.12).

Still more general solutions are obtained by applying the group conf(1,1) to the solution (5.10) with ξ as in (5.12). This amounts to replacing $x+t$ and $x-t$ by

$$t'+x' = (t+x+a_1)(1-c_1(t+x+a_1))^{-1}\exp\lambda,$$
$$t'-x' = (t-x+a_2)(1-c_2(t-x+a_2))^{-1}\exp\mu,$$
$$\lambda,\mu,a_1,a_2,c_1,c_1\in\mathbb{R}. \tag{5.14}$$

The results for other realizations of the conformal group are somewhat less clear-cut. Consider, for instance, Eq. (3.24′) for the realization K_2^0 of conf(1,1). The two invariants Σ_1 and Σ_2 are in this case on the same footing, in that they are both second-order differential operators. Hence we cannot expect to solve Eq. (3.25) without specifying the function $H(\Sigma_1,\Sigma_2)$. For definiteness we choose H so that Eq. (3.25) reduces to

$$\sum\nolimits_2 = A^2\sum\nolimits_1^2, \quad A\in\mathbb{R}. \tag{5.15}$$

The Poincaré group p(1,1) still acts trivially on F in this realization. Subgroups of the Poincaré group will hence lead to ODE's via the reduction formula (5.2), as in the case of the realization K_1. The dilations and conformal transformations act nontrivially on F and the reduction formulas will be

$$F(x,t) = \alpha(x,t)f(\xi), \quad \xi = \xi(x,t). \tag{5.16}$$

The function $\alpha(x,t)$ must be specified in each case, since it depends on the subgroup chosen. The variables ξ can be chosen to be the same as for the realization K_1, since the action of conf(1,1) on space-time is the same in both cases (we can, of course, replace ξ by any function of ξ). Let us look at some of the subalgebras of Table I.

p_1(1): Taking $\alpha=1, \xi=x+t$ in (5.16) we find $\Sigma_1=0$, $\Sigma_2=0$. Hence $f(x+t)$ is a solution of Eq. (3.25) for any H satisfying $H(0,0)=0$ [for instance, Eq. (5.15)].

p_2 and p_3: We have $\alpha=1$ in (5.16) and $\xi=ax+bt$, $a^2-b^2\neq 0$. We obtain

$$\sum\nolimits_1 = (b^2-a^2)(f_{\xi\xi}f-f_\xi^2), \quad \sum\nolimits_2 = (b^2-a^2)^2f_{\xi\xi}^2f^2. \tag{5.17}$$

Equation (5.15) is thus reduced to

$$(1+\epsilon A)ff_{\xi\xi} - \epsilon A f_\xi^2 = 0,$$

and the corresponding solution is

$$F(x,t) = c_1(ax+bt+c_2)^{1+\epsilon A}, \quad \epsilon = \pm 1, \tag{5.18}$$

$c_1,c_2,a,b = \text{const}$.

p_4: We again have $\alpha=1$ and now we choose $\xi=t^2-x^2$. The invariants reduce to

$$\sum\nolimits_1 = 4(\xi(ff_{\xi\xi}-f_\xi^2)+ff_\xi), \quad \sum\nolimits_2 = 16\xi^2f^2f_{\xi\xi}^2. \tag{5.19}$$

The corresponding solution of Eq. (5.15) is

$$F(x,t) = (c_1(t^2-x^2)^{1/(1-\epsilon A)}+c_2)^{1-\epsilon A}, \quad c_1,c_2=\text{const}. \tag{5.20}$$

As a further example, consider the subalgebra $s_5\in$sim(1,1). In this case, we take

$$F = xf(\xi), \quad \xi = t/x, \tag{5.21}$$

where Eq. (5.15) implies that $f(\xi)$ satisfies

$$(1-\epsilon A)(1-\xi^2)ff_{\xi\xi} + \epsilon A(1-\xi^2)f_\xi^2 + 2\epsilon A\xi ff_\xi - \epsilon A f^2 = 0, \quad \epsilon^2=1. \tag{5.22}$$

For $1-\epsilon A\neq 0$, we put

$$f = v^{1-\epsilon A}, \quad \epsilon A\neq 1, \tag{5.23}$$

and obtain a linear equation

$$(1-\epsilon A)^2(1-\xi^2)v_{\xi\xi} + 2\epsilon A(1-\epsilon A)\xi v_\xi - \epsilon A v = 0. \tag{5.24}$$

The solution of (5.24) is

$$v(\xi) = {}_2F_1(\alpha,\beta;\gamma;z), \quad z=(1-\xi)/2=(x-t)/2x,$$
$$\alpha\beta = \epsilon A(1-\epsilon A)^{-2}, \quad \alpha+\beta+1 = -2\epsilon A/(1-\epsilon A),$$
$$\gamma = -\epsilon A/(1-\epsilon A), \tag{5.25}$$

where ${}_2F_1(\alpha,\beta;\gamma;z)$ is a hypergeometric function.

In the special case $\epsilon A=1$, Eq. (5.22) has the solutions

$$f = c(1+\epsilon\xi), \quad c=\text{const}, \quad \epsilon=\pm 1. \tag{5.26}$$

As a final example, let us consider the subalgebra $k_8\in$conf(1,1). In this case we choose the invariants so as to have

$$F(x,t) = tf(\xi), \quad \xi = (1-t^2+x^2)/t. \tag{5.27}$$

We obtain

$$\sum\nolimits_1 = (4+\xi^2)(ff_{\xi\xi}-f_\xi^2) - f^2 + 2\xi ff_\xi,$$
$$\sum\nolimits_2 = (4+\xi^2)^2f^2f_{\xi\xi}^2. \tag{5.28}$$

Equation (5.15) reduces to

$$(1-\epsilon A)(4+\xi^2)ff_{\xi\xi} + \epsilon A(4+\xi^2)f_\xi^2 - 2\epsilon A\xi ff_\xi + \epsilon A f^2 = 0. \tag{5.29}$$

Equation (5.29) can again be reduced to a linear equation by the point transformation

$$f = u^{1-\epsilon A}, \quad \epsilon A\neq 1, \tag{5.30}$$
$$(1-\epsilon A)^2(4+\xi^2)u_{\xi\xi} - 2\epsilon A(1-\epsilon A)\xi u_\xi + \epsilon A u = 0, \tag{5.31}$$

which is reduced to a hypergeometric equation by putting

$$u = F(\alpha,\beta;\gamma;z), \quad z=(2+i\xi)/4,$$
$$\alpha\beta = \epsilon A(1-\epsilon A)^{-2}, \quad \alpha+\beta+1 = -2\epsilon A(1-\epsilon A)^{-1},$$
$$\gamma = -\epsilon A(1-\epsilon A)^{-1}. \tag{5.32}$$

For $\epsilon A=1$, we obtain the solutions

$$f = c(\xi\pm 2i), \quad c=\text{const}. \tag{5.33}$$

Reductions of Eq. (5.15) coming from invariance under the subgroups corresponding to $s_6,s_9,k_9,\ldots,k_{12}$ of Table I can be treated in the same manner, but we shall not present the results here. The same goes for the other realizations of the conformal Lie algebra and the corresponding equations involving (3.25)–(3.27).

VI. CONCLUDING REMARKS

We have shown that a classification of equations invariant under some space-time group involves as a first step the classification of possible realizations of the corresponding Lie algebra. In the classification of realizations of p(1,1),

sim(1,1), and conf(1,1), performed in Sec. II, we did not take parity P or time reversal T into account. Let us just point out that the realizations L_1 of p(1,1), S_1 and S_2 of sim(1,1), and K_1 and K_2^a of conf(1,1) are invariant under P and T. On the other hand, the realizations L_2, S_3^{ab}, $K_3^{\lambda ab}$, K_4^{bce}, and K_5^{ace} are not invariant under either of these transformations. A consequence of this is that the corresponding equations of Sec. III are appropriately invariant, or not invariant under parity and time reversal. We have not insisted on imposing these discrete symmetries, since the corresponding P or T invariance violating equations may also be of physical interest, e.g., in the context of weak interactions.

Sections IV and V are not intended to provide a complete treatment of their respective topics. Thus in Sec. IV we make contact with the theory of nonlinear representations of the Poincaré group p(1,1) only and give some examples. The theory of nonlinear representations of similitude and conformal groups has yet to be developed.

In Sec. V, we have applied the method of symmetry reduction to two of the five types of conformally invariant equations obtained in Sec. III. We find it remarkable that for Eq. (3.22) symmetry under any one-dimensional subgroup of conf(1,1) leads to an ODE that can be solved by successive quadratures. For the realization K_2^0 of conf(1,1) in which the dilation generator actually has the generic form $D = t\partial_t + x\partial_x + \lambda F\partial_F$ ($\lambda \neq 0$; we used the classification diffeomorphism to set $\lambda = 1$), we restricted ourselves to a specific invariant equation, namely, (5.15). Remarkably, all considered one-dimensional subgroups of conf(1,1) lead to equations that can either be integrated by quadratures, or at least be transformed, by a point transformation, into a linear equation.

The results of Sec. V point to a relation between conformal invariance and integrability, in some sense of the word.

An obvious continuation of the present investigation is to study invariant equations in higher dimensions and also to study spinor and tensor equations, rather than only scalar ones. We note here that the dimensions of the corresponding Poincaré, similitude, and conformal groups are

$$\begin{aligned} d_{\mathrm{p}} &= \dim \mathrm{p}(n,1) = (n+1)(n+2)/2, \\ d_S &= \dim \mathrm{sim}(n,1) = d_P + 1, \\ d_C &= \dim \mathrm{conf}(n,1) = (n+3)(n+2)/2. \end{aligned} \tag{6.1}$$

The second prolongation on the space $X \otimes U$ of independent and dependent variables for a scalar equation has dimension

$$d = 1 + (n+1)(n+6)/2. \tag{6.2}$$

The number of fundamental invariants of the second prolongation of the group action is hence, respectively,

$$\begin{aligned} n_{\mathrm{P}} &= d - d_{\mathrm{P}} = 2n + 3, \\ n_S &= d - d_S = 2n + 2, \quad n_C = d - d_C = n + 1, \end{aligned} \tag{6.3}$$

and is seen to increase linearly with the dimension n of space. The results of a classification of invariant equations for the physically important cases of $n = 2$ and $n = 3$ can thus be expected to be both richer and more complicated than for $n = 1$.

As far as applications are concerned, we note that the realization S_2 of sim(1,1) [see Eq. (2.7)] represents all realizations with a nontrivial scaling weight. Thus if we leave p(1,1) as in L_1, but transform D into $\tilde{D} = x\partial_x + t\partial_t + 2(1-p)F\partial_F$, we find that the physically important equation

$$F_{tt} - F_{xx} = AF^p, \quad p \neq 1, \tag{6.4}$$

is invariant under this realization of sim(1,1).

A study of tensor equations invariant under various groups in $n+1$ dimensions is relevant for field theories and in particular gravitational theories in these dimensions.[18] We plan to return to these problems in the near future.

Note added in proof: A conceptually, but less systematic, similar approach to the construction and solution of relativistically invariant equations has been adopted by W. Fushchich *et al.* They usually make an Ansatz about the general form of an equation and then impose invariance. See Ref. 19. Also see Ref. 20 and references therein.

ACKNOWLEDGMENTS

One of the authors (P. W.) thanks the Université Paris 7 (UFR de Physique) for its hospitality during visits in 1988 and 1989.

The research of P. W. was partially supported by research grants from NSERC of Canada and FCAR du Gouvernement du Québec. This research is part of joint program in the framework of the Coopération Québec–France (Project 20022089).

[1] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, Berlin, 1986).
[2] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, Berlin, 1974).
[3] L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[4] A. M. Grundland, J. Harnad, and P. Winternitz, J. Math. Phys. **25**, 791 (1984).
[5] P. Winternitz, A. M. Grundland, and J. A. Tuszyński, J. Math. Phys. **28**, 2194 (1987); J. Phys. C **21**, 4931 (1988).
[6] G. Gagnon and P. Winternitz, J. Phys. A **21**, 1493 (1988); **22**, 469 (1989); Phys. Rev. A **39**, 296 (1989); Phys. Lett. A **134**, 276 (1989).
[7] L. Gagnon, B. Grammaticos, A. Ramani, and P. Winternitz, J. Phys. A **22**, 499 (1989).
[8] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. **34**, 211 (1948).
[9] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[10] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, Oxford, 1964).
[11] M. Flato, J. Pinczon, and J. Simon, Ann. Sci. Ec. Norm. Sup. **10**, 405 (1977).
[12] E. Taflin, J. Math. Phys. **25**, 765 (1984).
[13] G. Rideau, Lett. Math. Phys. **8**, 421 (1984); **9**, 337 (1985); **15**, 243 (1988); Pub. IRMA, Lille **1**(X) (1986).
[14] J. Bertrand and G. Rideau, Lett. Math. Phys. **10**, 325 (1985).
[15] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597, 1615 (1975); **17**, 717 (1976).
[16] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977 (1976); **18**, 2259 (1977); Can. J. Phys. **54**, 950 (1976).
[17] E. Goursat, Ann. Sci. Ec. Norm. Sup. (3) **6**, 9 (1889).
[18] R. Jackiw, Nucl. Phys. B **252**, 343 (1985).
[19] W. Fushchich, W. Shtelen, and N. Serov, *Symmetry Analysis and Exact Solutions of Non-linear Equations of Mathematical Physics* (Kluwer Academic, 1989).
[20] W. Fushchich, in *Proceedings of the Annual Seminar of Canadian Mathematical Society*, Montreal, August 1989, to be published.

New method for deriving nonlinear integrable systems

V. K. Mel'nikov
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Head Post Office, Box 79, Moscow 101000, USSR

(Received 15 November 1988; accepted for publication 18 October 1989)

A new approach is proposed to derive nonlinear integrable systems. It is used to obtain several new nonlinear integrable systems. The above results are relevant to some problems of hydrodynamics, plasma physics, solid-state physics, etc.

I. INTRODUCTION

A new approach to derive nonlinear integrable systems is expounded in the present paper. It consists of the following. We take a nonlinear integrable system having the Lax representation[1] of the form

$$\frac{\partial L}{\partial t}+[A,L]=0,$$

where L and A are linear differential operators. For a wide class of operators L and A, a linear differential operator

$$\Gamma=\sum_{n=1}^{N}\Gamma_n$$

is shown to exist such that the coefficients of the operator Γ_n depend on the solution φ_n of the equation

$$(L-\lambda_n)\varphi_n=0,$$

and the coupled system of equations

$$\frac{\partial L}{\partial t}+[A,L]=\Gamma,\quad (L-\lambda_n)\varphi_n=0,\quad n=1,...,N,$$

is integrated by the inverse scattering method for the operator L.

It turned out that not only known integrable systems but several new ones can be derived in this way. It is important that for all earlier known integrable systems this approach is more simple and natural.

II. SECOND-ORDER SCALAR OPERATOR

We start with the simplest case of a second-order scalar operator, i.e., let

$$L=\partial^2+u, \tag{1}$$

where ∂ is the differentiation operator with respect to the space variable x, and u is the scalar function of x. Consider the linear system of equations

$$(L-\lambda)f_0=0,\quad f_n'=\psi_n f_0,\quad n=1,...,N, \tag{2}$$

with respect to the function $f_0,f_1,...,f_N$. Here, the prime means differentiation with respect to x and the functions $\psi_1,...,\psi_N$ are thought to be indefinite yet. Then, we take the operator A of the form

$$A=\partial^{k_0+2}+\sum_{k=0}^{k_0}a_k\partial^k,\quad k_0\geqslant 0, \tag{3}$$

such that the operator $\Delta=[A,L]$ has zero order, i.e., is the multiplication operator by a function and assume

$$\begin{aligned} g_0&=\frac{\partial f_0}{\partial t}+cAf_0+\varkappa\sum_{n=1}^{N}\varphi_n f_n,\\ g_n&=\psi_n f_0'-\psi_n' f_0-(\lambda-\lambda_n)f_n,\quad n=1,...,N,\end{aligned} \tag{4}$$

where c and $\varkappa$ are the constants, and $\varphi_1,...,\varphi_N$ are certain (yet unknown) functions.

Let us clarify what requirements are to be imposed on the functions $u,\varphi_1,...,\varphi_N$, $\psi_1,...,\psi_N$ so that the quantities $g_0,g_1,...,g_N$ determined by (1)–(4) should satisfy the conditions

$$(L-\lambda)g_0=\varkappa\sum_{n=1}^{N}\varphi_n g_n,\quad \frac{g_n}{\partial x}\equiv 0,\quad n=1,...,N. \tag{5}$$

By simple calculations we find that for the validity of (5) it is necessary and sufficient to fulfill the relations

$$\begin{aligned} &\frac{\partial L}{\partial t}+c[A,L]=2\varkappa\sum_{n=1}^{N}\frac{\partial}{\partial x}(\varphi_n\psi_n),\\ &(L-\lambda_n)\varphi_n=(L-\lambda_n)\psi_n=0,\quad n=1,...,N.\end{aligned} \tag{6}$$

In the case when c and $\varkappa$ are real constants and the quantities λ_n satisfy the condition $\lambda_n=\bar{\lambda}_n$, the system (6) has an invariant manifold $u=\bar{u}$, $\psi_n=\epsilon_n\bar{\varphi}_n$, where $\epsilon_n^2=1$, $n=1,...,N$, and the bar means complex conjugation. Motion on this manifold is described by the system of equations

$$\begin{aligned} &\frac{\partial L}{\partial t}+c[A,L]=2\varkappa\sum_{n=1}^{N}\epsilon_n\frac{\partial}{\partial x}|\varphi_n|^2,\\ &(L-\lambda_n)\varphi_n=0,\quad n=1,...,N.\end{aligned} \tag{7}$$

At $k_0=1$, $c=4$, and $\varkappa=1$ the first equation of this system has the form

$$\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(3u^2+\frac{\partial^2 u}{\partial x^2}\right)=2\sum_{n=1}^{N}\epsilon_n\frac{\partial}{\partial x}|\varphi_n|^2,$$

i.e., differs from the Kortweg–de Vries equation[2] by the expression in the right-hand side of this equality.

Reference 3 represents the scheme of integration of the system (7) in the class of rapidly decreasing x functions by the inverse scattering method for the operator L of the form (1).

III. THIRD-ORDER SCALAR OPERATOR

Consider now the case when the operator L has the form

$$L=\partial^3+u_1\partial+u_0. \tag{8}$$

Let $f_0,f_1,...,f_N$ be the solution of the linear system of equations

 0022-2488/90/051106-08$03.00

$$(L-\lambda)f_0=0,\quad f_n'=\psi_n f_0,\quad n=1,...,N, \tag{9}$$

where $\psi_1,...,\psi_N$ are certain (yet unknown) functions. Then, we use the operator A of the form (3) such that the operator $\Delta=[A,L]$ has the first order and assume

$$\begin{aligned} g_0&=\frac{\partial f_0}{\partial t}+cAf_0+\varkappa\sum_{n=1}^{N}\varphi_n f_n,\\ g_n&=\psi_n f_0''-\psi_n' f_0'+\psi_n'' f_0+u_1\psi_n f_0\\ &\quad-(\lambda-\lambda_n)f_n,\quad n=1,...,N, \end{aligned} \tag{10}$$

where c and $\varkappa$ are the constants, and $\varphi_1,...,\varphi_N$ are certain functions. Finally, we demand the quantities $g_0,g_1,...,g_N$ determined by (3) and (8)–(10) to satisfy the conditions

$$(L-\lambda)g_0=\varkappa\sum_{n=1}^{N}\varphi_n g_n,\quad \frac{\partial g_n}{\partial x}\equiv 0,\quad n=1,...,N. \tag{11}$$

By simple calculations we find that for the validity of (11) it is necessary and sufficient to fulfill the equalities

$$\begin{aligned} &\frac{\partial L}{\partial t}+c[A,L]=3\varkappa\sum_{n=1}^{N}\frac{\partial}{\partial x}\left(\frac{\partial\varphi_n}{\partial x}\psi_n+\varphi_n\psi_n\partial\right),\\ &(L-\lambda_n)\varphi_n=(\tilde{L}-\lambda_n)\psi_n=0,\quad n=1,...,N, \end{aligned} \tag{12}$$

where

$$\tilde{L}=-\partial^3-\partial\cdot u_1+u_0. \tag{13}$$

Assume now that $u_1=\frac{3}{2}u$, $u_0=\frac{3}{4}u'+ip$, where the functions u and p take only real values, i.e., choose the operator L of the form (8) so that the relation $L^*=-L$ is fulfilled (the asterisk means Hermitian conjugation). Then let the quantities λ_n satisfy the condition $\lambda_n+\bar{\lambda}_n=0$, i.e., be pure imaginary, $n=1,...,N$. In this case the functions ψ_n by virtue of (13) can be chosen so as to fulfill the equality $\psi_n=\epsilon_n\bar{\varphi}_n$, where $\epsilon_n^2=1$, $n=1,...,N$. Finally, we choose the operator A so as to fulfill the relation $A^*=(-1)^{k_0}A$ and the quantities c and $\varkappa$ are chosen under the requirements $\bar{c}=(-1)^{k_0+1}c$ and $\bar{\varkappa}=\varkappa$. Under these conditions the system (12) becomes

$$\begin{aligned} &\frac{\partial L}{\partial t}+c[A,L]=3\varkappa\sum_{n=1}^{N}\epsilon_n\frac{\partial}{\partial x}\left(\frac{\partial\varphi_n}{\partial x}\bar{\varphi}_n+|\varphi_n|^2\partial\right),\\ &(L-\lambda_n)\varphi_n=0,\quad n=1,...,N. \end{aligned}$$

At $k_0=0$, $c=i$, and $\varkappa=1$ the first relation of this system is equivalent to the equations

$$\begin{aligned} &3\frac{\partial u}{\partial t}-4\frac{\partial p}{\partial x}=6\sum_{n=1}^{N}\epsilon_n\frac{\partial}{\partial x}|\varphi_n|^2,\\ &4\frac{\partial p}{\partial t}-\frac{\partial}{\partial x}\left(3u^2+\frac{\partial^2 u}{\partial x^2}\right)\\ &\quad=6i\sum_{n=1}^{N}\epsilon_n\frac{\partial}{\partial x}\left(\varphi_n\frac{\partial\bar{\varphi}_n}{\partial x}-\frac{\partial\varphi_n}{\partial x}\bar{\varphi}_n\right). \end{aligned} \tag{14}$$

Assuming $\varphi_n\equiv 0$, $n=1,...,N$, in the system (14) and then eliminating the function p from it we get the known Boussinesq equation.[4]

IV. ARBITRARY-ORDER SCALAR OPERATOR

To emphasize some moments of the aforementioned, we consider the operator L of an arbitrary order, i.e., assume

$$L=\partial^{m_0+2}+\sum_{m=0}^{m_0}u_m\partial^m,\quad m_0>0. \tag{15}$$

Then we take the linear system of equations

$$(L-\lambda)f_0=0,\quad f_n'=\psi_n f_0,\quad n=1,...,N, \tag{16}$$

and using its solution $f_0,f_1,...,f_N$ determine the quantities $g_0,g_1,...,g_N$ by the equalities

$$\begin{aligned} g_0&=\frac{\partial f_0}{\partial t}+cAf_0+\varkappa\sum_{n=1}^{N}\varphi_n f_n,\\ g_n&=\sum_{m=0}^{m_0+1}(-1)^m\frac{\partial^m\psi_n}{\partial x^m}\frac{\partial^{m_0-m+1}f_0}{\partial x^{m_0-m+1}}\\ &\quad+\sum_{m=1}^{m_0}\sum_{r=0}^{m-1}(-1)^r\frac{\partial^r(u_m\psi_n)}{\partial x^r}\frac{\partial^{m-r-1}f_0}{\partial x^{m-r-1}}\\ &\quad-(\lambda-\lambda_n)f_n,\quad n=1,...,N, \end{aligned} \tag{17}$$

where c and $\varkappa$ are the constants and A is the operator of the form (3) such that the order of the operator $\Delta=[A,L]$ is m_0.

Now let us elucidate what requirements are to be imposed on the functions $u_0,...,u_{m_0}$, $\varphi_1,...,\varphi_N$, $\psi_1,...,\psi_N$, so that the quantities $g_0,g_1,...,g_N$ determined by (3) and (15)–(17) should satisfy the conditions

$$(L-\lambda)g_0=\varkappa\sum_{n=1}^{N}\varphi_n g_n,\quad \frac{\partial g_n}{\partial x}\equiv 0,\quad n=1,...,N. \tag{18}$$

By simple calculations we get that

$$\begin{aligned} &(L-\lambda)g_0-\varkappa\sum_{n=1}^{N}\varphi_n g_n\\ &\quad=\varkappa\sum_{n=1}^{N}f_n(L-\lambda_n)\varphi_n-\left\{\frac{\partial L}{\partial t}+c[A,L]-\varkappa\Gamma\right\}f_0, \end{aligned} \tag{19}$$

$$\frac{\partial g_n}{\partial x}=\psi_n(L-\lambda)f_0-f_0(\tilde{L}-\lambda_n)\psi_n,\quad n=1,...,N,$$

where

$$\tilde{L}=(-1)^{m_0}\partial^{m_0+2}+\sum_{m=0}^{m_0}(-1)^m\partial\cdot u_m, \tag{20}$$

and the action of the operator Γ on the function f_0 is determined by the equality

$$\begin{aligned} \Gamma f_0&=\sum_{n=1}^{N}\sum_{m=0}^{m_0+1}\frac{\partial^{m_0-m+1}}{\partial x^{m_0-m+1}}\left(\frac{\partial^m\varphi_n}{\partial x^m}\psi_n f_0\right)\\ &\quad-\sum_{n=1}^{N}\sum_{m=0}^{m_0+1}(-1)^m\varphi_n\frac{\partial^m\psi_n}{\partial x^m}\frac{\partial^{m_0-m+1}f_0}{\partial x^{m_0-m+1}}\\ &\quad+\sum_{n=1}^{N}\sum_{m=1}^{m_0}\sum_{r=0}^{m-1}u_m\frac{\partial^{m-r-1}}{\partial x^{m-r-1}}\left(\frac{\partial^r\varphi_n}{\partial x^r}\psi_n f_0\right)\\ &\quad-\sum_{n=1}^{N}\sum_{m=1}^{m_0}\sum_{r=0}^{m-1}(-1)^r\varphi_n\frac{\partial^r(u_m\psi_n)}{\partial x^r}\frac{\partial^{m-r-1}f_0}{\partial x^{m-r-1}}. \end{aligned} \tag{21}$$

One can easily see that the right-hand side of equality (21) does not in fact contain derivatives of the function f_0 higher than the m_0th order since expressions with a derivative of the (m_0+1)th order in the first and second terms cancel out. Thus the operator Γ determined by (21) is of an

order not higher than m_0. By virtue of (16) it follows from (19) that for the validity of the conditions (18) it is necessary and sufficient to fulfill the relations

$$\frac{\partial L}{\partial t} + c[A,L] = \varkappa\Gamma,$$
$$(L-\lambda_n)\varphi_n = (\tilde{L}-\lambda_n)\psi_n = 0, \quad n = 1,...,N. \tag{22}$$

If the parameters c, $\varkappa$, and λ_n entering into this system are taken under the conditions that $\bar{c} = (-1)^{k_0+1}c$, $\bar{\varkappa} = \varkappa$, $\bar{\lambda}_n = (-1)^{m_0}\lambda_n$, $n = 1,...,N$, then according to (20) the system (22) has an invariant manifold $L^* = (-1)^{m_0}L$, $\psi_n = \epsilon_n\bar{\varphi}_n$, where $\epsilon_n^2 = 1$, $n = 1,...,N$. In this case the operators A and Γ should satisfy the requirements $A^* = (-1)^{k_0}A$ and $\Gamma^* = (-1)^{m_0}\Gamma$.

V. FIRST-ORDER MATRIX OPERATOR

The procedure described above for deriving nonlinear integrable systems can also be used, with some alterations, in the case when the coefficients of the operator L are square matrices of an arbitrary order $r_0 > 1$. Let us proceed to this version considering the simplest case of a first-order operator, i.e., assume

$$L = \Lambda\partial + U, \tag{23}$$

where Λ is the diagonal matrix with constant elements on the principal diagonal, and U is the square matrix with zero elements on the principal diagonal. Then let A be a differential operator of the form

$$A = \sum_{k=0}^{k_0+1} A_k\partial^k, \quad k_0 \geqslant 0, \tag{24}$$

such that the coefficients $A_0,...,A_{k_0+1}$ are the square matrices of the same order r_0 and the operator $\Delta = [A,L]$ is of the zero order, i.e., is the multiplication operator by a matrix. It is well known that in this case on the principal diagonal the matrix Δ has zeros.

Consider now the linear system of equations

$$(L-\eta)F_0 = 0, \quad F_n' = \Psi_n F_0, \quad n = 1,...,N, \tag{25}$$

with respect to unknown matrices $F_0,F_1,...,F_N$. Moreover, assume that F_0 is the square matrix of the order r_0 but $F_1,...,F_N$ and $\Psi_1,...,\Psi_N$ are the rectangular matrices with r_1 rows and r_0 columns. Using the solutions $F_0,F_1,...,F_N$ of the system (25) we determine the quantities $G_0,G_1,...,G_N$ by the equalities

$$G_0 = \frac{\partial F_0}{\partial t} + cAF_0 + \varkappa\sum_{n=1}^{N}\Phi_n F_n,$$
$$G_n = \Psi_n\Lambda F_0 - (\eta-\eta_n)F_n, \quad n = 1,...,N, \tag{26}$$

where c and $\varkappa$ are the constants, and $\Phi_1,...,\Phi_N$ are the rectangular matrices with r_0 rows and r_1 columns, and consequently, G_0 is the square matrix of the order r_0 and $G_1,...,G_N$ are the rectangular matrices with r_1 rows and r_0 columns. Now let us find out what requirements are to be imposed on the matrices U, $\Phi_1,...,\Phi_N$, $\Psi_1,...,\Psi_N$ in order that the matrices $G_0,G_1,...,G_N$ determined by (23)–(26) should satisfy the conditions

$$(L-\eta)G_0 = \varkappa\sum_{n=1}^{N}\Phi_n G_n, \quad \frac{\partial G_n}{\partial x} \equiv 0, \quad n = 1,...,N. \tag{27}$$

By simple calculations we get that for the validity of the conditions (27) it is necessary and sufficient to fulfill the relations

$$\frac{\partial L}{\partial t} + c[A,L] = \varkappa\sum_{n=1}^{N}[\Lambda,\Phi_n\Psi_n],$$
$$\Lambda\Phi_n' + U\Phi_n - \eta_n\Phi_n = \Psi_n'\Lambda - \Psi_n U + \eta_n\Psi_n = 0, \quad n = 1,...,N. \tag{28}$$

The system (28) contains several interesting particular cases, some of which will be considered in detail below.

VI. NONLINEAR SCHRÖDINGER EQUATION WITH A SELF-CONSISTENT SOURCE

Let $r_0 = 2$. Take the operator L of the form (23) with the matrices Λ and U determined by the equalities

$$\Lambda = \operatorname{diag}(1,-1), \quad U = \begin{vmatrix} 0 & u \\ \bar{u} & 0 \end{vmatrix}. \tag{29}$$

As is known,[5] in this case at any $k_0 > 0$ there exists an operator A of the form

$$A = \Lambda\partial^{2k_0} + \sum_{k=0}^{2k_0-1} A_k\partial^k, \tag{30}$$

such that the operator $\Delta = [A,L]$ is the skew-Hermitian matrix with zero elements on the principal diagonal, i.e.,

$$\Delta = \begin{vmatrix} 0 & q \\ -\bar{q} & 0 \end{vmatrix}. \tag{31}$$

In this case the quantity q is the polynomial of the functions u, $\bar{u}$ and their derivatives with respect to x up to an order of $2k_0$, and the matrices $A_0,...,A_{2k_0-1}$ satisfy the condition

$$A_0 = \cdots = A_{2k_0-1} \equiv 0 \quad \text{at } u \equiv 0. \tag{32}$$

One can easily be convinced that if the vector column $\Phi = \Phi_n$ with the components $\varphi = \varphi_n$ and $\psi = \psi_n$ satisfies the equation

$$\Lambda\Phi' + U\Phi - \eta\Phi = 0 \tag{33}$$

at $\eta = \eta_n$, then the vector row $\Psi = \Psi_n$ with the components $\psi = \psi_n$ and $\varphi = \varphi_n$ satisfies the equation

$$\Psi'\Lambda - \Psi U + \eta\Psi = 0 \tag{34}$$

at the same value of the parameter $\eta = \eta_n$, i.e., the solution Ψ_n of Eq. (34) is connected with the solution Φ_n of Eq. (33) at the same values of the parameter $\eta = \eta_n$ by

$$\Psi_n = \tilde{\Phi}_n\sigma, \quad \sigma = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \tag{35}$$

where the tilde sign means transposition, i.e., in particular, transition from the vector column to the vector row. Then, it can easily be verified that if the vector column $\Phi = \Phi_n$ satisfies Eq. (33) at $\eta = \eta_n$, the vector column $\Phi = \sigma\Lambda\bar{\Phi}_n$ satisfies the same equation at $\eta = -\bar{\eta}_n$ and the vector row $\Psi = \Phi_n^*\Lambda$ by virtue of (35) satisfies Eq. (34) at $\eta = -\bar{\eta}_n$. Assume finally that the integer N is even, i.e., $N = 2N_0$, the quantities η_n satisfy the conditions $\eta_{N_0+n} = -\bar{\eta}_n$, $\eta_n + \bar{\eta}_n \neq 0$, $n = 1,...,N_0$, and the constants c and $\varkappa$ are pure imaginary. From the aforementioned we assume that at $n = 1,...,N_0$

$$\Psi_n = \epsilon_n \tilde{\Phi}_n \sigma, \quad \Phi_{N_0+n} = \sigma \Lambda \bar{\Phi}_n, \quad \Psi_{N_0+n} = \epsilon_n \Phi_n^* \Lambda, \tag{36}$$

where $\epsilon_n^2 = 1$. Then, at $n = 1,...,N_0$ the equalities

$$[\Lambda, \Phi_n \Psi_n] = 2\epsilon_n \begin{vmatrix} 0 & \varphi_n^2 \\ -\psi_n^2 & 0 \end{vmatrix},$$

$$[\Lambda, \Phi_{N_0+n} \Psi_{N_0+n}] = 2\epsilon_n \begin{vmatrix} 0 & \bar{\psi}_n^2 \\ -\bar{\varphi}_n^2 & 0 \end{vmatrix}$$

are valid, where φ_n and ψ_n are the vector column Φ_n components. Hence, it follows from (29) and (31) that in the case under consideration the system (28) has the form

$$\frac{\partial u}{\partial t} + cq = 2\varkappa \sum_{n=1}^{N_0} \epsilon_n (\varphi_n^2 + \bar{\psi}_n^2),$$

$$\varphi_n' + u\psi_n - \eta_n \varphi_n = \psi_n' - \bar{u}\varphi_n + \eta_n \psi_n = 0, \quad n = 1,...,N_0. \tag{37}$$

At $k_0 = 1$, $c = -2i$, and $\varkappa = -i$ the first equation of this system has the form

$$i\frac{\partial u}{\partial t} + 2|u|^2 u + \frac{\partial^2 u}{\partial x^2} = 2 \sum_{n=1}^{N_0} \epsilon_n (\varphi_n^2 + \bar{\psi}_n^2),$$

i.e., differs from the nonlinear Schrödinger equation[6] by the expression in the right-hand side of this equality.

In Sec. VII, we shall present the scheme of integrating the system (37) in the class of rapidly decreasing x functions.

VII. INTEGRATION OF THE SYSTEM (37)

It will be shown that in the class of rapidly decreasing x functions the system (37) can be integrated by the inverse scattering method for the operator L of the form (23) with the matrices Λ and U determined by (29). More exactly, let $u_0 = u_0(x)$ be an arbitrary complex-valued function of x rapidly enough decreasing with its derivatives with respect to x (up to a certain final order dependent on k_0) as $x \to \pm\infty$. Assume that the system of equations

$$\varphi' + u\psi - \eta\varphi = 0, \quad \psi' - \bar{u}\varphi + \eta\psi = 0 \tag{38}$$

at $u = u_0(x)$ has exactly $2N_0$ points of the discrete spectrum $\eta_1,...,\eta_{N_0}, -\bar{\eta}_1,..., -\bar{\eta}_{N_0}$. Let further $A_n(t)$ be arbitrary continuous functions of t, $n = 1,...,N_0$. Below we shall indicate the conditions under which the system (37) has the solution $u = u(x,t)$, $\varphi_n = \varphi_n(x,t)$, $\psi_n = \psi_n(x,t)$, $n = 1,...,N_0$, such that

$$u(x,0) = u_0(x), \tag{39}$$

$$\int_{-\infty}^{\infty} \varphi_n(x,t)\psi_n(x,t)dx = A_n(t), \quad n = 1,...,N_0. \tag{40}$$

To solve this problem we should apparently analyze how the data on the scattering of the system (38) change in time if the potential $u = u(x,t)$ satisfies the system (37) and the conditions (39) and (40). For this purpose we assume $\eta = i\zeta$, where $\zeta \in (-\infty, \infty)$ and take the solutions φ_-, ψ_- and φ_+, ψ_+ of the system (38) satisfying the requirements

$$\varphi_- \sim 0, \quad \psi_- \sim \exp(-i\zeta x), \quad x \to -\infty,$$
$$\varphi_+ \sim \exp(i\zeta x), \quad \psi_+ \sim 0, \quad x \to \infty. \tag{41}$$

Apart from the solution φ, ψ, the functions $\bar{\psi}, -\bar{\varphi}$ also satisfy the system (38). Therefore, the matrices F_0^- and F_0^+ of the form

$$F_0^- = \begin{vmatrix} \bar{\psi}_- & \varphi_- \\ -\bar{\varphi}_- & \psi_- \end{vmatrix}, \quad F_0^+ = \begin{vmatrix} \varphi_+ & -\bar{\psi}_+ \\ \psi_+ & \bar{\varphi}_+ \end{vmatrix} \tag{42}$$

are the fundamental matrices of the solutions of the system (38). Hence, the equality

$$F_0^+ = F_0^- S, \quad S = \begin{vmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix} \tag{43}$$

is valid where the elements $S_{\alpha\beta}$ of the matrix S are independent of x, $\alpha,\beta, = 1,2$. By virtue of the equality $\det F_0^- = \det F_0^+ = 1$ it follows that

$$S_{11} = \varphi_+\psi_- - \varphi_-\psi_+, \quad S_{12} = -\varphi_-\bar{\varphi}_+ - \psi_-\bar{\psi}_+,$$
$$S_{21} = \varphi_+\bar{\varphi}_- + \psi_+\bar{\psi}_-, \quad S_{22} = \bar{\varphi}_+\bar{\psi}_- - \bar{\varphi}_-\bar{\psi}_+, \tag{44}$$

i.e.,

$$S_{22} = \bar{S}_{11}, \quad S_{21} = -\bar{S}_{12},$$
$$\det S = |S_{11}|^2 + |S_{12}|^2 = 1.$$

According to (25) we assume at $n = 1,...,N$

$$F_n^- = \int_{-\infty}^{x} \Psi_n(z) F_0^-(z,\zeta)dz,$$
$$F_n^+ = -\int_{x}^{\infty} \Psi_n(z) F_0^+(z,\zeta)dz. \tag{45}$$

Then, we determine the matrices $G_0^-, G_0^+, G_1^-, G_1^+,...,$ G_N^-, G_N^+ by substituting the matrices $F_0^-, F_0^+, F_1^-, F_1^+,...,$ F_N^-, F_N^+ of the form (42) and (45) into equality (26). With (27) one can easily be convinced that at $n = 1,...,N$ the equalities $G_n^- = G_n^+ \equiv 0$ are valid. Taking this fact into account and using (26), (27), (30), (32), (41), and (42) we get the equalities

$$G_0^- = c(i\zeta)^{2k_0} F_0^- \Lambda, \quad G_0^+ = c(i\zeta)^{2k_0} F_0^+ \Lambda. \tag{46}$$

Now we take the equality following from (26) and (46):

$$\frac{\partial F_0^+}{\partial t} + cAF_0^+ + \varkappa \sum_{n=1}^{N} \Phi_n F_n^+ = c(i\zeta)^{2k_0} F_0^+ \Lambda,$$

change in it F_0^+ by $F_0^- S$ and pass to the limit as $x \to -\infty$. As a result of simple calculations we get the equality

$$\frac{\partial S}{\partial t} + c(i\zeta)^{2k_0}[\Lambda, S] = 0 \tag{47}$$

defining the evolution in time of the matrix S of the form (43). It is seen from this equality that the evolution of the S matrix is by no means influenced by the expression in the right-hand side of the first equation of the system (37).

As is known, the solutions φ_-, ψ_- and φ_+, ψ_+ of the system (38) determined by (41) admit an analytical continuation in ζ into the upper half-plane. Hence, it follows according to (44) that the function S_{11} also admits an analytical continuation in ζ into the upper half-plane. Further, the zeros $\zeta = \zeta_n$, $n = 1,...,N_0$, of the function S_{11} in the upper half-plane correspond to the points of the discrete spectrum of the operator L since by virtue of (44) the following equalities are valid:

$$\varphi_+(x,\zeta_n) = B_n\varphi_-(x,\zeta_n),$$
$$\psi_+(x,\zeta_n) = B_n\psi_-(x,\zeta_n), \tag{48}$$

where the quantities B_n are independent of x, $n = 1,...,N_0$. With (38) one can verify that at any $m,n = 1,...,N_0$ the following equalities are valid:

$$\int_{-\infty}^{\infty} \{\varphi_+(x,\zeta_m)\psi_+(x,\zeta_n) + \varphi_+(x,\zeta_n)\psi_+(x,\zeta_m)\}dx = 0, \quad m \neq n,$$
$$\int_{-\infty}^{\infty} \{\varphi_+(x,\zeta_m)\bar{\varphi}_+(x,\zeta_n) - \psi_+(x,\zeta_m)\bar{\psi}_+(x,\zeta_n)\}dx = 0. \tag{49}$$

Moreover, according to (38) and (44) we get that at any $n = 1,...,N_0$ the following relation holds:

$$\int_{-\infty}^{\infty} \{\varphi_+(x,\zeta_n)\psi_-(x,\zeta_n) + \varphi_-(x,\zeta_n)\psi_+(x,\zeta_n)\}dx = i\frac{\partial S_{11}}{\partial\zeta}\bigg|_{\zeta=\zeta_n}. \tag{50}$$

Let $I_- = \mathrm{diag}(0,1)$, $I_+ = \mathrm{diag}(1,0)$. According to the aforesaid, it follows from (42) that the matrices $f_0^- = F_0^- I_-$ and $f_0^+ = F_0^+ I_+$ admit an analytical continuation in ζ into the upper half-plane. From (45) it follows that the matrices $f_n^- = F_n^- I_-$ and $f_n^+ = F_n^+ I_+$ at $n = 1,...,N$ also admit an analytical continuation in ζ into the upper half-plane. Finally, by virtue of (26) it follows that the matrices $g_n^- = G_n^- I_-$ and $g_n^+ = G_n^+ I_+$ at $n = 0,1,...,N$ also admit an analytical continuation in ζ into the upper half-plane. Moreover, it follows from the second of the equalities (26) that at $\eta = \eta_n = i\xi_n$ the identities

$$\Psi_n(x)\Lambda f_0^-(x,\xi_n) = \Psi_n(x)\Lambda f_0^+(x,\xi_n) \equiv 0, \quad n = 1,...,N_0,$$

take place. Using the latter we get from (36) that the components φ_n and ψ_n of the vector column Φ_n must have the form

$$\varphi_n = C_n\varphi_+(x,\zeta_n), \quad \psi_n = C_n\psi_+(x,\zeta_n), \tag{51}$$

where the quantities C_n are independent of x, $n = 1,...,N_0$.

The equalities following from (26),

$$\frac{\partial f_0^-}{\partial t} + cAf_0^- + \varkappa\sum_{n=1}^{N}\Phi_n f_n^- = g_0^-, \tag{52}$$

$$\frac{\partial f_0^+}{\partial t} + cAf_0^+ + \varkappa\sum_{n=1}^{N}\Phi_n f_n^+ = g_0^+, \tag{53}$$

based on the aforesaid, admit an analytical continuation in ζ into the upper half-plane. Moreover, according to (42) and (46) the following equalities are valid:

$$g_0^- = -c(i\xi)^{2k_0}f_0^-, \quad g_0^+ = c(i\xi)^{2k_0}f_0^+.$$

Taking account of (48) we find that $\zeta = \zeta_m$ the relations

$$f_0^+(x,\zeta_m) = B_m f_0^-(x,\zeta_m)\sigma,$$
$$g_0^+(x,\zeta_m) = -B_m g_0^-(x,\zeta_m)\sigma, \quad m = 1,...,N_0,$$

are fulfilled. Now we multiply equality (52) from the right by the matrix $B_m\sigma$ and subtract the obtained result from equality (53). Hence, we get that at $\zeta = \zeta_m$ the following matrix equality holds:

$$\frac{\partial B_m}{\partial t}f_0^-(x,\zeta_m)\sigma + \varkappa\sum_{n=1}^{N}\Phi_n(x)\{f_n^+(x,\zeta_m) - B_m f_n^-(x,\zeta_m)\sigma\} = -2B_m g_0^-(x,\zeta_m)\sigma, \quad m = 1,...,N_0.$$

We pass in this equality to the limit as $x \to -\infty$. From (45) we get at $m = 1,...,N_0$ and $n = 1,...,N$:

$$\lim_{x\to-\infty} f_n^-(x,\zeta_m) = 0,$$
$$\lim_{x\to-\infty} f_n^+(x,\zeta_m) = -\int_{-\infty}^{\infty}\Psi_n(z)f_0^+(z,\zeta_m)dz.$$

By virtue of (36) and (48)–(51) at $n \neq m$ the equality

$$\int_{-\infty}^{\infty}\Psi_n(z)f_0^+(z,\zeta_m)dz = 0$$

is valid and at $n = m$ we have

$$\int_{-\infty}^{\infty}\Psi_n(z)f_0^+(z,\zeta_m)dz = H_m,$$

where H_m is the vector row with the components $h_{m,1}$ and $h_{m,2}$ of the form

$$h_{m,1} = i\epsilon_m B_m C_m\frac{\partial S_{11}}{\partial\xi}\bigg|_{\xi=\xi_m}, \quad h_{m,2} = 0, \quad m = 1,...,N_0.$$

Using these equalities we get the relation

$$\frac{\partial B_m}{\partial t} - [\varkappa h_{m,1}C_m + 2c(i\xi_m)^{2k_0}]B_m = 0, \quad m = 1,...,N_0. \tag{54}$$

In the case when all zeros $\zeta = \zeta_m$ of the function S_{11} in the upper half-plane are simple, the inequality

$$\frac{\partial S_{11}}{\partial\xi}\bigg|_{\xi=\xi_m} \neq 0, \quad m = 1,...,N_0,$$

is valid. Consequently, choosing the quantities C_m entering into (51) so as to fulfill the relation

$$iB_m C_m^2\frac{\partial S_{11}}{\partial\xi}\bigg|_{\xi=\xi_m} = 2A_m, \quad m = 1,...,N_0, \tag{55}$$

we can finally satisfy the condition (40) and Eq. (54) can be written in the form

$$\frac{\partial B_m}{\partial t} - 2[c(i\xi_m)^{2k_0} + \varkappa\epsilon_m A_m(t)]B_m = 0, \quad m = 1,...,N_0. \tag{56}$$

Thus, from the given functions $A_m(t)$, according to this equation we determine the quantities B_m and then on the basis of (55) we find C_m, $m = 1,...,N_0$.

Equations (47) and (56) completely determine the evolution of the scattering data of the system (38). In the case when all the zeros of the function S_{11} in the upper half-plane are simple and the S_{11} function itself does not vanish at any real ζ, the obtained data are used to determine in a known way the kernel of the integral Gel'fand–Levitan equation and the problem is thus reduced to the solution of this equation. The solution $u = u(x,t)$ of the system (37) satisfying the condition (39) is derived by solving the Gel'fand–Levitan equation in a known way. Then, according to (47), limitations on the S_{11} function pointed out above will be fulfilled to any $t > 0$ if they are fulfilled at $t = 0$. This imposes certain

additional requirements on the function $u_0 = u_0(x)$, which have already been mentioned at the beginning of this section.

VIII. MODIFIED KORTEWEG–DE VRIES EQUATION WITH A SELF-CONSISTENT SOURCE

Let us give one more interesting example of the system (28). Again, let $r_0 = 2$, $\Lambda = \mathrm{diag}(1, -1)$ and the matrix U have the form

$$U = \begin{vmatrix} 0 & u \\ u & 0 \end{vmatrix}, \tag{57}$$

where the function u satisfies the condition $\bar{u} = u$. As is known,[5] in this case at any $k_0 > 0$ there exists operator A of the form

$$A = \partial^{2k_0+1} + \sum_{k=0}^{2k_0-1} A_k \partial^k$$

such that the operator $\Delta = [A,L]$ is the symmetric matrix with zero elements on the principal diagonal, i.e.,

$$\Delta = \begin{vmatrix} 0 & q \\ q & 0 \end{vmatrix}, \tag{58}$$

In this case the quantity q is the polynomial of the function u and its derivatives with respect to x up to the $(2k_0+1)$th order, and the matrices $A_0,...,A_{2k_0-1}$ satisfy the condition (32).

Consider now Eqs. (33) and (34) with the potential U of the form (57). One can easily be convinced that if the vector column $\Phi = \Phi_n$ with the components φ_n and ψ_n satisfies Eq. (33) at $\eta = \eta_n$, the vector row $\Psi = \tilde{\Phi}_n \sigma$ with the components ψ_n and φ_n satisfies Eq. (34) at $\eta = \eta_n$. Then, one can easily verify that if the vector column $\Phi = \Phi_n$ satisfies Eq. (33) at $\eta = \eta_n$, the vector column $\Phi = \Lambda\sigma\Phi_n$ satisfies the same equation at $\eta = -\eta_n$ and the vector row $\Psi = \tilde{\Phi}_n \Lambda$ satisfies Eq. (34) at $\eta = -\eta_n$. Finally, note that if the vector column $\Phi = \Phi_n$ satisfies Eq. (33) at $\eta = \eta_n$, the vector column $\Phi = \bar{\Phi}_n$ satisfies the same equation at $\eta = \bar{\eta}_n$, the vector row $\Psi = \Phi_n^* \sigma$ satisfies Eq. (34) at $\eta = \bar{\eta}_n$, the vector column $\Phi = \Lambda\sigma\bar{\Phi}_n$ satisfies Eq. (33) at $\eta = -\bar{\eta}_n$ and the vector row $\Psi = \Phi_n^* \Lambda$ satisfies Eq. (34) at $\eta = -\bar{\eta}_n$. Assume that integer N is multiple to 4, i.e., $N = 4N_0$, the quantity η_n satisfies the conditions

$$\eta_{N_0+n} = -\eta_n, \quad \eta_{2N_0+n} = \bar{\eta}_n, \eta_{3N_0+n} = -\bar{\eta}_n,$$
$$\eta_n + \bar{\eta}_n \neq 0, \quad n = 1,...,N_0,$$

and the constants c and $\varkappa$ are real. With due regard to the aforesaid we assume that at $n = 1,...,N_0$:

$$\Psi_n = \epsilon_n \tilde{\Phi}_n \sigma, \quad \Phi_{N_0+n} = \Lambda\sigma\Phi_n, \quad \Psi_{N_0+n} = \epsilon_n \tilde{\Phi}_n \Lambda,$$
$$\Phi_{2N_0+n} = \bar{\Phi}_n, \quad \Psi_{2N_0+n} = \epsilon_n \Phi_n^* \sigma,$$
$$\Phi_{3N_0+n} = \Lambda\sigma\bar{\Phi}_n, \quad \Psi_{3N_0+n} = \epsilon_n \Phi_n^* \Lambda,$$

where $\epsilon_n^2 = 1$. Then, at $n = 1,...,N_0$ the equalities

$$[\Lambda, \Phi_n \Psi_n] = 2\epsilon_n \begin{vmatrix} 0 & \varphi_n^2 \\ -\psi_n^2 & 0 \end{vmatrix},$$

$$[\Lambda, \Phi_{N_0+n} \Psi_{N_0+n}] = 2\epsilon_n \begin{vmatrix} 0 & -\psi_n^2 \\ \varphi_n^2 & 0 \end{vmatrix},$$

$$[\Lambda, \Phi_{2N_0+n} \Psi_{2N_0+n}] = 2\epsilon_n \begin{vmatrix} 0 & \bar{\varphi}_n^2 \\ -\bar{\psi}_n^2 & 0 \end{vmatrix},$$

$$[\Lambda, \Phi_{3N_0+n} \Psi_{3N_0+n}] = 2\epsilon_n \begin{vmatrix} 0 & -\bar{\psi}_n^2 \\ \bar{\varphi}_n^2 & 0 \end{vmatrix}$$

are valid where φ_n and ψ_n are the components of the vector column Φ_n. Hence, it follows from (57) and (58) that in the case considered the system (28) has the form

$$\frac{\partial u}{\partial t} + cq = 2\varkappa \sum_{n=1}^{N_0} \epsilon_n (\varphi_n^2 + \bar{\varphi}_n^2 - \psi_n^2 - \bar{\psi}_n^2),$$
$$\varphi_n' + u\psi_n - \eta_n \varphi_n = \psi_n' - u\varphi_n + \eta_n \psi_n = 0, \quad n = 1,...,N_0. \tag{59}$$

Note that in obtaining this system we did not assume anywhere that $\eta_n \neq \bar{\eta}_n$, $n = 1,...,N_0$. In the case when the equality $\eta_n = \bar{\eta}_n$ really takes place at some $n = m$, this results in that the components φ_m and ψ_m of the vector column Φ_m can be assumed real, and consequently, some terms in the right-hand side of the first equation of the system (59) are met twice. At $k_0 = 1$, $c = 4$, and $\varkappa = 1$ the first equation of (59) is

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(2u^3 + \frac{\partial^2 u}{\partial x^2}\right) = 2 \sum_{n=1}^{N_0} \epsilon_n (\varphi_n^2 + \bar{\varphi}_n^2 - \psi_n^2 - \bar{\psi}_n^2),$$

i.e., differs from the modified Korteweg–de Vries equation[7] by the expression in the right-hand side of this equality.

Integration of the system (59) is very similar to that of the system (37) and, therefore, is omitted.

IX. SECOND-ORDER MATRIX OPERATOR

The results obtained in Sec. V admit generalization to the case of an arbitrary-order matrix operator. However, avoiding cumbersome formulas we restrict our consideration to the second-order matrix operator. Thus let

$$L = \Lambda\partial^2 + U_1\partial + U_0, \tag{60}$$

where Λ, U_1, and U_0 are the square matrices of an arbitrary order $r_0 > 1$. In this case Λ is the diagonal matrix with constant elements on the principal diagonal, and the matrix U_1 has zeros on the principal diagonal. Let then A be a differential operator of the form (24) such that the operator $\Delta = [A,L]$ is of the first order, i.e., $\Delta = \Delta_1\partial + \Delta_0$ where Δ_1 and Δ_0 are the square matrices of the order r_0. It is well known that on the principal diagonal, the matrix Δ_1 has zeros.

Consider the linear system of equations

$$(L - \eta)F_0 = 0, \quad F_n' = \Psi_n F_0, \quad n = 1,...,N, \tag{61}$$

with respect to unknown matrices $F_0, F_1,...,F_N$. We assume that F_0 is the square matrix of an order of r_0 and the matrices $F_1,...,F_N$ and $\Psi_1,...,\Psi_N$ have r_1 rows and r_0 columns. Suppose that

$$G_0 = \frac{\partial F_0}{\partial t} + cAF_0 + \varkappa \sum_{n=1}^{N} \Phi_n F_n,$$
$$G_n = \Psi_n \Lambda F_0' - \Psi_n' \Lambda F_0 + \Psi_n U_1 F_0 - (\eta - \eta_n) F_n, \quad n = 1,...,N, \tag{62}$$

where the matrices $\Phi_1,...,\Phi_N$ have r_0 rows and r_1 columns, and consequently, G_0 is the square matrix of an order of r_0

and the matrices $G_1,...,G_N$ have r_1 rows and r_0 columns.

Now we require the matrices $G_0,G_1,...,G_N$ determined by (24) and (60)–(62) to satisfy the conditions

$$(L-\eta)G_0 = \varkappa \sum_{n=1}^{N} \Phi_n G_n, \quad \frac{\partial G_n}{\partial x} \equiv 0, \quad n = 1,...N. \tag{63}$$

By simple calculations we get that for the validity of the conditions (63) it is necessary and sufficient to fulfill the relations

$$\begin{aligned} &\frac{\partial L}{\partial t} + c[A,L] = \varkappa \Gamma, \\ &\Lambda \Phi_n'' + U_1 \Phi_n' + U_0 \Phi_n = \eta_n \Phi_n, \\ &\Psi_n'' \Lambda - (\Psi_n U_1)' + \Psi_n U_0 = \eta_n \Psi_n, \quad n = 1,...,N, \end{aligned} \tag{64}$$

where the operator Γ has the form

$$\begin{aligned} \Gamma = \sum_{n=1}^{N} \{ &[\Lambda, \Phi_n \Psi_n \partial - \Phi_n \Psi_n'] \\ &+ 2\Lambda(\Phi_n \Psi_n)' - [U_1, \Phi_n \Psi_n]\} \end{aligned} \tag{65}$$

Let $r_0 = 2$ and $k_0 = 0$. Assume

$$U_1 = \begin{vmatrix} 0 & p \\ -\bar{p} & 0 \end{vmatrix}, \quad U_0 = \begin{vmatrix} u & q + \frac{1}{2}p' \\ \bar{q} - \frac{1}{2}\bar{p}' & v \end{vmatrix},$$

$$c_0 = \frac{a_1 - a_2}{\lambda_1 - \lambda_2}, \quad \Lambda = \operatorname{diag}(\lambda_1, \lambda_2),$$

$$A_1 = \operatorname{diag}(a_1, a_2), \quad A_0 = c_0 U_1,$$

where $\lambda_1 \neq \lambda_2$ and $a_1 \neq a_2$ are arbitrary real constants. In this case the system (64) in accordance with (65) becomes

$$\begin{aligned} &\frac{\partial p}{\partial t} + c\left(aq + b\frac{\partial p}{\partial x}\right) = \varkappa\lambda \sum_{n=1}^{N} \epsilon_n \varphi_n \bar{\psi}_n, \\ &\frac{\partial u}{\partial t} + a_1 c \frac{\partial u}{\partial x} + cc_0 \left(\frac{1}{2}\frac{\partial}{\partial x}|p|^2 + p\bar{q} + \bar{p}q\right) \\ &\quad = \varkappa \sum_{n=1}^{N} \epsilon_n \left(2\lambda_1 \frac{\partial}{\partial x}|\varphi_n|^2 + p\bar{\varphi}_n \psi_n + \bar{p}\varphi_n \bar{\psi}_n\right), \\ &\frac{\partial v}{\partial t} + a_2 c \frac{\partial v}{\partial x} + cc_0 \left(\frac{1}{2}\frac{\partial}{\partial x}|p|^2 - p\bar{q} - \bar{p}q\right) \\ &\quad = \varkappa \sum_{n=1}^{N} \epsilon_n \left(2\lambda_2 \frac{\partial}{\partial x}|\psi_n|^2 - p\bar{\varphi}_n \psi_n - \bar{p}\varphi_n \bar{\psi}_n\right), \\ &\frac{\partial q}{\partial t} + ca_0 \frac{\partial q}{\partial x} - \frac{1}{4} ca \frac{\partial^2 p}{\partial x^2} - cc_0(u - v)p \\ &\quad = \varkappa \sum_{n=1}^{N} \epsilon_n \left\{2\lambda_0 \frac{\partial}{\partial x}(\varphi_n \bar{\psi}_n) \right. \\ &\qquad + \frac{1}{2}\lambda \left(\frac{\partial \varphi_n}{\partial x}\bar{\psi}_n - \varphi_n \frac{\partial \bar{\psi}_n}{\partial x}\right) \\ &\qquad \left. - (|\varphi_n|^2 - |\psi_n|^2)p\right\}, \\ &\lambda_1 \varphi_n'' + p\psi_n' + u\varphi_n + (q + \tfrac{1}{2}p')\psi_n = \eta_n \varphi_n, \\ &\lambda_2 \psi_n'' - \bar{p}\varphi_n' + (\bar{q} - \tfrac{1}{2}\bar{p}')\varphi_n + v\psi_n = \eta_n \psi_n, \quad n = 1,...,N, \end{aligned}$$

where

$$\begin{aligned} &\lambda = \lambda_1 - \lambda_2, \quad \lambda_0 = \tfrac{1}{2}(\lambda_1 + \lambda_2), \\ &a = a_1 - a_2, \quad a_0 = \tfrac{1}{2}(a_1 + a_2), \\ &b = \tfrac{3}{2}a_1 - \tfrac{1}{2}a_2 - 2c_0\lambda_1 = -\tfrac{1}{2}a_1 + \tfrac{3}{2}a_2 - 2c_0\lambda_2. \end{aligned}$$

Moreover, u and v are real functions, c, $\varkappa$, and $\eta_1,...,\eta_N$ are real parameters, and $\epsilon_n^2 = 1$, $n = 1,...,N$.

X. THE KADOMTSEV–PETVIASHVILI EQUATION WITH A SELF-CONSISTENT SOURCE

The above-mentioned procedure of deriving nonlinear integrable systems is also applicable to the case of two space variables. To illustrate this statement let us consider the following example. Let L and A be linear differental operators of the form

$$L = \partial_x^2 - i\partial_y + u, \quad A = 4\partial_x^3 + 3(u\partial_x + \partial_x \cdot u) + ip, \tag{66}$$

where ∂_x and ∂_y are the differentiation operators with respect to the space variables x and y, respectively, and u and p be the scalar functions. Consider a linear system of equations of the form

$$(L - \lambda)f_0 = 0, \quad \frac{\partial f_n}{\partial x} = \psi_n f_0, \quad n = 1,...,N, \tag{67}$$

with respect to the functions $f_0, f_1,..., f_N$. The functions $\psi_1,...,\psi_N$ will be determined below. Then, with the solutions $f_0, f_1,..., f_N$ of the system (67) we determine the quantities $g_0,g_1,...,g_N$ by the equalities

$$\begin{aligned} g_0 &= \frac{\partial f_0}{\partial t} + Af_0 + 4\sum_{n=1}^{N} \varphi_n f_n, \\ g_n &= \psi_n \frac{\partial f_0}{\partial x} - \frac{\partial \psi_n}{\partial x} f_0 \\ &\quad - i\frac{\partial f_n}{\partial y} - (\lambda - \lambda_n)f_n, \quad n = 1,...,N, \end{aligned} \tag{68}$$

where $\varphi_1,...,\varphi_N$ are certain (yet undetermined) functions.

Now we shall find out what requirements are to be imposed on the functions $u, p, \varphi_1,...,\varphi_N, \psi_1,...,\psi_N$ in order that the quantities $g_0,g_1,...,g_N$ determined by (66)–(68) should satisfy the conditions

$$(L - \lambda)g_0 = 4\sum_{n=1}^{N} \varphi_n g_n, \quad \frac{\partial g_n}{\partial x} \equiv 0, \quad n = 1,...,N. \tag{69}$$

By simple calculations we get that for the validity of the conditions (69) it is necessary and sufficient to fulfill the equalities

$$\begin{aligned} &\frac{\partial L}{\partial t} + [A,L] = 8\sum_{n=1}^{N} \frac{\partial}{\partial x}(\varphi_n \psi_n), \\ &(L - \lambda_n)\varphi_n = (\tilde{L} - \lambda_n)\psi_n = 0, \quad n = 1,...,N, \end{aligned} \tag{70}$$

where

$$\tilde{L} = \partial_x^2 + i\partial_y + u. \tag{71}$$

With the allowance made for (66) the first of the equalities (70) is equivalent to the system of equations

$$\begin{aligned} &\frac{\partial u}{\partial t} + 6u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} - 8\sum_{n=1}^{N} \frac{\partial}{\partial x}(\varphi_n \psi_n) = \frac{\partial p}{\partial y}, \\ &\frac{\partial p}{\partial x} = 3\frac{\partial u}{\partial y}. \end{aligned}$$

Eliminating p we get

$$3\frac{\partial^2 u}{\partial y^2}-\frac{\partial}{\partial x}\left[\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(3u^2+\frac{\partial^2 u}{\partial x^2}\right.\right.$$
$$\left.\left.-8\sum_{n=1}^{N}\varphi_n\psi_n\right)\right]=0.$$

In the case when all λ_n are real values we may assume, by virtue of (71), that $\bar{u}=u$ and $\psi_n=\epsilon_n\bar{\varphi}_n$, where $\epsilon_n^2=1$, $n=1,\ldots,N$. Thus the system (70) takes its final form

$$3\frac{\partial^2 u}{\partial y^2}-\frac{\partial}{\partial x}\left[\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(3u^2+\frac{\partial^2 u}{\partial x^2}\right.\right.$$
$$\left.\left.-8\sum_{n=1}^{N}\epsilon_n|\varphi_n|^2\right)\right]=0, \tag{72}$$

$$i\frac{\partial\varphi_n}{\partial y}=\frac{\partial^2\varphi_n}{\partial x^2}+(u-\lambda_n)\varphi_n,\quad n=1,\ldots,N.$$

At $\varphi_n\equiv 0$ the first equation of this system turns into the well-known Kadomtsev–Petviashvili equation.[8] Note that in Ref. 9 the operator representation has been found for this system, in Ref. 10 a multisoliton solution has been constructed and in Refs. 11 and 12 the dynamics of a multisoliton solution is investigated. However, the approach expounded in this section allows one to look differently upon the problem of integration of the system (72).

XI. CONCLUSION

In conclusion, we should like to emphasize that all the systems obtained with the help of the procedure expounded in this paper have one remarkable property: They can describe the soliton capture and confinement processes. This means that all these systems have solutions describing solitons that come from infinity, then are captured into different oscillatory regimes and remain in them during all subsequent time moments. For the first time this phenomenon has been observed in Ref. 13 and then investigated in detail in Refs. 3, 12, and 14.

[1]P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).
[2]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[3]V. K. Mel'nikov, JINR preprint P2-88-668, Dubna, 1988.
[4]V. E. Zakhorov, Sov. Phys. JETP **38**, 108 (1974).
[5]V. K. Mel'nikov, Particles Nuclei **11**, 1224 (1980).
[6]V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972).
[7]M. J. Ablowitz, D. J. Kaup, A. S. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973).
[8]B. B. Kadomtsev and V. I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970).
[9]V. K. Mel'nikov, Lett. Math. Phys. **7**, 129 (1983).
[10]V. K. Mel'nikov, Commun. Math. Phys. **112**, 639 (1987).
[11]V. K. Mel'nikov, J. Math. Phys. **28**, 2603 (1987).
[12]V. K. Mel'nikov, JINR preprint P2-88-232, Dubna, 1988.
[13]V. K. Mel'nikov, JINR preprint P2-88-55, Dubna, 1988.
[14]V. K. Mel'nikov, JINR preprint P2-88-114, Dubna, 1988.

Comment on "Some results from a Mellin transform expansion for the heat kernel" [J. Math. Phys. 30, 1226 (1989)]

A. P. C. Malbouisson[a)]
Centre de Physique Théorique, Ecole Polytechnique, 91128 Palaiseau, France

F. R. A. Simào
Centro Brasileiro de Pesquisas Fisicas-CBPF/CNPq, Rua Dr. Xavier Sigaud, 150, 22290 Rio de Janeiro, RJ, Brazil

A. Ferraz de Camargo Fo
Instituto de Fisica Têorica-Universidade Estadual Paulista, Rua Pamplona, 145, 01405 Sao Paulo, SP, Brazil

(Received 2 August 1989; accepted for publication 13 December 1989)

In this note the inaccuracies contained in the above mentioned paper are corrected and some of the results therein are modified and extended. In particular, formulae are given for directly calculating the coefficients of the heat kernel expansion, not making appeal to recurrence in the heat equation. The nonexistence of a class of anomalies in odd dimension is proved.

I. INTRODUCTION

Our starting point is the Mellin transform expansion for the solution of the "heat equation"—the so-called *heat kernel* associated to an elliptic operator H of order m, acting on a D-dimensional compact manifold without boundary, which has been obtained in a previous paper.[1] We consider in this note only differential operators of order m *even*. The diagonal part of the heat kernel has, for $t \ll 1$, the asymptotic expansion,

$$F(t;x,x) = -(4\pi t)^{-D/m} \times \left\{ \sum_{l=0}^{\infty} (4\pi)^{D/m} \frac{(-1)^l}{1!} K(l;x,x) t^{l+(D/m)} + \sum_{j}^{\infty} \Gamma\left(\frac{D-j}{m}\right) R_j(x)\,(4\pi)^{D/m} t^{j/m} \right\}, \quad (1)$$

where $R_j(x)$ is the residue of $K(1;x,x)$ at the pole situated at $s=(j-D)/m$, $K(s;x,x)$ is the diagonal part of the Seeley's analytic extension to the complex s plane of the kernel $K(s;x,y)$ of the power operator H^s (see Ref. 2), and the sum over j *excludes* the values of j such that $(j-D)/m = 0,1,2,\ldots$.

Concerning our paper in the title (henceforth refered to as Paper I) we first observe that we have not taken into account the vanishing of the residues of the poles of $K(s;x,x)$ at *all* integers s values $\geqslant 0$ in the best way, in the derivation of our asymptotic expansion. When this is done, we need neither introduce the function $\phi(l)$ in formula (1.4) of Paper I, nor calculate its derivatives at the points $s=l$, so the coefficients of the sum over l in (1) are obtained in a much simpler way. The simplified derivation of the asymptotic expansion (1) is given in Ref. 1. It follows that the formula (4.3) in Paper I is unnecessary [by the way, there is a misprint in that formula—the factor (-2) must be dropped out] and the expression for the anomaly in dimension D is given directly by

$$A = q\,\mathrm{Tr}\{(X+Y)\,[K(0;x,x)+P_o(x)]\}. \quad (2)$$

Equation (2) replaces Eqs. (4.2) and (4.5) of Paper I.

As it was remarked in Ref. 1 the series (1) looks different from the de Witt ansatz[3] currently used for even D and operators of order $m=2$,

$$F(t;x,x) = (4\pi t)^{-D/2} \sum_{n=0}^{\infty} c_n(x)\, t^n. \quad (3)$$

For our expansion (1) to coincide formally with (3) it would be necessary, for example in even D and $m=2$, that all the residues $R_j(x)$ vanish for odd values of j and $(j-D)/m \neq 0,1,2,\ldots$, which is not explicitly stated in Seeley's paper.[2]

II. THE VANISHING OF SOME RESIDUES OF THE POLES OF $K(s;x,x)$

To further investigate this point we remember that $R_j(x)$ is given by

$$R_j(x) = \frac{1}{\mathrm{im}(2\pi)^{D+1}} \int_{|\xi|=1} d\xi \times \int_{\Gamma} d\lambda\, \lambda^{(j-D)/m} b_{-m-j}(x,\xi,\lambda), \quad (4)$$

where Γ is a curve coming from ∞ along a ray of minimal growth, clockwise on a small circle around the origin and then going back to ∞. The quantities b_{-m-j} are obtained from the coefficients a_{m-k} of the symbol of H, $\sigma(H)(x,\xi) = \Sigma_0^m a_{m-k}(x,\xi)$, by the set of equations

$$b_{-m}(a_m - \lambda) = 1, \quad l=0, \quad (5a)$$

$$b_{-m-l} = -(a_m-\lambda)^{-1} \sum \left(\frac{\partial}{\partial \xi}\right)^{\alpha} b_{-m-j} \times D_x^{\alpha} \frac{1}{\alpha!} a_{m-k}, \quad l>0, \quad (5b)$$

where multi-index notation for α is used and the sum is taken for

$$j<l, \quad j+k+|\alpha| = l, \quad |\alpha| \equiv \alpha_1 + \cdots + \alpha_D.$$

Now, from the definition of the a_{m-k}'s,

[a)] On leave of absence from Centro Brasileiro de Pesquisas Fisicas-CBPF/CNPq-Post doctoral fellow of CAPES (Brazil).

 0022-2488/90/051114-02$03.00

$$a_{m-k}(x,\xi) = \sum_{|\alpha| = m-k} H_\alpha \xi^\alpha,$$

where the H_α are the coefficients of the operator H in Seeley's notation, we see that the parity in the ξ-variables of a_{m-k} is $(-1)^k$, and from (5a) that the parity in ξ of b_{-m} is always equal to 1. Then by induction (the recurrence hypothesis is easily verified directly for the two first steps) we obtain from (5b) that the parity in ξ of $b_{-m-l}(x,\xi,\lambda)$ equals $(-1)^l$ for any $l \geqslant 0$. Thus from Eq. (4) since the integration over ξ is constrained to the unit sphere (in the cotangent space) $|\xi| = 1$, we see that $R_j(x)$ *vanishes for* j *odd*.

To see the implications of the vanishing of the residues of these poles of $K(s;x,x,)$, we consider separately the two situations of the manifold dimension D being even and being odd.

(1) D *even*: $D = 2p$; $m = 2q$ (remember we consider m even). In this case, since the sum over j in (1) is such that $(j-D)/m \neq 0,1,2,...$, we must have $j \neq 2(qk+p)$, $k = 0,1,2,...$, i.e., the only allowed even values of j are those smaller than $2p-2$; otherwise j is odd. But if j is odd, $R_j(x)$ vanishes, therefore for even D the asymptotic expansion is given by the first sum (sum over l) in (1) and a piece of the sum over j, corresponding to $j = 0,2,...,2p-2$.

(2) D *odd*: $D = 2p+1$; $m = 2q$. All the $K(l;x,x)$ vanish: they are given by the integral[2]

$$K(l;x,x) = \frac{1}{(-1)^l \cdot l (2\pi)^D} \int_{|\xi|=1} d\xi \times \int_\Gamma d\lambda\, \lambda^l b_{-m-lm-D}(x,\xi,\lambda), \tag{6}$$

which is zero by the parity argument above, since $lm + D = 2(lq+p) + 1$ is odd. The asymptotic expansion is given by the second sum (sum over j) in Eq. (1), if their coefficients do not vanish. This is indeed the case, for we must have $(j-D)/m \neq k$, $k = 0,1,2,...$; so, $j \neq 2(qk+p)+1$, i.e., j must be *even*. But, if j is even, the $R_j(x)$'s do not vanish by the same parity argument above.

III. CONCLUSIONS

In the case of D *even*, the expansion coincides formally exactly with the de Witt ansatz, in the sense that the series remaining after factorization of $(4\pi t)^{-D/m}$ is a series of integer powers of t, if p is a multiple of q ($D/2$ is a multiple of $m/2$). Otherwise the sum over l in (1) contains only fractionary powers of t, although of course the series as a whole is made up of integer powers of t.

For *odd* D, the series has, after factorization of $(4\pi t)^{-D/m}$, integer or fractionary powers of t if, respectively, $j/2$ is or is not a multiple of $q = m/2$. The global powers of t are in any case fractionary.

In the special but important case of a differential operator H of order $m = 2$, the asymptotic expansion (1) has the form of the de Witt ansatz,

$$F(t;x,x) = (4\pi t)^{-D/2} \sum_{l=0}^{\infty} \alpha_l(x)\, t^l, \tag{7}$$

where the coefficients $\alpha_l(x)$ are:

for *even* D,

$$\alpha_l(x) = -\Gamma\left(\frac{D-2l}{2}\right) R_{2l}(x) (4\pi)^{D/2};$$

$$l = 0,1,2,..., \frac{D}{2} - 1,$$

$$\alpha_l(x) = -\frac{(-1)^l}{l!} K(l;x,x) (4\pi)^{D/2};$$

$$l = \frac{D}{2}, \frac{D}{2}+1, ..., \tag{8}$$

and for *odd* D,

$$\alpha_l(x) = -\Gamma\left(\frac{D-2l}{2}\right) R_{2l}(x) (4\pi)^{D/2}. \tag{9}$$

Another implication of the vanishing of $R_j(x)$ for odd j is that the residues of the new poles of the Hawking's zeta function $\xi(s)$, which we had claimed to exist in Paper I, vanish in even dimension D. On the other side, those poles are already present with nonvanishing residues for *odd* D and situated at the values $s = (D-j)/m$ for $j = 0,2,4,...$.

Also, replacing (6) in (2) we see that all the class of anomalies described by Cognola and Zerbini[4] *cannot exist in odd dimension*, due to the vanishing of $K(0;x,x)$. This is to be compared with previous results from both the mathematical and physical literature, such as those in the papers by Greiner,[5] Gilkey,[6] and Romanov and Schwartz,[7] which explicitly exhibit the connection between the Seeley's kernel and the coefficients of the heat kernel expansion, in the case of a differential operator. As a consequence, an elliptic differential operator acting on an odd-dimensional compact manifold without boundary has zero index. From a physical point of view this may be interpreted as the absence of anomalies. Here we recover this fact.

ACKNOWLEDGMENTS

One of us (A.P.C.M.) wants to thank his colleagues from the Centre de Physique Théorique-Ecole Polytechnique for the kind hospitality since January 1989. He is particularly indebted to C. de Calan and P. Renouard for very fruitful discussions.

[1] A. P. C. Malbouisson, F. R. A. Simao, and M. A. do Rego Monteiro, J. Math Phys. **30**, 2016 (1989).

[2] R. T. Seeley, Am. Math. Soc. Proc. Symp. Pure Math **10**, 288 (1967).

[3] B. de Witt, *The Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); M. D. Birrel and P. C. Davies, *Quantum Fields in Curved Spaces* (Cambridge U.P., Cambridge, 1984).

[4] G. Cognola and S. Zerbini, Phys. Lett. B **195**, 435 (1988); Mod. Phys. Lett. A **3**, 599 (1988).

[5] P. Greiner, Arch. Rat. Mech. Anal. **41**, 163 (1971).

[6] P. Gilkey, J. Differential Geom. **10**, 601 (1975).

[7] V. N. Romanov and A. S. Schwartz, Theor. Math. Phys. **41**, 190 (1979) (English translation).

Averaging principle and systems of singularly perturbed stochastic differential equations

J. Golec
Department of Mathematics, Fordham University, Bronx, New York 10458

G. Ladde
The University of Texas at Arlington, Arlington, Texas 76019

(Received 26 April 1989; accepted for publication 13 December 1989)

By developing certain auxiliary results, a modified version of the stochastic averaging principle is developed to investigate dynamical systems consisting of fast and slow phenomena. Moreover, an attempt is made to establish a relationship between the averaging assumption and certain ergodic-type properties of the random process determined by an auxiliary system of stochastic differential equations. Finally, an example is given to illustrate the scope of the results.

I. INTRODUCTION

In physical and engineering sciences several dynamical systems possess slow and fast phenomena. A mathematical model for such systems is described by a system of singularly perturbed differential equations. For example, singularly perturbed systems very often occur due to the presence of small "parasitic" parameters (masses, capacitors, etc.) multiplying time derivatives of some of the state variables. A simplified model of such systems is often based on the assumption that during the fast transients, the slow state variables remain constant and by the time their changes become noticeable, the fast transients have already reached their quasisteady states. Then the only state variables used for short term studies are the fast variables while the slow state variables are considered as constants. In long term studies the model is formed by the slow state variables. This common practice is mathematically inconsistent because it treats the time-varying quasisteady state variables as constants, that is it neglects their derivatives. Instead, a rigorous approach to this type of a model simplification is to treat it as a multitime scale singular perturbation problem. The difficulties in studying such a problem are due to cross interactions of slow and fast modes.

The main objective of the present work is to formulate sufficient conditions for the exact mathematical model of a dynamical system consisting of slow and fast state variables to be close to its simplified form as mentioned above. The present result extends and generalizes the work in Ref. 1. The paper is organized as follows. Section II deals with certain notions and definitions. In Sec. III some auxiliary results are developed. These results will be utilized in the subsequent discussion. In Sec. IV we state and prove a stochastic version of the averaging principle.[1] This result is applicable for systems of singularly perturbed stochastic differential equations with coefficients that admit certain growth conditions. Furthermore, sufficient conditions are given to express the drift of the limit process in terms of an ergodic measure. Finally, the result is illustrated by an example.

II. DEFINITIONS

Let us consider the following stochastic singularly perturbed system (SSPS):

$$dX_t^\epsilon = b_1(X_t^\epsilon, Y_t^\epsilon)dt + \sigma_1(X_t^\epsilon, Y_t^\epsilon)dW_t, \quad X_0^\epsilon = X_0,$$
$$dY_t^\epsilon = (1/\epsilon)b_2(X_t^\epsilon, Y_t^\epsilon)dt + (1/\sqrt{\epsilon})\sigma_2(X_t^\epsilon, Y_t^\epsilon)dW_t, \quad Y_0^\epsilon = Y_0, \tag{2.1}$$

for $t\in[0,T_0]$, where $\epsilon\in(0,1)$, $b_1 = b_1(x,y)$, and $b_2 = b_2(x,y)$ are n and l-dimensional vectors (respectively), $\sigma_2 = \sigma_1(x,y)$ and $\sigma_2 = \sigma_2(x,y)$ are $n\times m$ and $l\times m$ matrices (respectively) for $x\in R^n$, $y\in R^1$, W denotes the standard m-dimensional Wiener process, X_0 and Y_0 are initial random vectors. Let us assume that the classical conditions[2,3] for existence and uniqueness of L^2-solutions of system (2.1) are satisfied. In particular, from now on we assume that functions b_1, b_2, σ_1, σ_2 satisfy the global Lipschitz condition and the sublinear growth condition.[2,3]

Let us denote by $(\mathcal{F}_t)_{t\geqslant 0}$ the standard filtration corresponding to the Wiener process W. We will also use another m-dimensional standard Wiener process $\overline{W} = (\overline{W}_t)_{t\geqslant 0}$ defined on the same probability space and independent of W. The existence of such a process may be assumed without loss of generality by considering (if necessary) an augmentation of the underlying probability space. For any $t\in[0,T_0]$ and any $\mathcal{F}_t$-measurable n and l-dimensional random vectors Z and V (respectively) we introduce a random process $(Y_s^{ZV})_{s\geqslant 0}$ which is defined by means of the following system of stochastic differential equations:

$$dY_s^{ZV} = b_2(Z, Y_s^{ZV})ds + \sigma_2(Z, Y_s^{ZV})d\overline{W}_s, \quad Y_0^{ZV} = V. \tag{2.2}$$

Under the above assumptions on b_2, σ_2, and the Wiener processes W, $\overline{W}$, the L^2 solution of system (2.2) exists and is uniquely determined on the infinite time-interval $[0,\infty)$.

For any real numbers α,β $(\alpha<\beta)$ and for a natural number d let us denote by $M^2([\alpha,\beta];R^d)$ the vector space of all adapted, measurable d-dimensional random vectors $(g_t)_{t\in[\alpha,\beta]}$ for which $\int_\alpha^\beta E\|g_t\|^2\,dt$ is finite.

III. AUXILIARY RESULTS

In this section we present some results that will be used in the subsequent discussion. First two simple lemmas deal

 0022-2488/90/051116-08$03.00

with shifting and rescaling of the time interval in a stochastic integral.

Lemma 3.1: Let $M=(M_t)_{t\in[\beta,\beta+\Delta]}$ be a random process with values in the vector space of $p\times m$ matrices ($p\in N$, $\beta>0, \Delta>0$). We also assume that $M\in M^2([\beta,\beta+\Delta];R^{pm})$. Then

$$\int_\beta^{\beta+\Delta} M_t\, dW_t = \int_0^{\Delta} M_{\beta+t}\, dW_t^*,$$

where $W^*=(W_t^*)_{t\geqslant 0}$ is a "shifted" Wiener process defined as $W_t^*=W_{t+\beta}-W_\beta$ for $t\geqslant 0$.

Proof: The stochastic integral on the right-hand side of the formula exists. Indeed, it is easy to see that $\mathcal{G}_t=\mathcal{F}_{\beta+t}$ for $t\geqslant 0$, where $\mathcal{G}=(\mathcal{G}_t)_{t\geqslant 0}$ is the standard filtration determined by the shifted process W^*. The measurability requirement for a random process in stochastic integral follows from this. It is easy to prove assertion of the lemma for M being a step process. In the general case the proof follows from the standard limiting procedure.

Lemma 3.2: Let $M=(M_t)_{t\in[0,\beta/\epsilon]}$ be a matrix-valued stochastic process such that $M\in M^2([0,\beta/\epsilon];R^{pm})$, where $p\in N$, $\beta>0$, $\epsilon\in(0,1)$. Then

$$\int_0^{\beta/\epsilon} M_t\, dW_t = \frac{1}{\sqrt{\epsilon}}\int_0^{\beta} M_{t/\epsilon}\, dW_t^{**},$$

where $W^{**}=(W_t^{**})_{t\geqslant 0}$ is a "rescaled" Wiener process defined as $W_t^{**}=\sqrt{\epsilon}W_{t/\epsilon}$ for $t\geqslant 0$.

Proof: We note that because of the form of the rescaled process W^{**} and assumed measurability conditions for M the stochastic integral on the right hand of the formula exists. It is easy to prove assertion of the lemma for a step process M. The proof in general case follows from passing into the limit.

Lemma 3.3: Let us suppose that in addition to the standing assumptions on the coefficients of SSPS (2.1) the following conditions are also satisfied:

(i) $xb_1(x,y)+yb_2(x,y)\leqslant -\alpha(|x|^2+|y|^2)$ for $x\in R^n$, $y\in R^l$, where $\alpha>0$ and where for any vectors A and B we denote by AB their scalar product,

(ii) σ_i is globally bounded ($i=1,2$).

Then there exists a positive constant C such that for any $\epsilon\in(0,1)$

$$\sup_{t\geqslant 0} E|X_t^\epsilon|^2\leqslant C \text{ and } \sup_{t\geqslant 0} E|Y_t^\epsilon|^2\leqslant C/\epsilon,$$

where X^ϵ and Y^ϵ are the solution processes of SSPS (2.1) defined on the infinite time-interval.

Proof: By applying the Itô formula for the function $f(t,z)=\exp(2\alpha t)|z|^2$ ($t\geqslant 0$, $z\in R^n$ or $z\in R^l$) and the solution processes of system (2.1) we obtain

$$\begin{aligned} d\{e^{2\alpha t}|X_t|^2\} &= [2\alpha|X_t|^2+2X_t b_1(X_t,Y_t) \\ &\quad +|\sigma_1(X_t,Y_t)|^2]e^{2\alpha t}\,dt \\ &\quad +2X_t^*\sigma_1(X_t,Y_t)e^{2\alpha t}\,dW_t, \end{aligned} \tag{3.1}$$

and also

$$\begin{aligned} d\{e^{2\alpha t}|Y_t|^2\} &= \Big[2\alpha|Y_t|^2+\frac{2}{\epsilon}Y_t b_2(X_t,Y_t) \\ &\quad +\frac{1}{\epsilon}|\sigma_2(X_t,Y_t)|^2\Big]e^{2\alpha t}\,dt \\ &\quad +\frac{2}{\sqrt{\epsilon}}Y_t^*\sigma_2(X_t,Y_t)e^{2\alpha t}\,dW_t, \end{aligned} \tag{3.2}$$

where * denotes the transposition and for notational simplicity we have supressed the superscript ϵ. After multiplying (3.2) by ϵ, adding integral forms of (3.1) and (3.2), taking expected values on both sides of the resulting equality, and using assumptions, we obtain

$$\begin{aligned} &E(|X_t|^2+\epsilon|Y_t|^2)e^{2\alpha t} \\ &\leqslant E(|X_0|^2+\epsilon|Y_0|^2)+\int_0^t [2\alpha(E|X_s|^2+\epsilon E|Y_s|^2) \\ &\quad -2\alpha(E|X_s|^2+E|Y_s|^2)+C_1]e^{2\alpha s}\,ds \\ &\leqslant E(|X_0|^2+\epsilon|Y_0|^2)+\int_0^t C_1 e^{2\alpha s}\,ds \\ &\leqslant C_2+C_1/2\alpha)e^{2\alpha t}, \end{aligned} \tag{3.3}$$

for $t\geqslant 0$, where C_1, C_2 are positive constants. Inequality (3.3) implies that

$$E|X_t|^2\leqslant C_2 e^{-2\alpha t}+C_1/2\alpha\leqslant C \text{ and also } \epsilon E|Y_t|^2\leqslant C$$

for $t\geqslant 0$, where C is a constant independent of ϵ and t. The proof of the lemma is complete.

Next, we introduce some auxiliary random functions. Let us fix a positive number Δ and a natural number k such that $k\Delta<T_0$. For any $\epsilon\in(0,1)$ we define a random process $\hat{Y}^\epsilon$ by means of the following relation:

$$\begin{aligned} \hat{Y}_t^\epsilon &= Y_{k\Delta}^\epsilon+\frac{1}{\epsilon}\int_{k\Delta}^t b_2(X_{k\Delta}^\epsilon,\hat{Y}_s^\epsilon)\,ds \\ &\quad +\frac{1}{\sqrt{\epsilon}}\int_{k\Delta}^t \sigma_2(X_{k\Delta}^\epsilon,\hat{Y}_s^\epsilon)\,dW_s \end{aligned} \tag{3.4}$$

for $t\in[k\Delta,\min((k+1)\Delta,T_0))$, where $X_{k\Delta}^\epsilon$ and $Y_{k\Delta}^\epsilon$ are slow and fast solution processes at time $k\Delta$. Let us also define the following process $\hat{X}^\epsilon$:

$$\hat{X}_t^\epsilon = X_0+\int_0^t b_1(X_{[s/\Delta]\Delta}^\epsilon,\hat{Y}_s^\epsilon)\,ds+\int_0^t \sigma_1(X_s^\epsilon,Y_s^\epsilon)\,dW_s \tag{3.5}$$

for $t\in[0,T_0]$, where $[*]$ denotes the greatest integer function. For simplicity, we do not indicate explicitly the dependence of processes $\hat{X}^\epsilon$ and $\hat{Y}^\epsilon$ on the grid size Δ.

By applying Lemma 3.3 we will prove the following fact:

Lemma 3.4: Let us suppose that the assumptions of Lemma 3.3 hold. Then

$$\sup_{t\in[0,T_0]} E|X_t^\epsilon-X_{[t/\Delta]\Delta}^\epsilon|^2\leqslant C\frac{\Delta^2}{\epsilon}+D\Delta,$$

where $\Delta\in(0,T_0]$, C, D are positive constants, and X^ϵ is the slow solution process.

Proof: Let us fix any $t\in[0,T_0]$ and let $t\in[k\Delta,(k+1)\Delta)$. We can write

$$X_t^\epsilon - X_{k\Delta}^\epsilon = \int_{k\Delta}^t b_1(X_s^\epsilon, Y_s^\epsilon)ds + \int_{k\Delta}^t \sigma_1(X_s^\epsilon, Y_s^\epsilon)dW_s.$$

Using the Schwarz inequality we obtain

$$E|X_t^\epsilon - X_{k\Delta}^\epsilon|^2 \leqslant 2\Delta \int_{k\Delta}^t E|b_1(X_s^\epsilon, Y_s^\epsilon)^2 ds + 2\int_{k\Delta}^t E|\sigma_1(X_s^\epsilon, Y_s^\epsilon)|^2 ds. \tag{3.6}$$

By applying Lemma 3.3 and using assumptions for b_1 and σ_1 we conclude that

$$E|b_1(X_s^\epsilon, Y_s^\epsilon)|^2 \leqslant C_1/\epsilon \text{ and also } E|\sigma_1(X_s^\epsilon, Y_s^\epsilon)|^2 \leqslant C_2, \tag{3.7}$$

for $s \in [0, T_0]$, where C_1, C_2 are positive constants independent of ϵ and s. Combining (3.6) and (3.7) we obtain the desired estimate for $t \in [k\Delta, (k+1)\Delta)$. The proof of the lemma is complete by noting that the estimate is independent of k.

Let us formulate a remark that will be used in the next lemma.

Remark 3.1. For any positive number L there exists a positive real-valued function $\Delta = \Delta(\epsilon)$ defined for $\epsilon \in (0,1)$ such that

$$\Delta(\epsilon) \to 0, \quad \Delta(\epsilon)/\epsilon \to \infty, \quad (\Delta(\epsilon))^2/\epsilon \to 0,$$
$$(\Delta(\epsilon)/\epsilon)^3 \Delta(\epsilon) \exp\{L(\Delta(\epsilon)/\epsilon)^2\} \to 0 \text{ as } \epsilon \to 0^+.$$

Indeed, it is easy to check that given a positive number L the following function:

$$\Delta(\epsilon) = \epsilon[-(1/2(L+2)) \ln \epsilon]^{1/2}, \quad \epsilon \in (0,1),$$

satisfies the properties of Remark 3.1.

The next lemma establishes mean-square convergence of the auxiliary process (3.4) to the fast solution process Y^ϵ.

Lemma 3.5: Let the assumptions of Lemma 3.3 hold. Then it is possible to choose an ϵ-dependent grid size $\Delta = \Delta(\epsilon)$ in such a way that

$$\sup_{t \in [0, T_0]} E|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \to 0 \text{ as } \epsilon \to 0^+,$$

where Y^ϵ is the fast solution process and $\hat{Y}^\epsilon$ is the auxiliary process (3.4).

Proof: For any $t \in [0, T_0]$ with $t \in [k\Delta, (k+1)\Delta)$ we can write

$$Y_t^\epsilon - \hat{Y}_t^\epsilon = \frac{1}{\epsilon}\int_{k\Delta}^t [b_2(X_s^\epsilon, Y_s^\epsilon) - b_2(X_{k\Delta}^\epsilon, \hat{Y}_s^\epsilon)]ds + \frac{1}{\sqrt{\epsilon}}\int_{k\Delta}^t [\sigma_2(X_s^\epsilon, Y_s^\epsilon) - \sigma_2(X_{k\Delta}^\epsilon, \hat{Y}_s^\epsilon)]dW_s.$$

This implies that

$$\begin{aligned} E|Y_t^s - \hat{Y}_t^\epsilon|^2 &\leqslant C_1 \frac{\Delta}{\epsilon^2}\int_{k\Delta}^t [E|X_s^\epsilon - X_{k\Delta}^\epsilon|^2 + E|Y_s^\epsilon - \hat{Y}_s^\epsilon|^2]ds + C_2\frac{1}{\epsilon}\int_{k\Delta}^t [E|X_s^\epsilon - X_{k\Delta}^\epsilon|^2 + E|Y_s^\epsilon - \hat{Y}_s^\epsilon|^2]ds \\ &\leqslant C_3\left(\frac{\Delta}{\epsilon^2} + \frac{1}{\epsilon}\right)\int_{k\Delta}^t [E|X_s^\epsilon - X_{d\Delta}^\epsilon|^2 + E|Y_s^\epsilon - \hat{Y}_s^\epsilon|^2]ds, \end{aligned} \tag{3.8}$$

where C_1, C_2, C_3 are positive constants independent of ϵ, Δ, and k. Using Lemma 3.4 and inequality (3.8) we obtain the following estimate:

$$E|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leqslant C_4\left[\left(\frac{\Delta}{\epsilon}\right)^3 + 2\left(\frac{\Delta}{\epsilon}\right)^2 + \frac{\Delta}{\epsilon}\right]\Delta + C_3\left(\frac{\Delta}{\epsilon^2} + \frac{1}{\epsilon}\right)\int_{k\Delta}^t E|Y_s^\epsilon - \hat{Y}_s^\epsilon|^2 ds \tag{3.9}$$

for $t \in [k\Delta, (k+1)\Delta)$, where C_4 is a positive constant. Applying Gronwell's inequality it follows that

$$E|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leqslant C_4\left[\left(\frac{\Delta}{\epsilon}\right)^3 + 2\left(\frac{\Delta}{\epsilon}\right)^2 + \frac{\Delta}{\epsilon}\right]\Delta \times \exp\left\{C_3\left[\left(\frac{\Delta}{\epsilon}\right)^2 + \frac{\Delta}{\epsilon}\right]\right\}, \tag{3.10}$$

for $t \in [k\Delta, (k+1)\Delta)$. The right-hand side of inequality (3.10) does not depend on k, so we can conclude that the estimate remains valid for $t \in [0, T_0]$. Now, the assertion of the lemma follows from Remark 3.1. The proof is complete.

We conclude the present section by formulating the following lemma. The proof is an immediate consequence of Lemma 3.4, Lemma 3.5 and Remark 3.1.

Lemma 3.6: Let the assumptions of Lemma 3.3 hold. Then it is possible to choose an ϵ-dependent grid size $\Delta = \Delta(\epsilon)$ in such a way that

$$\sup_{t \in [0, T_0]} E|X_t^\epsilon - \hat{X}_t^\epsilon|^2 \to 0 \text{ as } \epsilon \to 0$$

where X^ϵ is the slow solution process and $\hat{X}^\epsilon$ is the auxiliary process (3.5).

IV. GENERALIZED STOCHASTIC AVERAGING PRINCIPLE

We are ready to present a generalized version of the stochastic averaging principle. In the following we assume that the function σ_1 in SSPS (2.1) does not depend on y, that is $\sigma_1 = \sigma_1(x)$. In the context of system (2.1) this means that the fast process Y^ϵ does not interfere with the diffusion part of the slow process X^ϵ. In the course of the proof of the averaging principle whenever Δ and ϵ appear together we will always assume that Δ is ϵ-dependent and that the function $\Delta = \Delta(\epsilon)$ satisfies properties of Remark 3.1.

Theorem 4.1 *(stochastic averaging principle)*: Let the assumptions of Lemma 3.3 be satisfied. Let us suppose that for some function $\bar{b} = \bar{b}(x)$ and a number $\eta \in (0,1)$ the following averaging assumption holds:

$$\sup_{t \in [0, T_0], \epsilon \in (0, \eta)} \left\{E\left|\frac{1}{T}\int_0^T b_1(X_t^\epsilon, Y_s^{X_t^\epsilon Y_t^\epsilon})ds - \bar{b}(X_t^\epsilon)\right|^2\right\} \to 0 \text{ as } T \to \infty,$$

where $Y^{X_t^\epsilon Y_t^\epsilon}$ denotes the random process defined by (2.2), with $Z = X_t^\epsilon$ and $V = Y_t^\epsilon (t \in [0, T_0])$. Furthermore, we assume that $\bar{b}$ is a Lipschitz continuous function and let us denote by $\bar{X}$ the random process defined by means of the following system of stochastic differential equations:

$$d\bar{X}_t = \bar{b}(\bar{X}_t)dt + \sigma_1(\bar{X}_t)dW_t, \quad \bar{X}_0 = X_0. \tag{4.1}$$

Under the above assumptions

$$\sup_{t\in[0,T_0]} E|X_t^\epsilon - \bar{X}_t|^2 \to 0 \text{ as } \epsilon \to 0^+.$$

Proof: In view of Lemma 3.6 it is enough to prove that

$$\sup_{t\in[0,T_0]} E|\hat{X}_t^\epsilon - \bar{X}_t|^2 \to 0 \text{ as } \epsilon \to 0^+, \tag{4.2}$$

where $\hat{X}^\epsilon$ is the auxiliary process (3.5) and $\bar{X}$ is the solution process of (4.1). For any $t\in[0,T_0]$ we have

$$\begin{aligned}\hat{X}_t^\epsilon - \bar{X}_t = &\int_0^t [b_1(X^\epsilon_{[s/\Delta]\Delta}, \hat{Y}_s^\epsilon) - \bar{b}(X_s^\epsilon)]ds + \int_0^t [\bar{b}(X_s^\epsilon) \\ &- \bar{b}(\hat{X}_s^\epsilon)]ds + \int_0^t [\bar{b}(\hat{X}_s^\epsilon) - \bar{b}(\bar{X}_s)]ds \\ &+ \int_0^t [\sigma_1(X_s^\epsilon) - \sigma_1(\hat{X}_s^\epsilon)]dW_s \\ &+ \int_0^t [\sigma_1(\hat{X}_s^\epsilon) - \sigma_1(\bar{X}_s)]dW_s. \end{aligned} \tag{4.3}$$

Let us define

$$M_\epsilon = \sup_{s\in[0,T_0]} E|X_s^\epsilon - \hat{X}_s^\epsilon|^2 \text{ and } m_\epsilon(t) = E|\hat{X}_t^\epsilon - \bar{X}_t|^2,$$

for $t\in[0,T_0]$. Using the Schwartz inequality we obtain

$$\begin{aligned}&E\left|\int_0^t [\bar{b}(X_s^\epsilon) - \bar{b}(\hat{X}_s^\epsilon)]ds\right|^2 \\ &\leqslant T_0 \int_0^{T_0} E|\bar{b}(X_s^\epsilon) - \bar{b}(\hat{X}_s^\epsilon)|^2 ds \leqslant C_1 M_\epsilon, \end{aligned} \tag{4.4}$$

$$\begin{aligned}&E\left|\int_0^t [\bar{b}(\hat{X}_s^\epsilon) - \bar{b}(\bar{X}_s)]ds\right|^2 \\ &\leqslant C_2 \int_0^t m_\epsilon(s)ds, \end{aligned} \tag{4.5}$$

for $t\in[0,T_0]$, where C_1, C_2, are positive constants. We also have

$$E\left|\int_0^t [\sigma_1(X_s^\epsilon) - \sigma_1(\hat{X}_s^\epsilon)]dW_s\right|^2 \leqslant C_3 M_\epsilon, \tag{4.6}$$

$$E\left|\int_0^t [\sigma_1(\hat{X}_s^\epsilon) - \sigma_1(\bar{X}_s)]dW_s\right|^2 \leqslant C_4 \int_0^t m_\epsilon(s)ds, \tag{4.7}$$

for $t\in[0,T_0)$, where C_3, C_4 are positive constants. Dealing with the first term in (4.3) we use the averaging assumption. For any $t\in[0,T_0]$ with $t\in[k\Delta,(k+1)\Delta)$ we can write

$$\begin{aligned}&\int_0^t [b_1(X^\epsilon_{[s/\Delta]\Delta}, \hat{Y}_s^\epsilon) - \bar{b}(X_s^\epsilon)]ds \\ &= \sum_{p=0}^{k-1} \int_{p\Delta}^{(p+1)\Delta} [b_1(X_{p\Delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{b}(X_{p\Delta}^\epsilon)]ds \\ &\quad + \sum_{p=0}^{k-1} \int_{p\Delta}^{(p+1)\Delta} [\bar{b}(X_{p\Delta}^\epsilon) - \bar{b}(X_s^\epsilon)]ds \\ &\quad + \int_{k\Delta}^t [b_1(X_{k\Delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{b}(X_s^\epsilon)]ds. \end{aligned} \tag{4.8}$$

Let us denote by T_1, T_2, T_3 the three terms in (4.8). We observe that

$$\begin{aligned}E|T_2|^2 &= E\left|\int_0^{k\Delta} [\bar{b}(X^\epsilon_{[s/\Delta]\Delta}) - \bar{b}(X_s^\epsilon)]ds\right|^2 \\ &\leqslant C_5 T_0 \int_0^{T_0} E|X^\epsilon_{[s/\Delta]\Delta} - X_s^\epsilon|^2 ds. \end{aligned}$$

In view of Lemma 3.3 and the special choice of the function $\Delta = \Delta(\epsilon)$ it follows from the last inequality that

$$E|T_2|^2 \to 0 \text{ as } \epsilon \to 0^+, \text{ uniformly in } t\in[0,T_0]. \tag{4.9}$$

Next, because of the Lipschitz continuity of b_1 and $\bar{b}$ we have

$$\begin{aligned}E|T_3|^2 \leqslant C_6\Delta \int_{k\Delta}^t [&C_7 + E|X_{k\Delta}^\epsilon|^2 + E|\hat{Y}_s^\epsilon|^2 \\ &+ E|X_s^\epsilon|^2]ds. \end{aligned}$$

From Lemma 3.3 and Lemma 3.5 the above inequality reduces to

$$E|T_3|^2 \leqslant C_8\Delta + C_9(\Delta^2/\epsilon),$$

with positive constants C_8, C_9. Using again the special choice of Δ we conclude that

$$E|T_3|^2 \to 0 \text{ as } \epsilon \to 0^+, \text{ uniformly in } t\in[0,T_0]. \tag{4.10}$$

Now, let us observe that

$$\begin{aligned}E|T_1|^2 &\leqslant k \sum_{p=0}^{k-1} E\left|\int_{p\Delta}^{(p+1)\Delta} [b_1(X_{p\Delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{b}(X_{p\Delta}^\epsilon)]ds\right|^2 \\ &= k\Delta^2 \sum_{p=0}^{k-1} E\left|\frac{1}{\Delta}\int_{p\Delta}^{(p+1)\Delta} b_1(X_{p\Delta}^\epsilon, \hat{Y}_s^\epsilon)ds \right. \\ &\quad \left. - \bar{b}(X_{p\Delta}^\epsilon)\right|^2 \\ &= k\Delta^2 \sum_{p=0}^{k-1} E\left|\frac{1}{\Delta}\int_0^{\Delta} b_1(X_{p\Delta}^\epsilon, \hat{Y}^\epsilon_{p\Delta+s})ds \right. \\ &\quad \left. - \bar{b}(X_{p\Delta}^\epsilon)\right|^2. \end{aligned} \tag{4.11}$$

For any fixed p and $s\in[0,\Delta)$ let us define

$$Z_s = \hat{Y}^\epsilon_{p\Delta+s} \text{ and } V_s = Y^{X^\epsilon_{p\Delta} Y^\epsilon_{p\Delta}}_{s/\epsilon}.$$

Applying Lemma 3.1 we can write

$$\begin{aligned}Z_s &= Y_{p\Delta}^\epsilon + \frac{1}{\epsilon}\int_{p\Delta}^{p\Delta+s} b_2(X_{p\Delta}^\epsilon, \hat{Y}_u^\epsilon)du \\ &\quad + \frac{1}{\sqrt{\epsilon}}\int_{p\Delta}^{p\Delta+s} \sigma_2(X_{p\Delta}^\epsilon, \hat{Y}_u^\epsilon)dW_u \\ &= Y_{p\Delta}^\epsilon + \frac{1}{\epsilon}\int_0^s b_2(X_{p\Delta}^\epsilon, Z_u)du \\ &\quad + \frac{1}{\sqrt{\epsilon}}\int_0^s \sigma_2(X_{p\Delta}^\epsilon, Z_u)dW_u^*, \end{aligned} \tag{4.12}$$

for $s\in[0,\Delta)$, where W^* is the shifted Wiener process defined as in Lemma 3.1. Similarly, by using definition of the process $Y^{X^\epsilon_{p\Delta} Y^\epsilon_{p\Delta}}$ and Lemma 3.2 we can write

$$V_s = Y^{\epsilon}_{p\Delta} + \frac{1}{\epsilon}\int_0^s b_2(X^{\epsilon}_{p\Delta}, V_u)du + \frac{1}{\sqrt{\epsilon}}\int_0^s \sigma_2(X^{\epsilon}_{p\Delta}, V_u)d\overline{W}^{**}_u, \tag{4.13}$$

for $s\in[0,\Delta)$, where $\overline{W}^{**}$ is the rescaled Wiener process $\overline{W}$, defined as in Lemma 3.2. By comparing (4.12) and (4.13) we come to the conclusion that the process

$$(X^{\epsilon}_{p\Delta}, \hat{Y}^{\epsilon}_{p\Delta+s}) \text{ and } (X^{\epsilon}_{p\Delta}, Y^{X^{\epsilon}_{p\Delta}Y^{\epsilon}_{p\Delta}}_{s/\epsilon})$$

have the same distributions for any $s\in[0,\Delta)$. Therefore, using (4.11) we obtain

$$\begin{aligned} E|T_1|^2 &\leqslant k\Delta^2 \sum_{p=0}^{k-1} E\left|\frac{1}{\Delta}\int_0^{\Delta} b_1(X^{\epsilon}_{p\Delta}, Y^{X^{\epsilon}_{p\Delta}Y^{\epsilon}_{p\Delta}}_{s/\epsilon})ds - \bar{b}(X^{\epsilon}_{p\Delta})\right|^2 \\ &= k\Delta^2 \sum_{p=0}^{k-1} E\left|\frac{\epsilon}{\Delta}\int_0^{\Delta/\epsilon} b_1(X^{\epsilon}_{p\Delta}, Y_s^{X^{\epsilon}_{p\Delta}Y^{\epsilon}_{p\Delta}})ds - \bar{b}(X^{\epsilon}_{p\Delta})\right|^2. \end{aligned} \tag{4.14}$$

Let ρ be an arbitrary positive number. Applying the averaging assumption we can find a constant $\tau_1 > 0$ such that for $\Delta/\epsilon > \tau_1$ all terms under the summation sign in (4.14) are less than ρ. Therefore we have

$$E|T_1|^2 \leqslant k\Delta^2 kp \leqslant T_0^2\rho, \tag{4.15}$$

for sufficiently small ϵ. Since the estimate (4.15) is k-independent we conclude that

$$E|T_1|^2 \to 0 \text{ as } \epsilon\to 0^+, \text{ uniformly in } t\in[0,T_0]. \tag{4.16}$$

From (4.9), (4.10), (4.16), and (4.8) we observe that the first term in (4.3) satisfies the following property:

$$E\left|\int_0^t \left[b_1(X_{[\frac{s}{\Delta}]\Delta}, \hat{Y}^{\epsilon}_s) - \bar{b}(X^{\epsilon}_s)\right]ds\right|^2 \to 0$$

as $\epsilon\to 0^+$, uniformly in $t\in[0,T_0]$. (4.17)

In view of (4.3), (4.4), (4.5), (4.6), (4.7), and (4.17) we obtain the following inequality:

$$m_{\epsilon}(t) \leqslant L_{\epsilon} + C_{10}\int_0^t m_{\epsilon}(s)ds, \tag{4.18}$$

for $t\in[0,T_0]$, where C_{10} is a positive constant and L_{ϵ} is an ϵ-dependent constant. It follows from Lemma 3.6 and from (4.17) that $L_{\epsilon}\to 0$ as $\epsilon\to 0^+$. Now, applying Gronwall's inequality we obtain

$$m_{\epsilon}(t) \leqslant L_{\epsilon}e^{C_{10}t} \leqslant L_{\epsilon}e^{C_{10}T_0}, \tag{4.19}$$

for $t\in[0,T_0]$. Finally, from (4.19) we have

$$\sup_{t\in[0,T_0]} m_{\epsilon}(t) \to 0 \text{ as } \epsilon\to 0^+.$$

The proof of the theorem is complete.

Remark 4.1: Let us note that if for any $x\in R^n$, $y\in R^l$

$$E\left|\frac{1}{T}\int_0^T b_1(x, Y^{xy}_s)ds - \bar{b}(x)\right|^2 \to 0 \text{ as } T\to\infty,$$

and the above limit is uniform with respect to $x\in R^n$ and $y\in R^l$ then the averaging assumption is satisfied.

Verification of Remark 4.1 follows immediately from the equality

$$\begin{aligned} &E\left|\frac{1}{T}\int_0^T b_1(X^{\epsilon}_t, Y_s^{X^{\epsilon}_t Y^{\epsilon}_t})ds - \bar{b}(X^{\epsilon}_t)\right|^2 \\ &\quad = EE_{X^{\epsilon}_t Y^{\epsilon}_t}\left|\frac{1}{T}\int_0^T b_1(X^{\epsilon}_t, Y_s^{X^{\epsilon}_t Y^{\epsilon}_t})ds - \bar{b}(X^{\epsilon}_t)\right|^2 \end{aligned}$$

where E with the subscript denotes conditional expectation with respect to the specified random variables.

The averaging assumption that plays a crucial role in the proof of the averaging principle may be difficult to verify in practice. Therefore any method for obtaining a candidate for the function $\bar{b}$ in the formulation of Theorem 4.1 is of a particular interest. Under certain ergodic type properties of the process Y^{xy}, a suitable choice of $\bar{b}$ will be exhibited in the following theorem.

Theorem 4.2: Let us suppose that $b_2(x,y)$ is twice continuously differentiable with respect to y, for each $x\in R^n$, and

$$\mu\left(\frac{\partial b_2}{\partial y}(x,\bar{y}(t,0,y))\right) \leqslant -\beta(x), \tag{4.20}$$

where $\beta(x)$ is an x-dependent positive number and $\bar{y}(t,0,y)$ is the solution of

$$\bar{y}' = b_2(x,\bar{y}), \quad \bar{y}(0) = y,$$

where μ is the logarithmic norm[2] of the Jacobian matrix function $(\partial b_2/\partial y)(x,\bar{y}(t,0,y))$. Furthermore, let us assume that for any $x\in R^n$ there exists a probability Borel measure μ^x on R^l such that

$$\left|E_y f(Y^{xy}_t) - \int_{R^l} f(z)d\mu^x(z)\right| \leqslant C(x)e^{-\lambda(x)t}, \tag{4.21}$$

for $y\in R^l$, $t\geqslant 0$ and for any continuous function f defined on R^l into R^n and having the following property:

$$\sup_{s>0} E|f(Y^{xy}_s)|^2 < \infty, \quad \text{for } x\in R^n, y\in R^l. \tag{4.22}$$

In (4.21) $C(x)$, $\lambda(x)$ are positive constants possibly depending on x. Let us define

$$\bar{b}(x) = \int_{R^l} b_1(x,z)d\mu^x(z), \quad (x\in R^n).$$

Then for any $x\in R^n$, $y\in R^l$

$$E\left|\frac{1}{T}\int_0^T b_1(x, Y^{xy}_s)ds - \bar{b}(x)\right|^2 \to 0 \text{ as } T\to\infty.$$

Proof: By applying the Itô formula with regard to the random process $Z_u = \bar{y}(s,u,Y^{xy}_u)$, using assumptions and the result in Ref. 2 one can prove that

$$\sup_{s>0} E|Y^{xy}_s|^2 < \infty, \tag{4.23}$$

for $x\in R^n$, $y\in R^l$. Now, we can immediately verify relation (4.22) for the function $f(z) = b_1(x,z)$ with a fixed $x\in R^n$. For any $x\in R^n$, $y\in R^l$ we can write

$$A_T = E\left|\frac{1}{T}\int_0^T [b_1(x,Y_s^{xy}) - \bar{b}(x)]ds\right|^2$$
$$= \frac{1}{T^2}\int_0^T\int_0^T E\{[b_1(x,Y_s^{xy}) - \bar{b}(x)]b_1(x,Y_t^{xy}) - \bar{b}(x)]\}dt\,ds$$
$$= \frac{2}{T^2}\int_0^T\int_s^T E\{[b_1(x,Y_t^{xy}) - \bar{b}(x)][b_1(x,Y_s^{xy}) - \bar{b}(x)]\}dt\,ds. \quad (4.24)$$

For $0\leqslant s\leqslant t$ the integrand in (4.24) may be estimated as follows:

$$E\{[b_1(x,Y_t^{xy}) - \bar{b}(x)][b_1(x,Y_s^{xy}) - \bar{b}(x)]\}$$
$$= E\{[b_1(x,Y_s^{xy}) - \bar{b}(x)] \times E[(b_1(x_1,Y_t^{xy}) - \bar{b}(x))|Y_s^{xy}]\}$$
$$= E\{[b_1(x,Y_s^{xy}) - \bar{b}(x)]E_{Y_s^{xy}} \times [b_1(x,Y_{t-s}^{xY_s^{xy}}) - \bar{b}(x)]\}$$
$$\leqslant M\left\{E\left|E_{Y_s^{xy}}\left[b_1(x,Y_{t-s}^{xY_s^{xy}}) - \int_{R^r} b_1(x,z)d\mu^x(z)\right]\right|^2\right\}^{1/2}$$
$$\leqslant MC(x)e^{-\lambda(x)(t-s)},$$

where M is a positive constant. The above estimate together with (4.24) implies that

$$A_T \leqslant \frac{2MC(x)}{T^2}\int_0^T\int_s^T e^{-\lambda(x)(t-s)}dt\,ds$$
$$= \frac{2MC(x)}{\lambda(x)T} - \frac{2MC(x)}{\lambda(x)^2T^2}(1 - e^{-\lambda(x)T}).$$

The assertion of the theorem follows from this. The proof is complete.

Remark 4.2: Since the constants C and λ in the formulation of Theorem 4.2 depend upon x the convergence in the assertion of this theorem may not be uniform in x and y, as required in Remark 4.1. However, since we are seeking only a candidate for the averaged function $\bar{b}$, Theorem 4.2 may suggest a choice of such an object. The problem of a formal verification of the averaging assumption with regard to the function $\bar{b}$ found in this way will be illustrated by the following example.

Example 4:1 Let us consider the following one-dimensional linear singularly perturbed system:

$$dX_t^\epsilon = (AX_t^\epsilon + BY_t^\epsilon)dt + \sigma_1(X_t^\epsilon)dW_t,$$
$$X_0^\epsilon = X_0,$$
$$dY_t^\epsilon = (1/\epsilon)(CX_t^\epsilon + DY_t^\epsilon)dt + (1/\sqrt{\epsilon})\sigma_2 dW_t,$$
$$Y_0^\epsilon = Y_0, \quad (4.25)$$

for $t\geqslant 0$. We assume that A, B, C, D, σ_2 are fixed constants, X_0, Y_0 are initial (square-integrable) random variables, $\sigma_1 = \sigma_1(x)$ is a real-valued, bounded and Lipschitz continuous function. It is easy to check that for system (4.25) the stability-type condition (i) in Lemma 3.3 is equivalent to the requirement that the quadratic form $Q = Q(x,y)$ generated by the matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is strictly negative definite. In the deterministic theory of singularly perturbed systems ($\sigma_1 = \sigma_2 = 0$) the following stability condition[4] is used:

$$D < 0 \text{ and } AD - BC > 0. \quad (4.26)$$

Remark 4.3: We note that the negative definiteness of the quadratic form Q implies the condition (4.26) and when $B = C$ the negative definiteness of the quadratic form Q and the condition (4.26) are equivalent.

In the following we assume that the quadratic form Q is strictly negative definite. Let us consider the auxiliary system (2.2) defined for (4.25)

$$dY_t^{xy} = (Cx + DY_t^{xy})dt + \sigma_2 d\bar{W}_t, \quad Y_0^{xy} = y, \quad (4.27)$$

where $x,y\in R$ and $\bar{W}$ is a Wiener process independent of W. The problem (4.27) may be solved explicitly:

$$Y_t^{xy} = ye^{Dt} - \frac{C}{D}x[1 - e^{Dt}] + \sigma_2\int_0^t e^{D(t-s)}d\bar{W}_s. \quad (4.28)$$

We can easily check that

$$Y_t^{xy} \to \mu^x \text{ weakly, as } t\to\infty,$$

where μ^x is a normally distributed probability measure with the mean $-Cx/D$ and the variance $C^2x^2/D^2 + \sigma_2^2/2D^2$. In view of Theorem 4.2 it is natural to present the following candidate for the function $\bar{b}$:

$$\bar{b}(x) = \int_{-\infty}^{+\infty}(Cx + Dz)d\mu^x(z) \quad (x\in R). \quad (4.29)$$

The integral in (4.29) may be computed explicitly:

$$\bar{b}(x) = [(AD - BC)/D]x \quad (x\in R). \quad (4.30)$$

This result is in agreement with a heuristic consideration commonly used in the deterministic theory of singular perturbations. Namely, if we multiply the second equation in (4.25) by ϵ and then formally set $\epsilon = 0$, then the resulting algebraic equation may be solved for Y_t. Substituting this formal solution into the first equation of system (4.25) we will obtain the following reduced problem:

$$dX_t = [(AD - BC)/D]X_t\,dt + \sigma_1(X_t)dW_t, \quad X_0 = X_0.$$

Let us note that the above reduced problem has the same form as problem (4.1) defined for system (4.25) with the function $\bar{b}$ given by (4.30).

In the following we will prove that for linear system (4.25) the averaging assumption is satisfied with the function $\bar{b}$ given by (4.30). First let us note that

$$E\left|\frac{1}{T}\int_0^T b_1(X_t^\epsilon, Y_s^{X_t^\epsilon Y_t^\epsilon})ds - \bar{b}(X_t^\epsilon)\right|^2$$
$$\leqslant 2B^2\left|\frac{1}{T}\int_0^T e^{Ds}ds\right|^2 E\left|\frac{C}{D}X_t^\epsilon + Y_t^\epsilon\right|^2$$
$$+ 2B^2\sigma_2^2 E\left|\frac{1}{T}\int_0^T\left[\int_0^s e^{D(s-u)}d\bar{W}_u\right]ds\right|^2 \quad (4.31)$$

for $t, T\geqslant 0$, $\epsilon\in(0,1)$. Using the next two lemmas we will show that the right-hand side of (4.31) converges to 0 as $T\to\infty$ and that the convergence is uniform in t and ϵ.

Lemma 4.1: For any m-dimensional Wiener process W we have

$$E\left|\frac{1}{T}\int_0^T\left[\int_0^s e^{D(s-u)}\,dW_u\right]ds\right|^2\to 0 \text{ as } T\to\infty.$$

Proof: Let us fix any $T>0$ let us define

$$I(s,u)=\begin{cases}1, & \text{if } s\geqslant u,\\ 0, & \text{if } s<u,\end{cases}$$

for $s,u\in[0,T]$. Using the Fubini-type theorem for stochastic integrals (Ref. 5, p. 116) we obtain

$$\begin{aligned}&\frac{1}{T}\int_0^T\left[\int_0^s e^{D(s-u)}\,dW_u\right]ds\\&=\frac{1}{T}\int_0^T\left[\int_0^T I(s,u)e^{D(s-u)}\,dW_u\right]ds\\&=\frac{1}{T}\int_0^T\left[\int_0^T I(s,u)e^{D(s-u)}\,ds\right]dW_u\\&=\frac{1}{T}\int_0^T\left[\int_u^T e^{D(s-u)}\,ds\right]dW_u.\end{aligned}$$

Hence

$$\begin{aligned}&E\left|\frac{1}{T}\int_0^T\left[\int_0^s e^{D(s-u)}\,dW_u\right]ds\right|^2\\&=\frac{m}{T^2}\int_0^T\left|\int_u^T e^{D(s-u)}\,ds\right|^2du\\&=\frac{m}{D^2T^2}\int_0^T\left[e^{2D(T-u)}-2e^{D(T-u)}+1\right]du\\&\leqslant 2m/D^2T.\end{aligned}$$

The assertion of the lemma follows from this. The proof is complete.

Lemma 4.2: Let us consider the fast solution process Y^ϵ of system (4.25). There exists an $\eta\in(0,1)$ such that

$$\sup_{t>0,\epsilon\in(0,\eta)}E|Y_t^\epsilon|<\infty.$$

Proof: We will first diagonalize the deterministic part of system (4.25). It is easy to see that for sufficiently small $\eta_1\in(0,1)$ and $\epsilon\in(0,\eta_1)$ both eigenvalues

$$\lambda_{1,2}(\epsilon)=\frac{1}{2}\left[A+\frac{D}{\epsilon}\pm\sqrt{\left(A-\frac{D}{\epsilon}\right)^2+\frac{4BC}{\epsilon}}\right] \tag{4.32}$$

of the coefficient matrix are negative. Using the corresponding eigenvectors we obtain

$$\Gamma(\epsilon)^{-1}M(\epsilon)\Gamma(\epsilon)=\begin{bmatrix}\lambda_1(\epsilon) & 0\\ 0 & \lambda_2(\epsilon)\end{bmatrix},$$

where $M(\epsilon)=\begin{bmatrix}A & B\\ \epsilon^{-1}C & \epsilon^{-1}D\end{bmatrix}$ is the coefficient matrix and a matrix corresponding to eigenvectors is given by

$$\Gamma(\epsilon)=\begin{bmatrix}2\epsilon B & 2\epsilon B\\ D-A\epsilon+\Delta & D-A\epsilon-\Delta\end{bmatrix},$$

where $\Delta=\sqrt{(A\epsilon-D)^2+4\epsilon BC}$.

For any $\epsilon\in(0,1)$ the transformation

$$\begin{bmatrix}u^\epsilon\\ v^\epsilon\end{bmatrix}=\Gamma(\epsilon)^{-1}\begin{bmatrix}x^\epsilon\\ y^\epsilon\end{bmatrix}$$

diagonalizes the deterministic part of system (4.25). It is also easy to see that the transformed system is of the form:

$$\begin{aligned}dU_t^\epsilon&=\lambda_1(\epsilon)U_t^\epsilon dt+\left[\frac{A_\epsilon-D+\Delta}{4\epsilon B\Delta}\sigma_1(\Gamma(\epsilon)(U_t^\epsilon,V_t^\epsilon))\right.\\&\quad\left.+\frac{\sigma_2}{2\sqrt{\epsilon\Delta}}\right]dW_t,\\dV_t^\epsilon&=\lambda_2(\epsilon)V_t^\epsilon dt+\left[\frac{D-A\epsilon+\Delta}{4\epsilon B\Delta}\sigma_1(\Gamma(\epsilon)(U_t^\epsilon,V_t^\epsilon))\right.\\&\quad\left.-\frac{\sigma_2}{2\sqrt{\epsilon\Delta}}\right]dW_t,\end{aligned}$$

with transformed initial random variables:

$$\begin{aligned}U_0^\epsilon&=\frac{A\epsilon-D+\Delta}{(4\epsilon B\Delta)}X_0+\frac{1}{2\Delta}Y_0,\\V_0^\epsilon&=\frac{D-A\epsilon+\Delta}{(4\epsilon B\Delta)}X_0-\frac{1}{2\Delta}Y_0.\end{aligned}$$

We note that

$$Y_t^\epsilon=(D-A\epsilon+\Delta)U_t^\epsilon+(D-A\epsilon-\Delta)V_t^\epsilon.$$

Therefore it is sufficient to prove the following facts:

$$\sup_{t>0,\epsilon\in(0,\eta)}E|(D-A\epsilon+\Delta)^2U_t^\epsilon|^2<\infty$$

and

$$\sup_{t>0,\epsilon\in(0,\eta)}E|V_t^\epsilon|^2<\infty \tag{4.33}$$

for some positive $\eta>0$. It is easy to check that

$$\begin{aligned}U_t^\epsilon&=e^{\lambda_1(\epsilon)t}\left[\frac{A\epsilon-D+\Delta}{4\epsilon B\Delta}X_0+\frac{1}{2\Delta}Y_0\right]\\&\quad+\int_0^t e^{\lambda_1(\epsilon)(t-s)}\left[\frac{A\epsilon-D+\Delta}{4\epsilon B\Delta}\sigma_1(\Gamma(\epsilon)\right.\\&\quad\left.\times(U_s^\epsilon,V_s^\epsilon))+\frac{\sigma_2}{2\sqrt{\epsilon}\Delta}\right]dW_s,\end{aligned}$$

for $t\geqslant 0$. This implies that

$$\begin{aligned}E|U_t^\epsilon|^2&\leqslant 2e^{2\lambda_1(\epsilon)t}E\left|\frac{A\epsilon-D+\Delta}{4\epsilon B\Delta}X_0+\frac{1}{2\Delta}Y_0\right|^2\\&\quad+2\int_0^t e^{2\lambda_1(\epsilon)(t-s)}E\left|\frac{A\epsilon-D+\Delta}{4\epsilon B\Delta}\right.\\&\quad\left.\times\sigma_1(\Gamma(\epsilon)(U_s^\epsilon,V_s^\epsilon))+\frac{\sigma_2}{2\sqrt{\epsilon\Delta}}\right|^2ds\end{aligned}$$

for $t\geqslant 0$. The first statement in (4.33) may be now obtained by multiplying both sides of the last inequality by $(D-A\epsilon+\Delta)^2$, using boundedness of the function σ_1 and applying condition (4.26) along with the fact that $\lambda_1(\epsilon)\to(AD-BC)/D$ as $\epsilon\to 0^+$. Similarly we have

$$\begin{aligned}E|V_t^\epsilon|^2&\leqslant 2e^{2\lambda_2(\epsilon)t}E\left|\frac{D-A\epsilon+\Delta}{4\epsilon B\Delta}X_0-\frac{1}{2\Delta}Y_0\right|^2\\&\quad+2\int_0^t e^{2\lambda_2(\epsilon)(t-s)}E\left|\frac{D-A\epsilon+\Delta}{4\epsilon B\Delta}\right.\\&\quad\left.\times\sigma_1(\Gamma(\epsilon)(U_s^\epsilon,V_s^\epsilon))-\frac{\sigma^2}{2\sqrt{\epsilon}\Delta}\right|^2ds,\end{aligned}$$

for $t \geqslant 0$. The first term on the right-hand side is globally bounded for $\epsilon \in (0,1)$ and $t \geqslant 0$, the second term may be estimated by

$$C_1 \int_0^t e^{2\lambda_2(\epsilon)(t-s)} \left| \frac{D - A\epsilon + \Delta}{4\epsilon B \Delta} \right|^2 ds + C_2 \int_0^t e^{2\lambda_2(\epsilon)(t-s)} \frac{\sigma_2^2}{4\epsilon \Delta^2} ds = T_1 + T_2,$$

for $t \geqslant 0$, where C_1, C_2 are positive constants. It easy to see that T_1 is bounded for $\epsilon \in (0,1)$ and $t \geqslant 0$. Finally, the term T_2 may be estimated as follows:

$$T_2 \leqslant \frac{C_3}{\epsilon} e^{2\lambda_2(\epsilon)t} \int_0^t e^{-2\lambda_2(\epsilon)s} ds = \frac{C_3}{2\epsilon\lambda_2(\epsilon)} [e^{2\lambda_2(\epsilon)t} - 1] \leqslant \frac{-C_3}{2\epsilon\lambda_2(\epsilon)},$$

for $t \geqslant 0$, $\epsilon \in (0, \eta_2)$, where η_2 is a positive number. Using the above estimates and the fact that $\epsilon\lambda_2(\epsilon) \to D < 0$ as $\epsilon \to 0^+$, we obtain the validity of statement (4.33) with $\eta = \min(\eta_1, \eta_2)$. The proof of the lemma is complete.

Remark 4.4: It follows immediately from Lemma 3.3 that the assertion of Lemma 4.2 remains also valid for the slow solution process X^ϵ. In view of inequality (4.31), Lemma 4.1 and Lemma 4.2 we conclude that the Averaging Principle (Theorem 4.1) is applicable for system (4.25). This may be regarded as a formal justification of the heuristic method of reduced problem for system (4.25).

ACKNOWLEDGMENTS

The authors would like to thank Dr. A. Korzeniowski for useful discussions and for his comments on the final form of the manuscript.

Research was partially supported by the U.S. Army Research Office under Grant No. DAAL 03-89-G-0107.

[1] M. I. Friedlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1984).
[2] G. S. Ladde and V. Lakshmikantham, *Random Differential Inequalities* (Academic, New York, 1980).
[3] A. Friedman, *Stochastic Differential Equations and Applications* (Academic, New York, 1975), Vol. I.
[4] D. D. Siljak, *Large-Scale Dynamical Systems: Stability and Structure* (North-Holland, New York, 1978).
[5] N. Ikeda, and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes* (North-Holland, Amsterdam, 1981).

A triangular property of the associated Legendre functions

S. Fineschi
Space Science Laboratory, NASA Marshall Space Flight Center, Huntsville, Alabama 35812

E. Landi Degl'Innocenti
Dipartimento di Astronomia e Scienza dello Spazio, Universitá di Firenze, Largo Enrico Fermi, 5-50125 Firenze, Italy

(Received 1 February 1989; accepted for publication 20 December 1989)

The property of the associated Legendre functions with non-negative integer indices, $P_n^m(z)$, described by the formula: $P_n^m(\cos\beta) = (-1)^m (a/c)^n \sum_{k=0}^{n-m} \binom{n+m}{k} (-b/a)^{n-k} \times P_{n-k}^m(\cos\gamma)$, where a,b,c are the sides of an assigned triangle and α,β,γ the respective opposite angles, is introduced. A useful application of this series in simplifying the calculation of collisional electron–atom cross sections higher than the dipole is mentioned. A proof of the stated identity by use of the Gegenbauer polynomials and of their generating function is given.

I. FORMULATION OF THE PROPERTY

The aim of the present paper is to introduce and to prove a mathematical formula that relates the associated Legendre functions with non-negative integral indices $P_n^m(\cos\beta)$ and $P_n^m(\cos\gamma)$, where β and γ are two different angles of an assigned triangle. We also mention an application of this formula in simplifying a problem of atomic physics. Referring to Fig. 1 for the notations, the formula at issue is

$$P_n^m(\cos\beta) = (-1)^m \left(\frac{a}{c}\right)^n \sum_{k=0}^{n-m} \binom{n+m}{k} \left(-\frac{b}{a}\right)^{n-k} \times P_{n-k}^m(\cos\gamma), \tag{1}$$

hereafter referred to as the "triangular property."

II. APPLICATION OF THE PROPERTY

We found it useful to apply the triangular property in solving—in an alternative way—a well-known problem of mathematical physics: The calculation of cross sections for electron–atom collisions.

We met this problem in developing a very general theoretical scheme able to describe the phenomenon of impact polarization in a wide variety of physical situations, particularly relevant for astrophysical applications. Starting from the general principles of quantum mechanics, we derived the statistical equilibrium equations for the density-matrix elements of an atomic system interacting with a beam of fast electrons (10 eV–10 keV). Our scheme, that is presented elsewhere,[1] reduces to the previous one developed by Oppenheimer[2] when only the diagonal density-matrix elements are accounted for. Obviously, the presence of collisional rates in the equilibrium equations requires the calculation of the electron–atom impact cross section.

Considering for example an inelastic electron–atom collision, the total cross section, $\sigma_{m\to n}$, corresponding to the transition between the atomic energy levels m and n, can be written in the following form:

$$\sigma_{m\to n} = \oint I_{m,n}(\mathbf{\Omega}) d\Omega, \tag{2}$$

where $I_{m,n}(\mathbf{\Omega})$ is the differential cross section for the atomic excitation from the mth to the nth atomic level, corresponding to the scattering of the incident electron along the direction $\mathbf{\Omega}$.

When the Born approximation is assumed, the integral in Eq. (2) is usually performed introducing the new variable $\mathbf{K}$, the momentum transferred to the electron during the collision:[3]

$$\sigma_{m\to n} = \int_{K_{\min}}^{K_{\max}} I'_{m,n}(K) K\, dK. \tag{3}$$

As the explicit expression for $I'_{m,n}(K)$ involves associated Legendre functions depending on the angle specifying the $\mathbf{K}$ direction, the total cross section is derived by integrating over this angle.

In our theoretical scheme for impact polarization, we found it more natural and convenient, especially for the calculation of cross sections corresponding to multipole transition higher than the dipole, to integrate over the angular coordinates of the scattered electron. This was achieved by making use of Eq. (1). We give a proof of this equation in the following section.

III. PROOF

To shorten our notations, we introduce the following definitions:

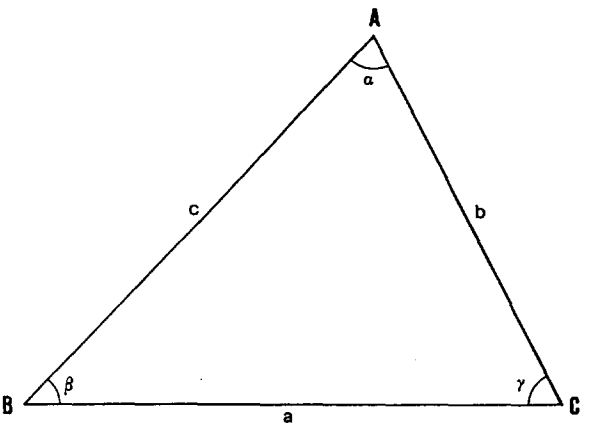


FIG. 1. Notations employed for labeling the angles and sides of the triangle ABC.

 0022-2488/90/051124-02$03.00

$$c_\beta \equiv \cos\beta, \tag{4a}$$

$$c_\gamma \equiv \cos\gamma, \tag{4b}$$

$$x \equiv b/a, \tag{4c}$$

$$q \equiv c/a = (1 + x^2 - 2xc_\gamma)^{1/2}. \tag{4d}$$

If the associate Legendre functions are expressed as Gegenbauer polynomials

$$P_n^m(z) = (-1)^m\,[(2m)!/m!2^m] \times (1-z^2)^{m/2}\,C_{n-m}^{\lambda}(z), \tag{5}$$

where $\lambda = m + \frac{1}{2}$, then Eq. (1) is equivalent to

$$C_n^{\lambda}(c_\beta)q^n = \sum_{k=0}^{n} \frac{(2\lambda+k)_{n-k}}{(n-k)!}(-x)^k\,C_k^{\lambda}(c_\gamma), \tag{6}$$

where we have introduced the Pochhammer symbol: $(a)_m = a(a+1)\cdots(a+m-1)$.

The proof is by use of the generating function. Let s satisfy $|xs/(1-s)| < 1$, multiply the left side of the formula by s^n, and sum over n:

$$\begin{aligned}\sum_{n=0}^{\infty} C_n^{\lambda}(c_\beta)q^n s^n &= \sum_{n=0}^{\infty} C_n^{\lambda}\left(\frac{1-xc_\gamma}{q}\right)q^n s^n \\ &= [1 - 2(1-xc_\gamma)s + s^2q^2]^{-\lambda} \\ &= [(1-s)^2 + 2c_\gamma xs(1-s) + x^2s^2]^{-\lambda} \\ &= \{(1-s)^2[1 - 2c_\gamma(-xs/(1-s)) \\ &\quad + (-xs/(1-s))^2]\}^{-\lambda}\end{aligned}$$

and by the generating function, again:

$$\begin{aligned}\cdots &= (1-s)^{-2\lambda}\sum_{k=0}^{\infty} C_k^{\lambda}(c_\gamma)\cdot[-xs/(1-s)]^k \\ &= \sum_{k=0}^{\infty} C_k^{\lambda}(c_\gamma)\,(-x)^k s^k (1-s)^{-2\lambda-k},\end{aligned}$$

and, finally, by the negative binomial series:

$$\cdots = \sum_{k=0}^{\infty} C_k^{\lambda}(c_\gamma)\,(-x)^k s^k \sum_{j=0}^{\infty} \frac{(2\lambda+k)_j}{j!}s^j. \tag{7}$$

The coefficient of s^n in this double sum is exactly the right-hand side of Eq. (6), the equivalent form of the triangular property. Implicitly, the Chu–Vandermonde sum was used.

As a final remark, it is worth noting that Eq. (1) can also be written in the more compact form:

$$P_n^m(\cos\beta) = \sum_{k=0}^{n-m}\left[(-1)^{n+m-k}\binom{n+m}{k}\frac{a^k b^{n-k}}{c^n}\right] \cdot P_{n-k}^m(\cos\gamma). \tag{8}$$

ACKNOWLEDGMENTS

The authors wish to thank the referee for having suggested the proof above, much shorter than the one they originally provided.

This paper was partially completed while S. F. held a NRC-NAS research associateship.

[1] S. Fineschi, "Polarizzazione da Impatto di Elettroni Veloci in Presenza di Campi Magnetici," Tesi di laurea, Universitá di Firenze, Italy, 1988; S. Fineschi and E. Landi Degl'innocenti, submitted to Astrophys. J.

[2] J. R. Oppenheimer, Phys. Rev. **32**, 316 (1928).

[3] I. C. Percival and M. J. Seaton, Phil. Trans. R. Astron. Soc. London, Ser. A **990**, 113 (1958), see Eq. (6.13).

Bäcklund transformations, soliton solutions and wave functions of Kaup–Newell and Wadati–Konno–Ichikawa systems

Abbas A. Rangwala
Department of Physics, University of Bombay, Bombay 400 098, India

Jyoti A. Rao
Department of Physics, Ramnarain Ruia College, Bombay 400 019, India

(Received 6 October 1988; accepted for publication 19 July 1989)

Using the Bargmann–Darboux method, the Bäcklund transformations, n-soliton solutions and corresponding wave functions of the Kaup–Newell and Wadati–Konno–Ichikawa systems are obtained. These results culminate in an algebraic recursive procedure for the determination of multisoliton solutions and their wave functions of the derivative and mixed derivative nonlinear Schrödinger equations $iQ_t + Q_{xx} \mp i\alpha(|Q|^2Q)_x \pm \beta|Q|^2Q = 0$, $\alpha > 0$, $\beta \geqslant 0$.

I. INTRODUCTION

The nonlinear equation

$$iQ_t + Q_{xx} \mp i\alpha(|Q|^2Q)_x \pm \beta|Q|^2Q = 0 \tag{1.1}$$

with $\alpha, \beta > 0$, describes nonlinear propagation of the Alfvén wave with a small nonvanishing wave number.[1] It is also relevant in the discussion of deformed continuous Heisenberg ferromagnet[2] and in the study of two-photon self-induced transparency and ultrashort light pulse propagation in an optical fiber.[3] For $\alpha = 0$, Eq. (1.1) becomes the nonlinear Schrödinger (NLS) equation

$$iq_t + q_{xx} \mp 2|q|^2q = 0 \tag{1.2}$$

whose soliton solutions were first obtained, using the inverse scattering transform, by Zakharov and Shabat[4] (ZS), who also formulated an associated eigenvalue problem for this equation analogous to the one previously formulated for the Korteweg–de Vries (KdV) equation by Gardner–Greene–Kruskal–Miura.[5] The ZS eigenvalue problem was enlarged by Ablowitz–Kaup–Newell–Segur[6] (AKNS) which additionally encompassed the nonlinear evolution equations like the KdV and sine–Gordon, among others. The NLS equation occurs in a large variety of physical situations.[7] For $\beta = 0$, Eq. (1.1) reduces to the derivative nonlinear Schrödinger (DNLS) equation

$$iq_t + q_{xx} \mp i(|q|^2q)_x = 0\,, \tag{1.3}$$

which was discussed by Kaup and Newell[8] (KN). We shall call Eq. (1.1) the mixed DNLS (MDNLS) equation since it has both the nonlinear derivative term of Eq. (1.3) and the nonderivative term of Eq. (1.2).

In Sec. II we state the Wadati–Konno–Ichikawa (WKI) scattering problem[1] which results in the MDNLS equation (1.1). The KN scattering problem[8] is then seen to be a special case of the WKI problem. The aim of the present paper is to obtain the Bäcklund transformations (BT's), soliton solutions, and their wave functions for both the above systems by extending our earlier work,[9,10] wherein we had obtained the BT's, the multisoliton solutions, and the corresponding wave functions of the nonlinear evolution equations envisaged by the ZS/AKNS scattering problem.[6] The organization of the paper is as follows. In Sec. III we discuss the KN scattering problem. We start by writing n-soliton wave functions in terms of $(n-1)$ soliton wave functions and, using an idea due to Bargmann,[11] obtain both the space and time parts of the BT's for arbitrary q and r in Sec. III A. In Sec. III B, we obtain the space and time Riccati equations and using Darboux's method[12] obtain the particular solutions of these equations which provide the basic equation for the determination of soliton solutions. In Sec. III C, we briefly indicate how the results obtained can be converted into a recursive method for solving the soliton problem for the DNLS equation. Finally, in Sec. III D, we display another evolution equation to which also the considerations of Secs. III A–III C apply. In Sec. IV we take up the WKI system for which, instead of following the above procedure, we take the cue from Kundu's work[13] to construct a gauge transformation to map the KN problem into the WKI problem. This enables us to obtain the complete soliton solution of the WKI problem and in particular that of the MDNLS equation. We explicitly exhibit the soliton solutions up to one soliton case of the nonlinear equations considered in this paper in the Appendix.

II. THE WKI SCATTERING PROBLEM

The WKI scattering problem is defined by the eigenvalue equations

$$\psi_{1x} + iF(\zeta)\psi_1 = G(\zeta)Q(x,t)\psi_2\,, \tag{2.1a}$$

$$\psi_{2x} - iF(\zeta)\psi_2 = G(\zeta)R(x,t)\psi_1\,, \tag{2.1b}$$

where

$$F(\zeta) = \alpha\zeta^2 + i\sqrt{2\beta}\,\zeta\,, \tag{2.2a}$$

$$G(\zeta) = \alpha\zeta + i\sqrt{\beta/2}\,, \tag{2.2b}$$

are functions of the eigenvalue ζ. The time evolution of the wave functions, ψ_1 and ψ_2, is determined by

$$\psi_{1t} = \mathscr{A}(\zeta;Q,R)\psi_1 + \mathscr{B}(\zeta;Q,R)\psi_2\,, \tag{2.3a}$$

$$\psi_{2t} = \mathscr{C}(\zeta;Q,R)\psi_1 - \mathscr{A}(\zeta;Q,R)\psi_2\,, \tag{2.3b}$$

where $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ are functions of x and t through their dependence on Q and R. Equation (1.1) with α, $\beta \geqslant 0$ is the integrability condition for Eqs. (2.1) and (2.3) if we choose

 0022-2488/90/051126-07$03.00

$$\mathscr{A}(\zeta;Q,R) = -2i\alpha^2\zeta^4 + 4\alpha\sqrt{2\beta}\,\zeta^3 + i(4\beta - \alpha^2 QR)\zeta^2 + \sqrt{2\beta}\,\alpha QR\zeta + i(\beta/2)QR\,, \tag{2.4a}$$

$$\mathscr{B}(\zeta;Q,R) = 2\alpha^2 Q\zeta^3 + 3i\sqrt{2\beta}\,\alpha Q\zeta^2 + (-2\beta Q + i\alpha Q_x + \alpha^2 Q^2 R)\zeta + (-(\sqrt{\beta/2})Q_x + i\alpha(\sqrt{\beta/2})Q^2 R)\,, \tag{2.4b}$$

$$\mathscr{C}(\zeta;Q,R) = 2\alpha^2 R\zeta^3 + 3i\sqrt{2\beta}\,\alpha R\zeta^2 + (-2\beta R - i\alpha R_x + \alpha^2 QR^2)\zeta + ((\sqrt{\beta/2})R_x + i\alpha(\sqrt{\beta/2})QR^2)\,, \tag{2.4c}$$

with

$$R = \pm Q^*\,.$$

As is to be expected, the KN[8] and the AKNS[6] problems are special cases of the above WKI system and are obtained from it if we, respectively, let

$$\text{KN: } \alpha = 1,\quad \beta = 0;\quad Q = q,\quad R = r\,, \tag{2.5}$$

$$\text{AKNS: } \alpha = 0,\quad \sqrt{2\beta} = -i;\quad Q = 2q,\quad R = 2r\,, \tag{2.6}$$

in Eqs. (2.1)–(2.3). The DNLS equation (1.3) and the NLS equation (1.2) are obtained as integrability conditions if we use Eqs. (2.5) and (2.6) in Eqs. (2.4) with $r = \pm q^*$.

From the above results it would be natural to first obtain the Bäcklund transformations (BT's), the soliton solutions, and corresponding wave functions for the WKI system and then specialize these to obtain those for the KN system. While this can be done, the algebra becomes formidable. It so turns out that knowing the BT's and the soliton solutions of the KN system, one can use a gauge transformation due to Kundu[13] to obtain in a simple manner the entire soliton solution of the WKI system. We thus look first at the KN system.

III. THE KAUP–NEWELL SYSTEM

A. Bäcklund transformations

The KN system[8] is defined by the eigenvalue equations

$$v_{1x} + i\zeta^2 v_1 = \zeta q v_2\,, \tag{3.1a}$$

$$v_{2x} - i\zeta^2 v_2 = \zeta r v_1\,, \tag{3.1b}$$

giving the space evolution of the two-component wave function $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Here q and r are functions of x and t satisfying some suitable nonlinear evolution equations which we do not need to specify here and ζ is the eigenvalue. To obtain the BT's and the soliton solutions of this system, we follow the method developed by us in the context of the AKNS system.[9,10] A part of the method leading up to the BT's was developed independently by Kundu.[13]

It is well-known that the BT's can be interpreted as introducing an additional soliton to the existing set.[14] Thus it seems possible to obtain the BT's by considering a connection between $n-1$ and n soliton problems. We denote the quantities referring to $n-1$ soliton problem by $v_1 = v_1(n-1)$, $v_2 = v_2(n-1)$, $q = q(n-1)$, $r = r(n-1)$, etc. This $n-1$ soliton problem is then defined by Eqs. (3.1). We denote the n soliton quantities by placing primes over the quantities. Thus $v_1' = v_1(n)$, $v_2' = v_2(n)$, $q' = q(n)$, $r' = r(n)$, etc. The n soliton problem is specified by the eigenvalue equations similar to Eqs. (3.1) with all quantities, except the eigenvalue ζ, replaced by the primed ones.

We concretize the expected relationship between n and $n-1$ soliton by expanding n soliton wave functions in terms of $n-1$ soliton wave functions:

$$v_1' = Av_1 + Bv_2\,, \tag{3.2a}$$

$$v_2' = Cv_1 + Dv_2\,, \tag{3.2b}$$

where $A,\ldots,D$ are functions of x and t to be determined. Using Eqs. (3.2) in Eqs. (3.1) and in the latter's primed counterparts we find that $A,\ldots,D$ satisfy the equations

$$\begin{aligned} A_x &= -\zeta rB + \zeta q'C\,,\\ B_x &= -\zeta qA - 2i\zeta^2 B + \zeta q'D\,,\\ C_x &= \zeta r'A + 2i\zeta^2 C - \zeta rD\,,\\ D_x &= \zeta r'B - \zeta qC\,. \end{aligned} \tag{3.3}$$

We now invoke an idea due to Bargmann[11,15] who showed that for the Schrödinger equation, $-d^2\psi/dx^2 + V(x)\psi = k^2\psi$, and for a potential capable of giving n bound states, the wave function can be written in the form $e^{ikx}\chi(k,x)$, where $\chi(k,x)$ is an nth degree polynomial in k. It is well-known that the Schrödinger equation is the associated eigenvalue equation for the application of the inverse scattering transform method to the KdV equation.[5] These considerations have been extended[9,10,13] to the entire AKNS problem. We here extend it further to the KN system identifying the n soliton solution of this system with a potential giving n bound states. The Bargmann idea then suggests that for the KN system v and v' must differ by a quadratic power of the eigenvalue ζ since the eigenvalue equations (3.1) contain powers of ζ up to the second degree. We thus try the ansatz

$$\begin{aligned} A &= \sum_{n=0}^{2} a_n\zeta^n,\quad B = \sum_{n=0}^{2} b_n\zeta^n\,,\\ C &= \sum_{n=0}^{2} c_n\zeta^n,\quad D = \sum_{n=0}^{2} d_n\zeta^n\,, \end{aligned} \tag{3.4}$$

where $a_n,\ldots,d_n$ are generally functions of x and t. We use Eqs. (3.4) in Eqs. (3.3) and equate equal powers of ζ on both sides to obtain a set of algebraic and differential relations involving the expansion coefficients. These equations can be easily solved and we display the solution in Eqs. (3.9) below. In addition to the above two further differential equations, which are in fact the BT's, are also required to be satisfied

$$q_x' e^{-i\chi} - q_x e^{i\chi} = -2i\alpha_0'(\epsilon q' + q) - (1/2i)(q'r' - qr)(q'e^{-i\chi} + qe^{i\chi})\,, \tag{3.5a}$$

$$r_x' e^{i\chi} - r_x e^{-i\chi} = -2i\alpha_0'(r' + \epsilon r) + (1/2i)(q'r' - qr)(r'e^{i\chi} + re^{-i\chi})\,, \tag{3.5b}$$

where $\chi(x,t) \equiv \chi(n,n-1)$ is given by

$$\frac{\partial \chi}{\partial x} = \frac{1}{2}(q'r' - qr), \tag{3.6}$$

and where α_0' and ϵ are essentially constants (they can at most be functions of t). Equations (3.5) were obtained earlier by Kundu[13] as well as by Boiti *et al.*[16]

There is a simpler way of writing Eqs. (3.5) which is essential for our later discussion. Using Eq. (3.6), we can rewrite Eqs. (3.5) as

$$\omega_x' e^{-i\chi} - \omega_x e^{i\chi} = -2i\alpha_0'(\epsilon\omega' + \omega), \tag{3.7a}$$

$$\eta_x' e^{i\chi} - \eta_x e^{-i\chi} = -2i\alpha_0'(\eta' + \epsilon\eta), \tag{3.7b}$$

where we have introduced ω and η through $q = -\omega_x$ and $r = -\eta_x$. These are the BT's in terms of the primitives (ω,η) of (q,r). What yet remains to be determined is the dependence of χ on (ω,η). To this end we note that simple manipulations with Eqs. (3.7) yield expressions for $(\partial/\partial x)(e^{\mp i\chi})$ in terms of (ω,η) and (ω',η'). The equations resulting on adding and subtracting these latter expressions are seen to be integrable only if we set $\epsilon = \mp 1$. This gives the possible values of ϵ and χ, in turn determined by

$$\epsilon = +1: \ \sin\chi = -(i\alpha_0'/2)(\omega' + \omega)(\eta' + \eta) + p', \tag{3.8a}$$

$$\epsilon = -1: \ \cos\chi = (\alpha_0'/2)(\omega' - \omega)(\eta' - \eta) + p', \tag{3.8b}$$

where p' is a constant of integration. The two values of ϵ reflect the possibility that we may have nonlinear equations such that if (q,r) is a solution so is $(-q,-r)$. In such a case both values of ϵ are permitted.

The determination of the expansion coefficients in Eqs. (3.4) gives the following expressions for the coefficients $A,\ldots,D$ in Eqs. (3.2):

$$\begin{aligned} A &= \sqrt{K'}(\alpha_0' + e^{i\chi}\zeta^2), \\ B &= (\sqrt{K'}/2i)(q'e^{-i\chi} - qe^{i\chi})\zeta \\ &= \alpha_0'\sqrt{K'}(\epsilon\omega' + \omega)\zeta, \\ C &= (\sqrt{K'}/2i)(-r'e^{i\chi} + re^{-i\chi})\zeta \\ &= -\alpha_0'\sqrt{K'}(\eta' + \epsilon\eta)\zeta, \\ D &= \sqrt{K'}(-\epsilon\alpha_0' + e^{-i\chi}\zeta^2), \end{aligned} \tag{3.9}$$

where K' is another constant of integration. We see from Eqs. (3.9) and (3.8) that for each choice of ϵ, the solution is given in terms of two parameters α_0', p'; the overall constant $\sqrt{K'}$ can be chosen at will. For the pure soliton solutions these parameters turn out to be constants independent of time since, as we shall see, they are related to the soliton parameters representing a pole in the complex ζ^2 plane.

To complete our discussion of the BT's, we still need the time BT's. This requires that we specify the time counterparts of Eqs. (3.1) which would determine the time evolution of the wave functions. These time evolution equations have the same form as Eqs. (2.3), except that $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ would now be appropriate to nonlinear evolution equations generated by the KN system and ψ_1, ψ_2 should be replaced by v_1, v_2, respectively. As in the AKNS case, we require the eigenvalue to be time independent, $\zeta_t = 0$. The cross differentiation of Eqs. (3.1) and the KN versions of Eqs. (2.3), mentioned above, and their primed counterparts yield the relations

$$\begin{aligned} A_t &= A(\mathscr{A}' - \mathscr{A}) - B\mathscr{C} + C\mathscr{B}', \\ B_t &= B(\mathscr{A}' + \mathscr{A}) - A\mathscr{B} + D\mathscr{B}', \\ C_t &= -C(\mathscr{A}' + \mathscr{A}) + A\mathscr{C}' - D\mathscr{C}, \\ D_t &= -D(\mathscr{A}' - \mathscr{A}) + B\mathscr{C}' - C\mathscr{B}, \end{aligned} \tag{3.10}$$

where $\mathscr{A}' \equiv \mathscr{A}(\zeta;q',r')$, etc. In view of the second and third of Eqs. (3.9), the second and third of Eqs. (3.10) give the time part of the BT's. In general, Eqs. (3.10) would determine both the time evolutions of (q',r') given that of (q,r) and also the possible time dependence of α_0', p'. As noted before, for the soliton solutions, α_0', p' are, however, expected to be time independent.

With both the space and time parts of the BT's at our disposal, we could attempt to determine the n soliton solutions (q',r') [or (ω',η')] given the $n-1$ soliton solutions of nonlinear equations generated by the KN system. This is a very difficult, if not well nigh impossible, task. Looking for an alternate method we turn to Riccati equations whose particular solutions help us to develop a recursive method for obtaining soliton solutions of some special nonlinear equations, like the DNLS equation, without an explicit recourse to the above BT's.

B. Riccati analysis

Defining

$$\Gamma = v_1/v_2, \tag{3.11}$$

we obtain, from Eqs. (3.1) and (2.3) suitably rewritten for the KN problem, the space and time Riccati equations as

$$\Gamma_x = \zeta q - 2i\zeta^2\Gamma - \zeta r\Gamma^2, \tag{3.12a}$$

$$\Gamma_t = \mathscr{B} + 2\mathscr{A}\Gamma - \mathscr{C}\Gamma^2. \tag{3.12b}$$

Analogously we also define $\Gamma' = v_1'/v_2'$ appropriate to the n soliton situation which satisfies equations similar to Eqs. (3.12) with all quantities, except the eigenvalue ζ, replaced by the primed ones.

We are interested in obtaining the bound states of Eqs. (3.1). This requires an analytic continuation of the eigenvalue ζ^2 to the upper half complex plane.[8] We thus let

$$\zeta^2 \to \lambda' = \xi' + i\eta', \quad \xi',\eta' \text{ real with } \eta' > 0. \tag{3.13}$$

The real and imaginary parts of λ' give the soliton parameters. Besides requiring quantities as functions of the eigenvalue ζ, we shall also need quantities to be evaluated at λ'. We indicate the latter by placing bars overhead. Thus, for instance, Eqs. (3.12) become

$$\bar{\Gamma}_x = \sqrt{\lambda'}q - 2i\lambda'\bar{\Gamma} - \sqrt{\lambda'}r\bar{\Gamma}^2, \tag{3.14a}$$

$$\bar{\Gamma}_t = \bar{\mathscr{B}} + 2\bar{\mathscr{A}}\bar{\Gamma} - \bar{\mathscr{C}}\bar{\Gamma}^2. \tag{3.14b}$$

A feature of Eqs. (3.14) that should be noted is that while Γ, q, and r refer to the $n-1$ soliton problem, the functions $\Gamma(\zeta)$ and ζ in these equations are, however, evaluated at the n soliton parameters $\lambda' = \xi' + i\eta'$. This step is crucial in the development of the recursive method.

We now follow Darboux[12,17] to write down the particular solutions of Eq. (3.14). We have from Eqs. (3.2)

$$\bar{v}_1' = \bar{A}\bar{v}_1 + \bar{B}\bar{v}_2 , \tag{3.15a}$$

$$\bar{v}_2' = \bar{C}\bar{v}_1 + \bar{D}\bar{v}_2 . \tag{3.15b}$$

If we set $\bar{v}_1' = 0 = \bar{v}_2'$, we obtain

$$\bar{\Gamma} = -\bar{B}/\bar{A} = -\bar{D}/\bar{C} . \tag{3.16}$$

The second equality requires that $\bar{A}\,\bar{D} = \bar{B}\,\bar{C}$ and this is not automatically guaranteed. However, from Eqs. (3.9), we find that we would have $\bar{A}\,\bar{D} = \bar{B}\,\bar{C}$ if we choose the constants α_0', p' appearing in Eqs. (3.8) and (3.9) in terms of $\lambda' = \xi' + i\eta'$ through

$$[\lambda' - \epsilon i^{(1+\epsilon)/2}\alpha_0' p']^2 - \epsilon\alpha_0'^2(1-p'^2) = 0 . \tag{3.17}$$

In all our subsequent discussion whenever we use the barred quantities we shall require that (α_0', p') be related to (ξ', η') by Eq. (3.17) so that the validity of the second equality in Eq. (3.16) is guaranteed. It is easy to see that Eq. (3.16) indeed provides particular solutions of Eqs. (3.14).

To obtain the particular solutions of the primed version of Eqs. (3.14), we note that Eqs. (3.15) give $\bar{C}\bar{v}_1' - \bar{A}\bar{v}_2' = 0$ and $\bar{D}\bar{v}_1' - \bar{B}\bar{v}_2' = 0$ in view of $\bar{A}\,\bar{D} = \bar{B}\,\bar{C}$. These yield

$$\bar{\Gamma}' = \bar{B}/\bar{D} = \bar{A}/\bar{C} . \tag{3.18}$$

C. Soliton solutions of DNLS equation

For the DNLS equation (1.3),

$$r = \tau q^* \text{ or } \eta = \tau\omega^*, \quad \tau = \pm 1 , \tag{3.19}$$

in Eqs. (3.1). The time evolution of the wave functions is given by Eqs. (2.3) with the ψ's replaced by the v's and $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ obtained from Eqs. (2.4) on setting $\alpha = 1$, $\beta = 0$, $Q = q$, and $R = r$. The equations of the preceding sections specialize appropriately in view of Eq. (3.19). Knowing the $n-1$ soliton wave functions, $v_{1,2}$, which enable us to obtain $\bar{\Gamma}$, it is now easy to see that using $\bar{A},\ldots,\bar{D}$ from Eqs. (3.9) in Eq. (3.16), noting Eq. (3.19) and on eliminating $e^{\pm ix}$, we can obtain ω' in terms of the n soliton parameters λ' and $n-1$ soliton solution and wave functions, ω and $\bar{\Gamma}$, respectively. Once ω and ω' are known we can obtain $e^{ix(n,n-1)}$ from Eq. (3.16) say, and these quantities in turn enable us to determine $v_{1,2}'$ from Eqs. (3.2) on using Eqs. (3.9). This forms the basis of our recursive method. We shall not carry out this program explicitly here. We do this later in Sec. IV D for the MDNLS case.

We, however, display the expressions for the DNLS equations up to one soliton case in the Appendix.

D. Another evolution equation

Besides the DNLS equation we consider another nonlinear evolution equation belonging to the class, $\eta = \tau\omega^*$, in Eq. (3.19). This is

$$\omega_{xt} + \omega \pm i|\omega|^2\omega_x = 0 , \tag{3.20}$$

which can be obtained as the integrability condition for Eqs. (3.1) and the KN equivalent of Eqs. (2.3) with

$$\begin{aligned}
\mathscr{A}(\zeta;\omega,\tau\omega^*) &= -(i/2)\tau|\omega|^2 - i/4\zeta^2 , \\
\mathscr{B}(\zeta;\omega,\tau\omega^*) &= -i\omega/2\zeta , \\
\mathscr{C}(\zeta;\omega,\tau\omega^*) &= i\tau\omega^*/2\zeta .
\end{aligned} \tag{3.21}$$

The difference between the soliton solutions and wave functions of the DNLS equation and those for Eq. (3.20) arises from different dispersion relations, $\omega(\zeta) \equiv i\mathscr{A}(\zeta;0,0)$, satisfied by the two. We also display the solutions of this equation up to one soliton in the Appendix.

IV. THE WADATI–KONNO–ICHIKAWA SYSTEM

A. Gauge transformation

The space and time evolutions of the WKI wave function are given by Eqs. (2.1)–(2.3). The MDNLS equation results when $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ are chosen as in Eqs. (2.4) with $R = \pm Q^*$.

To obtain the soliton solution of the WKI problem, we could follow the same method as in the preceding section. This, however, is not necessary since for the present system there is another way of obtaining the solution which depends on the observation made by Kundu[13] that there exists a simple gauge transformation that maps the KN system into the WKI system. The Kundu transformation is a two-step transformation.

(1) A Galilean transformation

$$x \to x' = x - ut, \quad t \to t' = t , \tag{4.1}$$

and a relabeling

$$(x',t') \to (x,t) \tag{4.2}$$

followed by the mapping.

(2)
$$\begin{aligned}
&\zeta \to z = \sqrt{\alpha}\zeta + i(\sqrt{\beta/2\alpha}) , \\
&q(x+ut,t) \to \sqrt{\alpha}Q(x,t)e^{i\varphi(x,t)} , \\
&r(x+ut,t) \to \sqrt{\alpha}R(x,t)e^{-i\varphi(x,t)} , \\
&v_1(x+ut,t) \to \psi_1(x,t)e^{if(x,t)} , \\
&v_2(x+ut,t) \to \psi_2(x,t)e^{ig(x,t)} ,
\end{aligned} \tag{4.3}$$

with

$$\varphi(x,t) = (u/2)(x + \tfrac{1}{2}ut) . \tag{4.4}$$

It is easy to show that under (1) and (2) above, Eqs. (3.1) go into Eqs. (2.1), provided we require that

$$f(x,t) = (\beta/2\alpha)x + a(t) , \tag{4.5a}$$

$$g(x,t) = -(\beta/2\alpha)x + b(t) , \tag{4.5b}$$

with

$$g - f = -(u/2)x - \tfrac{1}{4}u^2 t . \tag{4.5c}$$

Equations (4.5) in turn give

$$u = 2\beta/\alpha \tag{4.6a}$$

and

$$a - b = \tfrac{1}{4}u^2 t = \beta^2/\alpha^2 t . \tag{4.6b}$$

For the time evolution of the wave functions the above mapping shows that the $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ of the WKI system in Eqs. (2.3) can be related to the corresponding functions of the KN system through

$$\mathscr{A}(\zeta;Q,R) = \mathscr{A}^K(z;\sqrt{\alpha}Qe^{i\varphi},\sqrt{\alpha}Re^{-i\varphi}) - i\frac{da}{dt} - iuz^2,$$
$$= \mathscr{A}^K(z;\sqrt{\alpha}Qe^{i\varphi},\sqrt{\alpha}Re^{-i\varphi}) + i\frac{db}{dt} - iuz^2, \tag{4.7a}$$

$$\mathscr{B}(\zeta;Q,R) = \mathscr{B}^K(z;\sqrt{\alpha}Qe^{i\varphi},\sqrt{\alpha}Re^{-i\varphi})e^{-i\varphi} + u\sqrt{\alpha}zQ, \tag{4.7b}$$

$$\mathscr{C}(\zeta;Q,R) = \mathscr{C}^K(z;\sqrt{\alpha}Qe^{i\varphi},\sqrt{\alpha}Re^{-i\varphi})e^{i\varphi} + u\sqrt{\alpha}zR. \tag{4.7c}$$

In the above superscript K indicates the corresponding KN quantities. From Eqs. (4.7a) and (4.6b) it follows that

$$a = \tfrac{1}{8}u^2t = (\beta^2/2\alpha^2)t, \quad b = -\tfrac{1}{8}u^2t = -(\beta^2/2\alpha^2)t. \tag{4.8}$$

These then determine the remaining phase factors in Eqs. (4.3) and the gauge transformation from the KN system to the WKI, Eq. (4.3), is now completely established.

B. Bäcklund transformations

Using the above mapping the space part of the BT's for the WKI system can be trivially obtained from those for the KN system, Eqs. (3.7), as

$$\Omega_x' e^{-iX} - \Omega_x e^{iX} = -2i\alpha_0'(\epsilon\Omega' + \Omega), \tag{4.9a}$$

$$H_x' e^{iX} - H_x e^{-iX} = -2i\alpha_0'(H' + \epsilon H), \tag{4.9b}$$

where Ω and H are defined by

$$Q(x,t) = -\Omega_x(x,t)e^{-i\varphi(x,t)}, \tag{4.10a}$$

$$R(x,t) = -H_x(x,t)e^{i\varphi(x,t)}, \tag{4.10b}$$

and where the function $X(x,t) = X(n,n-1)$ is given by

$$X_x = (\alpha/2)(Q'R' - QR). \tag{4.11}$$

C. Relation between n and $n-1$ soliton wave functions

We stipulate, as in the KN case, that the n soliton wave functions, ψ_1', ψ_2' are given in terms of the $n-1$ soliton wave functions, ψ_1, ψ_2, by

$$\psi_1' = A\psi_1 + B\psi_2, \quad \psi_2' = C\psi_1 + D\psi_2. \tag{4.12}$$

The expansion functions, $A,\ldots,D$, above can be obtained from the corresponding KN quantities in the same manner as we obtained Eqs. (4.7). The results, obtained from Eqs. (3.9) and the correspondence established above, are

$$A(\zeta;Q,R) = \sqrt{K'}(\alpha_0' + e^{iX}z^2), \tag{4.13a}$$

$$B(\zeta;Q,R) = (\sqrt{K'\alpha}/2i)(Q'e^{-iX} - Qe^{iX})z$$
$$= \sqrt{K'\alpha}\alpha_0'(\epsilon\Omega' + \Omega)e^{-i\varphi}z, \tag{4.13b}$$

$$C(\zeta;Q,R) = -(\sqrt{K'\alpha}/2i)(R'e^{iX} - Re^{-iX})z$$
$$= -\sqrt{K'\alpha}\alpha_0'(H' + \epsilon H)e^{i\varphi}z, \tag{4.13c}$$

$$D(\zeta;Q,R) = \sqrt{K'}(-\epsilon\alpha_0' + e^{-iX}z^2). \tag{4.13d}$$

D. Recursive determination of soliton solutions of the MDNLS equation

With the BT's and the expansion functions $A,\ldots,D$ determined, for the Riccati analysis and the recursive procedure we may follow the procedure of Secs. III B and III C. We here only quote the main results as applied to the MDNLS equation (1.1). This time we analytically continue $F(\zeta) = z^2 + u/4$ in the upper half complex plane and let $F(\zeta) \to \lambda' = \xi' + i\eta', \eta' > 0$. A particular solution of the Riccati variable, $\Gamma = \psi_1/\psi_2$ is still given by $\Gamma = -B/A = -D/C$ with $\bar{A},\ldots,\bar{D}$, obtained from Eqs. (4.13) on letting $z^2 \to \lambda' - u/4$. The condition that $\bar{A}\bar{D} = \bar{B}\bar{C}$ be valid now requires that, in place of Eq. (3.17), we must have

$$(\lambda' - u/4 - \epsilon i^{(1+\epsilon)/2}\alpha_0'p')^2 - \epsilon\alpha_0'^2(1-p'^2) = 0, \quad \epsilon = \pm 1. \tag{4.14}$$

From its definition, Eq. (4.11), we see that $X(x,t) = X(n,n-1)$ is a real function for the MDNLS case. This implies, from the MDNLS analog of Eqs. (3.8), that $\alpha_0' = -i\nu'$ for $\epsilon = +1$ and $\alpha_0' = -\nu'$ for $\epsilon = -1$, with ν' and p' real. This enables the following parametrization[8]:

$$\epsilon = \pm 1, \quad \xi' - u/4 = \nu'p',$$
$$\eta' = \nu'(1-p'^2)^{1/2}, \quad -1 < p' < 1$$
$$\nu' = \Delta'^2 > 0, \quad p' = -\tau\cos\gamma', \quad 0 < \gamma' < \pi. \tag{4.15}$$

Here $\tau = \pm 1$ and arises from the MDNLS requirement that $R = \tau Q^*$.

Obtaining $\bar{A},\ldots,\bar{D}$ from Eqs. (4.13), using these in $\Gamma = -B/A = -D/C$ and on eliminating $e^{iX(n,n-1)}$ we find that the n and $n-1$ soliton solutions of the MDNLS equation are related by the recursion relation

$$\Omega(n) + \epsilon\Omega(n-1) = -\frac{2i\epsilon\tau e^{i\varphi}\bar{g}(n-1)\sin\gamma_n}{\sqrt{\alpha}\left[(\Delta_n\bar{g}^*(n-1))(\Delta_n\bar{g}(n-1)) + e^{i\tau\gamma_n}\right]}, \tag{4.16}$$

where we have introduced

$$\bar{g}(n-1) = \bar{\Gamma}(n-1)/\sqrt{\lambda_n - u/4} \tag{4.17}$$

with

$$\bar{\Gamma}(n-1) = \Gamma|_{F(\zeta)\to\lambda_n = \xi_n + i\eta_n}. \tag{4.18}$$

The phase factor, $e^{iX(n,n-1)}$, is given by

$$e^{iX(n,n-1)} = \frac{i^{(1+\epsilon)/2}\Delta_n^2}{(\lambda_n - u/4)} \times\left[1 + \frac{\sqrt{\alpha}(\epsilon\Omega(n) + \Omega(n-1))e^{-i\varphi}}{\bar{g}(n-1)}\right]. \tag{4.19}$$

The procedure for determining the solution is as follows. The zero soliton solution of the MDNLS equation (1.1) is $Q(0) = 0$ or $\Omega(0) = 0$. Using this in Eqs. (2.1) we determine the zero soliton wave functions $\psi_{1,2}(0)$. From these we construct $\bar{\Gamma}(0)$ or $\bar{g}(0)$. Use of this in Eq. (4.16) enables us to obtain $\Omega(1)$ and hence $Q(1)$ on using Eq. (4.10a). We then determine $\psi_{1,2}(1)$ from Eqs. (4.12) and (4.13). It should now be clear that this process can be con-

tinued and the complete multisoliton solution of the MDNLS problem obtained. It is essential here to note that in the construction of $\bar{\Gamma}$ general solutions, $\psi_{1,2}$, are to be used and not the fundamental solutions. We give the explicit expressions up to one soliton case in the Appendix.

Finally we note that except for the zeroth soliton situation, where one is required to solve a trivial differential equation, the above recursive procedure is entirely algebraic.

E. Another evolution equation

The WKI counterpart of the KN equation (3.20) is

$$\Omega_{xt} - (2\beta/\alpha)\Omega_{xx} + \Omega + i\tau\alpha|\Omega|^2\Omega_x = 0\,, \tag{4.20}$$

with

$$\begin{aligned}\mathscr{A}(\zeta;\Omega,\tau\Omega^*) &= -(i/2)\alpha\tau|\Omega|^2 - i/4z^2 \\ &\quad - i(u^2/8) - iuz^2\,, \\ \mathscr{B}(\zeta;\Omega,\tau\Omega^*) &= \sqrt{\alpha}e^{-i\varphi}\left[-(i/2z)\Omega - uz\Omega_x\right], \\ \mathscr{C}(\zeta;\Omega,\tau\Omega^*) &= \tau\sqrt{\alpha}e^{i\varphi}\left[(i/2z)\Omega^* - uz\Omega_x^*\right]. \end{aligned} \tag{4.21}$$

The results obtained in Sec. IV D apply also to this evolution equation except that the time dependence of this equation is different from that for the MDNLS equation due to the difference in their dispersion relations. We display the soliton solutions up to one soliton case for this equation in the Appendix.

ACKNOWLEDGMENTS

One of us (AAR) would like to thank Professor M. Lakshmanan and Mr. K. Porsezian for bringing this problem to his attention.

He would also like to thank the Board of Research in Nuclear Studies, Department of Atomic Energy, India for a grant.

APPENDIX: SOLITON SOLUTIONS AND WAVE FUNCTIONS

1. MDNLS equation

a. Zero soliton case

The zero soliton solution of the MDNLS equation (1.1) is

$$Q(0) = 0 \text{ or } \Omega(0) = 0\,. \tag{A1}$$

The corresponding zero soliton wave functions obtained from Eqs. (2.1)–(2.4) are

$$\begin{aligned}\psi_1(0) &= \alpha_0 e^{-iF(\zeta)x + \mathscr{A}(\zeta;0,0)t} \\ &= \alpha_0 e^{-iF(\zeta)[x + 2F(\zeta)t]}\,, \end{aligned}$$

$$\begin{aligned}\psi_2(0) &= \beta_0 e^{iF(\zeta)x - \mathscr{A}(\zeta;0,0)t} \\ &= \beta_0 e^{iF(\zeta)[x + 2F(\zeta)t]}\,, \end{aligned} \tag{A2}$$

where $F(\zeta)$ and $\mathscr{A}(\zeta;0,0)$ are given, from Eqs. (2.2a) and (2.4a), respectively, by

$$F(\zeta) = \alpha\zeta^2 + i\sqrt{2\beta}\,\zeta\,, \tag{A3a}$$

$$\omega(\zeta) = i\mathscr{A}(\zeta;0,0) = 2[F(\zeta)]^2\,, \tag{A3b}$$

where $\omega(\zeta)$ is the dispersion function.

b. One soliton case

We first obtain $\Omega(1)$ from Eq. (4.16) and then obtain $\psi_1(1)$, $\psi_2(1)$ from Eqs. (4.12), (4.13), and (4.19).

To this end, we find from Eqs. (A2) that

$$\bar{\Gamma}(0) = (\alpha_0/\beta_0)e^{-2i\lambda_1(x + 2\lambda_1 t)}\,. \tag{A4}$$

The choice

$$(\alpha_0/\beta_0) = -(\tau\sqrt{\lambda_1 - u/4})/\Delta_1 e^{-i\tau\gamma_1} \tag{A5}$$

results in

$$\bar{g}(0) = -(\tau/\Delta_1)e^{2\theta_1 - 2i\sigma_1 - i\tau\gamma_1}\,, \tag{A6}$$

where

$$\theta_1(x,t) = \eta_1(x + 4\xi_1 t)\,, \tag{A7a}$$

$$\sigma_1(x,t) = \xi_1 x + 2(\xi_1^2 - \eta_1^2)t \tag{A7b}$$

$$\xi_1 = u/4 - \tau\Delta_1^2\cos\gamma_1\,, \tag{A7c}$$

and

$$\eta_1 = \Delta_1^2\sin\gamma_1\,. \tag{A7d}$$

Using Eq. (A6) in Eq. (4.16), we get the one soliton solution of the MDNLS equation as

$$\Omega(1) = \frac{2i\epsilon\sin\gamma_1}{\Delta_1\sqrt{\alpha}}\cdot\frac{e^{2\theta_1 - 2i\sigma_1 - i\tau\gamma_1}}{e^{4\theta_1} + e^{i\tau\gamma_1}}\,e^{i\varphi}\,, \tag{A8a}$$

or

$$\begin{aligned}Q(1) &= -\Omega_n(1)e^{-i\varphi} \\ &= \frac{4\epsilon\tau\Delta_1\sin\gamma_1}{\sqrt{\alpha}}\cdot\frac{e^{2\theta_1 - 2i\sigma_1}(e^{4\theta_1} + e^{-i\tau\gamma_1})}{(e^{4\theta_1} + e^{i\tau\gamma_1})^2}\,, \end{aligned} \tag{A8b}$$

where φ is given by Eq. (4.4). To obtain the one soliton wave functions, we find from Eq. (4.19) that

$$e^{iX(1,0)} = -\tau i^{(1+\epsilon)/2}\cdot\frac{\cosh(2\theta_1 + i\tau\gamma_1/2)}{\cosh(2\theta_1 - i\tau\gamma_1/2)}\,. \tag{A9}$$

Using Eqs. (A8a) and (A9) in Eq. (4.13), we finally obtain from Eqs. (4.12) the two linearly independent one soliton wave functions as

$$\left\{\begin{pmatrix} -i^{(1+\epsilon)/2}\left[\Delta_1^2 + \tau\left(F(\zeta) - \frac{u}{4}\right)\frac{\cosh(2\theta_1 + i\tau\gamma_1/2)}{\cosh(2\theta_1 - i\tau\gamma_1/2)}\right] \\ -2i^{(3+\epsilon)/2}\epsilon\tau\Delta_1\left(F(\zeta) - \frac{u}{4}\right)^{1/2}\sin\gamma_1\frac{e^{2\theta_1 + 2i\sigma_1 + i\tau\gamma_1}}{e^{4\theta_1} + e^{-i\tau\gamma_1}} \end{pmatrix} e^{-iF(\zeta)[x + 2F(\zeta)t]}\,,\right.$$

$$\left.\begin{pmatrix} -2i^{(3+\epsilon)/2}\Delta_1\left(F(\zeta) - \frac{u}{4}\right)^{1/2}\sin\gamma_1\frac{e^{2\theta_1 - 2i\sigma_1 - i\tau\gamma_1}}{e^{4\theta_1} + e^{i\tau\gamma_1}} \\ i^{(1+\epsilon)/2}\epsilon\left[\Delta_1^2 + \tau\left(F(\zeta) - \frac{u}{4}\right)\frac{\cosh(2\theta_1 - i\tau\gamma_1/2)}{\cosh(2\theta_1 + i\tau\gamma_1/2)}\right] \end{pmatrix} e^{iF(\zeta)[x + 2F(\zeta)t]}\right\}. \tag{A10}$$

It should be noted that $\psi_1(1)$ is a linear combination of the upper two entries in Eq. (A10) and $\psi_2(1)$ is the *same* linear combination of the lower two entries and it is these $\psi_1(1)$, $\psi_2(1)$ that must be used to obtain $\bar{\Gamma}(1)$ necessary for the construction of the two soliton solution.

2. Equation (4.20)

The dispersion relation here is different from the MDNLS case. This alters the time dependence of the soliton solutions and the wave functions from those given above. The functions θ_1 and σ_1 above are now given by

$$\theta_1(x,t) = \eta_1\left(x + ut - \frac{t}{4[(\xi_1 - u/4)^2 + \eta_1^2]}\right),$$
$$\sigma_1(x,t) = \xi_1(x + ut) + \left\{\frac{(\xi_1 - u/4)}{4[(\xi_1 - u/4)^2 + \eta_1^2]} - \frac{u^2}{8}\right\}t, \tag{A11}$$

while the phase factors $\exp[\pm iF(\zeta)(x + 2F(\zeta)t)]$ are now replaced by

$$\exp\{\pm i[F(\zeta)x + (uF(\zeta) - u^2/8 + 1/4(F(\zeta) - u/4))t]\}. \tag{A12}$$

3. DNLS equation

The relevant expressions in this case can be obtained by setting $\alpha = 1, \beta = 0$ in the MDNLS equations (A1)–(A10) along with appropriate notational changes like $\psi_{1,2} \to v_{1,2}$, $Q \to q$, $\Omega \to \omega$, $X \to x$, etc.

4. Equation (3.20)

The relevant expressions in this case can be obtained by setting $\alpha = 1, \beta = 0$ in those for Eq. (4.20) along with notational changes indicated in Sec. 3 above.

[1] M. Wadati, K. Konno, and Y. Ichikawa, J. Phys. Soc. Jpn. **46**, 1965 (1979).
[2] K. Porsezian, K. M. Tamizhmani, and M. Lakshmanan, Phys. Lett. A **124**, 159 (1987).
[3] A. A. Zabolotskii, Phys. Lett. A **124**, 500 (1987).
[4] V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972).
[5] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Commun. Pure Appl. Math. **27**, 97 (1974).
[6] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[7] See, for instance, R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, New York, 1982), Chap. 8.
[8] D. J. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
[9] A. A. Rangwala and J. A. Rao, Phys. Lett. A **112**, 188 (1985).
[10] J. A. Rao and A. A. Rangwala, in *Solitons: Introduction and Applications*, edited by M. Lakshmanan, Springer Series in Nonlinear Dynamics Preceedings, 176 (1987).
[11] V. Bargmann, Rev. Mod. Phys. **21**, 488 (1949).
[12] G. Darboux, *Lecôns Sur la Théorie Générale des Surfaces et les Applications Geometriques du Calcul Infinitesimal* (Gauthier-Villars, Paris, 1915), Vol. 2 (2nd ed.), p. 210; Also E. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), p. 132.
[13] A. Kundu, J. Phys. A: Math. Gen. **20**, 1107 (1987).
[14] H. D. Wahlquist and F. B. Eastabrook, Phys. Rev. Lett. **31**, 1386 (1973).
[15] G. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980), p. 96.
[16] M. Boiti, C. Laddomada, F. Pempinelli, and G. Z. Tu, Physica D **9**, 425 (1983).
[17] See Lamb, Ref. 15, p. 38.

A Hamiltonian-free description of single particle dynamics for hopelessly complex periodic systems

Étienne Forest
Exploratory Studies Group, Accelerator and Fusion Research Division, Lawrence Berkeley Laboratory, 1 Cyclotron Road, MS 47-112, Berkeley, California 94720

(Received 27 July 1989; accepted for publication 10 January 1990)

A picture of periodic systems that does not rely on the Hamiltonian of the system, but on maps between a finite number of time locations, is developed. Moser or Deprit-like normalizations are done directly on the maps, thereby avoiding the complex time-dependent theory. Linear and nonlinear Floquet variables are redefined entirely in terms of maps. This approach relies heavily on the Lie representation of maps introduced by Dragt and Finn [J. Math. Phys. **20**, 2649 (1979); J. Geophys. Res. **81**, 13 (1976)]. One might say that although the Hamiltonian is not used in the normalization transformation, Lie operators are used, which are themselves, in some sense, pseudo-Hamiltonians for the maps they represent. The techniques find application in accelerator dynamics or in any field where the Hamiltonian is periodic, but hopelessly complex, such as magnetic field design in stellarators.

I. INTRODUCTION

The study of dynamical systems in traditional branches of classical mechanics uses the Hamiltonian as a starting point, which means that the numerical, as well as the analytical work, is done directly on the Hamiltonian of the system. In certain applications, such as dynamics of charged particle beams in accelerators, this approach may lead to a dead end. To illustrate this point, one needs only to compare celestial mechanics to accelerator physics. In accelerator physics, we try to understand the potential behavior of circular machines whose Hamiltonian is a very complex position-dependent function (the timelike variable is actually a length denoted by s). For example, even a small synchrotron radiation ring can consist of several dozens of bending magnets, quadrupoles, sextupoles, orbit correction dipoles, and rf cavities. The necessary inclusion of random errors in the simulation of such a device implies that the Hamiltonian will be a horrible periodic s-dependent function with the number of parameters ranging in the thousands. On the other hand, the problems of celestial mechanics, which might have a higher phase space dimensionality, are parametrized by a relatively small number of variables. In addition, accelerator Hamiltonians are very discontinuous in the timelike variable, which adds to the complexity of using a formalism devised for smooth time dependence.

Analytical computations (such as normalization procedures) have emphasized the "flow" (i.e., the Hamiltonian) instead of the study of a one-period "map" (i.e., a turn around a circular storage ring). Accelerator theorists have tried to adapt these tools to the study of circular machines, but have always had to restrict themselves to less than realistic problems. For this reason, a new approach for understanding our systems has been developed in recent years: It emphasizes the computation and analysis of large time maps. We believe that the tools that have been developed (software and theory) can be of use in other fields.

A. A bit of history

This new approach did not develop overnight. In fact, the approach has its roots deep in the field of accelerator physics. To orient the reader, we will present a subjective (not exhaustive) historical perspective of the use of flow and maps in accelerator theory and simulation. Since pioneering work in accelerator physics has often been obscure, a historical perspective may be viewed as an attempt at the proper recognition of such work.

Originally, the use of maps entered in the design of circular accelerators as paraxial (linear) representations of the ray propagation. The theory was derived from the light optics equivalent and consequently, the periodic structure of the systems was not properly exploited. Eventually, Courant and Snyder,[1] in their seminal paper on strong focusing, parametrized the motion around a linear ring in terms of invariant quantities, taking full advantage of the pseudoharmonic oscillator structure of the motion. In fact, as we will see in this paper, the Lie operator associated to the so-called Courant–Snyder invariant is proportional to the Lie operator of the linear one-turn map.

On another front, the inclusion of sextupoles in a ring worried a few people in the late 1950's. Because computers were not very powerful, Meier and Symon and Laslett *et al.*[2] used simple one-dimensional maps to guess at the potential harm caused by nonlinearities in a circular machine. As computers improved, maps disappeared from simulation and were replaced by the so-called kick codes, which are in fact second-order explicit symplectic integrators. In these codes, each time step of the integrator is derivable from a Hamiltonian. These codes are still the main ingredients of brute force simulations.[3]

On the nonlinear theoretical front, accelerator theorists tried to compute various relevant quantities such as frequency shifts (shear terms known as tune shifts in accelerator physics) and distortions of the invariants with the help of canonical perturbation theory. The theorists adapted to ac-

 0022-2488/90/051133-12$03.00

celerator problems the algorithms of Poincaré (-Von Zeippel)[4]; Moser–Birkhoff[2]; and later, Deprit.[5] The inherent complexity of our Hamiltonians leads to great technical difficulties in the application of such algorithms to realistic problems.

Meanwhile, the use of maps kept creeping into the linear theory. This probably culminated in a series of papers by Chao[6] in the late 1970's, where he computed the equilibrium emittances and spin polarization in a circular electron ring with the help of a map-based theory instead of a flow or Hamiltonian approach. One might ask why accelerator theorists continued to write papers on the evaluation of nonlinear quantities, relying entirely on the usual canonical perturbation theory, while linear calculations often used the more suitable map approach: We venture to suggest that the answer is the Lie representation of the map. Although it is obvious after a little thought that the quantities obtained by standard canonical perturbation theory are present in a power series expansion of the one-turn map,[7] the expansion of the final position and momentum (+ spin if you care about it) in terms of their initial components is very different in form from the central object of canonical perturbation theory: the Hamiltonian. Consequently, without at least an awareness of the Lie representation, accelerator theorists could not have been expected to rephrase the nonlinear theory in a way suitable for the circular machine, such as they did for the linear case.

At this point enters Dragt. In the 1970's Dragt and Finn worked on some version of the Deprit algorithm[8] and applied it to various problems of plasma physics.[9,8] Dragt became involved with accelerator theory and with the help of Douglas, they wrote the first version of a code for beam optics (MARYLIE)[10] which parametrizes the Taylor series maps in terms of their Lie generators. In collaboration with the present author, a normal form algorithm was first introduced in the code MARYLIE by Dragt *et al.*[8] Finally, Dragt, in an obscure report, introduced for the first time the concept of phase advance from a map point of view[11]: His ideas were not complete, but they planted the right seed in the present author's mind. At this point it became clear that the extraction of maps and their subsequent analysis (normalization) provided a powerful approach to numerical and analytic computation in accelerator theory. Remarkably, in 1959, Meier and Symon (Ref. 2) used a Lie representation of the map without knowing it. Meier and Symon were studying a map consisting of a rotation followed by a sextupolar kick. By writing a time-dependent pseudo-Hamiltonian that generates the exact same map, Meier and Symon were able to compute a canonical transformation in order to simplify it: Their pseudo-Hamiltonian was the factorized Lie representation of the map proposed by Dragt and Finn.[8] Meier and Symon's only error was in not trying to extirpate from the theory and normalization the bogus time dependence.

Most recently, software development has increased the numerical power of the map approach enormously. Indeed, the extraction of Taylor series representation of maps from simulation codes can be very tedious. Chao and, later, the present author restricted themselves to codes where the individual magnets had a simple representation (thin lenses): At most we could extract fifth-degree polynomial maps in six-dimensional phase space.[12] Fortunately, Berz has created a software package, the Differential Algebra Package, which permits exact automatic differentiation of any quantities integrated on the computer, in particular, the position and momentum vector that is evaluated in our simulation codes.[13] The same tools of Berz permit a user to manipulate the resulting power series map into any type of representation and in particular the factored Lie representation suggested by Dragt and Finn.[8] It suffices to say that the theoretical concepts discussed in this paper have *all* been implemented for the power series representation thanks to Berz's package.[14]

To set the tone, we will review a few concepts concerning symplectic maps.

B. A few words about symplectic maps[15–17]

A symplectic map $\mathbf{M}$ transforms a differentiable function $f(\mathbf{z}_0)$ of the initial phase space into another function $(\mathbf{M}f)(\mathbf{z}_0)$, where $\mathbf{z}_0 = (q_{01},p_{01},\ldots,q_{0N},p_{0N})$. We say that $\mathbf{M}$ is symplectic if it preserves the Poisson bracket of two functions f and g:

$$\mathbf{M}[f,g] = [\mathbf{M}f,\mathbf{M}g], \tag{1a}$$

$$\mathbf{z}_0 = (q_{01},p_{01},\ldots,q_{0N},p_{0N}). \tag{1b}$$

Now consider the motion generated by a Hamiltonian K from location s_0 to s. We know that any function $f(\mathbf{z}_0)$ will be transformed at location s into a new function $f_s(\mathbf{z}_0)$. The two functions are connected by a symplectic map $\mathbf{M}(s_0,s)$. Using the properties of Hamilton's equations, one can show that $\mathbf{M}(s_0,s)$ obeys[17]

$$\frac{d}{ds}\mathbf{M}(s_0,s) = \mathbf{M}(s_0,s):-K(\mathbf{z}_0;s):, \tag{2a}$$

$$:f:g = [f,g], \tag{2b}$$

$$\mathbf{M}(s_0,s_0) = \mathbf{E} = \text{identity map}. \tag{2c}$$

Here we follow Dragt's notation for the Lie operator [Eq. (2b)]. From Eqs. (2) we deduce that any map of the form $\mathbf{M}(\mathbf{z}_0) = \exp(:f(\mathbf{z}_0):)$ is a symplectic map. Equation (2a) is very similar to Schrödinger's equation for the unitary transformation in quantum mechanics. However, here the resulting map will be symplectic. Notice that the generator $:-K(\mathbf{z}_0;s):$ depends only on the initial phase space variable $\mathbf{z}_0$. It is also easy to show that the differential properties of any Lie operator such as $:-K:$ imply that

$$f_s(\mathbf{z}_0) = (\mathbf{M}(s_0,s)f)(\mathbf{z}_0) = f(\mathbf{M}(s_0,s)\mathbf{z}_0)$$
$$\equiv f(\mathbf{z}_s(\mathbf{z}_0)). \tag{3}$$

Finally, it is worth remembering that symplectic maps act in the reverse order from matrix multiplication when expressed in terms of the initial phase space variables. To show this property, we imagine a two-step process:

$$0 \overset{\mathbf{M}}{\rightarrow} s' \overset{\mathbf{N}}{\rightarrow} s,$$

$$\mathbf{M}(\mathbf{z}_0)\mathbf{z}_0 = \mathbf{z}_{s'}, \tag{4}$$
$$\mathbf{N}(\mathbf{z}_{s'})\mathbf{z}_{s'} = \mathbf{z}_s.$$

Here the map $\mathbf{M}$ transforms functions of the initial phase space at $s=0$, while $\mathbf{N}$ acts on functions of the phase space at s' [throughout this paper, the notation $\mathbf{M}(\mathbf{z}_0)$ indicates that the Lie operators of $\mathbf{M}$ are expressed in terms of $\mathbf{z}_0$].

Clearly, we can propagate a function $f(\mathbf{z}_0)$ to a location s by the composition of functions rule:

$$f_s(\mathbf{z}_0) = f(\mathbf{z}_s), \tag{5a}$$

but

$$f(\mathbf{z}_s) = \mathbf{N}(\mathbf{z}_{s'})f(\mathbf{z}_{s'}) = \mathbf{N}(\mathbf{z}_{s'})\mathbf{M}(\mathbf{z}_0)f(\mathbf{z}_0), \tag{5b}$$
$$= \mathbf{N}(\mathbf{M}(\mathbf{z}_0)\mathbf{z}_0)\mathbf{M}(\mathbf{z}_0)f(\mathbf{z}_0), \tag{5c}$$
$$= \mathbf{M}(\mathbf{z}_0)\mathbf{N}(\mathbf{z}_0)\mathbf{M}^{-1}(\mathbf{z}_0)\mathbf{M}(\mathbf{z}_0)f(\mathbf{z}_0), \tag{5d}$$
$$= \mathbf{M}(\mathbf{z}_0)\mathbf{N}(\mathbf{z}_0)f(\mathbf{z}_0). \tag{5e}$$

Equation (5a) is the result of simple composition. In (5b) we apply the definition of the two maps to (5a). Finally, (5c) and (5d) are the results of the differential properties of the Lie operators associated to the maps.

Indeed, it can be shown using properties (1) and (3) that

$$\exp(:f(\mathbf{M}\mathbf{z}_0):) = \exp(:\mathbf{M}f(\mathbf{z}_0):)$$
$$= \exp(\mathbf{M}:f(\mathbf{z}_0):\mathbf{M}^{-1})$$
$$= \mathbf{M}\exp(:f(\mathbf{z}_0):)\mathbf{M}^{-1}. \tag{6}$$

The reverse ordering is also present in Eq. (2a), as seen by integrating it from s to $s+ds$.

We are now ready to introduce our map description of complex periodic systems. In Sec. II and Appendix A we review in very general terms the basic Hamiltonian and canonical transformations used in circular machine theory and simulations. In Sec. III A and B we present the equivalent map description of our system by viewing the ring as an ordered set of maps. In Sec. III C we sketch the one-turn map normalization. In Sec. IV we use the one-turn map normalization to define the Floquet ring. In Sec. V these concepts are applied to the second-order normalization of a perturbed Floquet ring. In Sec. VI we apply the map approach to linear systems in the continuous limit (i.e., in the Hamiltonian limit): A set of well-known results follows explicitly. In Appendix B we sketch the proof of a few theorems.

II. THE PROBLEM OF CIRCULAR MACHINES

As mentioned in Sec. I, the need for a map-based theory is most apparent in accelerator physics. Therefore, we will describe in very general terms the central problem of accelerator design.

Consider a Hamiltonian $H(\mathbf{x},\Delta;s)$, where $\mathbf{x}$ is a phase space vector of dimension $2N$ and Δ is the set of N_p parameters describing the departure of our system from its design value (i.e., by definition, the ideal machine is described by the case $\Delta=0$). We also assume that H is periodic in s with period $s=1$:

$$H(\mathbf{x},\Delta;s+1) = H(\mathbf{x},\Delta;s). \tag{7}$$

In addition, we can select a new set of canonical variables $\mathbf{z}(\epsilon)$ that is generated by a periodic Lie operator associated to the function $w(\mathbf{z},\Delta;s;\epsilon)$:

$$\frac{d\mathbf{z}}{d\epsilon} = [\mathbf{z},w] = :-w:\mathbf{z}, \quad \mathbf{z}(\epsilon=0) = \mathbf{x}. \tag{8}$$

Ultimately the parameter ϵ is set to 1. It has been shown that the variable $\mathbf{z}(\epsilon=1)$ is propagated by a Hamiltonian $K(s)$ obeying:[18,19]

$$K(\mathbf{z},\Delta;s) = \mathbf{A}(\mathbf{z},\Delta;s)\Big(H(\mathbf{z},\Delta;s) + \int_0^{\epsilon=1} d\epsilon$$
$$\times \mathbf{A}^{-1}(\mathbf{z},\Delta;s)\frac{\partial}{\partial s}w(\mathbf{z},\Delta;s;\epsilon)\Big),$$
$$\frac{d\mathbf{z}}{ds} = [\mathbf{z},K]. \tag{9}$$

Here $\mathbf{A}^{-1}$ is the periodic canonical map that transforms $\mathbf{x}$ into $\mathbf{z}$; it is generated by w. Equations (8) and (9) can be viewed as the fundamental equations of an accelerator in the absence of collective or dissipative effects. The *fundamental problem* of accelerator dynamics is to study the stability of the motion generated by K or H as one iterates n turns around the machine from $s=s_0$ to $s=s_0+n$ $(n\rightarrow\infty)$.

Often the theorist attempts to select the generator w in a way that will simplify the structure of K. We refer to this kind of process as a normalization process. In general, the computation of K is extremely difficult because it requires a knowledge of $\mathbf{A}$ for every value of s! [For completeness, see Appendix A for the map equivalent of Eqs. (7)–(9) and a derivation of $K(\mathbf{z},\Delta;s)$.]

Typically, no attempt is made to simplify the Hamiltonian and one integrates the motion generated by H with the help of a symplectic integrator.[20] The phase space data are then examined at a finite number of surface of sections (often only one!) and all hope of analytical understanding is abandoned.

The map description of dynamics described in this paper was developed as a direct consequence of the impossibility of normalizing a realistic accelerator Hamiltonian by blindly applying a Deprit-type algorithm.

III. THE HAMILTONIAN-FREE OR MAP DESCRIPTION

A. The motivation

The map approach is based on a redefinition of the system in terms of a finite number of maps. In accelerator theory, we are motivated in redefining the problem by the following facts.

(i) Most simulations are performed by symplectic integrators.[21] Only a finite set of location $\{s_i\}$ are examined during this process. The maximum number of locations ever to be examined is the number of integration steps around the ring.

(ii) The form of the equations of motion for a computer simulation may not and will not in general be suitable for a normal form analysis.

(iii) Although symplectic maps act on the infinite-dimensional space of functions, the property displayed in Eq.

(3) permits us to restrict ourselves to the coordinate representation of the map $\mathbf{z}_s(\mathbf{z}_0) = \mathbf{M}\mathbf{z}_0$.

Statement (i) simply says that most simulations involve very discontinuous Hamiltonians in s. While it is cumbersome to fold a discontinuous Hamiltonian into a Deprit style algorithm, the production of maps and their subsequent analysis are ideally suited for discontinuous systems.

Statement (ii) points to the necessary discrepancies between the world of a computer and the world of theoretical analysis. The process we will describe allows for a total decoupling. One may extract maps using a noncanonical set of coordinates and later transform these maps into canonical variables quite independent of the system that produced them. For example, the calculation of the motion through some complex fringe field may be easiest in noncanonical variables, in some cases using even time as the Hamiltonian parameter.

Statement (iii) is extremely important: It implies that our efforts should be in the direction of extracting a representation of $\mathbf{z}_s$. For example, in the case of a Taylor series representation of $\mathbf{z}_s$, we mentioned that Berz has developed powerful software tools (the Differential Algebra Package) that perform automatic differentiation to arbitrary order on a computer,[13] making it possible to extract $\mathbf{z}_s$ as a power series around some trajectory in phase space (usually the periodic closed orbit). The same tools used by Berz allowed Irwin and the present author to write the necessary software for the normalization of the one-turn map.[14] Needless to say, a Taylor series representation may not be always suitable. Presently, nonpower series representations are being studied by Warnock *et al.* who have also developed methods to normalize the map. Warnock *et al.*'s representation can permit the study of very nonlinear processes and they succeed in many cases at finding numerically approximate invariants of the motion near chaotic regions.[22] Unfortunately, Warnock *et al.* do not have tools as flexible as the Differential Algebra Package of Berz. For this reason, thanks to Berz's tools, the power series representation of $\mathbf{z}_s$ is the only representation for which all the concepts presented in this paper are and have been implemented.

B. The redefinition of the system

1. Definition of the ring

A ring is an ordered m-tuple $\mathfrak{R} = (\mathbf{N}_{\mathbf{i}\,\mathbf{i}+\mathbf{1}})$ of m maps connecting m surfaces of sections or observation points (see Fig. 1). Here, the index $\mathbf{i}$ runs from 1 to m with the convention $\mathbf{i}+m=\mathbf{i}$. The maps in $\mathfrak{R} = (\mathbf{N}_{\mathbf{i}\,\mathbf{i}+\mathbf{1}})$ are symplectic. Without loss of generality, we assume that these maps transform the origin of phase space into itself:

$$\forall i \quad \mathbf{N}_{\mathbf{i}\,\mathbf{i}+\mathbf{1}}\mathbf{z}_0|_{\mathbf{z}_0=\mathbf{0}} = \mathbf{0}. \tag{10}$$

Clearly, the number of maps will depend on the particular aspect of the problem being studied: Again, we emphasize that it cannot exceed the number of steps in our symplectic integrator.

We are now in a position to define standard concepts on our ring. The reader must remember that the underlying assumption in this paper is our ability to extract and manipulate maps to any order in the perturbation [i.e., in the case of a power series this would be the degree in the vector $(\mathbf{z}_0,\delta)$, where δ is a subset of the total parameter set Δ of Eq. (7)].[13,14]

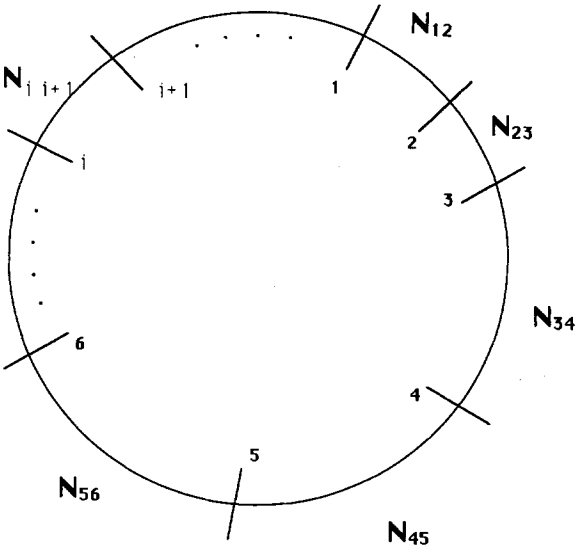


FIG. 1. Schematic view of the ring $\mathfrak{R}$.

C. The one-turn maps and their normalization[23]

In the following we sketch the steps of the normalization of a one-turn map. This formal procedure is explained in detail in Ref. 14. Given the ring $\mathfrak{R}$, we can easily compute the one-turn map from location s_i back to s_i. This map is simply given by a left to right product:

$$\mathbf{M}_\mathbf{i} = \prod_{k=i}^{i+m-1} \mathbf{N}_{\mathbf{k}\,\mathbf{k}+\mathbf{1}} \tag{11a}$$

and

$$\mathfrak{I} = (\mathbf{M}_\mathbf{i}). \tag{11b}$$

The new m-tuple $\mathfrak{I}$ that is created out of $\mathfrak{R}$ is not equivalent to $\mathfrak{R}$, as we will see in Sec. IV when we define the Floquet ring.

The first property we will assign to our map is linear stability. To define linear stability we expand the coordinate representation of $\mathbf{M}_\mathbf{i}$ around the origin:

$$\mathbf{z}_{s_\mathbf{i}}(\mathbf{z}_0) = L_i\mathbf{z}_0 + \cdots. \tag{12}$$

We assume that the $2N \times 2N$ matrix L_i has $2N$ *distinct* eigenvalues on the unit circle:

$$\lambda_j^{\pm} = \exp(\pm i2\pi\nu_j), \quad 0<\nu_j<1, \quad j=1,N. \tag{13}$$

Associated to the matrix L_i is a Lie map $\mathbf{L}_\mathbf{i}$. Clearly, the eigenvalues of L_i do not depend on location since $\mathbf{L}_\mathbf{j}$ is obtained from $\mathbf{L}_\mathbf{i}$ by a similarity transformation (the same is true for the full nonlinear map):

$$\mathbf{L}_\mathbf{j} = \mathbf{L}_{\mathbf{ij}}{}^{-1}\mathbf{L}_\mathbf{i}\mathbf{L}_{\mathbf{ij}}, \text{ where } \mathbf{L}_{\mathbf{ij}} = \prod_{k=i}^{j-1} \mathbf{L}_{\mathbf{k}\,\mathbf{k}+\mathbf{1}}. \tag{14}$$

For the linear part of the map, we claim that their exists a linear transformation $\mathbf{A}_{\mathbf{Li}}$ such that

$$\mathbf{A}_{\mathbf{Li}}\mathbf{L}_\mathbf{i}\mathbf{A}_{\mathbf{Li}}{}^{-1} = \exp(:-\boldsymbol{\mu}\cdot\mathbf{J}:) = \mathbf{R}_\mathbf{L}, \tag{15a}$$

$$\mu_j = 2\pi\nu_j, \tag{15b}$$

$$J_j = ((z_{0;2j-1})^2 + (z_{0;2j})^2)/2. \tag{15c}$$

The transformation $\mathbf{A}_{\mathrm{Li}}$ depends on the location s_i. However, the map $\mathbf{R}_{\mathrm{L}}$ is universal once a tune ν_j has been assigned a given plane j. The transformation done in Eqs. (15) is always possible for a stable linear map with distinct eigenvalues. Let us go back to the nonlinear map $\mathbf{M}_i$. Following Dragt and Finn,[16] we express it in a factored form:

$$\mathbf{M}_i = \mathbf{L}_i(\mathbf{L}_i^{-1}\mathbf{M}_i) = \mathbf{L}_i \prod_{k=1}^{N_0} \exp(:f_k(\mathbf{z}_0):) \tag{16}$$

In the case of a power series representation of $\mathbf{M}_i$, $N_0 - 1$ is the degree of the polynomial approximation of the function $\mathbf{z}_{s_i}(\mathbf{z}_0)$ and $k+2$ is the degree in $\mathbf{z}_0$ of the Lie exponent. For a different type of approximation the reader can view N_0 and k as the degree of some smallness parameter. Our ultimate goal is to normalize $\mathbf{M}_i$ partially or totally. In a total normalization, we must find a transformation $\mathbf{A}_i$ (analytic around the origin) such that

$$\mathbf{A}_i\mathbf{M}_i\mathbf{A}_i^{-1} = \exp(:-\boldsymbol{\mu}\cdot\mathbf{J} + D(\mathbf{J}):) = \mathbf{R}. \tag{17}$$

For analytic $\mathbf{A}_i$ one can show that $\mathbf{R}$ is unique for a given ordering of the planes and independent of the location (see Appendix B).

Obviously, using Eqs. (15), we factor $\mathbf{A}_i$ into linear and nonlinear parts:

$$\mathbf{A}_i = \mathbf{A}_{\mathrm{Ni}}\mathbf{A}_{\mathrm{Li}}, \tag{18a}$$

$$\mathbf{A}_{\mathrm{Ni}} = \mathbf{A}_{\mathrm{N},i}\cdots\mathbf{A}_{1i}, \tag{18b}$$

$$\mathbf{A}_{ki} = \exp(:F_k:), \quad k = 1, N_0. \tag{18c}$$

Applying (18) on $\mathbf{M}_i$ gives us

$$\mathbf{A}_i\mathbf{M}_i\mathbf{A}_i^{-1} = \mathbf{A}_{\mathrm{Ni}}\mathbf{R}_{\mathrm{L}} \prod_{k=1}^{N_0} \exp(:g_{ki}(\mathbf{z}_0):)\mathbf{A}_{\mathrm{Ni}}^{-1}, \tag{19a}$$

where

$$g_{ki}(\mathbf{z}_0) = \mathbf{A}_{\mathrm{L}i}f_{ki}(\mathbf{z}_0) = f_{ki}(\mathbf{A}_{\mathrm{L}i}\mathbf{z}_0). \tag{19b}$$

The normalization of the nonlinear map starts with Eqs. (19). To see the type of operators involved, let us compute the effect of $\mathbf{A}_{1i}$ on the second-degree map $\mathbf{R}_{\mathrm{L}}\exp(:g_{1i}:)$:

$$\begin{aligned}
&\mathbf{A}_{1i}\mathbf{R}_{\mathrm{L}}\exp(:g_{1i}:)\mathbf{A}_{1i}^{-1} \\
&\quad = \exp(:F_{1i}:)\mathbf{R}_{\mathrm{L}}\exp(:g_{1i}:)\exp(:-F_{1i}:) \\
&\quad = \mathbf{R}_{\mathrm{L}}\mathbf{R}_{\mathrm{L}}^{-1}\exp(:F_{1i}:)\mathbf{R}_{\mathrm{L}}\exp(:g_{1i}:)\exp(:-F_{1i}:) \\
&\quad = \mathbf{R}_{\mathrm{L}}\exp(:\mathbf{R}_{\mathrm{L}}^{-1}F_{1i}:)\exp(:g_{1i}:)\exp(:-F_{1i}:).
\end{aligned} \tag{20}$$

To first order in the Lie exponents, we can rewrite (20) as

$$\begin{aligned}
&\mathbf{R}_{\mathrm{L}}\exp(:\mathbf{R}_{\mathrm{L}}^{-1}F_{1i}:)\exp(:g_{1i}:)\exp(:-F_{1i}:) \\
&\quad = \mathbf{R}_{\mathrm{L}}\exp(:-(\mathbf{E}-\mathbf{R}_{\mathrm{L}}^{-1})F_{1i} + g_{1i}:) + \mathrm{order}(g_2) \\
&\quad = \mathbf{R}_{\mathrm{L}}\exp(:-\mathbf{T}F_{1i} + g_{1i}:) + \mathrm{order}(g_2)\cdots,
\end{aligned} \tag{21a}$$

$$\mathbf{T} = \mathbf{E} - \mathbf{R}_{\mathrm{L}}^{-1}, \tag{21b}$$

where $\mathbf{E}$ is identity map.

From the Eqs. (21), we see that the operator $\mathbf{T}$ is central to the understanding of the effect of any similarity transformation. Since $\mathbf{T}$ is essentially $\mathbf{R}_{\mathrm{L}}$, we must study the Lie operator $:\boldsymbol{\mu}\cdot\mathbf{J}:$. As pointed out by Cushman *et al.*,[24] as well as Dragt and Finn,[8] the operator $:\boldsymbol{\mu}\cdot\mathbf{J}:$ is a semisimple endomorphism of the space P_k of homogeneous polynomials of degree $k \geqslant 1$ in $\mathbf{z}_0$; hence it is true that

$$P_k = \mathrm{Im}\,\mathbf{T} \oplus \mathrm{Ker}\,\mathbf{T}. \tag{22}$$

In fact, it is easy to derive Eq. (22) by simply constructing the linear eigenfunctions of $\mathbf{T}$ (or $:\boldsymbol{\mu}\cdot\mathbf{J}:$) (Ref. 8):

$$:\boldsymbol{\mu}\cdot\mathbf{J}:h_j^{\pm} = \pm i\mu_j h_j^{\pm}, \tag{23a}$$

$$h_j^{\pm} = z_{2j-1} \pm iz_{2j} = \sqrt{2J_j}\exp(\mp i\varphi_j),$$

where

$$[\varphi_j, J_j] = \delta_{ij} \tag{23b}$$

are the only nonzero brackets,

$$J_j = \tfrac{1}{2}h_j^{+}h_j^{-}, \quad j = 1, N. \tag{23c}$$

For completeness, in (23b) we displayed the connection between our eigenfunctions and the usual set of action-angle variables (J_j, φ_j).

Using these linear eigenvectors, we can construct an eigenbasis for P_k:

$$\begin{aligned}
&|\mathbf{m},\mathbf{n}\rangle = (h_1^{+})^{m_1}(h_1^{-})^{n_1}\cdots(h_N^{+})^{m_N}(h_N^{-})^{n_N} \\
&:\boldsymbol{\mu}\cdot\mathbf{J}:|\mathbf{m},\mathbf{n}\rangle = i(\mathbf{m}-\mathbf{n})\cdot\boldsymbol{\mu}|\mathbf{m},\mathbf{n}\rangle.
\end{aligned} \tag{24}$$

Since $\mathbf{T}$ is diagonal in the $|\mathbf{m},\mathbf{n}\rangle$ basis, it follows that P_k decomposes into the direct sum of its image and its kernel. In fact, if the μ's are irrational among each other, the kernel is given by

$$|\mathbf{m},\mathbf{n}\rangle \in \mathrm{Ker}\,\mathbf{T} \Rightarrow \mathbf{m}-\mathbf{n} = \mathbf{0}, \quad \text{i.e.,} \quad |\mathbf{m}-\mathbf{n}| = 0. \tag{25}$$

In a partial normalization, we decide to leave in the final map terms for which $\mathbf{m}-\mathbf{n} \neq \mathbf{0}$ [or often $(\mathbf{m}-\mathbf{n})\cdot\boldsymbol{\mu} \cong 0$]. This allows us to study islands produced by a resonance.

We say that $\mathbf{M}_i$ is partially normalized into the map $\mathbf{R}$ if

$$\mathbf{A}_i\mathbf{M}_i\mathbf{A}_i^{-1} = \exp\Big(:-\boldsymbol{\mu}\cdot\mathbf{J} + D_i(\mathbf{J}) + \sum_{\mathbf{m}-\mathbf{n}\in\mathbf{I}_r} D_{\mathbf{m},\mathbf{n};i}|\mathbf{m},\mathbf{n}\rangle:\Big) = \mathbf{R}_i, \tag{26}$$

where $I_r = \{\mathbf{k} \in Z^N \mid \mathbf{k} = \text{selected resonances}\}$. As indicated in (26) by the index i, in a partial normalization the final map will depend on the location in the ring. According to Eqs. (21), we obtain $\mathbf{R}$ by inverting $\mathbf{T}$. In fact we can redefine $\mathbf{T}^{-1}$ using a projection operator:

$$\begin{aligned}
\mathbf{T}_{\mathrm{im}}^{-1}g &= \mathbf{T}^{-1}\mathbf{P}_{\mathrm{im}}g \\
&= \sum_{\mathbf{m}-\mathbf{n}\in\mathbf{I}_r\cup\{0\}} \frac{\mathbf{A}_{\mathbf{m},\mathbf{n}}}{1-\exp(i(\mathbf{m}-\mathbf{n})\cdot\boldsymbol{\mu})}|\mathbf{m},\mathbf{n}\rangle,
\end{aligned} \tag{27a}$$

$$\mathbf{P}_{\mathrm{im}}g = \sum_{\mathbf{m}-\mathbf{n}\in\mathbf{I}_r\cup\{0\}} \mathbf{A}_{\mathbf{m},\mathbf{n}}|\mathbf{m},\mathbf{n}\rangle, \tag{27b}$$

$$g = \sum_{\mathbf{m},\mathbf{n}} \mathbf{A}_{\mathbf{m},\mathbf{n}}|\mathbf{m},\mathbf{n}\rangle. \tag{27c}$$

Provided that one knows how to compose the maps involved in the normalization and extract their leading order Lie representation, the maps $\mathbf{A}_{ki}$ can be computed by iteration using $\mathbf{T}_{\mathrm{im}}^{-1}$ as defined in Eqs. (27). This procedure was first implemented to third order in the Taylor series by the present author and Dragt in the context of the code MAR-

YLIE.[10] Later, Neri and Dragt pushed the process to fifth order using the same code. Recently, the present author and Irwin, in close collaboration with Berz,[14] developed the algorithm and software necessary to extend the map normalization to an arbitrary order. The process is semianalytic since in practice only a small number of components of Δ [see Eqs. (7)–(9)] can be retained. This number can vary depending on the order of the normalization, the phase space dimension, and the power of the computer used.

For our purpose, it suffices to know that one can define (exactly for linear maps and formally in the nonlinear case) a normalized map $\mathbf{R}$. Although we concentrate in this article on a normal form algorithm based on the semisimple operator $:\boldsymbol{\mu}\cdot\mathbf{J}:$, it is possible and sometimes desirable to study systems that are not semisimple.[14,24] What can be done on the Hamiltonian can also be done on the map.

IV. THE FLOQUET RING

Let us assume that we have achieved a complete normalization of the ring. As described by Eq. (17), we have

$$\begin{aligned} &\rho(\mathbf{M}_i)=\mathbf{A}_i, \quad \rho:\mathfrak{I}\rightarrow\text{symplectic maps},\\ &\mathbf{A}_i\mathbf{M}_i\mathbf{A}_i^{-1}=\exp(:-\boldsymbol{\mu}\cdot\mathbf{J}+D(\mathbf{J}):)=\mathbf{R}. \end{aligned} \tag{28}$$

The transformation ρ introduced in (28) can be viewed as a map over the set of one-turn maps $\mathfrak{I}$ defined over $\mathfrak{R}$. Using ρ we can define a new ring: the Floquet ring. We first proceed by mapping the m-tuple $\mathfrak{I}$:

$$\mathfrak{I}_\rho=\rho(\mathfrak{I})=(\rho(\mathbf{M}_i)\mathbf{M}_i\rho(\mathbf{M}_i)^{-1})=(\mathbf{R},\mathbf{R},\ldots,\mathbf{R}). \tag{29}$$

The fact that $\mathfrak{I}_\rho$ contains only $\mathbf{R}$ is demonstrated in Appendix B by generalizing Eq. (14) to the maps $\mathbf{N}_{k\,k+1}$ and using the assumed analyticity of the maps that are involved.

More important, we must find out what happens to $\mathfrak{R}$. We first state the result.

(i) The new ring $\rho(\mathfrak{R})$ (or $\mathfrak{R}_\rho$) is made out of amplitude-dependent rotations (called phases). The angles of these rotations reduce to the so-called linear phase advance in the linear regime.

(ii) Two different normalizations ρ and ρ' can only differ by a phase for a given ordering of the tunes.

Corollary: The phase advance between two matched locations ($\mathbf{M}_i=\mathbf{M}_j$) is the same for any definition of ρ.

Proof: We now prove the above results.

Property (i): Using the normalization ρ, we conclude from Eq. (29) that

$$\rho(\mathbf{M}_i)\mathbf{M}_i\rho(\mathbf{M}_i)^{-1}=\rho(\mathbf{M}_j)\mathbf{M}_j\rho(\mathbf{M}_j)^{-1}=\mathbf{R}. \tag{30}$$

Using the definition of $\mathfrak{I}$, we may write

$$\mathbf{M}_j=\mathbf{N}_{ij}^{-1}\mathbf{M}_i\mathbf{N}_{ij}. \tag{31}$$

We can substitute (31) into (30):

$$\begin{aligned} &\rho(\mathbf{M}_i)\mathbf{M}_i\rho(\mathbf{M}_i)^{-1}\\ &\quad=\rho(\mathbf{M}_j)\mathbf{N}_{ij}^{-1}\mathbf{M}_i\mathbf{N}_{ij}\rho(\mathbf{M}_j)^{-1}\\ &\quad\Rightarrow\mathbf{M}_i=\rho(\mathbf{M}_i)^{-1}\rho(\mathbf{M}_j)\mathbf{N}_{ij}^{-1}\mathbf{M}_i\mathbf{N}_{ij}\rho(\mathbf{M}_j)^{-1}\rho(\mathbf{M}_i)\\ &\quad\Rightarrow\mathbf{R}=\rho(\mathbf{M}_j)\mathbf{N}_{ij}^{-1}\rho(\mathbf{M}_i)^{-1}\mathbf{R}\rho(\mathbf{M}_i)\mathbf{N}_{ij}\rho(\mathbf{M}_j)^{-1}\\ &\quad\Rightarrow\mathbf{R}=\mathbf{B}_{ij}^{-1}\mathbf{R}\mathbf{B}_{ij}, \end{aligned}$$

where

$$\mathbf{B}_{ij}=\rho(\mathbf{M}_i)\mathbf{N}_{ij}\rho(\mathbf{M}_j)^{-1}. \tag{32}$$

We now take advantage of the Lie algebraic representation of $\mathbf{R}$:

$$\begin{aligned} \mathbf{R}&=\mathbf{B}_{ij}^{-1}\exp(:-\boldsymbol{\mu}\cdot\mathbf{J}+D(\mathbf{J}):)\mathbf{B}_{ij}\\ &\Rightarrow\exp(:-\boldsymbol{\mu}\cdot\mathbf{J}+D(\mathbf{J}):)\\ &=\exp(:\mathbf{B}_{ij}^{-1}(-\boldsymbol{\mu}\cdot\mathbf{J}+D(\mathbf{J})):). \end{aligned} \tag{33}$$

Using the assumed analyticity of the various maps involved in Eq. (33), one can show that $\mathbf{B}_{ij}$ can depend only on $\mathbf{J}$ (see Appendix B); hence it can be written with a single Lie operator Φ_{ij}:

$$\mathbf{B}_{ij}=\exp(:-\Phi_{ij}(\mathbf{J}):). \tag{34}$$

The angle of the rotation produced by $\mathbf{B}_{ij}$ is simply

$$\Delta\Phi_{ij}=-\frac{\partial\Phi_{ij}}{\partial\mathbf{J}}. \tag{35}$$

Using Eqs. (32) and (35), we can define the Floquet ring $\mathfrak{R}_\rho$ associated to ρ to be the m-tuple

$$\begin{aligned} &\mathfrak{R}_\rho=(\mathbf{B}_{k\,k+1}) \text{ such that}\\ &\mathbf{B}_{k\,k+1}=\rho(\mathbf{M}_k)\mathbf{N}_{k\,k+1}\rho(\mathbf{M}_{k+1})^{-1}. \end{aligned} \tag{36}$$

Property (ii): Finally, from Eq. (30) and the uniqueness of $\mathbf{R}$, we obtain a relation identical to Eq. (33) in the presence of two different normalizations ρ and ρ':

$$\begin{aligned} \mathbf{R}&=\exp(:-\boldsymbol{\mu}\cdot\mathbf{J}+D(\mathbf{J}):)\\ &=\rho(\mathbf{M}_i)\rho'(\mathbf{M}_i)^{-1}\exp(:-\boldsymbol{\mu}\cdot\mathbf{J}+D(\mathbf{J}):)\\ &\quad\times\rho'(\mathbf{M}_i)^{-1}\rho(\mathbf{M}_i). \end{aligned} \tag{37}$$

Hence $\rho(\mathbf{M}_i)\rho'(\mathbf{M}_i)^{-1}$ is a rotation and equivalent normalizations can only differ by a phase. It is a simple exercise to prove the corollary on matched locations.

V. PERTURBATION OF THE RING $\mathfrak{R}$

Often one perturbs a Hamiltonian at several locations. One would like to know how the ring $\mathfrak{R}$ and its Floquet counterpart $\rho(\mathfrak{R})$ are affected by perturbations, in particular Hamiltonian perturbations.

Let us assume that the ring is perturbed at the ith location by a Lie operator $\mathbf{C}_i=\exp(:-V_i:)$. In accelerator physics, this kind of question is often asked. For example, $\mathbf{C}_i$ could represent a nonlinear multipole error or a beam–beam kick: The list is endless. Clearly, the perturbed ring $\mathfrak{R}^p$ is just the m-tuple

$$\mathfrak{R}^p=(\mathbf{C}_i\mathbf{N}_{i\,i+1}). \tag{38}$$

More interesting, we would like to examine the perturbed Floquet ring:

$$\begin{aligned} \mathfrak{R}_f^p&=(\mathbf{A}_i\mathbf{C}_i\mathbf{A}_i^{-1}\mathbf{B}_{i\,i+1})\\ &=(\exp(:-\mathbf{A}_iV_i:)\mathbf{B}_{i\,i+1})\\ &=(\exp(:-V_i(\mathbf{A}_i\mathbf{z}_0):)\mathbf{B}_{i\,i+1}). \end{aligned} \tag{39}$$

In Dragt's original paper on lattice functions,[11] he refers to $\mathbf{A}_i^{-1}$ as the "irritability": In some sense, it gives the true extent of the damage done on the Floquet ring.

As an example of the use of Eq. (39), let us completely normalize $\mathfrak{R}_f^p$ to second order in the perturbation V_i. This has practical application in the design of a large synchrotron ring, where one needs to keep the shear terms resulting from sextupoles under control. This process is schematically displayed in Fig. 2.

To proceed as before, we first compute the one-turn maps:

$$\mathbf{R}_i = \prod_{k=i}^{i+m-1} \exp(:-V_k(\mathbf{A}_k \mathbf{z}_0):)\mathbf{B}_{k\,k+1},$$
$$\mathfrak{I}_f^p = (\mathbf{R}_i). \tag{40}$$

We then isolate the perturbations on the rhs of the factored product of $\mathbf{R}_i$:

$$\mathbf{R}_i = \mathbf{R}\mathbf{R}^{-1}\left(\prod_{k=i}^{i+m-1} \exp(:-V_k(\mathbf{B}_{i\,k}\mathbf{A}_k \mathbf{z}_0):)\right)\mathbf{R},$$
$$\mathbf{R}_i = \mathbf{R}\left(\prod_{k=i}^{i+m-1} \exp(:-V_k(\mathbf{B}_{k\,i+m}{}^{-1}\mathbf{A}_k \mathbf{z}_0):)\right), \tag{41}$$
$$\mathbf{R}_i = \mathbf{R}\Pi_i.$$

To second order in the perturbation, we can factor Π_i:

$$\Pi_i = \exp(:W_{1i}:)\exp(:W_{2i}:)\cdots, \tag{42a}$$
$$W_{1i} = \sum_{k=i}^{i+m-1} -V_k(\mathbf{B}_{k\,i+m}{}^{-1}\mathbf{A}_k \mathbf{z}_0), \tag{42b}$$
$$W_{2i} = \frac{1}{2}\sum_{k=i}^{i+m-1}\sum_{k'=k+1}^{i+m-1}[-V_k(\mathbf{B}_{k\,i+m}{}^{-1}\mathbf{A}_k \mathbf{z}_0), - V_{k'}(\mathbf{B}_{k'i+m}{}^{-1}\mathbf{A}_{k'} \mathbf{z}_0)]. \tag{42c}$$

The expression for W_{2i} is the result of a simple application of the Campbell–Baker–Hausdorff formula.

The normalization starts with the application of $\mathbf{A}_{1i} = \exp(:F_{1i}:)$:

$$\mathbf{A}_{1i}\mathbf{R}_i\mathbf{A}_{1i}{}^{-1} = \mathbf{A}_{1i}\mathbf{R}\Pi_i\mathbf{A}_{1i}{}^{-1},$$
$$F_{1i} = \mathbf{T}_{\mathrm{im}}{}^{-1}W_{1i}, \quad D_{w1}(\mathbf{J}) = (\mathbf{E} - \mathbf{P}_{\mathrm{im}})W_{1i}. \tag{43}$$

In the case of a complete normalization, $D_{w1}(\mathbf{J})$ is often known as the average (or secular) term, which in the usual action-angle representation of (23b) has the form

$$D_{w1}(\mathbf{J}) = \langle W_1 \rangle = \frac{1}{(2\pi)^N}\int_0^{2\pi}\cdots\int_0^{2\pi} W_{1i}(\varphi,\mathbf{J})d^N\varphi. \tag{44}$$

The resulting map is given by

$$\mathbf{A}_{1i}\mathbf{R}_i\mathbf{A}_{1i}{}^{-1} = \mathbf{R}\exp(:D_{w1}(\mathbf{J}):)\exp(:W_{2i}^1:),$$
$$W_{2i}^1 = \tfrac{1}{2}[D_{w1},\mathbf{P}_{\mathrm{im}}W_{1i} - 2F_{1i}] + \tfrac{1}{2}[F_{1i},\mathbf{P}_{\mathrm{im}}W_{1i}] + W_{2i}. \tag{45}$$

We now proceed with the second-order calculation:

$$F_{2i} = \mathbf{T}_{\mathrm{im}}{}^{-1}W_{2i}^1$$
$$= \mathbf{T}^{-1}\tfrac{1}{2}[D_{w1},\mathbf{P}_{\mathrm{im}}W_{1i} - 2F_{1i}] + \mathbf{T}_{\mathrm{im}}{}^{-1}(\tfrac{1}{2}[F_{1i},\mathbf{P}_{\mathrm{im}}W_{1i}] + W_{2i}), \tag{46a}$$
$$D_{w2}(\mathbf{J}) = (\mathbf{E} - \mathbf{P}_{\mathrm{im}})(\tfrac{1}{2}[F_{1i},\mathbf{P}_{\mathrm{im}}W_{1i}] + W_{2i}). \tag{46b}$$

The first term of W_{2i}^1 is entirely in the range of the operator $\mathbf{T}$. This completes the second-order normalization process. To second order in the $\mathbf{C}_i$'s, the Floquet ring is given by

$$\mathfrak{R}_f = (\exp(:F_{2k}:)\exp(:F_{1k}:)\mathbf{A}_k\mathbf{C}_k\mathbf{A}_k{}^{-1}\mathbf{B}_{k\,k+1} \times \exp(:-F_{1\,k+1}:)\exp(:-F_{2\,k+1}:)), \tag{47a}$$
$$\mathfrak{I}_f = (\mathbf{R}\exp(:D_{w1}(\mathbf{J}) + D_{w2}(\mathbf{J}):)) = (\exp(:-\boldsymbol{\mu}\cdot\mathbf{J} + D(\mathbf{J}) + D_{w1}(\mathbf{J}) + D_{w_2}(\mathbf{J}):)). \tag{47b}$$

VI. THE LINEAR PHASE ADVANCE

When new techniques are introduced, it is instructive to compare the approach with the old techniques whenever they exist. The difficulty in doing so is proportional to the enhanced power the new methods provide over the old ones. Therefore, while the mathematical equivalence is not in doubt, it is hard to work out a nontrivial and nonlinear example which explicitly displays the mathematical equivalence. Therefore, we will settle for a linear example. The reader with a knowledge of accelerator theory will see here an explicit connection between the two methods by allowing our ring to become an "∞-tuple," i.e., by reverting to the Hamiltonian. Clearly, from Eq. (2a), the Hamiltonian picture corresponds to the maximum ring $\mathfrak{R}_\infty$:

$$\mathfrak{R}_{\rho\infty} = \lim_{\substack{ds\to 0 \\ s\in[0,1]}} (\mathbf{N}_{s\,s+ds}) = \lim_{\substack{ds\to 0 \\ s\in[0,1]}} (\mathbf{E} + ds:-H(\mathbf{x}_0;s):) \tag{48}$$

First, let us state a few well-known results. In the one-dimensional case, where the Hamiltonian is given by

$$H = \tfrac{1}{2}(p^2 + k(s)q^2), \quad \mathbf{x} = (q,p), \tag{49}$$

a stable one-turn map can be parametrized by the so-called Twiss parameters[15]:

$$\mathbf{L}_s = \exp(:-\mu\mathbf{I}:) = \exp(:-\tfrac{1}{2}\mu c:), \tag{50a}$$
$$c = \gamma(s)q_0{}^2 + 2\alpha(s)q_0p_0 + \beta(s)p_0{}^2, \quad 1 + \alpha^2 = \beta\gamma. \tag{50b}$$

It is easy to verify that the matrix representation L_s of $\mathbf{L}_s$ is given by[1,15]

$$L_s = \begin{pmatrix} \cos\mu + \alpha\sin\mu & \beta\sin\mu \\ -\gamma\sin\mu & \cos\mu - \alpha\sin\mu \end{pmatrix}. \tag{51}$$

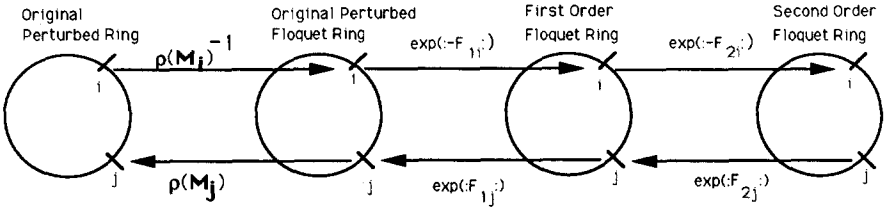


FIG. 2. Schematic view of the second-order normalization process.

The invariant c of Eqs. (50) is called the Courant–Snyder invariant in accelerator physics literature. From our previous discussion it is clear that the Twiss parameters (α,β,γ) are s dependent. Incidently, the maps $\mathbf{L}_s$ define our one-turn ∞-tuple, while H defines the ring

$$\mathfrak{R}_\infty = \lim_{\substack{ds\to 0 \\ s\in[0,1]}} \left(\mathbf{E} + ds{:}\, -\frac{1}{2}(p_0^2 + k(s)q_0^2){:}\right), \tag{52a}$$

$$\mathfrak{I}_\infty = (\mathbf{L}_s)_{s\in[0,1]}. \tag{52b}$$

To proceed further, we must define the map $\rho(\mathbf{L}_s)$. Following Courant–Snyder,[1] we define $\rho(\mathbf{L}_s)$ as

$$A_{L_s} = \rho(L_s) = \begin{pmatrix} \sqrt{\beta} & 0 \\ -\alpha/\sqrt{\beta} & 1/\sqrt{\beta} \end{pmatrix}. \tag{53}$$

Equation (53) uniquely defines $\mathbf{A}_{\mathbf{L}_s}$ given a linear map $\mathbf{L}_s$. Given (52) and (53), we will prove, using our concepts, that the Floquet ring and resulting phase advance are given by

$$\mathfrak{R}_{\rho\infty} = \lim_{\substack{ds\to 0 \\ s\in[0,1]}} \left(\mathbf{E} + ds{:}\, -\frac{1}{2\beta}(p_0^2 + q_0^2){:}\right)$$

$$= \lim_{\substack{ds\to 0 \\ s\in[0,1]}} \left(\mathbf{E} + ds{:}\, -\frac{1}{\beta} J{:}\right), \tag{54a}$$

$$\Delta\Phi_{s_0 s_1} = \int_{s_0}^{s_1} \frac{ds}{\beta}. \tag{54b}$$

Before proving a generalization of this result, we point out that the choice of Courant–Snyder was dictated by the kind of perturbation expected in an accelerator. In our machine, we expect the perturbation $\mathbf{C}_i = \exp(:-V_i:)$ of Sec. V to depend mostly on the position vector $\mathbf{q}$ because the leading contribution to the perturbed Hamiltonian is proportional to the longitudinal component of a magnetic vector potential. Therefore, a choice of $\rho(L_s)$ that minimizes the change in the functional form of V_i is best. We can generalize the Courant–Snyder choice to a higher dimensionality. The resulting phase advance formula is given for the Hamiltonian

$$H(\mathbf{x};s) = \frac{1}{2}\sum_{i=1,j=1}^{2N} \mathbf{H}_{ij}(s) X_i X_j \tag{55a}$$

and for $A = \rho(L)$ such that

$$A_{2i-1\,2i-1} > 0, \quad A_{2i-1\,2i} = 0, \quad i = 1,N; \tag{55b}$$

by [(55a) and (55b)]

$$\Rightarrow \frac{d\Phi_i}{ds} = \sum_{j=1}^{N} \frac{\mathbf{H}_{2ij}\mathbf{A}_{j2i}}{\mathbf{A}_{2i-1\,2i-1}}. \tag{55c}$$

Proof: We rewrite Eq. (32) for an infinitesimal change in s:

$$\mathbf{B}_{\Delta\Phi} = \exp(:-\Delta\Phi\cdot\mathbf{J}:)$$
$$= \mathbf{A}_s(\mathbf{E} + ds{:}H(\mathbf{x}_0;s){:})\mathbf{A}_{s+ds}^{-1}$$
$$\Rightarrow \mathbf{A}_{s+ds}$$
$$= (\mathbf{E} + :\Delta\Phi\cdot\mathbf{J}:)\mathbf{A}_s(\mathbf{E} + ds{:}\, -H(\mathbf{x}_0;s){:})$$
$$+ O(ds^2)\cdots. \tag{56}$$

Next we assume that $\mathbf{A}_s$ obeys (55b) and we impose on $\mathbf{A}_{s+ds}$ the same condition (j and k are summed over):

$$\begin{aligned}
\mathbf{A}_{s+ds}X_{0;2i-1} &= (\mathbf{E} + :\Delta\Phi\cdot\mathbf{J}:)\mathbf{A}_s(X_{0;2i-1} + ds[-H(\mathbf{x}_0;s), X_{0;2i-1}]) \\
&= (\mathbf{E} + :\Delta\Phi\cdot\mathbf{J}:)\mathbf{A}_s(X_{0;2i-1} + ds\,\mathbf{H}_{2ij}X_{0;j}) \\
&= (\mathbf{E} + :\Delta\Phi\cdot\mathbf{J}:)(A_{2i-1\,k}X_{0;k} + ds\,\mathbf{H}_{2ij}A_{jk}X_{0;k}) \\
&= (A_{2i-1\,k}X_{0;k} + (-\Delta\Phi_k A_{2i-1\,2k-1}X_{0;2k} + \Delta\Phi_k A_{2i-1\,2k}X_{0;2k-1} + ds\,\mathbf{H}_{2ij}A_{jk}X_{0;k})).
\end{aligned} \tag{57}$$

We extract from (57) the $(2i-1,2i)$ component of A_{s+ds} and set it to zero:

$$0 = ds\,\mathbf{H}_{2ij}A_{j2i} - \Delta\Phi_i A_{2i-1\,2i-1}$$

$$\Rightarrow \text{property (55c)}. \tag{58}$$

Q.E.D.

We can apply formula (58) to a problem already solved by Edwards and Teng[25], where they accidently chose the same definition for the transformation $\rho(L)$.

In Edwards and Teng's case, the Hamiltonian matrix $\mathbf{H}_{ij}$ was

$$\mathbf{H} = \begin{pmatrix} F & 0 & K & -L \\ 0 & 1 & L & 0 \\ K & L & G & 0 \\ -L & 0 & 0 & 1 \end{pmatrix}; \tag{59}$$

they parametrized the matrix A as

$$A = BC, \quad B = \begin{pmatrix} \mathbf{I}\cos\phi & \mathbf{D}^{-1}\sin\phi \\ -\mathbf{D}\sin\phi & \mathbf{I}\cos\phi \end{pmatrix}, \tag{60a}$$

$$\cos(\phi) > 0, \quad \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad |\mathbf{D}| = 1, \tag{60b}$$

$$C = \begin{pmatrix} \mathbf{b}_1 & 0 & 0 \\ & 0 & 0 \\ 0 & 0 & \\ 0 & 0 & \mathbf{b}_2 \end{pmatrix}, \mathbf{b}_i = \begin{pmatrix} \sqrt{\beta_i} & 0 \\ -\alpha_i/\sqrt{\beta_i} & 1/\sqrt{\beta_i} \end{pmatrix}. \tag{60c}$$

Because the map A obeys (55b), we can compute the phase advance using (55c):

$$\frac{d\Phi_1}{ds} = \frac{(1 - Lb\tan(\phi))}{\beta_1}, \tag{61a}$$

$$\frac{d\Phi_2}{ds} = \frac{(1 + Lb\tan(\phi))}{\beta_2}. \tag{61b}$$

Results (61) are exactly those obtained by Edwards and Teng through a totally different method.[25]

In a final exercise, we would like to explore the relationship between two definitions of the phase advance. Going back to the original one-dimensional problem of Courant–Snyder,[1] we can examine the following definition:

$$A_{L_s} = \rho'(L_s) = \begin{pmatrix} 1/\sqrt{\gamma} & -\alpha/\sqrt{\gamma} \\ 0 & \sqrt{\gamma} \end{pmatrix}. \tag{62}$$

If the world was made out of primarily velocity-dependent potentials, Courant and Snyder would have selected ρ'. The phase advance is obtained by a symmetry argument [canonically exchanging q and p and applying (55c)]:

$$\Delta\Phi_{s_0 s_1} = \int_{s_0}^{s_1} \frac{k(s)ds}{\gamma}. \tag{63}$$

Using the ring given by (48) and the underlying Hamiltonian given by (49), we can derive a famous set of rules for the evolution of the Twiss parameters (α,β,γ). This will allow us to relate the phase advances of ρ and and ρ' explicitly.

$$\begin{aligned}
\mathbf{L}_{s+ds} &= \mathbf{N}^{-1}_{s\,s+ds}\mathbf{L}_s\mathbf{N}_{s\,s+ds} \\
&= \mathbf{N}^{-1}_{s\,s+ds}\exp(:-\tfrac{1}{2}\mu c:)\mathbf{N}_{s\,s+ds} = \exp(:-\tfrac{1}{2}\mu\mathbf{N}^{-1}_{s\,s+ds}c:) \\
&= \exp(:-\tfrac{1}{2}\mu(\mathbf{E}-ds:-\tfrac{1}{2}(p_0^2+k(s)q_0^2):)(\gamma(s)q_0^2+2\alpha(s)q_0p_0+\beta(s)\rho_0^2):) \\
&= \exp(:-\tfrac{1}{2}\mu(c-ds[-\tfrac{1}{2}(p_0^2+k(s)q_0^2),\gamma(s)q_0^2+2\alpha(s)q_0p_0+\beta(s)p_0^2]):) \\
&\Rightarrow \frac{d\alpha}{ds} = \beta k - \gamma, \quad \frac{d\beta}{ds} = -2\alpha, \quad \frac{d\gamma}{ds} = 2\alpha k.
\end{aligned} \tag{64}$$

Using the relations of (64) and the relation $1+\alpha^2=\beta\gamma$, we obtain

$$\begin{aligned}
\Delta\Phi_{s_0 s_1} &= \int_{s_0}^{s_1}\frac{k(s)ds}{\gamma} = \int_{s_0}^{s_1}\frac{\gamma+d\alpha/ds}{\beta\gamma}ds \\
&= \int_{s_0}^{s_1}\frac{1}{\beta}ds + \tan^{-1}(\alpha)\Big|_{s=s_0}^{s=s_1}.
\end{aligned} \tag{65}$$

In Eq. (65) we explicitly demonstrated that matched $(\alpha_1=\alpha_2)$ locations are separated by the same phase advance. If we then multiply ρ'^{-1} by ρ [Eqs. (53) and (62)], we obtain another advertised result:

$$\begin{aligned}
&\begin{pmatrix} \sqrt{\gamma} & \alpha/\sqrt{\gamma} \\ 0 & 1/\sqrt{\gamma} \end{pmatrix}\begin{pmatrix} \sqrt{\beta} & 0 \\ -\alpha/\sqrt{\beta} & 1/\sqrt{\beta} \end{pmatrix} \\
&\quad = \begin{pmatrix} 1/\sqrt{1+\alpha^2} & \alpha/\sqrt{1+\alpha^2} \\ -\alpha/\sqrt{1+\alpha^2} & 1/\sqrt{1+\alpha^2} \end{pmatrix}.
\end{aligned} \tag{66}$$

As we stated previously, the matrix in (66) is a rotation; two definitions of ρ can only differ by a phase. In fact, the angle of the rotation in (66) is $\tan^{-1}(\alpha)$, in perfect agreement with (65).

We close the present discussion with a remark on Dragt's original definition of the phase advance. In Ref. 11, Dragt defined the canonical transformation ρ in terms of maps. However, Dragt artificially introduced the timelike variable s of the original Hamiltonian in the definition of ρ. As a result, it was not true that

$$\mathbf{M}_i = \mathbf{M}_j \Rightarrow \rho(\mathbf{M}_i) = \rho(\mathbf{M}_j). \tag{67}$$

Although it is conceivable to imagine cases where (67) should be discarded on the basis of connecting two different types of perturbations, it is unacceptable to do so at random using the Hamiltonian parameter s. In that sense, Dragt's treatment was not totally Hamiltonian-free.

VII. CONCLUSION

We would like to summarize the actual achievements of the Hamiltonian-free theory. First, as we emphasized throughout this paper, our apporach goes directly to the quantities of interest; this greatly simplifies the theory for any representation of the map.

Second, our ability to generate and analyze Taylor series maps allows us to study arbitrarily complex systems, in particular, circular accelerators. One can define a Floquet ring and perturb it by Hamiltonian and/or stochastic effects. For example, one can easily implement the stochastic calculation of the final emittances proposed by Chao in *any* tracking code.[6] We are no longer restricted to simple models: This could become important in understanding the behavior of small light sources because of the nontrivial fringe fields they generate.

Finally, other areas of physics could benefit from such an approach. For example, in the design of toroidal stellarators, one can show that the magnetic field line pattern is (in some variables) a two-dimensional symplectic map. The computation of this map is extremely complex and tedious since one must integrate the Biot–Savart law around the stellarator. Hanson and Cary, in a paper on the stochastic nature of this map,[26] did exactly that: Had they known of the automatic differentiation of Berz,[13] they could have attempted to compute a one-turn map with some dependence on the current parameters they used to reduce the stochasticity. In fact, the stellarator problem seems to typify a proper use of a map-based theory: the map is simple (two-dimensional), but the Hamiltonian generating it is extremely complex (i.e., Maxwell's equations). In addition, the field lines are best integrated using non-Hamiltonian variables. One can convert the two-dimensional map into canonical variables at the end of the calculation, just before feeding it into some canonical perturbation theory algorithm.

By this example, we just wanted to point out the generality of certain concepts. Since not all problems are identical, we are convinced that the greater the selections of tools, the more efficiently a researcher or designer can attack a complex problem.

VIII. PROSPECT FOR THE FUTURE

We began this paper by pointing out that the essential problem of accelerator dynamics is to study the long-

term stability of the one-turn map. In fact, from a strict analysis of error propagation, accelerator simulation integrates the motion of particles far beyond a rigorously reasonable limit. Given this fact, can we then use a one-turn map in our simulation? Experience has shown that truncated Taylor series produced nonsymplectic maps with vastly different long-term behavior. Indeed, the motion can settle on a fixed point in phase space after a relatively short number of turns, despite a highly accurate Taylor series representation. However, with our ability to extract maps and manipulate them, we can reexpress the Taylor series representation into various exactly symplectic representations. This is being extensively studied at the moment, driven by projects such as third-generation synchrotron light sources, small "pocket" light sources and large hadron rings such as the contemplated Superconducting Super Collider (SSC).

We are also trying to understand quasisymplectic maps. For example, in light sources and so-called "beauty factories," electrons radiate a substantial amount of energy. In the classical regime, this leads to a nonlinear map with damping. This map can be easily extracted with automatic differentiation techniques, but its analysis in the nonlinear regime will require new developments beyond those advertised in this paper. In particular, it will not be possible to express the one-turn maps using symplectic Lie generators, but it is hoped that a new expanded set of Lie generators can be found.

These problems and others are now within reach thanks to the type of rethinking introduced in this paper.

ACKNOWLEDGMENTS

I would like to thank Professor Alex Dragt, who has been my mentor and has initiated this vastly expanded usage of maps in accelerator theory. More recently, I had the pleasure to collaborate with John Irwin of SLAC. In addition, I am extremely grateful to Martin Berz for developing the Differential Algebra Package and attaching to it extra subroutines to ease the development of map normalization algorithms. Berz's work on automatic differentiation reinforced the seriousness and generality of the map approach to accelerator design: The usefulness of his methods goes far beyond the domain of dynamical systems. Finally, I am also grateful to my co-workers and, in particular, to my supervisor Dr. Swapan Chattopadhyay for providing the necessary support.

This work was supported by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

APPENDIX A: A MAP DERIVATION OF THE NEW HAMILTONIAN USING THE ADJOINT REPRESENTATION OF THE LIE OPERATOR ALGEBRA

Assuming that we are interested in all surfaces of sections, the ring becomes an "∞-tuple," as described in Eqs. (2), (8), and (9). In terms of maps, (2), (8), and (9) take the form

$$\frac{d}{d\epsilon}\mathbf{A}^{-1}(\mathbf{z}_0;s;\epsilon)=\mathbf{A}^{-1}(\mathbf{z}_0;s;\epsilon):-w(\mathbf{z}_0;s;\epsilon):, \tag{A1a}$$

$$\frac{d}{ds}\mathbf{M}(\mathbf{z}_0;s)=\mathbf{M}(\mathbf{z}_0;s):-K(\mathbf{z}_0;s):, \tag{A1b}$$

$$K(\mathbf{z}_0;s)=\mathbf{A}(\mathbf{z}_0;s)\Big(H(\mathbf{z}_0;s)+\int_0^{\epsilon=1}d\epsilon\,\mathbf{A}^{-1}(\mathbf{z}_0;s)\frac{\partial}{\partial s}w(z_0;s;\epsilon)\Big), \tag{A1c}$$

$$\mathbf{M}(\mathbf{z}_0;s=s_0)=\mathbf{E}=\text{identity}. \tag{A1d}$$

Equations (A1) must be integrated from s_0 to s_0+1 if one is to obtain a one-turn map. For a complete normalization, (A1) must be accompanied by the boundary conditions

$$K(\mathbf{z}_0;s)=\frac{\partial}{\partial s}\Phi_s(\mathbf{J}), \tag{A2a}$$

$$\mathbf{A}^{-1}(\mathbf{z}_0;s;\epsilon=1)=\mathbf{A}^{-1}(\mathbf{z}_0;s+1;\epsilon=1). \tag{A2b}$$

The periodicity of (A2b) and the normalized form imposed on K in (A2a) renders the direct solution of Eqs. (A1) and (A2) very difficult unless the original Hamiltonian H is simple. The process involves the computation of various Green's functions upon which the proper boundary conditions are imposed (see Ref. 5).

In this Appendix, we will concentrate on Eq. (A1c) (first obtained by Dewar[18] and then Cary[19]): It can be derived very elegantly using homomorphic Lie algebras. Let us denote by $\mathbf{N}$ the map generated by H:

$$\frac{d}{ds}\mathbf{N}(\mathbf{z}_0;s)=\mathbf{N}(\mathbf{z}_0;s):-H(\mathbf{z}_0;s):. \tag{A3}$$

Using Lie properties of maps, we can write $\mathbf{M}(\mathbf{z}_0;s)$ as

$$\mathbf{M}_s=\mathbf{A}_0\mathbf{N}_s\mathbf{A}_s^{-1}. \tag{A4}$$

In Eq. (A4) all the maps transform functions of the initial coordinates $\mathbf{z}_0$. First, $\mathbf{A}_0(\mathbf{z}_0)$ brings us to the original variables; these are propagated with the help of $\mathbf{N}_s(\mathbf{z}_0,s_0)$ and are finally taken back into the new variables by $\mathbf{A}_s(\mathbf{z}_0)^{-1}$. To obtain the new Hamiltonian K, we take the derivative of (A4) with respect to s:

$$\begin{aligned}\frac{d}{ds}\mathbf{M}_s&=\mathbf{A}_0\Big(\frac{d}{ds}\mathbf{N}_s\Big)\mathbf{A}_s^{-1}+\mathbf{A}_0\mathbf{N}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}\Big)\\&=\mathbf{A}_0\mathbf{N}_s\mathbf{A}_s^{-1}\mathbf{A}_s\\&\quad\times:-H:\mathbf{A}_s^{-1}+\mathbf{A}_0\mathbf{N}_s\mathbf{A}_s^{-1}\mathbf{A}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}\Big)\\&=\mathbf{M}_s\Big(:-\mathbf{A}_sH:+\mathbf{A}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}\Big)\Big)\ .\end{aligned} \tag{A5}$$

Comparing (A5) with (A1b), we conclude that $\mathbf{A}_s((d/ds)\mathbf{A}_s^{-1})$ must be a Lie operator: To evaluate it, we use the equation of $\mathbf{A}_s$ and $\mathbf{A}_s^{-1}$:

$$\begin{aligned}&\frac{d}{d\epsilon}\mathbf{A}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}\Big)\\&=:w:\mathbf{A}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}\Big)+\mathbf{A}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}:-w:\Big)\\&=\Big\{:w:,\mathbf{A}_s\Big(\frac{d}{ds}\mathbf{A}_s^{-1}\Big)\Big\}+:-\frac{\partial}{\partial s}w:,\end{aligned}$$

$$\frac{\partial}{\partial s} w = \frac{d}{ds} w. \tag{A6}$$

Here $\{\ ,\ \}$ denotes the commutator of two Lie operators. Denoting by $\mathbf{G}$ the operator $\mathbf{A}_s((d/ds)\mathbf{A}_s^{-1})$, we rewrite (A6) as

$$\frac{d}{d\epsilon}\mathbf{G} - \#w\#\mathbf{G} = :-\frac{\partial}{\partial s} w:. \tag{A7}$$

Here $\#w\#$ is a super operator that acts on the space of Lie transforms by taking a commutator. These commutators form a Lie algebra. Notice the homomorphism between the Lie algebra of super operators, Lie operators, and Poisson brackets[15]:

$$\#f\#:g: = \{:f:,:g:\} = :[f,g]:. \tag{A8}$$

To solve (A7), we make use of (A8) by writing $\mathbf{G}$ in terms of an ϵ-dependent super operator:

$$\mathbf{G}(\epsilon) = \mathbf{P}\mathbf{G}_0. \tag{A9}$$

We first solve the homogeneous equation

$$\frac{d}{d\epsilon}\mathbf{P} - \#w\#\mathbf{P} = 0 \Rightarrow \mathbf{P} = \mathbf{A}(\#w(\epsilon)\#). \tag{A10}$$

To obtain $\mathbf{P}$ in (A10), we notice that the formal functional dependence of $\mathbf{P}$ on the super operator $\#w\#$ must be the same as the dependence of $\mathbf{A}$ on Lie operator $:w:$; similarly, $\mathbf{P}^{-1}$ must have the same functional dependence on $\# - w\#$ as $\mathbf{A}^{-1}$ has on $: - w:$.

To solve the nonhomogeneous equation, we allow $\mathbf{G}_0$ to depend on ϵ (variation of parameters). For the particular solution $\mathbf{G}_p$ we obtain

$$\mathbf{G}_p = \mathbf{A}(\#w(\epsilon)\#)\int_0^\epsilon d\epsilon'\, \mathbf{A}^{-1}(\# - w(\epsilon')\#) :-\frac{\partial}{\partial s} w(\epsilon'):. \tag{A11}$$

The general solution is the sum of the homogeneous and particular solutions:

$$\mathbf{G} = \mathbf{A}(\#w(\epsilon)\#)\left(\mathbf{G}_0 + \int_0^\epsilon d\epsilon'\, \mathbf{A}^{-1}(\# - w(\epsilon')\#) :-\frac{\partial}{\partial s} w(\epsilon'):\right). \tag{A12}$$

We impose the boundary condition at $\epsilon = 0$:

$$(\mathbf{G}(\epsilon = 0) = \mathbf{0} \Rightarrow \mathbf{G}_0 = 0);$$

therefore,

$$\mathbf{G} = \mathbf{A}(\#w(\epsilon)\#)\int_0^\epsilon d\epsilon'\, \mathbf{A}^{-1}(\# - w(\epsilon')\#): -\frac{\partial}{\partial s} w(\epsilon'):. \tag{A13}$$

Finally, here the homomorphism enters between the three Lie algebras of (A8):

$$\mathbf{A}(\#w(\epsilon)\#)\int_0^\epsilon d\epsilon'\, \mathbf{A}^{-1}(\# - w(\epsilon')\#): -\frac{\partial}{\partial s} w(\epsilon'): = :-\mathbf{A}(:w(\epsilon):)\int_0^\epsilon d\epsilon'\, \mathbf{A}^{-1}(: - w(\epsilon'):)\frac{\partial}{\partial s} w(\epsilon'):. \tag{A14}$$

Substitution of (A14) into (A5) gives the advertised result.

APPENDIX B: THE UNIQUENESS OF R AND THE PHASE ADVANCE

We first prove that given a map $\mathbf{M}$, the assumed analyticity of the similarity transformation insures the uniqueness of $\mathbf{R}$.

We start by postulating the existence of two normalized rotations:

$$\mathbf{A}_1\mathbf{M}\mathbf{A}_1^{-1} = \mathbf{R}_1, \quad \mathbf{A}_2\mathbf{M}\mathbf{A}_2^{-1} = \mathbf{R}_2, \tag{B1}$$

which in turn imply that

$$\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{R}_1\mathbf{A}_1\mathbf{A}_2^{-1} = \mathbf{R}_2. \tag{B2}$$

Equation (B2) is a generalization of Eq. (33) for the phase advance:

$$\exp(:\mathbf{A}_2\mathbf{A}_1^{-1}(-\boldsymbol{\mu}_1\cdot\mathbf{J} + D_1(\mathbf{J})):) = \exp(:-\boldsymbol{\mu}_2\cdot\mathbf{J} + D_2(\mathbf{J}):). \tag{B3}$$

Using analyticity, we follow Dragt and Finn[16] by factorizing $\mathbf{A}_2\mathbf{A}_1^{-1}$:

$$\mathbf{A}_2\mathbf{A}_1^{-1} = \cdots\mathbf{r}_k\mathbf{r}_{k-1}\cdots\mathbf{r}_1\mathbf{r}_L, \tag{B4}$$

where $\cdots\mathbf{r}_k = \exp(:p_{k+2}:)$ and p_{k+2} is a homogeneous polynomial of degree $k+2$ in the phase space variables.

Except for a mere relabeling of the planes, let us assume that the uniqueness of $\mathbf{R}$ is true in the linear regime (the proof would be quite different for linear maps and amounts to the uniqueness of eigenvalues!). Then (B3) takes the form

$$\exp(:\cdots\mathbf{r}_k\mathbf{r}_{k-1}\cdots\mathbf{r}_1(-\boldsymbol{\mu}\cdot\mathbf{J} + D_1(\mathbf{J})):) = \exp(:-\boldsymbol{\mu}\cdot\mathbf{J} + D_2(\mathbf{J}):). \tag{B5}$$

To go further, we proceed by induction. Assuming that for $k < j-1$, the $\mathbf{r}_k$'s are rotations, we collect the terms of degree $j+2$ and obtain an equation for p_{j+2}:

$$:\boldsymbol{\mu}\cdot\mathbf{J}:p_{j+2} = D_2(\mathbf{J}) - D_1(\mathbf{J})|_{j+2\text{ component}}. \tag{B6}$$

Using the direct sum decomposition (or the eigenbasis) and the mutual irrationality of the tunes [Eqs. (22), (23), and (25)], we conclude from (B6) that p_{j+2} cannot contain anything from $\mathrm{Im}(:\boldsymbol{\mu}\cdot\mathbf{J}:)$. In addition, since the lhs of (B6) must be in $\mathrm{Ker}(:\boldsymbol{\mu}\cdot\mathbf{J}:)$, the only consistent solution to (B6) is

$$p_{j+2} \in \mathrm{Ker}(:\boldsymbol{\mu}\cdot\mathbf{J}:), \quad D_2(\mathbf{J}) - D_1(\mathbf{J})|_{j+2\text{ component}} = 0. \tag{B7}$$

Q.E.D.

The rest follows by induction, starting with $j = 1$.

Equation (B7) also proves the statement on the phase advance because it is a special case of (B1).

Finally, by generalizing Eq. (14) to the nonlinear maps, we can easily see that $\mathbf{R}$ does not depend on the location:

$$\mathbf{M}_j = \mathbf{N}_{ij}^{-1}\mathbf{M}_i\mathbf{N}_{ij}, \text{ where } \mathbf{N}_{ij} = \prod_{k=i}^{} \mathbf{N}_{k\,k+1}, \tag{B8a}$$

$$\mathbf{A}_i\mathbf{M}_i\mathbf{A}_i^{-1} = \mathbf{R}_i, \quad \mathbf{A}_j\mathbf{M}_j\mathbf{A}_j^{-1} = \mathbf{R}_j, \tag{B8b}$$

$$(\text{B8a}) \text{ and } (\text{B8b}) \Rightarrow \mathbf{A}_j\mathbf{N}_{ij}^{-1}\mathbf{M}_i\mathbf{N}_{ij}\mathbf{A}_j^{-1} = \mathbf{R}_j. \tag{B8c}$$

Equations (B8a) and (B8b) violate the uniqueness of $\mathbf{R}$ for a given map $\mathbf{M}$.

[1] E. D. Courant and H. S. Snyder, Ann. Phys. (NY) 3, 1 (1958).
[2] H. Meier and K. R. Symon, in *Proceedings of the International Conference*

on High Energy Accelerators and Instrumentation, Cern, edited by L. Kowarski (CERN, Geneva, 1959). See, also, L. J. Laslett, E. M. MacMillan, and J. Moser, *Long Term Stability for Particle Orbits* (Courant Institute of Mathematical Science, New York Univ. NYO-1480-101, 1968): In this reference, one finds some comments by Moser on the Birkhoff normalization of the Hamiltonian.

[3]The list of kick codes is endless. For example, R. Talman and L. Schachinger, Part. Accel. **22**, 35 (1987) or A. Wrulich, DESY Internal Rep. 84-026 (1984).

[4]A. Schoch, CERN 57-21 (1958).

[5]L. Michelotti, Part. Accel. **16**, 233 (1985).

[6]A. Chao, J. Appl. Phys. **50**, 595 (1979).

[7]In single-pass systems such as spectrometers, the extraction of Taylor series maps is a time honored practice. Indeed, the aberrations of interest are directly the coefficients of the Taylor series. Again, the list of power series codes is endless.

[8]A. J. Dragt and J. M. Finn, J. Math. Phys. **20**, 2649 (1979).

[9]A. J. Dragt and J. M. Finn, J. Geophys. Res. **81**, 13 (1976).

[10]A. J. Dragt *et al.*, Ann. Rev. Nucl. Sci. **38**, 455 (1988).

[11]A. J. Dragt, in *Proceedings of the 1984 Summer Study on the Design and Utilization of the Superconducting Super Collider, Snowmass*, edited by R. Donaldson and J. Morfin (Am. Phys. Soc., New York, 1985).

[12]E. Forest, Part. Accel. **22**, 15 (1987).

[13]M. Berz, Part. Accel. **24**, 109 (1989).

[14]E. Forest, J. Irwin, and M. Berz, Part. Accel. **24**, 91 (1989).

[15]A. J. Dragt, *Lectures on Nonlinear Orbit Dynamics*, American Institute of Physics Conference Proceedings No. 87, edited by R. A. Carrigan, F. R. Huson, and M. Month (AIP, New York, 1982).

[16]A. J. Dragt and J. M. Finn, J. Math. Phys. **17**, 2215 (1976).

[17]A. J. Dragt and E. Forest, J. Math. Phys. **24**, 2734 (1983).

[18]R. L. Dewar, J. Phys. A **9**, 2043 (1976).

[19]J. Cary, Phys. Rep. **79**, 129 (1981).

[20]R. D. Ruth, IEEE Trans. Nucl. Sci. **NS-30**, 2669 (1983). In addition, Ruth and the present author are preparing a paper on symplectic integration and its connection with the Lie representation of maps near the identity.

[21]Because we use symplectic integrators, the maps we extract are Hamiltonian in nature to computer accuracy. The theory presented here still applies for nonsymplectic integrators. Obviously, the violation of "symplecticity" is of the same order as the error due to the finite size of the integration steps.

[22]R. L. Warnock, R. D. Ruth, W. Gabella, and K. Ecklund, Internal Rep. SLAC-PUB -4846 (1989).

[23]The normalization of a one-turn Taylor series map using the Differential Algebra Package of Berz is explained in detail in Ref. 14. Here we summarize the results, deemphasizing the Taylor series representation. One must realize that even in other representations, normalization procedures are formal manipulations. Most problems are not integrable and, therefore, the expressions for the invariants are asymptotic series in some smallness parameter.

[24]R. Cushman, A. Deprit, and R. Mosak, J. Math. Phys. **24**, 2102 (1983).

[25]D. A. Edwards and L. C. Teng, IEEE Trans. Nucl. Sci. **NS-20**, 885 (1983).

[26]J. D. Hanson and J. R. Cary, Phys. Fluids **27**, 767 (1984).

Lorentz action on the space of relative velocities and relativity on a three-manifold

D. K. Sen
Department of Mathematics, University of Toronto, Toronto, Ontario M5S 1A1, Canada

(Received 24 January 1989; accepted for publication 10 January 1990)

The space of relative velocities in special relativity has a three-dimensional hyperbolic structure. This provides not only a geometric interpretation of the Einstein velocity addition law, but also a purely three-dimensional reformulation of both special and general relativity on a three-manifold whose tangent bundle is endowed with a hyperbolic distance function on each fiber. Here the basic concepts are that of local physical observers and local time in terms of nonsingular vector fields and their local flows. The hyperbolic structure on the tangent space enables one to define a relative velocity function between two physical observers, as well as space and time measurements and inertial physical observers. It is possible to rederive Lorentz time dilation and gravitational and cosmological redshifts and reformulate Maxwell (or Yang–Mills) equations in a purely three-dimensional framework instead of the traditional four-dimensional space-time approach.

I. INTRODUCTION

The Lorentz group acts as an isometry group on the flat Minkowski space-time with signature $(-+++)$; it also acts as an isometry group on the space of relative velocities which is a three-dimensional hyperbolic [but Riemannian with signature $(+++)$] space of constant curvature equal to -1. This provides the motivation for a purely three-dimensional reformulation of both special and general relativity on a three-manifold whose tangent bundle is endowed with a hyperbolic distance function on each fiber.

In this paper we propose the framework of an underlying three-dimensional manifold providing a background for physical observers relative to which it is possible to define space and time measurements in conformity with the principle of relativity. The motivation for this comes from the hyperbolic structure of the space of relative velocities in special relativity. Relative to some reference inertial system, the velocities $\{u = (u^1,u^2,u^3)\}$ of all other equivalent inertial systems, with velocities less than the velocity of light, form a three-dimensional open disk D^3. Let G be the standard hyperbolic metric on D^3. Then $H^3 = (D^3,g)$ is the standard hyperbolic three-space of constant curvature equal to -1. H^3 is geodesically complete and has a hyperbolic distance function $d(u,v)$, where u,v are the velocities of two inertial systems relative to the reference inertial system. Then it can be shown that the relative velocity w between the two systems, as given by special relativity, can be expressed as $\|w\| = \tanh\{d(u,v)\}$. The isometry group of H^3 is known to be $\mathrm{PSL}(2,\mathbb{C})$, the projective special complex linear group in two dimensions.

The action of the Lorentz group $L^{\uparrow}_{+}$ on the space of relative velocities H^3 is next considered in some detail and the infinitesimal operators, that is, the Killing vector fields of $L^{\uparrow}_{+}$ as an isometry group of H^3, are worked out explicitly.

The transition from special to general relativity is accomplished, in general, by the passage to an arbitrary manifold whose tangent space has the structure of special relativity. We are thus led to consider a Riemannian three-manifold (M,g) whose metric g determines, in an obvious manner, a hyperbolic structure on the tangent space $T_m(M)$ at each point $m \in M$. A physical observer is then defined to be a nonsingular vector field X on M with $g(X,X) < 1$ (which takes into account the limiting velocity of light). With the help of the hyperbolic structure on $T_m(M)$ one can then define a relative velocity function $V(X,Y)$ between any two physical observers X,Y. Two observers X,Y are then inertially equivalent if $V(X,Y) = \text{const}$ on M.

We next consider the problem of constructing a class of inertial observers starting from a representative physical observer. In the general case where $(M,g) \neq (\mathbb{R}^3,\delta)$, the problem leads to the definition of a "generalized Lorentz matrix" with appropriate "generalized" properties, which reduce to the usual properties of Lorentz matrices when $g_{ik} = \delta_{ik}$.

If $c: I \to M$ is a particle path, i.e., a smooth curve on M, and X is a physical observer, we can define space and time intervals of $c(I)$ relative to X in such a way that if Y is another physical observer inertially equivalent to X, then the space and time intervals relative to Y are related to that of X by a Lorentz transformation.

The Maxwell equations for electric and magnetic fields as observed by an arbitrary physical observer X can be cast in a particularly concise and elegant form in terms of three-dimensional differential forms on M and are susceptible to obvious generalization to the non-Abelian Yang–Mills case by considering the $\mathrm{SU}(2)$ Lie-algebra-valued three-dimensional differential forms on M.

Finally, we indicate briefly how dynamical problems such as gravitational and cosmological redshifts can be handled in this three-dimensional framework.

It is hoped that the rich topological and geometrical structure of three-manifolds and their vector fields may provide a new tool or insight to problems in general relativity.

II. ACTION OF $L^{\uparrow}_{+}$ ON THE SPACE OF RELATIVE VELOCITIES

In the special theory of relativity, the velocity transformation formula relating the relative velocities $\mathbf{u},\mathbf{v},\mathbf{w}$ of three

 0022-2488/90/051145-07$03.00

equivalent inertial systems is given by (in units $c = 1$)

$$\|\mathbf{w}\|^2 = [(\mathbf{u}-\mathbf{v})^2 - (\mathbf{u}\times\mathbf{v})^2]/(1-\mathbf{u}\cdot\mathbf{v})^2 . \tag{1}$$

Equation (1) has the following geometrical significance. Let us choose some *reference* inertial system and consider the velocities $u = (u^1,u^2,u^3)$ of all other equivalent inertial systems relative to this reference inertial system. Since for physical inertial systems $\|u\| = [\Sigma_i (u^i)^2]^{1/2} < 1$, the space of relative velocities is topologically a three-dimensional open disk

$$D^3 = \{u \in \mathbb{R}^3 \mid \|u\| < 1\} .$$

(For tachyons the appropriate space to consider would be the complement of the closure of D^3, that is $\mathbb{R}^3 \setminus \bar{D}^3$.) Let us now introduce the standard (positive definite) hyperbolic metric G on D^3 by

$$ds^2 = \sum_{i,k} G_{ik}(u)\,du^i\,du^k = \frac{[1-\Sigma_i (u^i)^2][\Sigma_i (du^i)^2] + [\Sigma_i u^i\,du^i]^2}{[1-\Sigma_i (u^i)^2]^2} , \tag{2}$$

i.e.,

$$G_{ik}(u) = \frac{[1-\Sigma_i (u^i)^2]\delta_{ik} + u^i u^k}{[1-\Sigma_i (u^i)^2]^2} .$$

Then $H^3 = (D^3,G)$ is the standard hyperbolic three-space of constant curvature -1. This is the so-called Poincaré disk model of H^3 (Ref. 1). Alternatively, introduce homogeneous coordinates $\xi = (\xi_0,\xi_1,\xi_2,\xi_3)$ by $\xi_i = \xi_0 u^i$ and define for $\xi = (\xi_0,\xi_1,\xi_2,\xi_3)$, $\eta = (\eta_0,\eta_1,\eta_2,\eta_3)$ the inner product $\langle \xi,\eta \rangle = \xi_0\eta_0 - \Sigma_i \xi_i \eta_i$. Then

$$ds^2 = \frac{\langle d\xi,\xi\rangle\langle \xi,d\xi\rangle - \langle \xi,\xi\rangle\langle d\xi,d\xi\rangle}{\langle \xi,\xi\rangle^2} . \tag{3}$$

H^3 is geodesically complete and has a distance function d given by[2]

$$d(u,v) = \cosh^{-1}\left\{\frac{[1-\Sigma_i u^i v^i]}{[1-\Sigma_i (u^i)^2]^{1/2}[1-\Sigma_i (v^i)^2]^{1/2}}\right\} \tag{4}$$

or, in homogeneous coordinates,

$$d(\xi,\eta) = \cosh^{-1}\left\{\frac{|\langle \xi,\eta\rangle|}{\langle \xi,\xi\rangle^{1/2}\langle \eta,\eta\rangle^{1/2}}\right\} . \tag{5}$$

We expect that the distance $d(u,v)$ has something to do with the relative velocity w between the two inertial systems whose velocities relative to the reference inertial system are u and v. To see the actual relationship[3] note that the reference system is at rest relative to itself. Now from (4), $d(0,v) = \cosh^{-1}\{1/(1-\|v\|^2)^{1/2}\} = \tanh^{-1}\|v\|$. That is, $\|v\| = \tanh\{d(0,v)\}$. We can now reinterpret Eq. (1) as a relation between the relative velocity w and the hyperbolic distance $d(u,v)$:

$$\|w\| = \tanh\{d(u,v)\} . \tag{6}$$

The isometry group of H^3 is intimately connected with the Lorentz group, as can be seen from the following.

First consider an *abstract* $\mathbb{R}^4$, with the *Lorentz* metric $g_{\mathbb{R}^4}$ given by $ds^2 = (dx^2)^2 + (dx^2)^2 + (dx^3)^2 - (dx^4)^2$. Consider now the hypersurface M of $\mathbb{R}^4$ (see Fig. 1),

$$M = \{x = (x^1,x^2,x^3,x^4) \in \mathbb{R}^4 \mid (x^1)^2 + (x^2)^2 + (x^3)^2 - (x^4)^2 = -1\} ,$$

which has the components $M' = \{x \in M \mid x^4 \geqslant 1\}$ and $M'' = \{x \in M \mid x^4 \leqslant -1\}$, each of which is *diffeomorphic* to D^3. The diffeomorphism map is given by

$$(x^1,x^2,x^3,x^4) \mapsto (u^1,u^2,u^3) ,$$
$$u^i = x^i/x^4 \quad (i = 1,2,3) ,$$

with its inverse

$$(u^1,u^2,u^3) \mapsto (x^1,x^2,x^3,x^4) ,$$
$$x^i = u^i x^4, \quad x^4 = \pm\left(\frac{1}{1-\Sigma_i (u^i)^2}\right)^{1/2} .$$

Now let g_M be the induced *Riemannian* metric on M obtained from $g_{\mathbb{R}^4}$ on $\mathbb{R}^4$. In coordinates (u^i), g_M is given exactly by (2). Thus $H^3 = (D^3,G)$ is *isometric* to $(M',g_{M|M'})$ and $(M'',g_{M|M''})$.

The isometry group of H^3 was shown by Poincaré[4] to be PSL(2,C), the projective special linear group in two complex dimensions; it is thus isomorphic to $L^{\uparrow}_{+}$, the orthochonons Lorentz group. It is clear from (5) that the isometry group of H^3 leaves the inner product $\langle \xi,\eta\rangle = \xi_0\eta_0 - \Sigma_i \xi^i \eta^i$, in homogeneous coordinates, invariant: Thus let its action be given by (from now on we shall use the summation convention for repeated indices)

$$(\xi_0\xi_1\xi_2\xi_3) \mapsto (\bar{\xi}_0\bar{\xi}_1\bar{\xi}_2\bar{\xi}_3) ,$$

where

$$\bar{\xi}_\mu = L^\mu_\alpha \xi_\alpha, \quad L^\mu_\alpha \in L^{\uparrow}_{+} \quad (\mu,\alpha = 0,1,2,3) .$$

Since $\bar{u}^i = \bar{\xi}_i/\bar{\xi}_0$, $L^{\uparrow}_{+}$ acts on H^3 as a fractional linear transformation

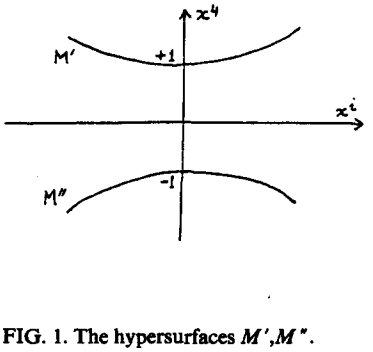


FIG. 1. The hypersurfaces M',M''.

$$u^i \mapsto \bar{u}^i = (L^i_0 + L^i_k u^k)/(L^0_0 + L^0_k u^k)\,. \tag{7}$$

Let us consider next the infinitesimal *operators* of this action. We first write L^μ_α in terms of its infinitesimal *generators* K_A and the parameters a_A $(A = 1,...,6)$ as $L^\mu_\alpha = L^\mu_\alpha(a) = (\exp a_A K_A)^\mu_\alpha$ and write (7) as $u^i \mapsto \bar{u}^i = f^i(u,a)$. Then the infinitesimal operators are the vector fields X_A on H^3 given by

$$X_A = \left(\frac{\partial f^i}{\partial a_A}\right)_{a=0} \frac{\partial}{\partial u^i}\,. \tag{8}$$

Now

$$\left(\frac{\partial f^i}{\partial a_A}\right)_{a=0} = \left[\frac{(\partial L^i_0/\partial a_A + (\partial L^i_j/\partial a_A)u^j)(L^0_0 + L^0_j u^j) - (L^i_0 + L^i_j u^j)(\partial L^0_0/\partial a_A + (\partial L^0_j/\partial a_A)u^j)}{(L^0_0 + L^0_j u^j)^2}\right]_{a=0}\,.$$

Thus

$$X_A = \{[(K_A)^i_0 + (K_A)^i_j u^j] - u^i[(K_A)^0_0 + (K_A)^0_j u^j]\}\frac{\partial}{\partial u^i}\,. \tag{9}$$

For example, in the case of the one-dimensional special Lorentz transformation with boost a, $f(u,a) = (u-a)/(1-ua)$, so that the infinitesimal operator (or vector field) on H^1 is $X = (u^2-1)\partial/\partial u$.

In general, let $\{K_1,K_2,K_3\}$ be the generators of rotations and $\{K_4,K_5,K_6\}$ that of boosts, i.e.,

$$K_1 = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&-1\\ 0&0&1&0 \end{pmatrix},\quad K_2 = \begin{pmatrix} 0&0&0&0\\ 0&0&0&1\\ 0&0&0&0\\ 0&-1&0&0 \end{pmatrix},$$

$$K_3 = \begin{pmatrix} 0&0&0&0\\ 0&0&-1&0\\ 0&1&0&0\\ 0&0&0&0 \end{pmatrix},\quad K_4 = \begin{pmatrix} 0&-1&0&0\\ -1&0&0&0\\ 0&0&0&0\\ 0&0&0&0 \end{pmatrix},$$

$$K_5 = \begin{pmatrix} 0&0&-1&0\\ 0&0&0&0\\ -1&0&0&0\\ 0&0&0&0 \end{pmatrix},$$

$$K_6 = \begin{pmatrix} 0&0&0&-1\\ 0&0&0&0\\ 0&0&0&0\\ -1&0&0&0 \end{pmatrix}.$$

Then from (9) we obtain

$$\begin{aligned} X_1 &= u^2\frac{\partial}{\partial u^3} - u^3\frac{\partial}{\partial u^2}, \quad X_2 = u^3\frac{\partial}{\partial u^1} - u^1\frac{\partial}{\partial u^3},\\ X_3 &= u^1\frac{\partial}{\partial u^2} - u^2\frac{\partial}{\partial u^1},\\ X_4 &= ((u^1)^2-1)\frac{\partial}{\partial u^1} + u^1u^2\frac{\partial}{\partial u^2} + u^1u^3\frac{\partial}{\partial u^3},\\ X_5 &= u^2u^1\frac{\partial}{\partial u^1} + ((u^2)^2-1)\frac{\partial}{\partial u^2} + u^2u^3\frac{\partial}{\partial u^3},\\ X_6 &= u^3u^1\frac{\partial}{\partial u^1} + u^3u^2\frac{\partial}{\partial u^2} + ((u^3)^2-1)\frac{\partial}{\partial u^3}, \end{aligned} \tag{10}$$

whose Lie brackets satisfy

$$\begin{aligned} &[X_1,X_2] = -X_3,\ [X_2,X_3] = -X_1,\ [X_3,X_1] = -X_2,\\ &[X_4,X_5] = X_3,\ [X_5,X_6] = X_1,\ [X_6,X_4] = X_2,\\ &[X_1,X_4] = 0,\ [X_2,X_5] = 0,\ [X_3,X_6] = 0,\\ &[X_1,X_5] = -X_6,\ [X_2,X_6] = -X_4,\ [X_3,X_4] = -X_5,\\ &[X_1,X_6] = X_5,\ [X_2,X_4] = X_6,\ [X_3,X_5] = X_4\,. \end{aligned} \tag{11}$$

Note that if $[K_A,K_B] = C^C_{AB}K_C$, then $[X_A,X_B] = -C^C_{AB}X_C$.

III. RELATIVITY ON A THREE-MANIFOLD

When one wants to make a transition from special to general relativity one considers a manifold whose tangent space has the structure of special relativity. This suggests the following general formalism of relativity on a three-manifold.

Let (M,g) be a three-manifold with a Riemannian metric g. Here M should be regarded as some sort of a universal *substratum* and not the absolute space of Newton. By an observer or particle path in M we shall mean a smooth curve, i.e., a smooth map c: $I \to M$ from a real interval I into M. Since to every such curve c one can associate *locally* a vector field c^* (Ref. 5) such that c is an integral curve of c^*, it is preferable to regard the set of all *nonsingular* vector fields on M as the set of all *observers* in M. We must demand nonsingularity in order to avoid an observer being at absolute rest relative to the substratum M. Therefore, each nonsingular vector field X on M represents an observer. To each such X corresponds a local flow χ_τ, i.e., a local one-parameter group of local diffeomorphisms of M. Here τ represents the "flow of time" as perceived by X. Time is thus a local concept since X is not necessarily complete unless M is compact.

Let us now consider only those nonsingular vector fields X such that $g(X,X) < 1$. Such observers will be called *physical observers* (corresponding to the fact that no such observer can attain the velocity of light) and let $O(M)$ be the set of all physical observers.

We now define a map V, called the *relative velocity map*, from a pair of physical observers to a smooth function on M as follows:

$$V: O(M)\times O(M)\mapsto C^\infty(M)$$
$$(X,Y)\mapsto V(X,Y)=\tanh\{\cosh^{-1}F(X,Y)\},$$

where

$$F(X,Y)=\frac{1-g(X,Y)}{[1-g(X,X)]^{1/2}[1-g(Y,Y)]^{1/2}}. \quad (12)$$

Note that if (M,g) is Euclidean, i.e., $M=\mathbb{R}^3$ with the Euclidean metric $g=\delta$, then $V(X,Y)(m)=\tanh\{d(X(m),Y(m))\}$, $m\in M$, where d is the hyperbolic distance function given by (4). Thus V defines a "hyperbolic" distance function on the tangent space $T_m(M)$.

We now introduce an equivalence relation $\sim$ in $O(M)$ by

$$X\sim Y \text{ iff } V(X,Y)=\text{constant function on } M. \quad (13)$$

In other words, the observers X and Y are said to be *inertially equivalent* if the relative velocity function $V(X,Y)$ between the two is a constant function on M.

We shall now give a theoretical definition of space and time intervals of a physical particle path relative to any physical observer X. Let $c: I\to M$, $\lambda\mapsto c(\lambda)$ be a particle path and c^* any representative corresponding vector field associated with it such that c is an integral curve of c^*. For a physical particle path, $0<g(c^*,c^*)<1$ on $c(I)$. From now on all observers and particle paths would be *physical*. The *time interval of c relative to observer X* will be denoted by $\Delta\tau_X$ and defined by an integral over I. (See Fig. 2.)

Definition:

$$\Delta\tau_X=\int_I F(X,c^*)d\lambda. \quad (14)$$

We now parametrize c by τ_X, where $d\tau_X/d\lambda=F(X,c^*)$, and define ΔD_X, the *space interval of c relative to observer X* by the following definition.

Definition:

$$\Delta D_X=\int_I V(X,c^*)d\tau_X=\int_I V(X,c^*)F(X,c^*)d\lambda. \quad (15)$$

Note that if $X=c^*$, i.e., c is an integral curve of X, then from (12), $F(X,c^*)=F(c^*,c^*)=1$ and $V(c^*,c^*)=0$, so that

$$\Delta\tau_X=\int_I d\lambda\equiv\Delta\lambda,$$
$$\Delta D_X=0. \quad (16)$$

In other words, X is at rest relative to c. *Thus an observer X is at rest relative to a particle if it traces out one of its own integral curves.*

If $(\Delta\tau_Y,\Delta D_Y)$ are time and space intervals of c relative to another observer Y, then $(\Delta\tau_X,\Delta D_X)\to(\Delta\tau_Y,\Delta D_Y)$ provides a space-time transformation from observer X to observer Y.

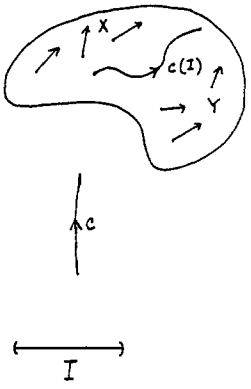


FIG. 2. The space and time interval of c relative to observers X, Y.

In general one has, for any particle path c and any pair of observers X, Y,

$$\left(\frac{d\tau_X}{d\lambda}\right)^2-\left(\frac{dD_X}{d\lambda}\right)^2=\left(\frac{d\tau_Y}{d\lambda}\right)^2-\left(\frac{dD_X}{d\lambda}\right)^2=1.$$

A. Lorentz time dilation

Consider two inertially equivalent physical observers X, Y such that

$$V(X,Y)=v=\text{const.}$$

Then

$$F(X,Y)=[1-V(X,Y)^2]^{-1/2}=(1-v^2)^{-1/2}\equiv\gamma.$$

Let c be an integral curve of X, i.e., $X=c^*$. Then, again from (12), $\Delta\tau_X=\int_I d\lambda\equiv\Delta\lambda$. However, relative to the observer Y the time interval of c is given as

$$\Delta\tau_Y=\int_I F(Y,c^*)d\lambda=\int_I F(Y,X)d\lambda=\int_I\gamma\,d\lambda=\gamma\,\Delta\lambda,$$

so that $\Delta\tau_Y=\gamma\,\Delta\tau_X$, which is the Lorentz time-dilation formula.

IV. A CLASS OF INERTIAL OBSERVERS

We shall now consider the problem of constructing a class of inertial observers starting from a representative physical observer: in other words, given a physical observer X, the problem of constructing an observer Y such that $V(X,Y)=\text{const}$.

The solution to this problem is obvious in the special case: $(M,g)=(\mathbb{R}^3,\delta)$. Let $X=X^i(x)(\partial/\partial x^i)$ in some global coordinates (x^i) in $\mathbb{R}^3$. Consider the Lorentz matrix with the pure boost $-\mathbf{X}(x)=(-X^1(x),-X^2(x),-X^3(x))$:

$$L^\mu_\lambda(-\mathbf{X})=\begin{pmatrix}\gamma_X & \gamma_X X^1 & \gamma_X X^2 & \gamma_X X^3\\ \gamma_X X^1 & 1+A_X X^1X^1 & A_X X^1X^2 & A_X X^1X^3\\ \gamma_X X^2 & A_X X^2X^1 & 1+A_X X^2X^2 & A_X X^2X^3\\ \gamma_X X^3 & A_X X^3X^1 & A_X X^3X^2 & 1+A_X X^3X^3\end{pmatrix}, \quad (17)$$

where $\gamma_{\mathbf{X}} = (1 - X^i X^i)^{-1/2}$, $A_{\mathbf{X}} = (\gamma_{\mathbf{X}} - 1)/X^i X^i$, and the vector field $Y = Y^i(x)(\partial/\partial x^i)$, where [$\mathbf{v} = (v^1, v^2, v^3)$, a constant vector]

$$Y^i(x) = \frac{L^i_0(-\mathbf{X}) + L^i_k(-\mathbf{X})v^k}{L^0_0(-\mathbf{X}) + L^0_k(-\mathbf{X})v^k}. \tag{18}$$

Then $V^2(X,Y) = v^i v^i$. This can be seen by writing (18) as $\mathbf{Y}(x) = L_{-\mathbf{X}(x)}\mathbf{v}$ and noting that $\mathbf{v} = L_{\mathbf{X}(x)} L_{-\mathbf{X}(x)}\mathbf{v}$. That is, the relative velocity between $\mathbf{X}(x)$ and $L_{-\mathbf{X}(x)}\mathbf{v}$ is $\pm\mathbf{v}$. More directly, one uses the well-known properties of Lorentz matrices

$$\begin{aligned} &L^i_0 L^i_0 = (L^0_0)^2 - 1, \\ &L^m_i L^m_j = L^0_i L^0_j + \delta_{ij}, \\ &L^m_0 L^m_i = L^0_0 L^0_i \end{aligned} \tag{19}$$

to show that

$$[1 - g(Y,Y)] = \frac{(L^0_0)^2 + (L^0_j v^j)^2 + 2L^0_0 L^0_j v^j - L^i_0 L^i_0 - 2L^i_0 L^i_m v^m - (L^i_m v^m)^2}{[L^0_0 + L^0_k v^k]^2} = \frac{1 - v^i v^i}{[L^0_0 + L^0_k v^k]^2}$$

and

$$[1 - g(X,Y)]^2 = \frac{[L^0_0 + L^0_k v^k - X^i L^i_0 - X^i L^i_k v^k]^2}{[L^0_0 + L^0_k v^k]^2} = [\gamma_{\mathbf{X}}(1 - X^i X^i)]^2/[L^0_0 + L^0_k v^k]^2,$$

so that

$$V^2(X,Y) = 1 - \frac{[1 - g(X,X)][1 - g(Y,Y)]}{[1 - g(X,Y)]^2} = v^i v^i.$$

In general, if $(M,g) \neq (\mathbb{R}^3, \delta)$, the above prescription does not work and has to be modified to take into account $g \neq \delta$.

Let X be given in some *local coordinates* by $X = X^i(x)(\partial/\partial x^i)$. Consider now a "generalized Lorentz matrix" corresponding to (17):

$$\tilde{L}^\mu_\lambda(-\mathbf{X}) = \begin{pmatrix} \tilde{\gamma}_{\mathbf{X}} & \tilde{\gamma}_{\mathbf{X}} X_1 & \tilde{\gamma}_{\mathbf{X}} X_2 & \tilde{\gamma}_{\mathbf{X}} X_3 \\ \tilde{\gamma}_{\mathbf{X}} X^1 & 1 + \tilde{A}_{\mathbf{X}} X^1 X_1 & \tilde{A}_{\mathbf{X}} X^1 X_2 & \tilde{A}_{\mathbf{X}} X^1 X_3 \\ \tilde{\gamma}_{\mathbf{X}} X^2 & \tilde{A}_{\mathbf{X}} X^2 X_1 & 1 + \tilde{A}_{\mathbf{X}} X^2 X_2 & \tilde{A}_{\mathbf{X}} X^2 X_3 \\ \tilde{\gamma}_{\mathbf{X}} X^3 & \tilde{A}_{\mathbf{X}} X^3 X_1 & \tilde{A}_{\mathbf{X}} X^3 X_2 & 1 + \tilde{A}_{\mathbf{X}} X^3 X_3 \end{pmatrix}, \tag{20}$$

where $\tilde{\gamma}_{\mathbf{X}} = (1 - X^i X_i)^{-1/2}$, $\tilde{A}_{\mathbf{X}} = (\tilde{\gamma}_{\mathbf{X}} - 1)/X^i X_i$, $X_i = g_{ik} X^k$. In other words, we use the metric g to lower (raise) indices and now distinguish between the contravariant and covariant indices. Thus

$$\begin{aligned} &\tilde{L}^0_0(-\mathbf{X}) = \tilde{\gamma}_{\mathbf{X}}, \quad \tilde{L}^k_j(-\mathbf{X}) = \delta^k_j + \tilde{A}_{\mathbf{X}} X^k X_j, \\ &\tilde{L}^0_k(-\mathbf{X}) = \tilde{\gamma}_{\mathbf{X}} X_k, \quad \tilde{L}^k_0(-\mathbf{X}) = \tilde{\gamma}_{\mathbf{X}} X^k. \end{aligned}$$

Then it is easily seen that the "generalized Lorentz matrix" (20) satisfies the following "generalized" properties:

$$\begin{aligned} &g_{ik}\tilde{L}^i_0 \tilde{L}^k_0 = (\tilde{L}^0_0)^2 - 1, \\ &g_{mn}\tilde{L}^m_i \tilde{L}^n_j = \tilde{L}^0_i \tilde{L}^0_j + \delta_{ij}, \\ &g_{mn}\tilde{L}^m_0 \tilde{L}^n_i = \tilde{L}^0_0 \tilde{L}^0_i. \end{aligned} \tag{21}$$

If one now defines $Y = Y^i(x)(\partial/\partial x^i)$ by

$$Y^i(x) = \frac{\tilde{L}^i_0(-\mathbf{X}) + \tilde{L}^i_k(-\mathbf{X})v^k}{\tilde{L}^0_0(-\mathbf{X}) + \tilde{L}^0_k(-\mathbf{X})v^k} \tag{22}$$

one can show, using the generalized properties (21), that $V^2(X,Y) = g(v,v) = g_{ik} v^i v^k$. Thus from a vector field v with $g(v,v) = \text{const}$, we can construct a class of inertial observers starting from X.

V. MAXWELL EQUATIONS

The Maxwell equations are usually expressed in terms of four-dimensional differential forms in order to exhibit their relativistic invariance, as well as their concise and elegant form; however, they can also be expressed in a concise form in the present three-dimensional formalism.

Relative to an observer X let the electromagnetic variables $\mathbf{E}, \mathbf{H}, \mathbf{B}, \mathbf{D}, \mathbf{j}, \rho$ (the usual electric, magnetic fields and inductions, current, and charges) be expressed (in some local coordinates) in terms of the following three-dimensional differential forms on our three-manifold (M,g):

one-forms $\alpha_X, \beta_X \in \Lambda^1(M)$:

$$\begin{cases} \alpha_X = \sum_i \alpha_{Xi}\, dx^i, & \mathbf{E} = (\alpha_{X1}, \alpha_{X2}, \alpha_{X3}), \\ \beta_X = \sum_i \beta_{Xi}\, dx^i, & \mathbf{H} = (\beta_{X1}, \beta_{X2}, \beta_{X3}), \end{cases}$$

two-forms $\gamma_X, \delta_X, \epsilon_X \in \Lambda^2(M)$:

$$\begin{cases} \gamma_X = \sum_{i<j} \gamma_{Xij}\, dx^i \Lambda dx^j, & \mathbf{B} = (\gamma_{X12}, \gamma_{X23}, \gamma_{X31}), \\ \delta_X = \sum_{i<j} \delta_{Xij}\, dx^i \Lambda dx^j, & \mathbf{D} = (\delta_{X12}, \delta_{X23}, \delta_{X31}), \\ \epsilon_X = \sum_{i<j} \epsilon_{Xij}\, dx^i \Lambda dx^j, & \mathbf{j} = (\epsilon_{X12}, \epsilon_{X23}, \epsilon_{X31}), \end{cases}$$

three-form $\eta_X \in \Lambda^3(M)$: $\eta_X = \rho\, dx^1 \Lambda dx^2 \Lambda dx^3$.

The index X specifies the observer X relative to which all physical variables are to be measured. In view of the way we have defined the "flow of time" as measured by X in terms of its local flow χ_τ, it is clear that the "time derivative" relative to X is determined by the "Lie derivative" $\mathcal{L}_X$ relative to the vector field X. Thus the Maxwell equations

$$\begin{aligned} &\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \\ &\text{curl } \mathbf{H} = 4\pi \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \\ &\text{div } \mathbf{D} = 4\pi\rho, \\ &\text{div } \mathbf{B} = 0 \end{aligned} \tag{23}$$

can be written as

$$d\alpha_X = -\mathcal{L}_X \gamma_X, \tag{24}$$

$$d\beta_X = 4\pi\epsilon_X + \mathcal{L}_X \delta_X, \tag{25}$$

$$d\delta_X = 4\pi\eta_X, \tag{26}$$

$$d\gamma_X = 0. \tag{27}$$

Since

$$\mathcal{L}_X = i_X \cdot d + d \cdot i_X, \tag{28}$$

where i_X is the contraction operator relative to X, we obtain, from (24) and (27),

$$d\alpha_X = (i_X \cdot d + d \cdot i_X)\gamma_X = -d(i_X \gamma_X)$$

or

$$d\Gamma_X = 0, \quad \text{where } \Gamma_X = \alpha_X + i_X \gamma_X. \tag{29}$$

Similarly, from (25) and (26),

$$\begin{aligned} d\beta_X &= 4\pi\epsilon_X + (i_X \cdot d + d \cdot i_X)\delta_X \\ &= 4\pi(\epsilon_X + i_X \eta_X) + d(i_X \delta_X) \end{aligned}$$

or

$$d\Omega_X = \Theta_X,$$

where

$$\Omega_X = \beta_X - i_X \delta_X, \quad \Theta_X = 4\pi(\epsilon_X + i_X \eta_X). \tag{30}$$

Thus the contents of the Maxwell equations (24)–(27) can be succinctly incorporated in the following:

$$\begin{aligned} &d\Gamma_X = 0, \\ &d\Omega_X = \Theta_X, \end{aligned} \tag{31}$$

where

$$\begin{aligned} &\Gamma_X = \alpha_X + i_X \gamma_X, \\ &\Omega_X = \beta_X - i_X \delta_X, \\ &\Theta_X = 4\pi(\epsilon_X + i_X \eta_X). \end{aligned} \tag{32}$$

In vacuum $\epsilon_X = \eta_X = \Theta_X = 0$; thus the vacuum Maxwell equations are

$$\begin{aligned} &d\Gamma_X = 0, \\ &d\Omega_X = 0. \end{aligned} \tag{33}$$

The continuity equation follows by taking the exterior derivative of the second equation in (31), i.e., $d\Theta_X = 0$. From (32) we obtain

$$0 = d(\epsilon_X + i_X \eta_X) = d\epsilon_X + (\mathcal{L}_X - i_X \cdot d)\eta_X.$$

Since η_X is a three-form on a three-manifold, $d\eta_X = 0$. Thus

$$\mathcal{L}_X \eta_X + d\epsilon_X = 0, \tag{34}$$

which is the continuity equation relative to the observer X if we keep in mind that the Lie derivative $\mathcal{L}_X$ corresponds to the time derivative $\partial/\partial t$ and that $d\epsilon_X$ is the divergence of the current $\mathbf{j}$ relative to X.

VI. GRAVITATIONAL RED SHIFT

We shall now indicate how dynamical models can be treated in this formalism.

Consider the problem of gravitational red shift in the field of a massive spherical object whose mass would determine the metric g of our three-manifold (M,g) according to some field equations, which we leave for the moment unspecified. Let us suppose only that $g \neq \delta$, the Euclidean metric.

Essential to this problem is the notion of *similar clocks*. By a clock we shall mean a periodic (or closed) curve c or a corresponding vector field c^* with *periodic* (or closed) integral curves. For example, in two dimensions in $\mathbb{R}^2$, a clock at the origin would be given by $c^* = -y(\partial/\partial x) + x(\partial/\partial y)$, with the integral curves c: $\lambda \to (x = a\cos\lambda,\ y = a\sin\lambda)$. Note that strictly speaking, a clock is not a physical observer since it is allowed to be a *singular* vector field (however, on $\mathbb{R}^2 - \{0\}$, c^* is nonsingular). Two clocks would be considered *similar* if there exists a differentiable transformation between the corresponding vector fields. For example, $c_1^* = y(\partial/\partial x) + (x - r_1)(\partial/\partial y)$ and $c_2^* = -y(\partial/\partial x) + (x - r_2)(\partial/\partial y)$ represent two similar clocks at $(r_1,0)$ and $r_2,0)$ in $\mathbb{R}^2$ (or in a two-dimensional M). Relative to some physical observer X, the time intervals of c_1 and c_2 would be given by

$$\Delta\tau_X^{(1)} = \int_0^{2\pi} F(X, c_1^*)\, d\lambda, \quad \Delta\tau_X^{(2)} = \int_0^{2\pi} F(X, c_2^*)\, d\lambda.$$

Since $g \neq \delta$, in general, $\Delta\tau_X^{(1)} \neq \Delta\tau_X^{(2)}$. More generally, one should perhaps consider a certain class $\{X\}$ of physical observers or vector fields of (M,g) representing, for example, the free-fall observers in a gravitational field determined by g, instead of similar clocks as defined above.

Some examples of special classes $\{X\}$ of vector fields in (M,g) are (where ∇ is the Levi-Civita connection determined by g and ∇_Z is the corresponding covariant derivative operator)

parallel vector fields: $\nabla_Z X = 0$, for all Z,

geodesic vector fields: $\nabla_X X = 0$,

Killing vector fields: $\mathcal{L}_X g = 0$,

Ricci-constant vector fields: $\mathcal{L}_X \text{Ricc}(g) = 0$.

Note that parallel vector fields are also inertially equivalent observers. For a Levi-Civita connection,

$$Z(g(X,Y)) = g(\nabla_Z X, Y) + g(\nabla_Z Y, X)$$

for all vector fields X, Y, Z. Hence, if X,Y are parallel, $g(X,X), g(Y,Y)$, and $g(X,Y)$ are constants and thus $V(X,Y)$ is also a constant function.

VII. COSMOLOGICAL RED SHIFT

We shall briefly sketch how problems of cosmology, for example, cosmological red shift can be treated in this formalism.

A cosmological model would consist of a three-manifold M with Riemannian three-metric g and a class $\{X\}$ of vector fields (see above), the so-called *cosmological observers*, determined by (M,g). The relative acceleration of two cosmological observers $X,Y \in \{X\}$ is given by $\mathcal{L}_X V(X,Y) = Y(V(X,Y))$ or $X(V(X,Y))$. If such a model is to predict a cosmological red shift a necessary condition would be, for example,

$$Y(V(X,Y)) = \alpha V(X,Y) F(X,Y), \quad \forall X,Y \in \{X\}. \tag{35}$$

Such a relation would imply a simple Hubble law of relative velocity strictly proportional to distance, with α as the constant of proportionality.

The problem then would be to determine whch (M,g) and for what class $\{X\}$ of vector fields relation (35) holds.

VIII. CONCLUSION

It thus seems possible to reformulate the essential ingredients of both special and general relativity in a purely three-dimensional framework instead of the traditional four-dimensional space-time approach.

The Maxwell equations (31) can be generalized to the non-Abelian Yang–Mills case by considering SU(2) Lie-algebra-valued three-dimensional differential forms on M.

We have not touched upon dynamics of particles or field equations for the metric g. Other topics for further study would be to consider Hamiltonian mechanics and subsequent quantization in this three-dimensional setup.

ACKNOWLEDGMENTS

This work was supported in part by Schweizerischer Nationalfonds Fellowship No. 86.536.0.87 and NSERC, Canada, Bilateral Exchange Program Grant No. 0039328.

[1] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1983), Vol. 1.

[2] G. D. Mostow, *Strong Rigidity of Locally Symmetric Spaces*, Ann. Math. Series (Princeton U.P., Princeton, NJ, 1973), p. 136.

[3] V. Variçak, Phys. Z. **11**, 93, 287, 586 (1910).

[4] J. Milnor, Bull. Am. Math. Soc. **6**, 9 (1982).

[5] S. Helgason, *Differential Geometry and Symmetric Space* (Academic, New York, 1972).

Solutions of an equation that lead to solutions of a modified Emden equation and a coupled Korteweg–deVries equation

Ashis Basu
Rishra High School, Rishra, Hooghly, West Bengal, India
Dipankar Ray
Department of Mathematics, Jadavpur University, Calcutta-700032, India

(Received 27 June 1989; accepted for publication 22 November 1989)

In a recent paper, Moreira [Res. Rep. IF/UFRJ/83/25, Universidade Federal do Rio de Janeiro Inst. de Fisica, Cidade Univ., Ilha do Fundao, Rio de Janeiro, Brazil] obtained a nonlinear second-order differential equation that leads to the first integral of a modified Emden equation. He also obtained two particular solutions of his equation. This paper completely integrates Moreira's equation and uses it to get a class of solutions of a coupled Korteweg–deVries (KdV) equation, recently studied by Guha Ray, Bagchi, and Sinha [J. Math. Phys. **27**, 2558 (1986)].

I. INTRODUCTION

In a recent paper, Moreira[1] has studied the modified Emden equation

$$q_{tt} + \alpha(t) q_t + q^n = 0 \tag{1.1}$$

and found the following first integral:

$$I = \exp\left(m \int^t \alpha(t')dt'\right)\{q_t^2 + \alpha q q_t (2-m) + [2/(n+1)]q^{n+1} + \tfrac{1}{2}q^2(m-2)[(m-1)\alpha^2 + \alpha_t]\}, \tag{1.2}$$

where $m = 2(n+1)/(n+3)$, provided the function $\alpha(t)$ satisfied the differential equation

$$\alpha_{tt} + (3m-2)\alpha\alpha_t + \alpha^3(m-1)m = 0, \tag{1.3}$$

where $\alpha_t \equiv d\alpha/dt, \alpha_{tt} \equiv d^2\alpha/dt^2$.

Moreira[1] did not find out the general solution of Eq. (1.3), but gave two particular solutions

$$\alpha(t) = 2/mt, \quad \alpha(t) = [(m-1)t]^{-1}. \tag{1.4}$$

The above results can also be found in a paper by Leach,[2] which gives some other particular solutions of Eq. (1.3). In the present paper, we intend to obtain the complete solutions of equation

$$\alpha_{tt} + a\alpha\alpha_t + b\alpha^3 = 0, \tag{1.5}$$

which is a more general solution of Eq. (1.3) where a and b are constants. Further, we intend to apply the solutions of Eq. (1.5) to obtain some new solutions of some coupled Korteweg–deVries (KdV) equations, recently studied by Guha Ray, Bagchi, and Sinha,[3] which are

$$u_t + \lambda v v_x + \beta u u_x + \delta u_{xxx} = 0, \tag{1.6a}$$

$$v_t + \gamma(uv)_x = 0, \tag{1.6b}$$

where $u_x \equiv \partial u/\partial x$ and so on and $\lambda, \beta, \delta, \gamma$ were kept arbitrary.

II. SOLUTIONS OF EQ. (1.5)

Equation (1.5) can be rewritten as

$$\frac{\alpha_t}{\alpha}\frac{d\alpha_t}{d\alpha} + a\alpha_t + b\alpha^2 = 0,$$

i.e.,

$$\frac{d(\alpha_t)}{d(\alpha^2)} + \frac{a\alpha_t + b\alpha^2}{2\alpha_t} = 0,$$

which gives

$$\frac{2\,ds}{s} = \frac{\theta\,d\theta}{\theta^2 - (a/2)\theta + b/2}, \tag{2.1}$$

where

$$s = 1/\alpha, \quad \theta = -\alpha_t/\alpha^2. \tag{2.2}$$

Equation (2.1) admits two solutions.

Solution 1: $\theta =$ constant given by $\theta^2 - (a/2)\theta + b/2 = 0$, i.e.,

$$\alpha = \left(\frac{4}{a \pm \sqrt{a^2 - 8b}}\right)\frac{1}{t}. \tag{2.3}$$

In particular when $a = 3m-2$ and $b = m(m-1)$ one gets the solutions obtained by Moreira.[1]

Solution 2: $\theta \neq$ constant. Equation (2.1) can be integrated to give

$$s^4 = f(\theta), \tag{2.4}$$

where

$$f(\theta) = A\left\{\left(\theta - \frac{a+\sqrt{a^2-8b}}{4}\right)^{a/\sqrt{a^2-8b}+1} \times \left[\left(\theta - \frac{a-\sqrt{a^2-8b}}{4}\right)^{a/\sqrt{a^2-8b}-1}\right]^{-1}\right\}, \tag{2.5}$$

and A is any integrating constant.

Differentiating (2.4) with respect to (w.r.t.) t and using (2.2), one can get

$$t + B = \frac{f^{1/4}}{\theta} + \int \frac{f^{1/4}}{\theta^2}\,d\theta, \tag{2.6}$$

where B is an integrating constant.

Also, from (2.4) and (2.2)

$$\alpha = 1/\sqrt[4]{f(\theta)}, \tag{2.7}$$

where $f(\theta)$ is given by (2.5).

Hence from (2.6) and (2.7) we can say α is a function of t through the parameter θ.

 0022-2488/90/051152-03$03.00

III. EQUATIONS (1.6) WITH $u_{xxx}=0$

In this section, we present a class of solutions of Eqs. (1.6) by introducing a simplifying assumption

$$u_{xxx}=0. \tag{3.1}$$

It can be shown that Eqs. (1.6) together with Eq. (3.1) imply

$$u_{xx}=0,$$

i.e.,

$$u=xl(t)+r(t), \tag{3.2}$$

provided $\beta\neq 0$, $\gamma\neq 0$.

The proof is given in the Appendix.

The solutions of Eqs. (1.6) will be done by using the solutions of Eqs. (1.5) obtained in the previous section.

Using (3.2), Eqs. (1.6) reduce to

$$l_t x + r_t + (\lambda/2)(v^2)_x + (\beta/2)(u^2)_x = 0, \tag{3.3a}$$

$$v_t + \gamma[(lx+r)v_x + lv] = 0. \tag{3.3b}$$

Integrating Eq. (3.3a) w.r.t. x,

$$\frac{x^2}{2}l_t + xr_t + \frac{\lambda}{2}v^2 + \frac{\beta}{2}u^2 + F(t) = 0, \tag{3.4}$$

where $F(t)$ is any arbitrary function of t.

Differentiating (3.4) w.r.t t, and using (3.3b)

$$(x^2/2)l_{tt} + xr_{tt} - \lambda\gamma[(lx+r)vv_x + lv^2] + \beta(lx+r)(l_t x + r_t) + F_t = 0. \tag{3.5}$$

Substituting the values of v^2 and vv_x given from (3.4) into Eq. (3.5) we get a second-order polynomial equation of x where its coefficients are function of t only. From this one can obtain

$$l_{tt} + 4\gamma\beta l^3 + (4\gamma+2\beta)ll_t = 0, \tag{3.6a}$$

$$r_{tt} + 4\gamma\beta l^2 r + (3\gamma+\beta)lr_t + (\beta+\gamma)rl_t = 0, \tag{3.6b}$$

$$2\gamma\beta l r^2 + (\gamma+\beta)rr_t + 2\gamma l F + F_t = 0. \tag{3.6c}$$

Equation (3.6a) is exactly the same form as Eq. (1.5). We can solve $l(t)$ from (3.6a). Substituting the values of $l(t)$ in (3.6b) we get $r(t)$ and finally $F(t)$ from (3.6c).

The solutions are presented in an article to follow.

IV. EXPLICIT SOLUTIONS OF EQS. (1.6)

A. Case I

Equation (3.6a) has two particular solutions

$$l = 1/\beta t, \tag{4.1a}$$

$$l = 1/2\gamma t. \tag{4.1b}$$

Using (4.1a), Eq. (3.6b) can be solved, and from it we get

$$r = c_1/t + c_2 t^{1-3\gamma/\beta}, \tag{4.2}$$

where c_1 and c_2 are constants.

Using (4.1a) and (4.2) one can obtain from Eq. (3.6c)

$$F = \frac{c_3\beta}{2(\gamma-\beta)}\frac{1}{t^2} + \frac{c_4\beta}{2(\beta-2\gamma)}t^{2-6\gamma/\beta} + \frac{c_5\beta}{\gamma}t^{-\gamma/\beta} + c_6, \tag{4.3}$$

where

$$c_3 = c_1^2(\beta-\gamma), \quad c_4 = (c_2^2/\beta)(3\gamma^2-\beta^2),$$

$$c_5 = (c_1c_2\gamma/\beta)(\beta-3\gamma),$$

c_6 = constant of integration,

and with the help of (4.1a), (4.2), and (4.3) we get solutions for Eqs. (1.6) through Eqs. (3.2) and (3.4).

Similarly, using (3.7b) we get another solution for Eqs. (1.6).

B. Case II

With the help of Eqs. (2.6) and (2.7) one can get from (3.6a)

$$l = 1/\sqrt[4]{g(\phi)} \tag{4.4a}$$

and

$$t + Q_1 = \frac{g^{1/4}}{\phi} + \int \frac{g^{1/4}}{\phi^2}\,d\phi, \tag{4.4b}$$

where Q_1 is an integrating constant;

and

$$g(\phi) = A_1[(\phi-K)^N/(\phi-M)^{N-2}], \tag{4.5}$$

where

$$K = [2\gamma+\beta \pm (2\gamma-\beta)]/2,$$

$$M = [2\gamma+\beta \mp (2\gamma-\beta)]/2,$$

$$N = (2\gamma+\beta)/\pm(2\gamma-\beta)+1,$$

A_1 = constant of integration.

Substituting $r=ly$ in Eq. (3.6c) we get a differential equation for y that can be solved, and from this one can get

$$r = \frac{2c}{\sqrt[4]{g(\phi)}}\int \frac{(\phi-K)^P}{(\phi-M)^{P+1/2}}\,d\phi, \tag{4.6}$$

where $g(\phi)$ is given by (4.5) and

c = constant of integration,

$$P = \beta/\pm 4(2\gamma-\beta) - \tfrac{1}{4}. \tag{4.7}$$

Using (4.4a) and (4.6) in Eq. (3.6c) one can get the following solution:

$$F = A_1^{-1/2}\left(\frac{\phi-K}{\phi-M}\right)^{\gamma/\mp(2\gamma-\beta)}\int\left[\frac{\gamma+\beta}{2}\frac{r^2}{l^2}\phi - \gamma\beta\frac{r^2}{l^2} - 2c(\gamma+\beta)\frac{r}{l}\frac{(\phi-K)^{P+1}}{(\phi-M)^{P-1/2}}\right] \times\frac{(\phi-K)^{\beta/\mp 2(2\gamma-\beta)-3/2}}{(\phi-M)^{\beta/\mp 2(2\gamma-\beta)+3/2}}\,d\phi, \tag{4.8}$$

where $l(t)$, $r(t)$, and P are given by (4.4a), (4.6), and (4.7).

Hence, the solutions of Eq. (1.6), i.e., u and v can be determined from Eqs. (3.2) and (3.4) with the help of Eqs. (4.4a), (4.6), and (4.8), respectively.

V. CONCLUSION

In summary, we state that Eq. (1.5) admits two types of solutions. The first type of solution is in explicit form given by Eqs. (2.3) and the second type is in parametric form given by (2.6) and (2.7). Equation (1.5) has two important uses. The first one is through a special case of Eq. (1.5),

which is Eq. (1.3). It has been shown by Moreira[1] that if $\alpha(t)$ of modified Emden equation (1.1) satisfied Eq. (1.3) then the first integral of Eq. (1.1) is given by Eq. (1.2). Moreira[1] also noted here that Eq. (1.2) admits a solution of the form given by Eq. (2.3). The second application of Eq. (1.5) is toward obtaining the particular solutions for coupled KdV equations (1.6). In Sec. III we solved Eqs. (1.6) for $u_{xxx} = 0$ with the help of the solutions of Eq. (1.5) in two cases. The solution is as follows: u and v are given by Eqs. (3.2) and (3.4), where $l(t)$, $r(t)$, and $F(t)$ satisfy Eqs. (3.6). Equation (3.6a) is of the same form as Eq. (1.5) and hence its solutions are known; $r(t)$ and $F(t)$ are then obtained from (3.6b) and (3.6c). These solutions are obtained explicitly in Sec. IV.

APPENDIX: DERIVATION OF EQ. (3.2)

Differentiating Eq. (1.6b) w.r.t. x and using (1.6a)

$$(u_t + \beta u u_x)_t + 2\gamma u_x (u_t + \beta u u_x) - \lambda\gamma v^2 u_{xx} + \gamma u_x (u_t + \beta u u_x) + \gamma u (u_t + \beta u u_x)_x = 0. \quad \text{(A1)}$$

Differentiating (A1) w.r.t. x twice and using Eqs. (3.1) and (1.6a) one can obtain, after simplification

$$(5\gamma + 3\beta) u_x u_{xxt} + (9\gamma + 3\beta) u_{xx} u_{xt} + 24\gamma\beta u_x^2 u_{xx} + 12\gamma\beta u u_{xx}^2 + u_{xxtt} = 0. \quad \text{(A2)}$$

Again differentiating (A2) w.r.t. x and using (3.1)

$$(14\gamma + 6\beta) u_{xx} u_{xxt} + 60\gamma\beta u_x u_{xx}^2 = 0,$$

which on differentiation gives

$$60\gamma\beta u_{xx} = 0.$$

Hence the result.

[1] I. de C. Moreira comments on "A direct approach to finding exact invariants for one-dimensional time dependent classical Hamiltonians," Res. Rep. IF/UFRJ/83/25, Universidade Federal do Rio de Janeiro Instituto de Física, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, Brazil.
[2] P. G. L. Leach, J. Maths. Phys. **26**, 2510 (1985).
[3] C. Guha Ray, B. Bagchi, and D. K. Sinha, J. Math. Phys. **27**, 2558 (1986).

Remarks on a system of coupled nonlinear wave equations

E. V. Krishnan
Department of Mathematics and Computing, Sultan Qaboos University, Al-Khod, P.O. Box 32486, Muscat, Oman

(Received 1 November 1988; accepted for publication 13 December 1989)

It is shown that a generalized system of coupled KdV-MKdV equation exhibits solitary wave solutions. It has been shown that an increase in nonlinearity in one variable in a particular fashion does not affect the existence of solitary wave solutions. The periodic traveling wave solutions of the system have also been investigated.

I. INTRODUCTION

Lots of work has been done on the solitary wave solutions of coupled nonlinear wave equations.[1–6] Recently Guha-Roy[7] has shown that a generalized system of combined KdV-MKdV equations in the form

$$u_t + \alpha v^2 v_x + \beta u^2 u_x + \lambda u u_x + \gamma u_{xxx} = 0, \tag{1.1}$$

$$v_t + \delta (uv)_x + \epsilon v v_x = 0, \tag{1.2}$$

where $\alpha, \beta, \lambda, \gamma, \delta$, and ϵ are arbitrary parameters, possesses solitary wave solutions. In this paper, we will now show that an equation of the form

$$u_t + \alpha v^2 v_x + \beta u^2 u_x + \lambda u u_x + \gamma u_{xxx} = 0, \tag{1.3}$$

$$v_t + \delta (uv)_x + \epsilon v^3 v_x = 0 \tag{1.4}$$

also exhibits solitary wave solutions.

It is remarkable that an increase in nonlinearity in the variable v in Eq. (1.3) with a corresponding increase in nonlinearity in the same variable in Eq. (1.4) does not affect the existence of solitary wave solutions. Also, we have found the periodic traveling wave solutions of the generalized system with increased nonlinearity.

II. THE COUPLED EQUATION

Considering traveling wave solutions of Eqs. (1.3) and (1.4) in the form

$$u = u(x - ct), \quad v = v(x - ct), \quad \xi = x - ct, \tag{2.1}$$

where c is the wave velocity, we obtain

$$-cu_\xi + \frac{\alpha}{3}(v^3)_\xi + \frac{\beta}{3}(u^3)_\xi + \frac{\lambda}{2}(u^2)_\xi + \gamma u_{\xi\xi\xi} = 0, \tag{2.2}$$

$$-cv_\xi + \delta (uv)_\xi + (\epsilon/4)(v^4)_\xi = 0. \tag{2.3}$$

Integrating (2.3), we get

$$u = (k/\delta v) + (c/\delta) - (\epsilon/4\delta)v^3, \tag{2.4}$$

where k is the integration constant.

Without loss of generality we can impose $k = 0$ to have a regular u everywhere so that (2.4) reduces to

$$v^3 = (4/\epsilon)(c - \delta u). \tag{2.5}$$

Substituting (2.5) in (2.2) and assuming the solitary wave boundary conditions that u, u_ξ, and $u_{\xi\xi} \to 0$ as $|\xi| \to \infty$, we arrive at

$$u_\xi^2 = a_2 u^2 + a_3 u^3 + a_4 u^4, \tag{2.6}$$

where the parameters a_2, a_3, and a_4 are given by

$$a_2 = (1/\gamma)(c + 4\alpha\delta/3\epsilon), \tag{2.7}$$

$$a_3 = -\lambda/3\gamma, \tag{2.8}$$

$$a_4 = -\beta/6\gamma. \tag{2.9}$$

Now we shall consider a system of equations with nonlinearity increased in v by one degree in both Eqs. (1.3) and (1.4) in the form

$$u_t + \alpha v^3 v_x + \beta u^2 u_x + \lambda u u_x + \gamma u_{xxx} = 0, \tag{2.10}$$

$$v_t + \delta (uv)_x + \epsilon v^4 v_x = 0. \tag{2.11}$$

Following the same procedure as above, we arrive at two equations:

$$-cu_\xi + \frac{\alpha}{4}(v^4)_\xi + \frac{\beta}{3}(u^3)_\xi + \frac{\lambda}{2}(u^2)_\xi + \gamma u_{\xi\xi\xi} = 0, \tag{2.12}$$

$$v^4 = (5/\epsilon)(c - \delta u). \tag{2.13}$$

Substituting (2.13) in Eq. (2.12), integrating twice, and rearranging the terms with the same boundary conditions, we get

$$u_\xi^2 = a_2 u^2 + a_3 u^3 + a_4 u^4, \tag{2.14}$$

where a_2, a_3, and a_4 are given by

$$a_2 = (1/\gamma)(c + 5\alpha\delta/4\epsilon), \tag{2.15}$$

$$a_3 = -\lambda/3\gamma, \tag{2.16}$$

$$a_4 = -\beta/6\gamma. \tag{2.17}$$

So, we notice that Eq. (2.14)–(2.17) have the same form as that of Eq. (2.6)–(2.9) which differ only in (2.15).

We may generalize the system in the form

$$u_t + \alpha v^r v_x + \beta u^2 u_x + \lambda u u_x + \gamma u_{xxx} = 0, \tag{2.18}$$

$$v_t + \delta (uv)_x + \epsilon v^{r+1} v_x = 0, \tag{2.19}$$

which will reduce to

$$u_\xi^2 = a_2 u^2 + a_3 u^3 + a_4 u^4 \tag{2.20}$$

with

$$a_2 = \frac{1}{\gamma}\left\{c + \frac{(r+2)\alpha\delta}{(r+1)\epsilon}\right\}, \tag{2.21}$$

and a_3 and a_4 will remain unchanged.

The investigation of solitary wave solution is a routine procedure that is already well known.[7] However, for the sake of completeness, we rederive solitary wave solution of Eq. (2.20).

 0022-2488/90/051155-02$03.00

III. SOLITARY WAVE SOLUTIONS

We put $u = 1/U$ and transform Eq. (2.20) into

$$U_\xi^2 = a_2U^2 + a_3U + a_4. \tag{3.1}$$

Equation (3.1) has a solution in the form

$$U(\xi) = \{D\cosh[\sqrt{a_2}(\xi+\xi_0)] - a_3\}/2a_2, \tag{3.2}$$

for $a_2 > 0$ and $a_3^2 > 4a_2a_4$. Here, $D = \sqrt{a_3^2 - 4a_2a_4}$ and ξ_0 is an integration constant.

Therefore, the system of Eqs. (2.18) and (2.19) have solutions in the form

$$u(\xi) = 2a_2/\{D\cosh[\sqrt{a_2}(\xi+\xi_0)] - a_3\}, \tag{3.3a}$$

$$v(\xi) = \left(\frac{r+2}{\epsilon}\right)^{1/(r+1)} \times\left(c - \frac{2\delta a_2}{D\cosh[\sqrt{a_2}(\xi+\xi_0)] - a_3}\right)^{1/(r+1)}. \tag{3.3b}$$

The denominator of (3.3a) may vanish for certain positive a_3 which could lead to a singularity, and so a physically admissible solution for (2.20) will be valid only for $a_3 < 0$. In other words, λ and γ should be of the same sign.

It is clear from the form of (3.3b) that ϵ should be nonzero for a nonsingular solution. Of course, it is evident from the form of a_2, a_3, and a_4 that γ should be nonzero for the existence of solitary wave solutions, which is even otherwise clear because $\gamma = 0$ is the case on nondispersive waves. Therefore, to ensure that u and v do not blow up, we have the conditions that ϵ is nonzero and that λ and γ are of the same sign.

For u and v to be real valued, $a_3^2 > 4a_2a_4$. This gives rise to certain restrictions on the coefficients. We can see that for a positive γ, β should always be greater than $-\lambda^2/6$ and for a negative γ, β should always be less than $\lambda^2/6$. Apart from these, from the form of (3.3b) one could see that ϵ and δ are positive with the condition that $\delta < c(D - a_3)$. Here, without loss of generality, we have assumed that c is positive.

Therefore the system (2.18) and (2.19) has solitary wave solutions that are bounded and real valued, given by equations (3.3a) and (3.3b) with the following constraints on the coefficients, viz.,

(i) λ and γ are of the same sign, $\epsilon \neq 0$,

(ii) $\beta > \dfrac{-\lambda^2}{6}$ for $\gamma > 0$,

$\beta < \dfrac{\lambda^2}{6}$ for $\gamma < 0$,

(iii) $\epsilon > 0$, $\delta > 0$, $\delta < c(D - a_3)$ for positive c.

IV. PERIODIC WAVE SOLUTIONS

Following the same procedure adopted in Sec. II, we arrive at two equations from the system (2.18) and (2.19) in the form

$$-cu_\xi + \frac{\alpha}{r+1}(v^{r+1})_\xi + \frac{\beta}{3}(u^3)_\xi + \frac{\lambda}{2}(u^2)_\xi + \gamma u_{\xi\xi\xi} = 0, \tag{4.1}$$

$$v^{r+1} = ((r+2)/\epsilon)(c - \delta u). \tag{4.2}$$

Substituting (4.2) in (4.1) and integrating twice, we get

$$u_\xi^2 = \frac{2}{\gamma}\left[K_2 + K_1u + \left\{\frac{c}{2} + \frac{(r+2)\alpha\delta}{(2r+2)\epsilon}\right\}u^2 - \frac{\lambda}{6}u^3 - \frac{\beta}{12}u^4\right], \tag{4.3}$$

where K_1 and K_2 are integration constants.

Now, (4.3) can be written in the form

$$u_\xi^2 = a_0 + a_1u + a_2u^2 + a_3u^3 + a_4u^4, \tag{4.4}$$

where

$$a_0 = \frac{2K_2}{\gamma},\quad a_1 = \frac{2K_1}{\gamma},\quad a_2 = \frac{c}{2} + \frac{(r+2)\alpha\delta}{(r+1)\epsilon},$$
$$a_3 = -\lambda/3\gamma,\quad a_4 = -\beta/6\gamma. \tag{4.5}$$

It can be shown by direct substitution that Eq. (4.4) has a real-valued periodic wave solution

$$u = \frac{u_0\,\mathrm{cn}(p\xi)}{1 - (a_1u_0/4a_0)\mathrm{cn}(p\xi)}, \tag{4.6}$$

where u_0 and p are positive real parameters and cn is the Jacobian cosine elliptic function with modulus k given by

$$k^2 = \frac{1}{2} + \left(\frac{a_2}{2} - \frac{3}{16}\frac{a_1^2}{a_0}\right)\Big/p^2. \tag{4.7}$$

We note that the condition $0 \leqslant k \leqslant 1$ imposes a restriction on p:

$$p^2 \geqslant a_2 - \tfrac{3}{8}(a_1^2/a_0). \tag{4.8}$$

It may also be noted that the constants of integration K_1 and K_2 should be of different signs for a nonsingular solution.

In the solitary wave limit when $k \to 1$, $\mathrm{cn}(p\xi) \to \mathrm{sech}(p\xi)$ and the governing equation will reduce to

$$u_\xi^2 = a_2u^2 + a_3u^3 + a_4u^4. \tag{4.9}$$

The solution of (4.9) will now be

$$u(\xi) = \frac{u_0\,\mathrm{sech}(p\xi)}{1 - E\,\mathrm{sech}(p\xi)}, \tag{4.10}$$

$$p = \sqrt{a_2},\quad u_0 = \frac{2a_2}{\sqrt{a_3^2 - 4a_2a_4}},\quad E = \frac{a_3}{\sqrt{a_3^2 - 4a_2a_4}}. \tag{4.11}$$

We can easily see that the solution (4.10) is exactly the same as (3.3a).

ACKNOWLEDGMENT

The author is greatly indebted to the referee for his valuable suggestions.

[1] M. Ito, Phys. Lett. A **91**, 335 (1982).
[2] E. V. Krishnan, J. Phys. Soc. Jpn. **51**, 2391 (1982).
[3] S. Kawamoto, J. Phys. Soc. Jpn. **53**, 1203 (1984).
[4] S. Kawamoto, J. Phys. Soc. Jpn. **54**, 1701 (1985).
[5] E. V. Krishnan, J. Phys. Soc. Jpn. **55**, 3753 (1986).
[6] C. Guha-Roy, B. Bagchi, and D. K. Sinha, J. Math. Phys. **27**, 2558 (1986).
[7] C. Guha-Roy, J. Math. Phys. **28**, 2087 (1987).

The free Dirac operator on compact and noncompact graphs

W. Bulla and T. Trenkler
Institut für Theoretische Physik, TU Graz, Petersgasse 16, A-8010 Graz, Austria

(Received 15 June 1989; accepted for publication 20 December 1989)

The free Dirac operator defined on composite one-dimensional structures consisting of finitely many half-lines and intervals is investigated. The influence of the connection points between the constituents is modeled by transition conditions for the wave functions or equivalently by different self-adjoint extensions of the Dirac operator. General relations between the parameters of the extensions and the eigenvalues *resp.* the scattering coefficients are derived and then applied to the cases of a bundle of half-lines, a point defect, a branching line, and an eye-shaped structure.

I. INTRODUCTION

The increasing interest in spatially one-dimensional quantum mechanical models during the last several years has two reasons: On the one hand one tries, with the aim of enlargement and completion of insights into the structure of quantum mechanics, apart from continuing general mathematical investigations, to construct further rigorously solvable models that can serve to illustrate already known statements of general character and to get hints for interrelations not yet noticed. One-dimensional models play an important role in this connection because of their good tractability.

On the other hand, methods have been found in the last several years to make structures with high electron mobility, which can be regarded as one-dimensional owing to the small diameter of their components. Metallic structures of this kind have been realized some years ago[1] and the quantum mechanics of the electrons moving therein has been studied by Exner and Šeba,[2,3] choosing a description of these devices by composite one-dimensional structures. They described the effects of the connection points by transition conditions (boundary conditions) for the wave functions and equivalently by different self-adjoint extensions of the free Schrödinger operators on these graphs and obtained results for the S matrix. They also considered problems of this kind on (partially) two-dimensional objects in Refs. 4–6.

Gerasimenko and Pavlov[7] treated arbitrary compact and noncompact graphs, starting with a general setting admitting potentials in the Schrödinger operators. They limited, although, the considered extensions to the particular ones given by the requirement of continuity of the wave functions at the connection points and the vanishing of the sums of their derivatives at these points and obtained general results on the Green's function, the spectrum, and the scattering coefficients of the problem.

Looking for further solvable models it is natural to study the Dirac operator on the same sort of geometrical objects as has been done for the Schrödinger operator. In the present work in Sec. II, using the extension theory for symmetric operators, boundary conditions in the form of linear equations characterizing different self-adjoint extensions of the free Dirac operator are derived for structures consisting of a finite number of half-lines and intervals. With the help of these equations we derive necessary and sufficient conditions for the eigenvalues and relations for the scattering coefficients of these extensions.

In Sec. III we apply the results obtained in Sec. II to certain special cases that can be expected to be the most relevant ones in practice and thus demonstrate the applicability of the general formalism to special problems. These are mostly explicitly solvable models for the one-dimensional Dirac operator in the sense that the eigenvalues and the scattering coefficients of the problem can be expressed by the parameters of the chosen extension of the Dirac operator (or equivalently of the transition conditions). We treat in detail the bundle of half-lines, then as special cases thereof a point defect in a line, and a branching line; furthermore an eye-shaped structure.

A physical model using the Dirac operator in one spatial dimension occurs in the treatment of electron transport in certain polymer molecules, taking into account the electron–phonon coupling. Its continuum limit results in a two-band model that is formulated by means of a Dirac operator.[8]

II. GENERAL THEORY

A. The Dirac operator on bounded intervals and on the half-line

Let I be an open interval in $\mathbb{R}$, bounded or unbounded, $\mathscr{W}_{2,1}(I)$ the Sobolev space of order 1 over I, and $\mathscr{W}^0_{2,1}(I)$ the subspace of it consisting of the functions vanishing on the boundary of I. The two Dirac operators $D_{0,I}$ and D_I are given by the differential expression

$$-i\frac{d}{dx}\otimes\sigma_1 + 1\otimes\sigma_3, \tag{2.1}$$

with domains $\mathscr{D}(D_{0,I}) = \mathscr{W}^0_{2,1}(I)\otimes\mathbb{C}^2$ and $\mathscr{D}(D_I) = \mathscr{W}_{2,1}(I)\otimes\mathbb{C}^2$. From these definitions we have the following lemma.

Lemma 2.1: $D_{0,I}$ is closed and symmetric with adjoint $D^*_{0,I} = D_I$.

In order to construct explicitly self-adjoint extensions of $D_{0,I}$ for the cases I being the half-line or $I = (-1,1)$ according to von Neumann's theory,[9,10] we must specify bases in the deficiency spaces. In what follows, double signs $\pm$ or $\mp$ always refer to $\mathrm{Ker}(D_I \pm i)$ in the corresponding positions. As linearly independent solutions in the form $\binom{f}{g}$ of the equation for the deficiency spaces we choose

 0022-2488/90/051157-07$03.00

$$f_1 = e^{-\sqrt{2}x}, \quad g_1 = [(i \mp 1)/\sqrt{2}]e^{-\sqrt{2}x},$$
$$f_2 = e^{+\sqrt{2}x}, \quad g_2 = [(\pm 1 - i)/\sqrt{2}]e^{+\sqrt{2}x}. \tag{2.2}$$

The deficiency index of $D_{0,I}$ being (1,1) in the case of the half-line and (2,2) in the other case, we introduce the notation

$$a^{\pm} = \begin{pmatrix} 1 \\ \dfrac{i \mp 1}{\sqrt{2}} \end{pmatrix}, \quad \varepsilon = (2\sinh(2\sqrt{2}))^{-1/2}, \tag{2.3}$$

and take as bases in the deficiency spaces

$$\left.\begin{aligned} f^{\pm} &= e^{-\sqrt{2}x} \otimes a^{\pm} \in \mathrm{Ker}(D_{\mathbb{R}^+} \pm i), \\ e_1^{\pm} &= \varepsilon e^{-\sqrt{2}x} \otimes a^{\pm} \\ e_2^{\pm} &= \varepsilon e^{\sqrt{2}x} \otimes \sigma_3 a^{\pm} \end{aligned}\right\} \in \mathrm{Ker}(D_{(-1,1)} \pm i). \tag{2.4}$$

Thus we obtain orthogonal bases whose elements are normalized to $2^{-1/4}$.

B. The Dirac operator on composite structures; self-adjoint extensions and boundary conditions

Turning to the treatment of composite one-dimensional structures consisting of n half-lines and m bounded intervals we take as the natural closed and symmetric operator whose extensions will be studied,

$$D = (\oplus_{j=1}^{n} D_{0,\mathbb{R}^+}) \oplus (\oplus_{j=1}^{m} D_{0,(-1,1)})$$

in

$$(\oplus_{j=1}^{n} \mathscr{L}_2(\mathbb{R}^+)^2) \oplus (\oplus_{j=1}^{m} \mathscr{L}_2((-1,1))^2), \tag{2.5}$$

as it has been done in Refs. 4–6 for the Schrödinger case. Adjoint and deficiency spaces of D are direct sums of the corresponding quantities of the summands, so as a basis for the deficiency space of D we take

$$d_j^{\pm} = (0,\ldots,0,\underset{j-1}{f^{\pm}},0,\ldots,0), \quad \text{for } j \in \{1,\ldots,n\},$$

$$\left.\begin{aligned} d_{n+2j-1}^{\pm} &= (0,\ldots,0,\underset{n+j-1}{e_1^{\pm}},0,\ldots,0) \\ d_{n+2j}^{\pm} &= (0,\ldots,0,\underset{n+j-1}{e_2^{\pm}},0,\ldots,0) \end{aligned}\right\} \quad \text{for } j \in \{1,\ldots,m\}, \tag{2.6}$$

both deficiency numbers of D being $n + 2m$. The self-adjoint extension D_U of D belonging to the matrix $U \in \mathscr{U}(n+2m)$ is then explicitly parametrized by

$$\mathscr{D}(D_U) = \mathscr{D}(D) + \left\{ \sum_{j=1}^{n+2m} \left(\alpha_j d_j^{-} + \sum_{i=1}^{n+2m} U_{ji} \alpha_i d_j^{+} \right) \middle| \alpha_i \in \mathbb{C} \right\},$$
$$D_U = D^*|_{\mathscr{D}(D_U)}. \tag{2.7}$$

In the following a characterization of the extensions by boundary conditions (transition conditions) fulfilled by the functions belonging to the domains of the D_U will be given, which is needed for the applications of this general setup. Noting that every vector ϕ in the domain of D_U is a direct sum of two component spinors we have

$$\phi = \oplus_{j=1}^{n+m} \begin{pmatrix} \phi_j^{(1)} \\ \phi_j^{(2)} \end{pmatrix} \in (\oplus_{j=1}^{n} \mathscr{W}_{2,1}(\mathbb{R}^+)^2) \oplus (\oplus_{j=1}^{m} \mathscr{W}_{2,1}((-1,1))^2), \tag{2.8}$$

where we choose the following splitting of a vector $\alpha \in \mathbb{C}^{n+2m}$ from the deficiency space of D:

$$\alpha_1 = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad \alpha_{1+l} = \begin{pmatrix} \alpha_{n+2l-1} \\ \alpha_{n+2l} \end{pmatrix}, \quad \text{for } l \in \{1,\ldots,m\}. \tag{2.9}$$

Using the matrix

$$\Delta = \begin{pmatrix} e^{\sqrt{2}} & e^{-\sqrt{2}} \\ e^{-\sqrt{2}} & e^{\sqrt{2}} \end{pmatrix}, \tag{2.10}$$

we introduce the following notation for the set of all boundary values of a function in $\mathscr{D}(D_U)$:

$$\left.\begin{aligned} \Phi_1^{(i)} &= \begin{pmatrix} \phi_1^{(i)}(0) \\ \vdots \\ \phi_n^{(i)}(0) \end{pmatrix}, \quad \Psi_1^{(i)} = \Phi_1^{(i)} \\ \Phi_{l+1}^{(i)} &= \begin{pmatrix} \phi_{n+l}^{(i)}(-1) \\ \phi_{n+l}^{(i)}(1) \end{pmatrix}, \quad \begin{aligned} \Psi_{l+1}^{(1)} &= (1/\epsilon)\Delta^{-1}\Phi_{l+1}^{(1)} \\ \Psi_{l+1}^{(2)} &= (1/\epsilon)\sigma_3\Delta^{-1}\Phi_{l+1}^{(2)} \end{aligned} \end{aligned}\right\}, \quad \text{for } \begin{matrix} i \in \{1,2\} \text{ and} \\ l \in \{1,\ldots,m\}. \end{matrix} \tag{2.11}$$

The decomposition (2.9) leads to a corresponding decomposition of the matrix U entering in (2.7)

$$U = \begin{pmatrix} \mathscr{U}_{1,1} & \mathscr{U}_{1,2} & \cdots & \mathscr{U}_{1,m+1} \\ \mathscr{U}_{2,1} & \mathscr{U}_{2,2} & \cdots & \mathscr{U}_{2,m+1} \\ \vdots & \vdots & \natural & \vdots \\ \mathscr{U}_{m+1,1} & \mathscr{U}_{m+1,2} & \cdots & \mathscr{U}_{m+1,m+1} \end{pmatrix}. \tag{2.12}$$

Here, $\mathscr{U}_{1,1}$ is a $n \times n$ matrix describing transitions between the half-lines, $\mathscr{U}_{1,k+1}$ are $2 \times n$ matrices describing transitions from the k th interval to the half-lines, $\mathscr{U}_{k+1,1}$ are $n \times 2$ matrices describing transitions from the half-lines to the k th interval, and the $\mathscr{U}_{k+1,l+1}$ are 2×2 matrices describing the transition from the l th to the k th interval.

Theorem 2.2: The following statements are equivalent:

(a) $\phi \in \mathscr{D}(D_U)$,

(b) $\phi \in (\oplus_{j=1}^{n} \mathscr{W}_{2,1}(\mathbb{R}^+)^2) \oplus [\oplus_{j=1}^{m} \mathscr{W}_{2,1}((-1,1))^2]$, (2.13)

where ϕ fulfills the boundary conditions

$$0=\Psi_k^{(2)}-\frac{1+i}{\sqrt{2}}\Psi_k^{(1)}+\sum_{j=1}^{m+1}\mathscr{U}_{k,j}\left(\Psi_j^{(2)}-\frac{i-1}{\sqrt{2}}\Psi_j^{(1)}\right), \tag{2.14}$$

for $k\in\{1,...,m+1\}$.

Proof: Because of (2.7), $\phi\in\mathscr{D}(D_U)$ is equivalent to

$$\begin{aligned}&\forall k\in\{1,...,n\} && \exists\phi_{k,0}\in\mathscr{W}^0_{2,1}(\mathbb{R}^+)^2,\\ &\forall k\in\{n+1,...,n+m\} && \exists\phi_{k,0}\in\mathscr{W}^0_{2,1}((-1,1))^2,\\ &\exists\alpha\in\mathbb{C}^{n+2m}\end{aligned} \tag{2.15}$$

such that

$$\begin{aligned}\phi_k&=\phi_{k,0}+\alpha_k f^-+\sum_{j=1}^{n+2m}U_{k,j}\alpha_j f^+, \quad \text{for } k\in\{1,...,n\},\\ \phi_{n+k}&=\phi_{n+k,0}+\alpha_{n+2k-1}e_1^-+\alpha_{n+2k}e_2^-\\ &\quad+\sum_{j=1}^{n+2m}\alpha_j(U_{n+2k-1,j}e_1^++U_{n+2k,j}e_2^+), \quad \text{for } k\in\{1,...,m\}.\end{aligned} \tag{2.16}$$

Since the $\phi_{k,0}$ vanish on the boundaries, using (2.4) we obtain by straightforward manipulations with the notation from (2.11) and (2.9)

$$\left.\begin{aligned}\Psi_k^{(1)}&=\alpha_k+\sum_{j=1}^{m+1}\mathscr{U}_{k,j}\alpha_j\\ \Psi_k^{(2)}&=\frac{1+i}{\sqrt{2}}\alpha_k+\frac{i-1}{\sqrt{2}}\sum_{j=1}^{m+1}\mathscr{U}_{k,j}\alpha_j\end{aligned}\right\} \quad k\in\{1,...,m+1\}. \tag{2.17}$$

This is equivalent to

$$\left.\begin{aligned}\Phi_k^{(1)}&=\alpha_k+\sum_{j=1}^{m+1}\mathscr{U}_{k,j}\alpha_j \qquad \text{(i)}\\ \alpha_k&=\frac{1-i}{2}\Psi_k^{(1)}+\frac{1}{\sqrt{2}}\Psi_k^{(2)} \qquad \text{(ii)}\end{aligned}\right\} \quad k\in\{1,...,m+1\}. \tag{2.18}$$

Inserting of (ii) into (i) yields (2.14). If we assume (2.14) we can choose $\alpha_k=[(1-i)/2]\Psi_k^{(1)}+(1/\sqrt{2})\Psi_k^{(2)}$ in order to obtain (2.18), which in turn together with the first condition in (2.14) is equivalent to (2.13) by the considerations above. □

C. Scattering states and eigenstates

The scattering states of D_U are solutions of

$$D\phi=E\phi, \quad E>1, \tag{2.19}$$

where D is now regarded as an operator on the set of absolutely continuous, locally square-integrable functions fulfilling (2.14). The solutions of (2.19) are

$$\begin{aligned}\phi&=\bigoplus_{j=1}^{n+m}\left[\beta_j e^{ikx}\otimes\begin{pmatrix}1\\k_0\end{pmatrix}+\gamma_j e^{-ikx}\otimes\begin{pmatrix}1\\-k_0\end{pmatrix}\right]\\ &\beta_j,\gamma_j\in\mathbb{C}, \quad k=\sqrt{E^2-1}, \quad k_0=\sqrt{(E-1)/(E+1)}.\end{aligned} \tag{2.20}$$

In order to describe an electron approaching on the sth halfline we set $\gamma_s=1$, and $\gamma_j=0$, for $j\in\{1,...,n\}\setminus\{s\}$. Writing

$$\mathbf{b}_1=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}, \quad \mathbf{e}_s=\begin{pmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix}\begin{matrix}\}s-1\\ \\ \}n-s,\end{matrix} \tag{2.21}$$

$$\mathbf{b}_{l+1}=\begin{pmatrix}\beta_{l+n}\\ \gamma_{l+n}\end{pmatrix} \quad \text{for } l\in\{1,...,m\},$$

$$\Gamma(E)=\begin{pmatrix}e^{-ik} & e^{ik}\\ e^{ik} & e^{-ik}\end{pmatrix},$$

and introducing the E-dependent matrices

$$\begin{aligned}\Lambda(E)&=(1/\epsilon)[\sqrt{2}k_0\sigma_3\Delta^{-1}\Gamma\sigma_3-(1+i)\Delta^{-1}\Gamma]\\ &=2\epsilon\begin{pmatrix}\kappa\sinh(\sqrt{2}-ik) & -\eta^*\sinh(\sqrt{2}+ik)\\ -\eta^*\sinh(\sqrt{2}+ik) & \kappa\sinh(\sqrt{2}-ik)\end{pmatrix},\\ \Omega(E)&=(1/\epsilon)[\sqrt{2}k_0\sigma_3\Delta^{-1}\Gamma\sigma_3+(1-i)\Delta^{-1}\Gamma]\\ &=2\epsilon\begin{pmatrix}\eta\sinh(\sqrt{2}-ik) & -\kappa^*\sinh(\sqrt{2}+ik)\\ -\kappa^*\sinh(\sqrt{2}+ik) & \eta\sinh(\sqrt{2}-ik)\end{pmatrix},\end{aligned} \tag{2.22}$$

where

$$\kappa(E)=\sqrt{2}k_0-1-i, \quad \eta(E)=\sqrt{2}k_0+1-i, \tag{2.23}$$

we obtain for the boundary conditions in this situation

$$\begin{aligned}[\eta^*+\kappa^*\mathscr{U}_{1,1}]\mathbf{e}_s&=[\kappa+\eta\mathscr{U}_{1,1}]\mathbf{b}_1+\sum_{j=1}^{m}\mathscr{U}_{1,j+1}\Omega\mathbf{b}_{j+1},\\ \kappa^*\mathscr{U}_{l,1}\mathbf{e}_s&=\eta\mathscr{U}_{l,1}\mathbf{b}_1+\Lambda\mathbf{b}_l+\sum_{j=1}^{m}\mathscr{U}_{l,j+1}\Omega\mathbf{b}_{j+1}, \quad \text{for } l\in\{2,...,m+1\}.\end{aligned} \tag{2.24}$$

Considering the eigenvalue problem we distinguish the three cases

(a) $|E|>1$,

(b) $|E|=1$,

(c) $|E|<1$.

(a) The general solution of the eigenvalue equation has been given in (2.20). Since in contrast to the case treated before the solutions must now be eigenvectors, we must take

$$\mathbf{b}_1=0 \text{ and } \gamma_j=0, \quad \text{for } j\in\{1,...,n\}. \tag{2.25}$$

The boundary conditions (2.14) turn into

$$\begin{aligned}0&=\sum_{j=1}^{m}\mathscr{U}_{1,j+1}\Omega\mathbf{b}_{j+1},\\ 0&=\Lambda\mathbf{b}_l+\sum_{j=1}^{m}\mathscr{U}_{l,j+1}\Omega\mathbf{b}_{j+1}, \quad l\in\{2,...,m+1\}.\end{aligned} \tag{2.26}$$

The solvability of this system of $2m+n$ equations for $2m$ unknown numbers β_{j+n} and γ_{j+n}, occurring in the $\mathbf{b}_{j+1}$ [cf. (2.21)] is the condition for possible eigenvalues $|E|>1$.

(b) The solutions of the eigenvalue equation are

$$E=-1,\quad \phi=\oplus_{j=1}^{m+n}\beta_j\begin{pmatrix}0\\1\end{pmatrix}+\gamma_j\begin{pmatrix}1\\-2ix\end{pmatrix},$$
$$E=+1,\quad \phi=\oplus_{j=1}^{m+n}\beta_j\begin{pmatrix}1\\0\end{pmatrix}+\gamma_j\begin{pmatrix}2ix\\1\end{pmatrix}. \tag{2.27}$$

As before we must take $\beta_j=\gamma_j=0$ for $j\in\{1,\ldots,n\}$. Setting

$$\Lambda_1=\sigma_3\Delta^{-1}\begin{pmatrix}0&1\\0&1\end{pmatrix}-\frac{1+i}{\sqrt{2}}\Delta^{-1}\begin{pmatrix}1&-2i\\1&2i\end{pmatrix},$$
$$\Omega_1=\sigma_3\Delta^{-1}\begin{pmatrix}0&1\\0&1\end{pmatrix}+\frac{1-i}{\sqrt{2}}\Delta^{-1}\begin{pmatrix}1&-2i\\1&2i\end{pmatrix},$$
$$\Lambda_{-1}=\sigma_3\Delta^{-1}\begin{pmatrix}1&2i\\1&-2i\end{pmatrix}-\frac{1+i}{\sqrt{2}}\Delta^{-1}\begin{pmatrix}0&1\\0&1\end{pmatrix}, \tag{2.28}$$
$$\Omega_{-1}=\sigma_3\Delta^{-1}\begin{pmatrix}1&2i\\1&-2i\end{pmatrix}+\frac{1-i}{\sqrt{2}}\Delta^{-1}\begin{pmatrix}0&1\\0&1\end{pmatrix},$$

we obtain again the form (2.26) for the boundary conditions. Formally all energies $|E|\geqslant 1$ can be treated in the same way. We remark however, that $\Lambda_{\pm 1}$ and $\Omega_{\pm 1}$ are not the limits of the matrix functions in (2.22) for $E\to\pm 1$.

(c) The appropriate form of (2.20) is now

$$\phi=\oplus_{j=1}^{n+m}\left[\beta_j e^{\bar{k}x}\otimes\begin{pmatrix}1\\-i\bar{k}_0\end{pmatrix}+\gamma_j e^{-\bar{k}x}\otimes\begin{pmatrix}1\\i\bar{k}_0\end{pmatrix}\right], \tag{2.29}$$

with

$$\beta_j,\gamma_j\in\mathbb{C}$$

and

$$\bar{k}=\sqrt{1-E^2},\quad \bar{k}_0=\sqrt{(1-E)/(1+E)}\in\mathbb{R}^+.$$

Now we must put $\beta_j=0$ for $j\in\{1,\ldots,n\}$. Adapting (2.21) and (2.23) to the case considered,

$$\gamma_1=\begin{pmatrix}\gamma_1\\\vdots\\\gamma_n\end{pmatrix},\quad \mathbf{b}_{l+1}=\begin{pmatrix}\beta_{l+n}\\\gamma_{l+n}\end{pmatrix},$$
$$\bar{\Gamma}(E)=\begin{pmatrix}e^{-\bar{k}}&e^{\bar{k}}\\e^{\bar{k}}&e^{-\bar{k}}\end{pmatrix}, \tag{2.30}$$

$\lambda=-1+(\sqrt{2}\cdot\bar{k}_0-1)i$ and $\mu=1+(\sqrt{2}\cdot\bar{k}_0-1)i$, and setting

$$\bar{\Lambda}(E)=(1/\epsilon)\left[-i\sqrt{2}\cdot\bar{k}_0\sigma_3\Delta^{-1}\Gamma\sigma_3-(1+i)\Delta^{-1}\Gamma\right]=2\epsilon\begin{pmatrix}-(2i+\mu)\sinh(\sqrt{2}-\bar{k})&-\lambda\sinh(\sqrt{2}+\bar{k})\\\lambda\sinh(\sqrt{2}+\bar{k})&-(2i+\mu)\sinh(\sqrt{2}-\bar{k})\end{pmatrix}, \tag{2.31}$$
$$\bar{\Omega}(E)=(1/\epsilon)\left[-i\sqrt{2}\cdot\bar{k}_0\sigma_3\Delta^{-1}\Gamma\sigma_3+(1-i)\Delta^{-1}\Gamma\right]=2\epsilon\begin{pmatrix}-(2i+\lambda)\sinh(\sqrt{2}-\bar{k})&\mu\sinh(\sqrt{2}+\bar{k})\\\mu\sinh(\sqrt{2}+\bar{k})&-(2i+\lambda)\sinh(\sqrt{2}-\bar{k})\end{pmatrix},$$

we obtain for the boundary conditions

$$0=(\lambda+\mu\mathscr{U}_{1,1})\gamma_1+\sum_{j=1}^{m}\mathscr{U}_{1,j+1}\bar{\Omega}\mathbf{b}_{j+1},$$
$$0=\mu\mathscr{U}_{l,1})\gamma_1+\bar{\Lambda}\mathbf{b}_l+\sum_{j=1}^{m}\mathscr{U}_{l,j+1}\bar{\Omega}\mathbf{b}_{j+1}, \tag{2.32}$$
$$\text{for } l\in\{2,\ldots,m+1\}.$$

III. EXAMPLES

A. Bundle of half-lines ($m=0$)

In this case the structure consists of n half-lines. The matrix U in (2.12) reduces to $\mathscr{U}_{1,1}$ and Eqs. (2.24), (2.26), and (2.32) simplify, allowing the deduction of specific statements.

Eigenvalues E with $|E|\geqslant 1$ cannot occur, because the corresponding eigenfunctions must vanish on all half-lines. In the case $|E|<1$ from (2.32) only

$$(\lambda+\mu U)\gamma_1=0 \tag{3.1}$$

is left. Because of $-\lambda=\mu^*$ and $\mu\neq 0$ [cf. (2.30)] we have the compatibility

$$|-\lambda/\mu|=1, \tag{3.2}$$

with the eigenvalue condition (3.1). Assuming that a matrix U defining a particular extension D_U is given, we establish the connection between the spectrum of U and the energy eigenvalues. For this purpose we analyze the function $-(\lambda/\mu)(E)$. From (2.30) we obtain, setting $\tau=\sqrt{2}\cdot\bar{k}_0-1$,

$$-\frac{\lambda}{\mu}=\frac{1-\tau^2}{1+\tau^2}+i\frac{-2\tau}{1+\tau^2}=f(\tau)+ig(\tau). \tag{3.3}$$

From $\bar{k}_0\in\mathbb{R}^+$ it follows that $\mathscr{D}(f)=\mathscr{D}(g)=(-1,\infty)$. The argument $\varphi(\tau)$ of $-\lambda/\mu$ now maps $(-1,\infty)$ in the following way into $[0,2\pi)$:

$$\varphi(-1,0]=\left[0,\frac{\pi}{2}\right),$$
$$\varphi(0,1]=\left[\frac{3\pi}{2},2\pi\right),\quad \varphi(1,\infty)=\left(\pi,\frac{3\pi}{2}\right). \tag{3.4}$$

Thus we obtain the following.

Theorem 3.1: To each eigenvalue of U not lying in the second quadrant is associated exactly one energy eigenvalue, given by the inverse of the function $\varphi(\tau(E))$:

$$E=(\sin\varphi+\cos\varphi)/(2-\sin\varphi+\cos\varphi). \tag{3.5}$$

Scattering states: Using that $\eta\neq 0$ for all $|E|>1$, (2.24) can be written as

$$(1/\eta)[\eta^*+\kappa^*U]\mathbf{e}_s=[\kappa/\eta+U]\mathbf{b}_1. \tag{3.6}$$

From (2.23) follows $|\kappa/\eta|\neq 1$ for all $|E|>1$, so $[\kappa/\eta+U]$ is always invertible, and (3.6) can be solved for $\mathbf{b}_1$:

$$\mathbf{b}_1=[\kappa+\eta U]^{-1}[\eta^*+\kappa^*U]\mathbf{e}_s. \tag{3.7}$$

Using

$$(\zeta + U)^{-1}(\xi + U) = 1 + (\xi - \zeta)(\zeta + U)^{-1}, \tag{3.8}$$

we get

$$\mathbf{b}_1 = (1/\eta)[\kappa^* + (|\eta|^2 - |\kappa|^2)(\kappa + \eta U)^{-1}]\mathbf{e}_s . \tag{3.9}$$

As a check we note that the matrix in (3.9) is unitary; so the sum of squared moduli of the reflection coefficient $|\mathbf{b}_{1,s}|^2$ and of the transition coefficients $|\mathbf{b}_{1,j}|^2$, ($j \in \{1,\ldots,n\} \setminus \{s\}$) equals 1, thus satisfying the elasticity condition for the scattering.

B. Defect point ($n=2, m=0$)

Two half-lines connected by a transition point can be regarded as a simple model for a defect in a polymer chain. The transition between the two half-lines is described by a unitary (2×2) matrix that is left in (2.12). Choosing the parametrization

$$U = e^{i\xi}\begin{pmatrix} \cos\beta e^{i(\alpha+\delta)} & \sin\beta e^{-i(\alpha-\delta)} \\ -\sin\beta e^{i(\alpha-\delta)} & \cos\beta e^{-i(\alpha+\delta)} \end{pmatrix}, \tag{3.10}$$

we obtain, introducing

$$\begin{aligned} q &= (|\eta|^2 - |\kappa|^2)/\det(\kappa + \eta U) \\ &= (|\eta|^2 - |\kappa|^2)/(\kappa^2 + 2\kappa\eta e^{i\xi}\cos\beta\cos(\alpha+\delta) \\ &\quad + \eta^2 e^{2i\xi}) \end{aligned} \tag{3.11}$$

for the scattering coefficients:

$$\begin{aligned} s=1,\quad \mathbf{b}_{1,1} &= \frac{1}{\eta}\left\{\kappa^*\begin{pmatrix}1\\0\end{pmatrix} + q\left[\kappa\begin{pmatrix}1\\0\end{pmatrix}\right.\right. \\ &\quad \left.\left. + \eta e^{i\xi}\begin{pmatrix}\cos\beta e^{-i(\alpha+\delta)}\\ \sin\beta e^{i(\alpha-\delta)}\end{pmatrix}\right]\right\}, \\ s=2,\quad \mathbf{b}_{1,2} &= \frac{1}{\eta}\left\{\kappa^*\begin{pmatrix}0\\1\end{pmatrix} + q\left[\kappa\begin{pmatrix}0\\1\end{pmatrix}\right.\right. \\ &\quad \left.\left. + \eta e^{i\xi}\begin{pmatrix}-\sin\beta e^{-i(\alpha-\delta)}\\ \cos\beta e^{i(\alpha+\delta)}\end{pmatrix}\right]\right\}. \end{aligned} \tag{3.12}$$

If we specialize by requiring that the two spinor components should be continuous at the defect point, we introduce a kind of point interaction that cannot be described by a δ interaction. In terms of (2.11) the continuity condition becomes

$$\Psi_1^{(1)} = \varphi_1\begin{pmatrix}1\\1\end{pmatrix}, \quad \Psi_1^{(2)} = \varphi_2\begin{pmatrix}1\\1\end{pmatrix}, \quad \varphi_j \in \mathbb{C}, \tag{3.13}$$

and putting $u_j = U_{j,1} + U_{j,2}$, $j \in \{1,2\}$, (2.14) gives

$$\varphi_2\begin{pmatrix}1+u_1\\1+u_2\end{pmatrix} = \frac{1}{\sqrt{2}}\varphi_1\begin{pmatrix}1+i+(i-1)u_1\\1+i+(i-1)u_2\end{pmatrix}. \tag{3.14}$$

Since at least one of the φ_j can be chosen different from zero (otherwise D_U would not be self-adjoint,), the two equations (3.14) coincide, so $u_1 = u_2$. This yields with (3.10) the equations

$$\cos\beta\sin(\alpha+\delta) = 0 \quad \text{and} \quad \sin\beta\cos(\alpha-\delta) = 0, \tag{3.15}$$

and for U the form

$$U = e^{i\xi}\begin{pmatrix}\cos\beta & i\sin\beta\\ i\sin\beta & \cos\beta\end{pmatrix}, \tag{3.16}$$

(3.12) specializes to

$$\begin{aligned} s=1,\quad \mathbf{b}_{1,1} &= (1/\eta)\left\{k^*\begin{pmatrix}1\\0\end{pmatrix} + q\left[\kappa\begin{pmatrix}1\\0\end{pmatrix}\right.\right. \\ &\quad \left.\left. + \eta e^{i\xi}\begin{pmatrix}\cos\beta\\ -i\sin\beta\end{pmatrix}\right]\right\}, \\ s=2,\quad \mathbf{b}_{1,2} &= \sigma_1\mathbf{b}_{1,1} . \end{aligned} \tag{3.17}$$

Remark: If we demand invariance of the situation with respect to an interchange of the half-lines instead of continuity, we obtain the same results for U, so both requirements are equivalent.

Finally we obtain for the transmission probability T

$$T = \frac{\sqrt{2}k_0\sin^2\beta}{|(1+2e^{i\xi}\cos\beta + e^{2i\xi})k_0^2 + \sqrt{2}[(1-i)e^{2i\xi} - i\cos\beta e^{i\xi} - (1+i)]k_0 + (i - e^{i\xi}\cos\beta - ie^{2i\xi})|^2}. \tag{3.18}$$

The reflection probability R is then given by $R = 1 - T$. Thus this example is explicitly solvable in the sense, that starting with some choice for the extension D_U the scattering data can be calculated analytically.

C. Three half-lines ($n=3, m=0$)

We treat the special case in which two of the half-lines are connected symmetrically to each other, whereas the third one has different transition parameters. Therefore we choose for U

$$U = \begin{pmatrix}a & b & c\\ b & a & c\\ c & c & d\end{pmatrix}, \tag{3.19}$$

which can be parametrized by

$$U = e^{i\xi/2}\begin{pmatrix} e^{i\alpha} - \sin\beta e^{i(\xi-\gamma)} & -e^{i\alpha} - \sin\beta e^{i(\xi-\gamma)} & \sqrt{2}\cos\beta \\ -e^{i\alpha} - \sin\beta e^{i(\xi-\gamma)} & e^{i\alpha} - \sin\beta e^{i(\xi-\gamma)} & \sqrt{2}\cos\beta \\ \sqrt{2}\cos\beta & \sqrt{2}\cos\beta & 2\sin\beta e^{i(\gamma-\xi)} \end{pmatrix}. \tag{3.20}$$

Now we make the ansatz $(\kappa+\eta U)^{-1}=c_1+c_2U+c_3U^*$. Inserting into the identity $(\kappa+\eta U)^{-1}(\kappa+\eta U)=1$ yields the c_j. With the abbreviations

$$C^{-1}(E)=\eta^2(2c^2-d(a+b))-\kappa\eta(a+b+d)-\kappa^2, \tag{3.21}$$

$$X(E)=\frac{\eta^2(c^2-ad)-\kappa\eta(a+d)-\kappa^2}{\kappa+\eta(a-d)}$$

and

$$Y(E)=\frac{\eta^2(bd-c^2)+\kappa\eta b}{\kappa+\eta(a-d)},$$

we obtain the result

$$(\kappa+\eta U)^{-1}=C(E)\begin{pmatrix}X(E) & Y(E) & \eta c\\ Y(E) & X(E) & \eta c\\ \eta c & \eta c & -\kappa-\eta(a+b)\end{pmatrix}, \tag{3.22}$$

and finally for the scattering coefficients

$$s=1,\quad \mathbf{b}_{1,1}=\frac{1}{\eta}\left\{\kappa^*\begin{pmatrix}1\\0\\0\end{pmatrix}+(|\eta|^2-|\kappa|^2)\times C(E)\begin{pmatrix}X(E)\\Y(E)\\\eta c\end{pmatrix}\right\}, \tag{3.23}$$

$$s=2,\quad \mathbf{b}_{1,2}=\frac{1}{\eta}\left\{\kappa^*\begin{pmatrix}0\\1\\0\end{pmatrix}+(|\eta|^2-|\kappa|^2)\times C(E)\begin{pmatrix}Y(E)\\X(E)\\\eta c\end{pmatrix}\right\}, \tag{3.24}$$

$$s=3\quad \mathbf{b}_{1,3}=\frac{1}{\eta}\left\{\kappa^*\begin{pmatrix}0\\0\\1\end{pmatrix}+(|\eta|^2-|\kappa|^2)\times C(E)\begin{pmatrix}\eta c\\\eta c\\-\kappa-\eta(a+b)\end{pmatrix}\right\}. \tag{3.25}$$

By this result it has been shown that even this case, with remarkably greater complexity as the foregoing one, gives us an explicitly solvable model at least with respect to scattering theory. It was not possible, however, to get manageable expressions for the eigenvalues of U in terms of the parameters introduced in (3.20), so we omit the treatment of the energy eigenvalues.

D. Eye ($n=m=2$)

In this case the geometry is given by two half-lines that are connected by two intervals joined at their end points. In order to avoid obstructive complexity we demand symmetry with respect to interchange as well of the half-lines as of the intervals. So we put

$$U=\begin{pmatrix}aE & A & A\\ B & C & D\\ B & D & C\end{pmatrix}, \tag{3.26}$$

where $a\in\mathbb{C}$, $|a|<1$, and $A,B,C,D\in\mathscr{B}(\mathbb{C}^2,\mathbb{C}^2)$. The conditions of unitarity of U expressed in terms of the block elements in (3.26) can be shown to be equivalent to

$$\sqrt{\frac{2}{1-|a|^2}}A,\quad \sqrt{\frac{2}{1-|a|^2}}B\in U(2),$$
$$A=-\frac{1}{a^*}B^*(C+D), \tag{3.27}$$

and

$$(1/|a|)(C+D),\quad C-D\in U(2).$$

Having in mind (3.27) the boundary conditions (2.26) for energy eigenvalues $|E|\geqslant 1$ reduce after two steps to

$$\begin{aligned}&\Lambda\mathbf{b}_2=-\Lambda\mathbf{b}_3,\\ &\Omega\mathbf{b}_2=-\Omega\mathbf{b}_3,\\ &(\Lambda+(C-D)\Omega)\mathbf{b}_2=0,\\ &(\Lambda+(C-D)\Omega)\mathbf{b}_3=0,\end{aligned} \tag{3.28}$$

while for energies $|E|<$ (2.32) yields

$$\begin{aligned}&\gamma_1=-[1/(\lambda+\mu a)]A\bar{\Omega}(\mathbf{b}_2+\mathbf{b}_3),\\ &0=[\bar{\Lambda}+(C-D)\bar{\Omega}](\mathbf{b}_2-\mathbf{b}_3),\\ &0=[\bar{\Lambda}+(C+D-[2\mu/(\lambda+\mu a)]BA)\bar{\Omega}](\mathbf{b}_2+\mathbf{b}_3).\end{aligned} \tag{3.29}$$

As $|\lambda/\mu|=1$ and $|a|<1$ no problems in the denominator occur. So in order to obtain nontrivial solutions we have to demand

$$|E|\geqslant 1,\quad \det[\Lambda+(C-D)\Omega]=0, \tag{3.30}$$

$$|E|<1:\quad \begin{cases}0=\det[\bar{\Lambda}+(C-D)\bar{\Omega}],\\ \text{or}\\ 0=\det[\bar{\Lambda}+(C+D-[2\mu/(\lambda+\mu a)]BA)\bar{\Omega}]\end{cases} \tag{3.31}$$

as can be shown easily.

Scattering states (for energies staying away from eigenvalues): In this case, (3.30) not being valid, we obtain from (2.24) the following conditions:

$$\begin{aligned}&\text{(i)}\quad \frac{|\kappa|^2-|\eta|^2}{\kappa+\eta a}B\mathbf{e}_s=\left[\Lambda+\frac{\kappa a^*+\eta}{\kappa+\eta a}\frac{1}{a^*}(C+D)\Omega\right]\mathbf{b}_2,\\ &\text{(ii)}\quad \mathbf{b}_1=-1/(\kappa+\eta a)\times[(\eta^*+\kappa^* a)\mathbf{e}_s-2A\Omega\mathbf{b}_2],\\ &\text{(iii)}\quad \mathbf{b}_3=\mathbf{b}_2,\end{aligned} \tag{3.32}$$

in the special case that $\kappa+\eta a\neq 0$. In order to get the scattering coefficients $\mathbf{b}_1$ from the system (3.32), (i) must be solvable. If $\kappa+\eta a=0$, (2.24) yields

$$\begin{aligned}&\text{(iv)}\quad \mathbf{b}_1=\frac{\kappa^*}{\eta}\mathbf{e}_s-[2/(1-|a|^2)\eta]\times B^*[\Lambda+(C+D)\Omega\mathbf{b}_2],\\ &\text{(v)}\quad [\eta^*+\kappa^* a]\mathbf{e}_s=2A\Omega\mathbf{b}_2,\end{aligned} \tag{3.33}$$

where v must be solvable to get scattering data.

Note: For the regular m-gon the condition for the existence of eigenvalues has been considered. The condition of the vanishing of the determinant of a $2m \times 2m$ matrix following from the general results in Sec. II C has been reduced to the condition of the vanishing of the determinants of certain 2×2 matrices. Because of the length of the deductions giving this result they are not exposed in this place.

ACKNOWLEDGMENT

The authors are indebted to the referee for drawing their attention to Ref. 7.

[1]J. D. Bishop, J. C. Licini, and G. J. Dolan, Lett. Appl. Phys. **46**, 1000 (1985).
[2]P. Exner and P. Šeba, "Quantum motion on a branching graph," preprint JINR E2-87-213, Dubna, 1987.
[3]P. Exner and P. Šeba, Phys. Lett. A **128**, 493 (1988).
[4]P. Exner and P. Šeba, J. Math. Phys. **28**, 386 (1987).
[5]P. Exner and P. Šeba, Lett. Math. Phys. **12**, 193 (1986).
[6]P. Exner and P. Šeba, Czech. J. Phys. B **38**, 1 (1988).
[7]N. J. Gerasimenko and B. S. Pavlov, Teor. Mat. Fiz. **74**, 345 (1988).
[8]W. B. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).
[9]J. Weidmann, *Lineare Operatoren in Hilberträumen* (Teubner, Stuttgart, 1976).
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Fourier Analysis, Self-Adjointness* (Academic, New York, 1975), Vol. II.

Nonspurious harmonic oscillator states in single particle coordinates

Akiva Novoselsky
Department of Nuclear Physics, The Weizmann Institute of Science, Rehovot 76100, Israel
Jacob Katriel
Department of Chemistry, Technion–Israel Institute of Technology, Haifa 32000, Israel

(Received 8 August 1989; accepted for publication 13 December 1989)

A procedure for the construction of nonspurious harmonic oscillator wave functions with arbitrary permutational symmetry has recently been proposed. The resulting wave functions are expressed in terms of normalized Jacobi coordinates, and involve a new type of harmonic oscillator coefficients of fractional parentage. A simple algorithm to transform these states from the Jacobi coordinates to the single particle coordinates is presented. This is a generalization to an arbitrary number of particles of the harmonic oscillator transformation from center-of-mass and relative coordinates to single particle coordinates.

I. INTRODUCTION

Harmonic oscillator wave functions have been widely used in molecular, atomic, and nuclear physics, and recently also in nonrelativistic quark calculations.[1] In all these applications the eigenvectors of a translationally invariant Hamiltonian are evaluated in terms of harmonic oscillator eigenstates. The harmonic oscillator states used in these calculations should have a separable dependence on the trivial center-of-mass motion: Spurious states, in which the center-of-mass is excited, must be eliminated.

In order to construct nonspurious states of n identical isotropic three-dimensional harmonic oscillators, we must use a set of coordinates consisting of the center-of-mass coordinate and $n-1$ internal coordinates. Among the various sets of coordinates satisfying this requirement, the normalized Jacobi coordinates

$$\boldsymbol{\rho}_1 = \mathbf{R} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbf{r}_j;$$

$$\boldsymbol{\rho}_i = \sqrt{\frac{i-1}{i}} \left(\mathbf{r}_i - \frac{1}{i-1} \sum_{j=1}^{i-1} \mathbf{r}_j \right), \quad i = 2,3,\ldots,n \qquad (1)$$

were found to be preferable because each internal coordinate $\boldsymbol{\rho}_i$ $i = 2,\ldots,n$ depends on the first i single particle coordinates only. This property enables the formulation of a recursive procedure for constructing the harmonic oscillator nonspurious states.[2]

In calculations involving clusters of identical particles, it is necessary to construct wave functions that belong to well-defined irreducible representations (irreps) of the symmetric group S_n. Interesting examples include the study of the characteristics of cluster knock-out reactions in nuclei (such as spectroscopic factors and distributions of momenta)[3] and the nonrelativistic description of few nucleon systems in terms of clusters of three quarks.[1]

The quantum states in these examples are labeled by multiple angular momentum quantum numbers (i.e., L–S, L–S–T, L–S color and flavor). In these cases it is very useful[4] to factor the space of the states into products of subspaces, one for each angular momentum quantum number. Wave functions belonging to the various irreps of S_n should be constructed within each subspace and ultimately coupled to form a totally antisymmetric state.

The harmonic oscillator center-of-mass coordinate and, consequently, the corresponding wave function, are totally symmetric with respect to permutations of the particle coordinates. On the other hand, the internal coordinates do not have simple symmetry properties with respect to permutation of particle indices.

The recursive algorithm recently proposed for constructing internal wave functions belonging to irreps of S_n (see Ref. 2) starts from $(n-1)$-particle internal states belonging to a given irrep Γ_{n-1} of S_{n-1}. A set of n-particle internal states with resultant angular momentum Λ_n and energy $\hbar\omega(\epsilon_n + \frac{3}{2}(n-1))$ is constructed. Finally, the two-cycle class operator of S_n is diagonalized in this basis. Although this operator is the sum of all the $(n-1)n/2$ transpositions $(i,i')i<i'=2,\ldots,n$, it was shown that only the matrix element of the transposition $(n-1,n)$ has to be evaluated. After diagonalization, the eigenvalues uniquely identify the n-particle irrep of S_n, and the eigenvectors are the appropriate harmonic oscillator coefficients of fractional parentage (hocfps).[2]

The resulting expression for the n-particle internal wave function, symmetry adapted to S_n, can be written in terms of the hocfps in the form[2]

$$|Y_n\Lambda_n\epsilon_n\alpha_n;\boldsymbol{\rho}_2\boldsymbol{\rho}_3\cdots\boldsymbol{\rho}_n\rangle = \sum_{\substack{\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1}N_nL_n \\ (\epsilon_n=\epsilon_{n-1}+2N_n+L_n)}} [\,Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1}N_nL_n\Lambda_n|\}Y_n\Lambda_n\epsilon_n\alpha_n\,]$$
$$\times [\,|Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1};\boldsymbol{\rho}_2\boldsymbol{\rho}_3\cdots\boldsymbol{\rho}_{n-1}\rangle|N_nL_n;\boldsymbol{\rho}_n\rangle\,]^{\Lambda_n}, \qquad (2)$$

 0022-2488/90/051164-03$03.00

where Y_n is a Yamanouchi symbol, which is equivalent to the set of good quantum numbers $\Gamma_2\Gamma_3\cdots\Gamma_n$ labeling the irreps of $S_2S_3\cdots S_n$. It satisfies $Y_n = Y_{n-1}\Gamma_n$. Here N_n and L_n are the individual harmonic oscillator radial and angular quantum numbers of the nth degree of freedom (dof), in Jacobi coordinates. The symbol α_n takes care of remaining degeneracies in the n-particle internal states. The sum in Eq. (2) is over the coupled products of the states of the nth dof and the $(n-1)$-particle internal states consistent with the energy relation written under the summation, and with the angular momentum coupling relation $\Lambda_{n-1} + L_n = \Lambda_n$.

The n-particle harmonic oscillator wave function is obtained by coupling the center-of-mass wave function to the internal wave function, obtaining

$$[|Y_n\Lambda_n\epsilon_n\alpha_n;\boldsymbol{\rho}_2\boldsymbol{\rho}_3\cdots\boldsymbol{\rho}_n\rangle|N_1L_1;\boldsymbol{\rho}_1\rangle]^{\mathscr{L}_n}. \tag{3}$$

The permutational symmetry is not affected by the center-of-mass state, which is totally symmetric. The total angular momentum $\mathscr{L}_n$ is different from Λ_n when $L_1 \neq 0$. For n-particle nonspurious states, the center-of-mass wave function is always in the ground state and therefore $L_1 = 0$.

II. TRANSFORMATION TO SINGLE-PARTICLE COORDINATES

The n-particle harmonic oscillator states [Eq. (3)] are expressed in terms of the normalized Jacobi coordinates [Eq. (1)]. However, in many calculations in atomic and nuclear physics it is useful to have expressions for the wave functions in terms of the single particle coordinates. This is particularly important when the harmonic oscillator states are used as a basis set in a calculation involving nonharmonic potentials, which are not easily expressible in Jacobi coordinates.

We propose a recursive algorithm to transform an n-particle harmonic oscillator state with a well-defined permutational symmetry from Jacobi into single particle coordinates. This method is based on using the hocfps defined in Eq. (2). The internal wave function in the state (3) is rewritten as in Eq. (2), obtaining for the state (3) the expression

$$\sum_{\substack{\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1}N_nL_n \\ (\epsilon_n=\epsilon_{n-1}+2N_n+L_n)}} [Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1}N_nL_n\Lambda_n|\}\, Y_n\Lambda_n\epsilon_n\alpha_n]$$
$$\times[[|Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1};\boldsymbol{\rho}_2\boldsymbol{\rho}_3\cdots\boldsymbol{\rho}_{n-1}\rangle\,|N_nL_n;\boldsymbol{\rho}_n\rangle]^{\Lambda_n}|N_1L_1;\boldsymbol{\rho}_1\rangle]^{\mathscr{L}_n}. \tag{4}$$

In order to separate the wave functions of the center-of-mass ($\boldsymbol{\rho}_1$) and of the nth dof ($\boldsymbol{\rho}_n$) from the $(n-1)$-particle internal state we use the change of coupling transformation,[5] obtaining

$$[[|Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1};\boldsymbol{\rho}_2\boldsymbol{\rho}_3\cdots\boldsymbol{\rho}_{n-1}\rangle|N_nL_n;\boldsymbol{\rho}_n\rangle]^{\Lambda_n}|N_1L_1;\boldsymbol{\rho}_1\rangle]^{\mathscr{L}_n}$$
$$=\sum_{\lambda}(-1)^{\Lambda_{n-1}+L_n+L_1+\mathscr{L}_n}\sqrt{(2\Lambda_n+1)(2\lambda+1)}\begin{Bmatrix}\Lambda_{n-1} & L_n & \Lambda_n\\ L_1 & \mathscr{L}_n & \lambda\end{Bmatrix}$$
$$\times[[|Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1};\boldsymbol{\rho}_2\boldsymbol{\rho}_3\cdots\boldsymbol{\rho}_{n-1}\rangle[|N_nL_n;\boldsymbol{\rho}_n\rangle|N_1L_1;\boldsymbol{\rho}_1\rangle]^{\lambda}]^{\mathscr{L}_n}, \tag{5}$$

where $\Lambda_{n-1} + \lambda = \mathscr{L}_n$.

The two harmonic oscillator functions coupled to a resultant angular momentum $\boldsymbol{\lambda}$, are expressed in terms of the coordinates $\boldsymbol{\rho}_n$ and $\boldsymbol{\rho}_1$ where

$$\boldsymbol{\rho}_1 = (1/\sqrt{n})(\mathbf{r}_1 + \cdots + \mathbf{r}_n),$$
$$\boldsymbol{\rho}_n = \sqrt{\frac{n-1}{n}}\left(\mathbf{r}_n - \frac{1}{n-1}(\mathbf{r}_1 + \cdots + \mathbf{r}_{n-1})\right). \tag{6}$$

These coordinates can be rewritten in terms of the normalized center-of-mass coordinate of the $(n-1)$ particles, $\tilde{\boldsymbol{\rho}}_1 = (1/\sqrt{n-1})(\mathbf{r}_1 + \cdots + \mathbf{r}_{n-1})$, and the nth particle coordinate $\mathbf{r}_n$ as

$$\boldsymbol{\rho}_1 = \sqrt{[(n-1)/n]}\,\tilde{\boldsymbol{\rho}}_1 + (1/\sqrt{n})\mathbf{r}_n,$$
$$\boldsymbol{\rho}_n = -(1/\sqrt{n})\,\tilde{\boldsymbol{\rho}}_1 + \sqrt{[(n-1)/n]}\,\mathbf{r}_n. \tag{7}$$

This transformation is actually a rotation of the coordinates by an angle β where $\cos\beta = \sqrt{n-1/n}$ and $\sin\beta = 1/\sqrt{n}$. Therefore, the rotation of the coordinates $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_n$ by the angle $-\beta$ yields the coordinates

$$\tilde{\boldsymbol{\rho}}_1 = \boldsymbol{\rho}_1\cos\beta - \boldsymbol{\rho}_n\sin\beta,$$
$$\mathbf{r}_n = \boldsymbol{\rho}_1\sin\beta + \boldsymbol{\rho}_n\cos\beta. \tag{8}$$

The harmonic oscillator wave functions expressed in terms of the coordinates $\boldsymbol{\rho}_n$ and $\boldsymbol{\rho}_1$ in Eq. (5) can be transformed into harmonic oscillator wave functions expressed in terms of the coordinates $\tilde{\boldsymbol{\rho}}_1$ and $\mathbf{r}_n$ [Eq. (8)] by using the harmonic oscillator brackets[6]:

$$[|N_nL_n;\boldsymbol{\rho}_n\rangle|N_1L_1;\boldsymbol{\rho}_1\rangle]^{\lambda}$$
$$=\sum_{\tilde{N}_1\tilde{L}_1\tilde{N}_n\tilde{L}_n}\langle(\tilde{N}_1\tilde{L}_1\tilde{N}_n\tilde{L}_n)\lambda|(N_nL_nN_1L_1)\lambda\rangle$$
$$\times[|\tilde{N}_1\tilde{L}_1;\tilde{\boldsymbol{\rho}}_1\rangle|\tilde{N}_n\tilde{L}_n;\mathbf{r}_n\rangle]^{\lambda}. \tag{9}$$

The sum in Eq. (9) is finite because of energy conservation, i.e., $2\tilde{N}_n + \tilde{L}_n + 2\tilde{N}_1 + \tilde{L}_1 = 2N_n + L_n + 2N_1 + L_1$ and of the angular momentum coupling condition $\tilde{L}_n + \tilde{L}_1 = \boldsymbol{\lambda}$.

A useful recursive relation for calculating harmonic oscillator brackets for particles with equal masses was proposed by Moshinsky[7]; extensive numerical tables were prepared by Brody and Moshinsky.[8] The harmonic oscillator

brackets in Eq. (9) involve two particles with different masses. Using group theoretical methods, Gal[9] showed that these brackets can be expressed as a sum over two equal mass transformation brackets, with a phase factor depending on the rotation angle. [Note that β as defined in Eq. (8) is half as large as that introduced in Gal's[9] Eq. (3.14).] Therefore, the calculation of the harmonic oscillator brackets in Eq. (9) is straightforward.

The harmonic oscillator transformation [Eq. (9)] enables us to express the wave functions on the right-hand side of Eq. (5) in terms of the wave functions

$$[|Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1};\rho_2\rho_3\cdots\rho_{n-1}\rangle \times[|\tilde{N}_1\tilde{L}_1;\tilde{\rho}_1\rangle|\tilde{N}_n\tilde{L}_n;\mathbf{r}_n\rangle]^{\lambda}]^{\mathscr{L}_n}. \tag{10}$$

Recalling that the coordinate $\tilde{\rho}_1$ is the normalized center-of-mass coordinate of the $(n-1)$-particle system we note that by separating the wave function of the nth particle in Eq. (10) we obtain an expression involving the $(n-1)$-particle total wave functions. This is carried out by using the change of coupling transformation,[5] which yields for the wave function (10) the expression

$$\sum_{\mathscr{L}_{n-1}} (-1)^{\Lambda_{n-1}+\tilde{L}_1+\tilde{L}_n+\mathscr{L}_n}\sqrt{(2\lambda+1)(2\mathscr{L}_{n-1}+1)} \times\begin{Bmatrix}\Lambda_{n-1} & \tilde{L}_1 & \mathscr{L}_{n-1}\\ \tilde{L}_n & \mathscr{L}_n & \lambda\end{Bmatrix} \times[[|Y_{n-1}\Lambda_{n-1}\epsilon_{n-1}\alpha_{n-1};\rho_2\rho_3\cdots\rho_{n-1}\rangle \times|\tilde{N}_1\tilde{L}_1;\tilde{\rho}_1\rangle]^{\mathscr{L}_{n-1}}|\tilde{N}_n\tilde{L}_n;\mathbf{r}_n\rangle]^{\mathscr{L}_n}. \tag{11}$$

The $(n-1)$-particle state with a total angular momentum $\mathscr{L}_{n-1}$ can be transformed into an expression in terms of $(n-2)$-particle wave functions and the wave function of the $(n-1)$th particle (expressed in terms of $\mathbf{r}_{n-1}$) by applying the same algorithm again. Thus by successive application of the same algorithm $(n-1)$ times we finally transform the nonspurious state (1) from the Jacobi coordinates to the single particle coordinates. This transformation constitutes a generalization of the harmonic oscillator transformation from center-of-mass and relative coordinates to single particle coordinates,[6] to an arbitrary number of particles.

We conclude by pointing out that the complete set of nonspurious harmonic oscillator wavefunctions is obtained most efficiently using a recursive procedure for the construction of the *internal* many-particle state in terms of Jacobi coordinates.[2] On the other hand, the expression for the nonspurious states in single particle coordinates is obtained by means of a recursive procedure starting from the *total* wavefunction in Jacobi coordinates, which includes the (totally symmetric) center-of-mass motion.

ACKNOWLEDGMENTS

Research by J.K. was supported by the Technion V.P.R. fund—Smoler Research Fund and the fund for the promotion of research at the Technion.

[1] M. Oka and K. Yazaki, "Baryon Baryon Interaction from Quark Model Viewpoint," in *Quarks and Nuclei*, edited by W. Weise (World Scientific, Singapore, 1985).
[2] A. Novoselsky and J. Katriel, Ann. Phys. (NY) **196**, 135 (1989).
[3] A. Arima, H. Horiuchi, K. Kubodera, and N. Takigawa, "Clustering in Light Nuclei," in *Advances in Nuclear Physics*, edited by M. Baranger and E. Vogt (Plenum, New York, 1972), Vol. 5.
[4] J.-Q. Chen, A. Novoselsky, M. Vallières, and R. Gilmore, Phys. Rev. C **39**, 1088 (1989).
[5] A. De-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[6] I. Talmi, Helv. Phys. Acta **25**, 185 (1952).
[7] M. Moshinsky, Nucl. Phys. **13**, 104 (1959).
[8] T. A. Brody and M. Moshinsky, *Tables of Transformation Brackets* (Gordon and Breach, New York, 1967).
[9] A. Gal, Ann. Phys. (NY) **49**, 341 (1968).

Generalized polyspheroidal periodic functions and the quantum inverse scattering method

Vadim B. Kuznetsov
Department of Mathematical and Computational Physics, Institute of Physics, Leningrad State University, Leningrad 198904, USSR

(Received 6 July 1989; accepted for publication 22 November 1989)

Bounded and π-periodic eigenfunctions of the operator $d^2/dv^2 + [2(\mu - \nu + (\mu + \nu + 1)\cos 2v)/\sin 2v]\,(d/dv) + 8\xi \cos 2v + 4\gamma^2 \sin^2 2v$ are studied. These solutions are new special functions, particular cases of which will be the spheroidal, the Coulomb spheroidal, and the hyperspheroidal harmonics. The algebraic technique due to the quantum inverse scattering method is applied to obtain the three-term recursion relations for the coefficients of eigenfunctions expansion at the Jacobi polynomials series and hence the suitable equation (the continued fraction is equal to zero) for eigenvalues. The special functions considered generalize the Truskova polyspheroidal periodic functions ($\gamma = 0$).

I. INTRODUCTION

Polyspheroidal periodic functions (PPF's) and modified PPF's were introduced and studied by Truskova in 1982.[1] Both these special functions arise in the solution of the Helmholtz equation in the coordinate systems of N-dimensional ($N \geqslant 4$) ellipsoids and hyperboloids of revolution and are necessary for a number of physical problems. Analytical expressions were found and the main properties of the new functions were considered (see Ref. 1). Here we deal with the generalized ($\gamma \neq 0$) PPF's that satisfy the differential equation

$$H\Phi = \left[-\frac{1}{4}\frac{d^2}{dv^2} - \frac{1}{2}\frac{\mu - \nu + (\mu + \nu + 1)\cos 2v}{\sin 2v}\frac{d}{dv} - 2\xi \cos 2v - \gamma^2 \sin^2 2v \right]\Phi = h\Phi, \quad \mu,\nu,\xi,\gamma \in \mathbb{R}. \tag{1}$$

They are bounded and π-periodic eigenfunctions of the Hermitian operator H. The solutions form a new class of special functions, particular cases of which are the spheroidal, the Coulomb spheroidal, and the hyperspheroidal harmonics.[2] For $\xi = \gamma = 0$, Eq. (1) turns into the equation for the Jacobi polynomials.[3]

We proceed from the main notions of the quantum inverse scattering method (QISM).[4] Let us consider the principal problem of quantum mechanics for integrable systems, which is to obtain the spectrum and eigenfunctions of the complete set of commuting integrals of motion. The method we apply here consists of the separation of variables, the R-matrix scheme, and the so-called "functional Bethe ansatz."[5]

In quantum mechanics one works with the Hermitian operators in Hilbert space. Let them belong to the enveloping algebra $U(g)$ of a certain Lie algebra g. Let g have the rank n. Then an integrable system is defined by n commuting operators from $U(g)$. In the future an important role will be played by the transformation of g into the direct sum $\oplus_{i=1}^{n} a_i$ of n Lie algebras a_i, each having rank 1. The scheme of construction of this transformation implies the quantum separation of variables, i.e., the fact of transition from the multidimensional spectral problem for commuting integrals of motion to n one-dimensional ones with the common spectrum. We will call the direct sum $A = \oplus_{i=1}^{n} a_i$ an "algebra of separated equations."

In this paper the expansions of generalized PPF's into a series of the Jacobi polynomials are obtained by applying the separated equations algebra technique. The generalized PPF's and their particular cases appear when one solves the principal problem of quantum mechanics for different integrable systems (Secs. II and III).

In Sec. II g is e(3) (the Euclid algebra) and $a_i = \mathrm{so}(2,1)$, the generalized hyperspheroidal harmonics (a particular case of PPF's, see Ref. 1), emerge as eigenfunctions of a certain top, when one considers the representation of e(3) through the Euler angles. The separated equations algebra technique depends on the transformation $\mathrm{e}(3) \to \mathrm{so}(2,1) \oplus \mathrm{so}(2,1)$ (an analog of the separation of variables in classical mechanics) but is invariant about the representations both of the former algebra and the latter one. By choosing a particular representation we are able to derive nonlinear equations for the spectrum of integrals of motion and three-term recursion relations for the coefficients of eigenfunctions expansion into the series of Jacobi polynomials by algebraic means only. The Jacobi polynomials themselves (in the representation chosen) form the Cartan basis of irreducible unitary representations of algebras $a_i = \mathrm{so}(2,1)$.

In Sec. III generalized PPF's are connected with the quantum one degree of freedom integrable system on the algebra $g = \mathrm{e}(2)$. Here $n = 1$, $a_1 = \mathrm{so}(2,1)$.

II. REPRESENTATION OF THE QISM I ALGEBRA AND GENERALIZED HYPERSPHEROIDAL HARMONICS

First, we give the definition of a QISM I algebra (see Refs. 4 and 5). It is an associative algebra defined by the generators $T_{\alpha\beta}(u)$ ($\alpha,\beta = 1,\ldots,d$; $u \in \mathbb{C}$) considered as the elements of the square matrix $T(u)$ with the commutation relation

 0022-2488/90/051167-08$03.00

$$R(u-v)T^{1}(u)T^{2}(v) = T^{2}(v)T^{1}(u)R(u-v), \quad (2)$$

where

$$T^{1}(u) = T(u) \otimes I_d, \quad T^{2}(v) = I_d \otimes T(v).$$

The matrix $R(u) \in \mathrm{Aut}(\mathbb{C}^d \times \mathbb{C}^d)$ is a solution of the quantum Yang–Baxter equation. The commuting integrals of motion are extracted from the trace of the matrix $T(u)$. In this paper we consider the simplest case $d = 2$ when $T(u)$ is the 2×2 matrix,

$$T(u) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}(u), \quad \tau(u) = \mathrm{tr}\, T(u) = A(u) + D(u), \quad (3)$$

and $R(u)$ is the R matrix of the XXX type given by

$$R(u) = u + i\varkappa P, \quad P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \varkappa \in \mathbb{C}. \quad (4)$$

In Ref. 6, a new representation of the QISM I algebra $T(u)$ (L operator) was constructed for a given R matrix (4). This new L operator is connected with a special integrable top. The Hamiltonian has the form ($\mu, b \in \mathbb{R}$)

$$H = J_1^2 + J_2^2 - b^2 x_3^2 + (\mu^2 - \tfrac{1}{4})/x_3^2 + \tfrac{1}{4}. \quad (5)$$

Variables J_i, x_i, $i = 1,2,3$, are the generators of the Lie algebra e(3) obeying the following commutation relations:

$$[J_i, J_j] = -i\varepsilon_{ijk}J_k, \quad [J_i, x_j] = -i\varepsilon_{ijk}x_k, \quad [x_i, x_j] = 0. \quad (6)$$

Everywhere below the values of the Casimir operators of this algebra are restricted by

$$\sum_{i=1}^{3} x_i^2 = 1, \quad l = \sum_{i=1}^{3} x_i J_i = 0. \quad (7)$$

The pair of commuting integrals of motion H and J_3 makes this top an integrable dynamical system on the algebra e(3). The variables J_i, x_i, and the Hamiltonian H are the Hermitian operators in some Hilbert space.

The L operator for the model is

$$L(u) = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} u^2 - 2J_3 u - J_1^2 - J_2^2 - \frac{1}{4} - (\mu^2 - \frac{1}{4})/x_3^2 & ib(x_+ u - \frac{1}{2}\{x_3, J_+\}) \\ ib(x_- u - \frac{1}{2}\{x_3, J_-\}) & b^2 x_3^2 \end{pmatrix}, \quad (8)$$

where $x_\pm = x_1 \pm ix_2$, $J_\pm = J_1 \pm iJ_2$, and the symbol $\{\cdot,\cdot\}$ signifies an anticommutator. One can directly verify that $L(u)$ satisfies the fundamental equation (2) of the QISM I algebra $T(u)$ with $\varkappa = 2i$. The trace $\tau(u)$ of $L(u)$ is the generating polynomial of the integrals of motion,

$$\tau(u) = u^2 - 2J_3 u - H. \quad (9)$$

Let us list the commutators that follow from (2):

$$[A,A] = [B,B] = [C,C] = [D,D] = 0; \quad (10)$$

$$[A,B] = [-i\varkappa/(u-v)](AB - \underline{A}\underline{B}), \quad (11a)$$

$$[B,A] = [-i\varkappa/(u-v)](BA - \underline{B}\underline{A}), \quad (11b)$$

$$[A,C] = [-i\varkappa/(u-v)](CA - \underline{C}\underline{A}), \quad (11c)$$

$$[C,A] = [-i\varkappa/(u-v)](AC - \underline{A}\underline{C}), \quad (11d)$$

$$[B,D] = [-i\varkappa/(u-v)](DB - \underline{D}\underline{B}), \quad (11e)$$

$$[D,B] = [-i\varkappa/(u-v)](BD - \underline{B}\underline{D}), \quad (11f)$$

$$[D,C] = [-i\varkappa/(u-v)](DC - \underline{D}\underline{C}), \quad (11g)$$

$$[C,D] = [-i\varkappa/(u-v)](CD - \underline{C}\underline{D}), \quad (11h)$$

$$\left.\begin{aligned} [A,D] &= [-i\varkappa/(u-v)](CB - \underline{C}\underline{B}), \\ [B,C] &= [-i\varkappa/(u-v)](DA - \underline{D}\underline{A}), \\ [C,B] &= [-i\varkappa/(u-v)](AD - \underline{A}\underline{D}), \\ [D,A] &= [-i\varkappa/(u-v)](BC - \underline{B}\underline{C}). \end{aligned}\right\} \quad (12)$$

Here we use the following notation for brevity: $[A,B]$ means the commutator $[A(u),B(v)]$, where the first parameter is u and the second one is v; DA stands for the noncommutative operator product $D(u)A(v)$; and $\underline{D}\underline{A}$ signifies $D(v)A(u)$ (here v is a first parameter). In Sec. II, $\varkappa = 2i$.

To find the eigenstates of $\tau(u)$ [(9)] we apply the separation of variables method worked out in Ref. 5. We briefly remind the reader of the main results of Ref. 6 as applied to the top considered. Our aim is to construct the above-mentioned algebra A of separated equations. The essential part of this algebra consists of the separated variables determined as mutually commuting [see (10)] roots of the operator equation $C(u) = 0$. This equation for the L operator (8) has one nilpotent root, in contrast to the Goryachev–Chaplygin gyrostat, its generalization, and the Toda lattice in Ref. 5. To overcome this difficulty we replace the L operator (8) with the $\tilde{L}$ operator

$$\tilde{L}(u) = \frac{\sigma_1 + \sigma_3}{2} L(u)(\sigma_1 + \sigma_3) = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix}(u), \quad (13)$$

where the σ_i are Pauli matrices. Here $\tilde{L}(u)$ satisfies Eqs. (10)–(12), the trace of $\tilde{L}(u)$ coincides with that of $L(u)$, and therefore $\tilde{L}(u)$ describes the same dynamical system.

The QISM I algebra (2) has an important property: if $T_1(u)$ and $T_2(u)$ are some representations of the algebra T in the spaces W_1 and W_2, respectively, then the product $T(u) = T_1(u)T_2(u)$ is also a representation of T in $W_1 \otimes W_2$ (comultiplication operation). In particular, one can choose $T_1(u) \equiv K$, where K is a representation of T in $\mathbb{C}^1$ that is simply a 2×2 matrix K satisfying the equality $[R(u), K \otimes K] = 0$, for any u. This property was used in

transformation (13).

From Eqs. (12) one can derive the quantum determinant $d(u)$ of $\tilde{L}(u)$ [equal to that of $L(u)$] generating the center of the QISM I algebra, which has the form

$$\begin{aligned} d(u) &= \tilde{A}(u+1)\tilde{D}(u-1) - \tilde{C}(u+1)\tilde{B}(u-1) \\ &= \tilde{A}(u-1)\tilde{D}(u+1) - \tilde{B}(u-1)\tilde{C}(u+1) \\ &= \tilde{D}(u-1)\tilde{A}(u+1) - \tilde{C}(u-1)\tilde{B}(u+1) \\ &= \tilde{D}(u+1)\tilde{A}(u-1) - \tilde{B}(u+1)\tilde{C}(u-1) \\ &= b^2(u^2-\mu^2). \end{aligned} \tag{14}$$

Commuting separated variables are now defined as roots of the square equation

$$\tilde{C}(u) = 0 \tag{15}$$

and are given by

$$\begin{aligned} u_{1,2} &= J_3 + bx_2 \\ &\pm [(J_3+bx_2)^2 + J_1^2 + \tfrac{1}{4} + (\mu^2 - \tfrac{1}{4})/x_3^2 \\ &+ (J_2 - bx_3)^2]^{1/2}. \end{aligned} \tag{16}$$

The variables $u_{1,2}$ are Hermitian operators.

Let us introduce two more operators $\lambda_j^{\pm}$, for each u_j:

$$\lambda_j^- = \tilde{A}(u \Rightarrow u_j), \quad \lambda_j^+ = \tilde{D}(u \Rightarrow u_j). \tag{17}$$

Here the left substitution of $u_{1,2}$ in $\tilde{A}(u)$ and $\tilde{D}(u)$ is chosen. The rewriting of the $\tilde{L}$ operator in terms of $u_j, \lambda_j^{\pm}$ is single-valued [the quantum analog of the Lagrange interpolation of operator polynomials $\tilde{A}(u)$ and $\tilde{D}(u)$ must be used]:

$$\begin{aligned} \tilde{A}(u) &= \tfrac{1}{2}(u-u_1)(u-u_2) \\ &\quad + \frac{u-u_1}{u_2-u_1}\lambda_2^- + \frac{u-u_2}{u_1-u_2}\lambda_1^-, \\ \tilde{D}(u) &= \tfrac{1}{2}(u-u_1)(u-u_2) \\ &\quad + \frac{u-u_1}{u_2-u_1}\lambda_2^+ + \frac{u-u_2}{u_1-u_2}\lambda_1^+, \\ \tilde{C}(u) &= \tfrac{1}{2}(u-u_1)(u-u_2), \end{aligned} \tag{18}$$

where the formula for $\tilde{B}(u)$ follows from any of the equalities (14).

The operators $u_j, \lambda_j^{\pm}$ possess the conjugation properties

$$u_j^* = u_j, \quad (\lambda_j^{\pm})^* = [1/(u_1-u_2)]\lambda_j^{\mp}(u_1-u_2), \tag{19}$$

and satisfy the following algebraic relations:

$$[\lambda_j^{\pm}, u_k] = \pm 2\delta_{jk}\lambda_k^{\pm}, \quad [\lambda_j^{\pm}, \lambda_k^{\pm}] = [u_j, u_k] = 0; \tag{20}$$

$$\lambda_j^-\lambda_j^+ = d(u_j - 1) = b^2(u_j - \mu - 1)(u_j + \mu - 1),$$

$$\lambda_j^+\lambda_j^- = d(u_j + 1) = b^2(u_j + \mu + 1)(u_j - \mu + 1); \tag{21}$$

where $d(u)$ is the quantum determinant (14). Equations (19)–(21) are derived from the fundamental relations (10)–(12) and (14) and the interpolation (18) by means of the technique developed by Sklyanin in Ref. 5.

Let us derive the commutation relation (20) between λ_j^- and u_k. Consider the equality (11c),

(i) $[\tilde{A}(u), \tilde{C}(v)]$

$$= [2/(u-v)](\tilde{C}(u)\tilde{A}(v) - \tilde{C}(v)\tilde{A}(u)),$$

and substitute $u \Rightarrow u_j$. Using the definition (17) and the fact that, by virtue of (10), $[\tilde{C}(v), u_j] = 0$ and $\tilde{C}(u \Rightarrow u_j) = 0$, one obtains

$$[\lambda_j^-, \tilde{C}(v)] = [-2/(u_j - v)]\tilde{C}(v)\lambda_j^-.$$

Substituting expression (18) for $\tilde{C}(v)$, one arrives at

(ii) $\lambda_1^- s(u_1,u_2) = s(u_1-2,u_2)\lambda_1^-$,

$\lambda_2^- s(u_1,u_2) = s(u_1,u_2-2)\lambda_2^-$,

which holds for any symmetric function $s(u_1,u_2)$. But, as the spectrum S of operators u_1,u_2 belongs to the lower right quadrant of $\mathbb{R}^2$ (since u_1 is positive definite and u_2 is negative definite), (ii) is valid for any function $s(u_1,u_2)$ including $s(u_1,u_2) = u_1$ and $s(u_1,u_2) = u_2$.

The rest of the relations (19) and (20) can be proved in a quite analogous way. As for the relations (21), their proof contains more tedious calculations (see the original proof of Sklyanin in Ref. 5).

Let us turn $\lambda_j^{\pm}$ into $m_j^{\pm}$ by

$$m_j^{\pm} = (u_1 - u_2)^{-1/2}\lambda_j^{\pm}(u_1 - u_2)^{1/2}. \tag{22}$$

Notice that the difference $u_1 - u_2$ is positive definite and the new operators obey another property $(m_j^{\pm})^* = m_j^{\mp}$ besides that for $\lambda_j^{\pm}$ [(19)]. Eventually we have six Hermitian operators

$$\begin{aligned} z_n^{(1)} &= (1/4b)(m_n^+ + m_n^-) \\ &= (1/2\sqrt{u_1-u_2})[bx_3^2 + u_n x_2 - \tfrac{1}{2}\{x_3,J_2\}]\sqrt{u_1-u_2}, \\ z_n^{(2)} &= (i/4b)(m_n^+ - m_n^-) \\ &= (1/2\sqrt{u_1-u_2})[u_n x_1 - \tfrac{1}{2}\{x_3,J_1\}]\sqrt{u_1-u_2}, \\ z_n^{(3)} &= u_n/2 = \tfrac{1}{2}[J_3 + bx_2 \pm [(J_3+bx_2)^2 \\ &\quad + J_1^2 + \tfrac{1}{4} + (\mu^2 - \tfrac{1}{4})/x_3^2 + (J_2 - bx_3)^2]^{1/2}] \end{aligned} \tag{23}$$

satisfying the standard commutation relations of the Lie algebra $A = \mathrm{so}(2,1) \oplus \mathrm{so}(2,1)$,

$$\begin{aligned} &[z_m^{(\alpha)}, z_n^{(\beta)}] \\ &\quad = -i\delta_{mn}\epsilon_{\alpha\beta\gamma}z_m^{(\delta)}\Delta_{\gamma\delta}, \quad \Delta = \mathrm{diag}(-1,-1,1). \end{aligned} \tag{24}$$

The Casimir operators are

$$C_n = \Delta_{\gamma\beta}z_n^{(\gamma)}z_n^{(\beta)} = \tfrac{1}{4}(\mu^2 - 1) = j(j+1), \tag{25}$$

where Eqs. (14) and (19)–(21) were used.

Consider discrete $D^{\pm}$ series of unitary irreducible representations of so(2,1). In accordance with (25) we are interested in the following "spin":

$$j = -\tfrac{1}{2} - \mu/2. \tag{26}$$

We bear in mind the projective representations where j is any negative real number ($j < 0$) and hence $\mu > -1$. We remind the reader that they are single-valued representations of the covering group of SU(1,1).[7]

Then choosing a representation in which compact gen-

erators $z_n^{(3)} = \frac{1}{2} u_n$ are diagonal, "spin" value j (26) defines the discrete spectrum S of operators $u_{1,2}$:

$$S = \{(u_1,u_2)\in\mathbb{R}^2: \quad (u_1,u_2) = (-2j+2n_1, 2j-2n_2), n_1,n_2 = 0,1,2,...\}, \tag{27}$$

where we put D^+ for u_1 and D^- for u_2 as $u_1 > u_2$ [(16)]. The eigenfunction space of the system can be realized as the space $\mathscr{L}_2(S)$ of square-summable functions on the spectrum S:

$$\mathscr{L}_2(S) = \left\{ f(u_1,u_2): \quad (u_1,u_2)\in S, \sum_{(u_1,u_2)\in S} |f(u_1,u_2)|^2 < \infty \right\}. \tag{28}$$

It is easy to check that the operators $m_n^{\pm}$, acting on eigenfunctions ϕ of u_n,

$$(u_n\phi)(u_1,u_2) = u_n\phi(u_1,u_2),$$

as

$$(m_1^{\pm}\phi)(u_1,u_2) = d^{1/2}(u_1 \pm 1)\phi(u_1 \pm 2,u_2),$$
$$(m_2^{\pm}\phi)(u_1,u_2) = d^{1/2}(u_2 \pm 1)\phi(u_1,u_2 \pm 2),$$

obey all the relations (20) and (21) or (24) and (25). Let us now consider the spectral problem for the generating function $\tau(u) = \tilde{A}(u) + \tilde{D}(u) = u^2 - 2J_3u - H$. It has the form [see (18) and (22)]

$$\begin{aligned} t(u)f(u_1,u_2) &= (u-u_1)(u-u_2)f(u_1,u_2) + (u-u_2) \\ &\times (u_1-u_2)^{-1/2}(m_1^+ + m_1^-)(u_1-u_2)^{-1/2} \\ &\times f(u_1,u_2) + (u_1-u)(u_1-u_2)^{-1/2} \\ &\times (m_2^+ + m_2^-)(u_1-u_2)^{-1/2} f(u_1,u_2), \end{aligned} \tag{29}$$

where $f\in\mathscr{L}_2(S)$ is the eigenfunction of $\tau(u)$ and $t(u) = u^2 - 2mu - h$, where m and h are eigenvalues of J_3 and H, respectively.

It appears that, for the separation of variables u_1 and u_2, we ought to write

$$f(u_1,u_2) = (u_1-u_2)^{1/2}\phi(u_1,u_2). \tag{30}$$

Now, putting $u = u_n$, $n = 1,2$, in (29), we get two separated one-dimensional equations:

$$t(u_n)\phi_n(u_n) = d^{1/2}(u_n-1)\phi_n(u_n-2) + d^{1/2}(u_n+1)\phi_n(u_n+2), \tag{31}$$

where the function $\phi(u_1,u_2)$ is factorized as

$$\phi(u_1,u_2) = \phi_1(u_1)\phi_2(u_2). \tag{32}$$

Such a form of the function $\phi(u_1,u_2)$ allows us to speak about separation of variables and reflects the structure of a direct sum of the algebra (24) $A = \mathrm{so}(2,1)\oplus\mathrm{so}(2,1)$. The two one-dimensional spectral problems are the three-term recursion relations (TTRR's) for the coefficients $\phi_n(u_n)$, where the variables $u_{1,2}$ belong to the lattice S depending on the value of μ.

The scalar product for the functions $\phi(u_1,u_2)$ [(30)] is

$$\langle\phi|\chi\rangle = \sum_{(u_1,u_2)\in S} (u_1-u_2)\bar{\phi}(u_1,u_2)\chi(u_1,u_2),$$

therefore the norm of eigenfunctions $\phi_n(u_n)$ (32) of Eqs. (31) has the form

$$\sum_{(u_n)} |u_n|\,|\phi_n(u_n)|^2 < \infty. \tag{33}$$

The numerical iterations of the TTRR's (31) require the knowledge of the initial values of m and h for $b\to 0$. We have

$$\begin{aligned} h|_{b=0} &= \langle J_1^2 + J_2^2 + \tfrac{1}{4} + (\mu^2 - \tfrac{1}{4})/x_3^2\rangle \\ &= -\langle u_1u_2\rangle|_{b=0} = 4(-j+n_1)(-j+n_2) \\ &= (2n_1+\mu+1)(2n_2+\mu+1), \\ m &= \langle J_3\rangle = \langle (u_1+u_2)/2\rangle|_{b=0} = n_1 - n_2, \\ & \qquad n_1,n_2 = 0,1,2,... , \end{aligned} \tag{34}$$

where the brackets $\langle\ \rangle$ stand for the eigenvalues of an operator. The eigenvalues h can be computed in complete analogy with the computations in the theory of spheroidal and Mathieu functions in Refs. 2 and 3. We note only that the TTRR's (31) imply that two convergent continued fractions are equal to zero and this gives nonlinear equations for the eigenvalues m and h. For $m = 0, \pm 1, \pm 2,...$ fixed it is enough to work with the one continued fraction with the additional condition

$$h|_{b=0} = 4(-j+n)(-j+n+|m|), \quad n = 0,1,2,... .$$

The standard way to solve the principal problem of quantum mechanics for the Hamiltonian H (5) is to express the generators J_i, x_i in the Euler angles ϕ, ψ, θ and the corresponding momenta

$$p_\phi = -i\frac{\partial}{\partial\phi}, \quad p_\psi = -i\frac{\partial}{\partial\psi}, \quad p_\theta = -i\frac{\partial}{\partial\theta}.$$

The explicit formulas are

$$\begin{aligned} J_1 &= p_\theta\cos\psi + p_\phi\frac{\sin\psi}{\sin\theta} - \frac{\cot\theta}{2}\{\sin\psi,p_\psi\}, \\ J_2 &= -p_\theta\sin\psi + p_\phi\frac{\cos\psi}{\sin\theta} - \frac{\cot\theta}{2}\{\cos\psi,p_\psi\}, \\ J_3 &= p_\psi, \quad x_1 = \sin\psi\sin\theta, \quad x_2 = \cos\psi\sin\theta, \quad x_3 = \cos\theta. \end{aligned} \tag{35}$$

Remember that $\langle p_\phi\rangle = l = 0$ [(7)] and $\langle J_3\rangle = m$ [(34)]; hence one can factorize the eigenfunction f of the Hamiltonian H [(5)] as

$$f(\phi,\psi,\theta) = \exp(-im\psi)\Phi(\theta), \tag{36}$$

where $\Phi(\theta)$ satisfies the equation

$$\left[\frac{d^2}{d\theta^2} - \frac{m^2-\frac{1}{4}}{\sin^2\theta} + b^2\cos^2\theta - \frac{\mu^2-\frac{1}{4}}{\cos^2\theta} + h + m^2\right]\Phi(\theta) = 0. \tag{37}$$

Thus $\Phi(\theta)$ are generalized hyperspheroidal harmonics (see Ref. 1). It is known that one must expand $\Phi(\theta)$ into a series of Jacobi polynomials to obtain the TTRR's for the expansion coefficients and the nonlinear equation (continued fraction is equal to zero) for the eigenvalues h (see Ref. 1). One can check the equivalence of this continued fraction to that given by either of the TTRR's (31). Further, from Eqs. (30) and (32) we see that the eigenfunction $f(u_1,u_2)$ was expand-

ed into a series of eigenfunctions of two commuting operators:

$$\frac{u_1+u_2}{2}=J_3+bx_2,$$

$$-u_1u_2=J_1^2+\frac{1}{4}+\frac{\mu^2-\frac{1}{4}}{x_3^2}+(J_2-bx_3)^2. \tag{38}$$

Under the automorphism of the e(3) algebra

$$J_3\to J_3'-bx_2,\quad J_2\to J_2'+bx_3,\quad J_1\to J_1',\quad x_i\to x_i',$$

the operators (38) transform into the following ones:

$$\frac{u_1+u_2}{2}=J_3',\quad -u_1u_2=J_1'^2+\frac{1}{4}+\frac{\mu^2-\frac{1}{4}}{x_3'^2}+J_2'^2. \tag{39}$$

Their spectrum coincides with that of the operators (38) and the eigenfunctions are a product of the Jacobi polynomials and the exponentials.

Thereby the algebra $A=\mathrm{so}(2,1)\oplus\mathrm{so}(2,1)$ of separated equations is exhibited for the generalized hyperspheroidal harmonics. In Ref. 6 it was called the dynamical algebra. We note that our algebraic method differs from the "dynamical algebra (group) scheme" in Ref. 8.

III. REPRESENTATION OF QISM II ALGEBRA AND GENERALIZED POLYSPHEROIDAL PERIODIC FUNCTIONS

First of all we give the definition of the QISM II algebra originally introduced by Sklyanin.[9] This algebra describes a new class of boundary conditions for quantum systems integrable by means of QISM. Two algebras $U_\pm(u)$ are analogs of the algebra $T(u)$ [(2)]:

$$R(u-v)\overset{1}{U}_-(u)R(u+v-i\varkappa)\overset{2}{U}_-(v)$$

$$=\overset{2}{U}_-(v)R(u+v-i\varkappa)\overset{1}{U}_-(u)R(u-v), \tag{40}$$

$$R(-u+v)\overset{1\ t_1}{U}_+(u)R(-u-v-i\varkappa)\overset{2\ t_2}{U}_+(v)$$

$$=\overset{2\ t_2}{U}_+(v)R(-u-v-i\varkappa)\overset{1\ t_1}{U}_+(u)R(-u+v), \tag{41}$$

where t_1 and t_2 are the matrix transpositions in the first and second spaces, respectively. The algebras U_- and U_+ are isomorphic: there exists an obvious isomorphism $X:U_-\to U_+$, $X\{U_-(u)\}=U_-^t(-u)$ (see Ref. 9). The matrix $R(u)$ is a solution of the quantum Yang–Baxter equation, too. As before we deal here with the R matrix of the XXX type $R(u)=u+i\varkappa P$ [(4)].

The quantities $\tau(u)=\mathrm{tr}\,U_+(u)U_-(u)$ defined in the direct product $U_+\times U_-$ form a commutative family $[\tau(u_1),\tau(u_2)]=0$, for any u_1,u_2 (Theorem 1 in Ref. 9). Hence $\tau(u)$ is the generating function of the integrals of motion.

Let us consider the quantum system with one degree of freedom given by the Hamiltonian

$$H=p^2+\frac{\mu^2-\frac{1}{4}}{4\sin^2(q/2)}+\frac{\nu^2-\frac{1}{4}}{4\cos^2(q/2)}$$

$$-\gamma^2\sin^2q-2\xi\cos q, \tag{42}$$

where $\mu,\nu,\gamma,\xi\in\mathbb{R}$, and p,q are the canonical variables $[p,q]=-i$. This model may be regarded as an integrable dynamical system on the e(2) algebra given by the generators p, $\cos(q/2)$, and $\sin(q/2)$. The variables p and q and the Hamiltonian H are the Hermitian operators in some Hilbert space.

The L operator for the system has the form

$$L(u)=\begin{pmatrix}A & B\\ C & D\end{pmatrix}=\begin{pmatrix}u\cdot\cos q-i\cdot\sin q\cdot p+\dfrac{\mu^2-\nu^2}{4(u-\frac{1}{2})} & -\sin^2q\\ u^2-p^2-\dfrac{\mu^2-\frac{1}{4}}{4\sin^2(q/2)}-\dfrac{\nu^2-\frac{1}{4}}{4\cos^2(q/2)} & u\cdot\cos q+i\cdot p\cdot\sin q+\dfrac{\mu^2-\nu^2}{4(u-\frac{1}{2})}\end{pmatrix}. \tag{43}$$

One can directly verify that $L(u)$ satisfies the fundamental equation (40) of the QISM II algebra $U_-(u)$ with $\varkappa=-i$.

As a representation of the algebra $U_+(u)$ we take that of the algebra $U_+(u)$ in $\mathbb{C}^1$, i.e., the c-number matrix. It has the form

$$U_+(u)\equiv K(u)=\begin{pmatrix}\xi & u+\frac{1}{2}\\ -\gamma^2(u+\frac{1}{2}) & \xi\end{pmatrix}. \tag{44}$$

It is easy to check that $K(u)$ satisfies Eq. (41) with $\varkappa=-i$. The trace $\tau(u)=\mathrm{tr}(K(u)\,L(u))$ is the generating function of the integrals of motion

$$\tau(u)=(u+\tfrac{1}{2})[u^2-H+\xi(\mu^2-\nu^2)/2(u^2-\tfrac{1}{4})]. \tag{45}$$

Let us list the commutators that follow from (40):

$$[A,A]=[-i\varkappa/(u+v)](BC-\underline{BC}),$$

$$[D,D]=[-i\varkappa/(u+v)](CB-\underline{CB}), \tag{46a}$$

$$[B,B]=0,\quad [C,C]=0; \tag{46b}$$

$$[A,B]=[-i\varkappa/(u-v)](AB-\underline{AB})$$
$$-[i\varkappa/(u+v-i\varkappa)](AB+BD)$$
$$+[\varkappa^2/(u-v)(u+v-i\varkappa)]$$
$$\times(AB+BD-\underline{AB}-\underline{BD}), \tag{47a}$$

$$[B,A] = [-i\varkappa/(u-v)](BA - \underline{B}\underline{A}) + [i\varkappa/(u+v-i\varkappa)](\underline{A}\underline{B} + \underline{B}\underline{D}), \quad (47b)$$

$$[A,C] = [-i\varkappa/(u-v)](CA - \underline{C}\underline{A}) + [i\varkappa/(u+v-i\varkappa)](\underline{C}\underline{A} + \underline{D}\underline{C}) + [\varkappa^2/(u-v)(u+v-i\varkappa)] \times (CA + DC - \underline{C}\underline{A} - \underline{D}\underline{C}), \quad (47c)$$

$$[C,A] = [-i\varkappa/(u-v)](AC - \underline{A}\underline{C}) - [i\varkappa/(u+v-i\varkappa)](CA + DC), \quad (47d)$$

$$[B,D] = [-i\varkappa/(u-v)](DB - \underline{D}\underline{B}) - [i\varkappa/(u+v-i\varkappa)](AB + BD), \quad (47e)$$

$$[D,B] = [-i\varkappa/(u-v)](BD - \underline{B}\underline{D}) + [i\varkappa/(u+v-i\varkappa)](\underline{A}\underline{B} + \underline{B}\underline{D}) + [\varkappa^2/(u-v)(u+v-i\varkappa)] \times (AB + BD - \underline{A}\underline{B} - \underline{B}\underline{D}), \quad (47f)$$

$$[C,D] = [-i\varkappa/(u-v)](CD - \underline{C}\underline{D}) + [i\varkappa/(u+v-i\varkappa)](\underline{C}\underline{A} + \underline{D}\underline{C}), \quad (47g)$$

$$[D,C] = [-i\varkappa/(u-v)](DC - \underline{D}\underline{C}) - [i\varkappa/(u+v-i\varkappa)](CA + DC) + [\varkappa^2/(u-v)(u+v-i\varkappa)] \times (CA + DC - \underline{C}\underline{A} - \underline{D}\underline{C}); \quad (47h)$$

$$[A,D] = \frac{-i\varkappa}{u-v}(CB - \underline{C}\underline{B})\left(1 + \frac{i\varkappa}{u+v}\right),$$

$$[D,A] = \frac{-i\varkappa}{u-v}(BC - \underline{B}\underline{C})\left(1 + \frac{i\varkappa}{u+v}\right),$$

$$[B,C] = \frac{-i\varkappa}{u-v}(DA - \underline{D}\underline{A}) \times \left(1 - \frac{i\varkappa}{u+v}\right) - \frac{i\varkappa}{u+v}(AA - \underline{D}\underline{D}), \quad (48)$$

$$[C,B] = \frac{-i\varkappa}{u-v}(AD - \underline{A}\underline{D}) \times \left(1 - \frac{i\varkappa}{u+v}\right) - \frac{i\varkappa}{u+v}(DD - \underline{A}\underline{A}).$$

From now on in Sec. III $\varkappa = -i$.

Our aim as formerly is to construct the algebra A of the separated equation (we now have one degree of freedom) in order to find the eigenstates of $\tau(u)$ [(45)].

Let us transform the L operator (43) into the $\tilde{L}$ operator according to the equality

$$\tau(u) = \operatorname{tr} K(u)L(u) = \operatorname{tr}(\xi \cdot I + i\gamma(u+\tfrac{1}{2})\cdot\sigma_3)\tilde{L}(u), \quad (49)$$

where

$$\tilde{L}(u) = V^{-1}LV = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix}(u), \quad V = \begin{pmatrix} 1 & 1 \\ i\gamma & -i\gamma \end{pmatrix}. \quad (50)$$

This $\tilde{L}(u)$ satisfies Eqs. (46)–(48) and describes the same dynamical system. Now we have the equation $\tau(u) = \operatorname{tr} \tilde{K}(u)\tilde{L}(u)$ (49) with the *diagonal* matrix $\tilde{K}(u)$ (and it is important in the future).

The QISM II algebra (40) has the following property: let $\tilde{U}_-(u)$ be some representation of the QISM II algebra U_- in the space $\tilde{\omega}_-$, and let $T_-(u)$ be that of the QISM I algebra T in $\mathscr{W}_-$. Then the $U_-(u)$ defined by

$$U_-(u) = T_-(u)\tilde{U}_-(u)\sigma_2 T'_-(-u)\sigma_2 \quad (51)$$

is a representation of the QISM II algebra U_- in $\tilde{\omega}_- \otimes \mathscr{W}_-$ (Proposition 2 of Ref. 9). This property was used in Eq. (50) with $T_-(u) \equiv V$.

From Eqs. (48) one can derive the quantum determinant $d(u)$ of $\tilde{L}(u)$ [equal to that of $L(u)$] generating the center of the QISM II algebra (40) that has the form

$$\begin{aligned} d(u) &= (\tilde{D}_- - (1/2u)(\tilde{D}_- + \tilde{A}_-))\tilde{A}_+ - \tilde{B}_-\tilde{C}_+ \\ &= (\tilde{A}_- - (1/2u)(\tilde{D}_- + \tilde{A}_-))\tilde{D}_+ - \tilde{C}_-\tilde{B}_+ \\ &= \tilde{A}_+(\tilde{D}_- - (1/2u)(\tilde{A}_- + \tilde{D}_-)) - \tilde{B}_+\tilde{C}_- \\ &= \tilde{D}_+(\tilde{A}_- - (1/2u)(\tilde{A}_- + \tilde{D}_-)) - \tilde{C}_+\tilde{B}_- \\ &= (u - (\mu^2 - \nu^2)/4u)^2 - \nu^2, \end{aligned} \quad (52)$$

where we use following notation: $\tilde{D}_+$ stands for $\tilde{D}(u+\frac{1}{2})$, $\tilde{B}_-$ for $\tilde{B}(u-\frac{1}{2})$, and so on.

Consider the operator roots of $\tilde{B}(u) = \frac{1}{2}(A - D - i\gamma B - iC/\gamma)$:

$$\tilde{B}(\pm u_1) = 0,$$

$$u_1^2 = (p - \gamma \sin q)^2 + \frac{\mu^2 - \frac{1}{4}}{4\sin^2(q/2)} + \frac{\nu^2 - \frac{1}{4}}{4\cos^2(q/2)}. \quad (53)$$

The Hermitian operator u_1^2 is positive definite. We choose u_1 to be positive definite, too.

Let us introduce two more operators $\lambda^{\pm}$. Put $\hat{D}(u) = 2u \times \tilde{D}(u) - \tilde{A}(u)$ and $\check{D}(u) = \hat{D}(u)/(2u+1)$. Then

$$\lambda^- = \tilde{A}(u \Rightarrow u_1), \quad \lambda^+ = \check{D}(u \Rightarrow u_1), \quad (54)$$

where the left substitution of u_1 into $\tilde{A}(u)$ and $\check{D}(u)$ is chosen. The rewriting of the $\tilde{L}$ operator in terms of u_1, $\lambda^{\pm}$ is single-valued:

$$\begin{aligned} \tilde{A}(u) &= \frac{\mu^2 - \nu^2}{4(u-\frac{1}{2})}\frac{u_1^2 - u^2}{u_1^2 - \frac{1}{4}} + \frac{u^2 - u_1^2}{2i\gamma} + \frac{u+u_1}{2u_1}\lambda^- + \frac{u-u_1}{2u_1}\lambda^+, \\ \tilde{D}(u) &= \frac{\mu^2 - \nu^2}{4(u-\frac{1}{2})}\frac{u_1^2 - u^2}{u_1^2 - \frac{1}{4}} - \frac{u^2 - u_1^2}{2i\gamma} + \frac{u-u_1+1}{2u_1}\lambda^- + \frac{u+u_1+1}{2u_1}\lambda^+, \\ \tilde{B}(u) &= (1/2i\gamma)(u^2 - u_1^2), \end{aligned} \quad (55)$$

where the formula for $\tilde{C}(u)$ follows from any of the equalities (52). To derive the interpolation (55) we used the identity

$$\check{D}(u) = -\tilde{A}(-u). \quad (56)$$

Operators $u_1, \lambda^{\pm}$ possess the following properties:

$$u_1^* = u_1, \quad (\lambda^{\pm})^* = (1/u_1)\lambda^{\mp}u_1; \tag{57}$$

$$[\lambda^{\pm}, u_1] = \pm\lambda^{\pm}; \tag{58}$$

$$\lambda^-\lambda^+ = d(u_1 - \tfrac{1}{2}) = [u_1 - \tfrac{1}{2} - (\mu^2 - \nu^2)/4(u_1 - \tfrac{1}{2})]^2 - \nu^2,$$

$$\lambda^+\lambda^- = d(u_1 + \tfrac{1}{2}) = [u_1 + \tfrac{1}{2} - (\mu^2 - \nu^2)/4(u_1 + \tfrac{1}{2})]^2 - \nu^2; \tag{59}$$

where $d(u)$ is the quantum determinant (52). To prove the commutation relation (58) between λ^- and u_1 we rewrite (47b) as

(i) $[\tilde{A}(u), \tilde{B}(v)]$
$= [-1/(u-v)](\tilde{B}(v)\tilde{A}(u) - \tilde{B}(u)\tilde{A}(v))$
$- [1/(u+v-1)](\tilde{A}(u)\tilde{B}(v) + \tilde{B}(u)\tilde{D}(v))$,

and substitute $u \Rightarrow u_1$. Using the definition (54) and the fact that, by virtue of interpolation (55) for $\tilde{B}(v)$, $[\tilde{B}(v), u_1] = 0$ and $\tilde{B}(u \Rightarrow u_1) = 0$, one arrives at

(ii) $\lambda^- s(u_1) = s(u_1 - 1)\lambda^-$,

which holds for any even function $s(u_1)$. But, since u_1 is positive definite, (ii) is valid for any function $s(u_1)$ including $s(u_1) = u_1$. As for the commutator of λ^+ and u_1 one must consider (47a), (47f), and (47b) to obtain

$$[\hat{D}(u), \tilde{B}(v)] = (2u+1)\left[-\frac{\tilde{B}(u)\hat{D}(v)}{2v(u-v)} + \frac{\tilde{B}(v)\hat{D}(u)}{u^2 - v^2} + \frac{(2v-1)\tilde{B}(u)\tilde{A}(v)}{2v(u+v)}\right],$$

instead of (i), and so on. The rest of the relations (57)–(59) can be proved in a quite analogous way using the fundamental equations (46)–(48) and the interpolation (55).

Let us replace $\lambda^{\pm}$ by $\tilde{\lambda}^{\pm}$:

$$\tilde{\lambda}^- = (u_1 - \tfrac{1}{2})\lambda^- = (u_1 - \tfrac{1}{2})(u_1 \cdot \cos q + ip \cdot \sin q - i\gamma \cdot \sin^2 q) + (\mu^2 - \nu^2)/4,$$

$$\tilde{\lambda}^+ = (u_1 + \tfrac{1}{2})\lambda^+ = (u_1 + \tfrac{1}{2})(u_1 \cdot \cos q - ip \cdot \sin q + i\gamma \cdot \sin^2 q) + (\mu^2 - \nu^2)/4. \tag{60}$$

Operators $u_1, \tilde{\lambda}^{\pm}$ obey Eqs. (57) and (58) and

$$\tilde{\lambda}^+\tilde{\lambda}^- = \Delta_+(u_1 + \tfrac{1}{2})\Delta_-(u_1 + \tfrac{1}{2}),$$
$$\tilde{\lambda}^-\tilde{\lambda}^+ = \Delta_+(u_1 - \tfrac{1}{2})\Delta_-(u_1 - \tfrac{1}{2}),$$
$$\Delta_{\pm}(u) = u^2 \pm \nu u - (\mu^2 - \nu^2)/4,$$
$$\Delta_+(u)\Delta_-(u) = d(u)\,u^2. \tag{61}$$

Let us define two more operators $\hat{\lambda}^{\pm}$,

$$\hat{\lambda}^{\pm} = \Delta_+^{-1/2}(u_1 \pm \tfrac{1}{2})\tilde{\lambda}^{\pm}, \tag{62}$$

which also obey Eqs. (57) and (58) and

$$\hat{\lambda}^+\hat{\lambda}^- = \Delta_-(u_1 + \tfrac{1}{2}), \quad \hat{\lambda}^-\hat{\lambda}^+ = \Delta_-(u_1 - \tfrac{1}{2}). \tag{63}$$

Now we have three Hermitian operators

$$z^{(3)} = u_1 - \nu/2,$$
$$z^{(1)} = (1/2\sqrt{u_1})(\hat{\lambda}^+ + \hat{\lambda}^-)\sqrt{u_1}, \tag{64}$$
$$z^{(2)} = (i/2\sqrt{u_1})(\hat{\lambda}^+ - \hat{\lambda}^-)\sqrt{u_1}$$

satisfying the standard commutation relations of the Lie algebra $A = \mathrm{so}(2,1)$:

$$[z^{(\alpha)}, z^{(\beta)}] = -i\epsilon_{\alpha\beta\gamma}\Delta_{\gamma\delta}z^{(\delta)}, \quad \Delta = \mathrm{diag}(-1,-1,1). \tag{65}$$

The Casimir operator is

$$C = \Delta_{\gamma\beta}z^{(\gamma)}z^{(\beta)} = \tfrac{1}{4}(\mu^2 - 1) = j(j+1). \tag{66}$$

Consider the discrete D^+ series ($\langle u_1 \rangle > 0$) of irreducible unitary representations of so(2,1). The "spin" is equal to

$$j = -\tfrac{1}{2} - \mu/2, \tag{67}$$

where $j < 0$; hence $\mu > -1$. Then choosing a representation in which the compact generator $z^{(3)} = u_1 - \nu/2$ is diagonal, the "spin" value j (67) defines the discrete spectrum of the operator $z^{(3)}$ (and u_1):

$$\langle u_1 \rangle = \nu/2 - j + n = (\mu + \nu + 1)/2 + n, \quad n = 0,1,2\ldots. \tag{68}$$

One can see that to fulfill $\langle u_1 \rangle > 0$ we must set $\nu \geqslant 0$. The eigenfunction space of the system can be realized as the space $\mathscr{L}_2$ of square-summable functions $f(u_1)$ on the spectrum of u_1.

It is easy to check that the operators $\hat{\lambda}^{\pm}$, acting on eigenfunctions ϕ of u_1,

$$(u_1\phi)(u_1) = u_1\phi(u_1),$$

as

$$(\hat{\lambda}^{\pm}\phi)(u_1) = \Delta_-^{1/2}(u_1 \pm \tfrac{1}{2})\phi(u_1 \pm 1), \tag{69}$$

obey all the relations (57), (58), and (63). Further, we consider the spectral problem for the generating function $\tau(u)$ [(45) and (49)]

$$\tau(u) = \mathrm{tr}\,\tilde{K}(u)\tilde{L}(u) = \mathrm{tr}\begin{pmatrix} \xi + i\gamma(u + \frac{1}{2}) & 0 \\ 0 & \xi - i\gamma(u + \frac{1}{2}) \end{pmatrix}\begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix}(u), \tag{70}$$

$$\tau(u)f(u_1) = t(u)f(u_1),$$

where $f \in \mathscr{L}_2$ is the eigenfunction of $\tau(u)$ and $t(u)$ is equal to $\tau(u)$ [(45)] with the Hamiltonian H replaced by corresponding eigenvalue h.

The function $f(u_1)$ is factorized as

$$f(u_1) = \sqrt{u_1}\phi(u_1) \tag{71}$$

and $\phi(u_1)$ satisfies the following TTRR's:

$$[(u_1^2 - \tfrac{1}{4})(u_1^2 - h) + \xi(\mu^2 - \nu^2)/2]\phi(u_1) = (1/u_1)(u_1^2 - \tfrac{1}{4})(i\gamma u_1 + \xi - i\gamma/2) \times d^{1/2}(u_1 - \tfrac{1}{2})\phi(u_1 - 1) + (1/u_1)(u_1^2 - \tfrac{1}{4}) \times (-i\gamma u_1 + \xi - i\gamma/2)d^{1/2}(u_1 + \tfrac{1}{2})\phi(u_1 + 1). \tag{72}$$

The TTRR's (72) are derived using Eqs. (70), (54), (60), (62), and (69). We remind the reader that $u_1 = n + (\mu + \nu + 1)/2$, $n = 0,1,2,...$, [(68)]. The norm of $\phi(u_1)$ has the form

$$\sum_{\{u_1\}} u_1 |\phi(u_1)|^2 < \infty. \tag{73}$$

In the special case $\mu = \nu$ the results of this section were already announced.[10] Complete numerical calculations of the eigenstates will be published elsewhere.

Equation (42) implies that in the q representation our spectral problem amounts to

$$Hf(q) = \left(-\frac{d^2}{dq^2} + \frac{\mu^2 - \frac{1}{4}}{4\sin^2(q/2)} + \frac{\nu^2 - \frac{1}{4}}{4\cos^2(q/2)} - \gamma^2 \sin^2 q - 2\xi \cdot \cos q \right) \cdot f(q) = h \cdot f(q). \tag{74}$$

If we set

$$f(q) = (\sin(q/2))^{\mu + 1/2} (\cos(q/2))^{\nu + 1/2} \Phi(v), \quad q = 2v, \tag{75}$$

then $\Phi(v)$ satisfies

$$\left[\frac{d^2}{dv^2} + 2\frac{\mu - \nu + (\mu + \nu + 1)\cos 2v}{\sin 2v} \frac{d}{dv} + 8\xi \cos 2v + 4\gamma^2 \sin^2 2v \right] \Phi(v) = -4h\Phi(v). \tag{76}$$

Thus the generalized polyspheroidal functions considered are the bounded and π-periodic solutions of Eq. (76). For $\gamma = 0$ we get an equation for PPF's in Ref. 1. We note that our generalization ($\gamma \neq 0$) corresponds to the solution in the coordinate systems of N-dimensional ($N \geqslant 4$) ellipsoids and hyperboloids of revolution of the N-dimensional oscillator turned upside-down:

$$(\Delta_N + k^2 + \tilde{\gamma}^2 \cdot |\mathbf{x}|^2) W = 0.$$

[1] N. F. Truskova, Yad. Fiz. **36**, 790 (1982).

[2] For physical applications, see, for example, J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer, Berlin, 1954); C. Flammer, *Spheroidal Wave Functions* (Stanford U.P., Stanford, CA, 1957); see, also, I. V. Komarov, L. I. Ponomarev, and S. Yu. Slavjanov, *Spheroidal and Coulomb Spheroidal Functions* (Nauka, Moscow, 1976) (in Russian).

[3] A. Erdelyi, *High Transcendental Functions* (McGraw-Hill, New York, 1953–1955), Vols. I–III.

[4] L. D. Faddeev, *Les Houches Lectures*, edited by J. B. Zuber and R. Stora (North-Holland, Amsterdam, 1984), pp. 719–756; P. P. Kulish and E. K. Sklyanin, *Integrable Quantum Field Theories, Lecture Notes in Physics*, Vol. 151, edited by J. Hietarinta and C. Montonen (Springer, Berlin, 1982), pp. 61–119.

[5] E. K. Sklyanin, J. Sov. Math. **31**, 3417 (1985); *Non-linear Equations in Classical and Quantum Field Theory, Lecture Notes in Physics*, Vol. 226, edited by H. Sanchez (Springer, Berlin, 1985), pp. 196–233; I. V. Komarov and V. B. Kuznetsov, Zap. Nauch. Semin. LOMI **164**, 134 (1987).

[6] V. B. Kuznetsov and A. V. Tsiganov, J. Phys. A **22**, L73 (1989).

[7] V. Bargmann, Ann. Math. **48**, 568 (1947); L. C. Biedenharn, in *Non-Compact Groups in Particle Physics*, edited by Y. Chow (Benjamin, New York, 1966), p. 23.

[8] A. O. Barut, A. Inomata, and R. Wilson, J. Math. Phys. **28**, 605 (1987); A. O. Barut, A. Inomata and R. Wilson, J. Phys. A **20**, 4075 (1987); M. Kibler and P. Winternitz, *ibid.* **20**, 4097 (1987); C. Quesne, *ibid.* **21**, 4487 (1988).

[9] E. K. Sklyanin, J. Phys. A **21**, 2375 (1988).

[10] V. B. Kuznetsov and A. V. Tsiganov, Zap. Nauch. Semin. LOMI **172**, 88 (1989).

Ground-state energy of $H=P^2+2X^2+2\sqrt{2\lambda}X^3+\lambda X^4$ for $\lambda<<<1$

Biswanath Rath[a)]
Physics Department, S. B. Women's College, Cuttack-753001, India

(Received 16 August 1988; accepted for publication 8 November 1989)

A self-consistent perturbation theory is developed to compute the ground-state energy for the Hamiltonian $H=P^2+2X^2+2\sqrt{2\lambda}X^3+\lambda X^4$ in the weak coupling regime $\lambda<<<1$. Energy corrections up to third order are given. Reasonably good agreement is obtained with previous computations.

I. INTRODUCTION

Perturbation theory is believed to be a useful tool for investigating different physical problems for which an exact solution of the Schrödinger equation has not yet been found. There have been many studies on the nature of the perturbation theory with the hope of understanding both its limitations[1] and of finding methods to best utilize the perturbative results we do have. In particular, the single-well oscillator in the context of perturbation theory[1] has drawn considerable attention over a wide range of investigations.[1,2] It is surprising that the two-well oscillator[3] in the context of perturbation theory has not received the attention it deserves. Hence the aim of the present paper is to calculate the ground-state energy of the two-well oscillator using perturbation theory.

II. THEORY

The Hamiltonian considered is of the form

$$H=T+CB(q)+H_1(q), \tag{1}$$

where $H_0=T+CB(q)$ is the unperturbed part of the Hamiltonian and H_1 is the perturbed part. Now the Hamiltonian in Eq. (1) is rewritten[2] as

$$H=H_{C_1}+H_2(q), \tag{2}$$

where

$$H_{C_1}=T+CB(q)+F(q) \tag{3}$$

and

$$H_2(q)=H_1(q)-F(q). \tag{4}$$

Further $F(q)$ is considered[4] as a function of $B(q)$ and $H_1(q)$:

$$F(q)=\alpha\,(\langle 0|H_1|0\rangle_{C_1}/\langle 0|B|0\rangle_{C_1})B(q), \tag{5}$$

where the value of α is determined by the condition

$$\sum_{p\neq 0}[\langle 0|H_1|p\rangle_{C_1}-\langle 0|F|p\rangle_{C_1}]=0. \tag{6}$$

Hence the ground-state energy of H_{C_1} is E_{C_1} and is a function of C_1. The value of C_1 is determined by the condition

$$C_1=C+\alpha(\langle 0|H_1|0\rangle_{C_1}/\langle 0|B|0\rangle_{C_1}). \tag{7}$$

As H_0 is a function of C, the above definition of C_1 can be used to calculate E_{C_1} in terms of C. Further $H_2(q)$ is divided into two parts as

$$H_2(q)=H_{2D}(q)+H_{2N}(q), \tag{8}$$

where $H_{2D}(q)$ is the diagonal part and $H_{2N}(q)$ is the nondiagonal part of $H_2(q)$. Now the diagonal part $H_{2D}(q)$ is added to H_{C_1} to construct the new unperturbed Hamiltonian $H(C_1)$ as

$$H(C_1)=H_{C_1}+H_{2D}(q). \tag{9}$$

Hence the term H_{2N} contains the only contribution of the nondiagonal part and is considered as the new perturbed part. The formal perturbation theory is applied to H_{2N} using $H(C_1)$ as the unperturbed part.

The above development is applied to compute the ground-state energy of the oscillator characterized by the Hamiltonian[3]

$$H=P^2+2X^2+2\sqrt{2\lambda}X^3+\lambda X^4. \tag{10}$$

The expression for H_W is

$$H_W=P^2+2X^2+\alpha\lambda(\langle 0|X^4|0\rangle_W/\langle 0|X^2|0\rangle_W)X^2 \tag{11}$$

and for $H_2(X)$ is

$$H_2(X)=\lambda X^4-\lambda\alpha(\langle 0|X^4|0\rangle_W/\langle 0|X^2|0\rangle_W)X^2+2\sqrt{2\lambda}X^3. \tag{12}$$

The value of α is found to be 2 and the parameter W satisfies the cubic equation

$$W^3-2W-3\lambda=0. \tag{13}$$

Using the coordinate transformation

$$X=(a+a^\dagger)/\sqrt{2W}, \tag{14}$$

$$P=-i[(a-a^\dagger)\sqrt{W}/\sqrt{2}], \tag{15}$$

and the commutation relation

$$[a,a^\dagger]=1, \tag{16}$$

where $a^\dagger$ stands for the creation operator and a stands for the annihilation operator, the expressions for H_{2N} and $H(W)$ are given below:

$$H_{2N}=2\sqrt{2\lambda}\,[a+a^\dagger/\sqrt{2W}]_N^3+(\lambda/4W^2)\,[a^4+(a^\dagger)^4+4a^\dagger a^3+4(a^\dagger)^3a] \tag{17}$$

and

$$H(W)=(2a^\dagger a+1)\left(\frac{W}{2}+\frac{1}{W}\right)+\frac{\lambda}{4W^2}[6(a^\dagger)^2a^2+12a^\dagger a+3]. \tag{18}$$

[a)] Address for correspondence: Biswanath Rath, c/o Satyapir Kar, at Sriram Nagar, Bhubaneswar-751002 (Orissa), India.

 0022-2488/90/051175-02$03.00

TABLE I. Ground state energy of the $H = P^2 + 2X^2 + 2\sqrt{2\lambda}\,X^3 + \lambda X^4$ oscillator.

λ	$E_0^{(0)}$	$E_0^{(1)}$	$E_0^{(2)}$	$E_0^{(3)}$	E_{total}	Previous[a]
0.0001	1.414 251	0	−0.000 13	0.000 000 002	1.414 1136	⋯
0.001	1.414 588	0	−0.001 37	0.000 000 26	1.413 216	1.413 211

[a]See Ref. 3.

In Eq. (17), $[\ ]_N$ stands for normal order. The explicit expressions for energy corrections for the ground state are

$$E_0^{(0)} = (8W + 9\lambda)/4W^2, \tag{19}$$

$$E_0^{(1)} = 0, \tag{20}$$

$$E_0^{(2)} = -\frac{9\lambda}{2W^4} - \frac{6\lambda}{W(12W + 27\lambda)} - \frac{3\lambda^2}{2W^2(16W + 42\lambda)}, \tag{21}$$

$$E_0^{(3)} = 18\lambda^2/W^4(12W + 27\lambda). \tag{22}$$

For weak coupling $\lambda << 1$, one can use Eqs. (19)–(22) for ground-state energy computation. However, for strong coupling one has to take more orders. For weak coupling $\lambda << 1$, the value of W used for energy calculation is positive, as it yields the correct limiting condition $\lambda \to 0$ and is given[5] by

$$W = (\tfrac{8}{3})^{1/2} \cos(\alpha_0/3), \tag{23}$$

where

$$\alpha_0 = \cos^{-1}(3\lambda/\sqrt{32/27}). \tag{24}$$

In Table I the ground-state energy is computed by $\lambda << 1$ and compared with the previous result. It is seen that for weak coupling $\lambda << 1$, present computation yields reasonably good agreement with previous computation.[3] For strong coupling the solutions of W can be found from the literature.[5]

Last, we want to test this approximation for an exactly solvable oscillator. The Hamiltonian considered here is given by

$$H = P^2/2 + \lambda X^4. \tag{25}$$

In this case only final expressions are given below. The expressions for $H(W)$ and H_{2N} are

$$H(W) = (2a^\dagger a + 1)(W/4) + (\lambda/4W^2)\,[6(a^\dagger)^2 a^2 + 12a^\dagger a + 3], \tag{26}$$

$$H_{2N} = (\lambda/4W^2)[a^4 + (a^\dagger)^4 + 4(a^\dagger)^3 a + 4a^\dagger a^3]. \tag{27}$$

The value of W is $(6\lambda)^{1/3}$, and α is found to be 2. The ground-state energy corrections up to third order are

$$E_0^{(0)} = 0.681\,42\lambda^{1/3}, \tag{28}$$

$$E_0^{(1)} = 0, \tag{29}$$

$$E_0^{(2)} = -0.010\,81\lambda^{1/3}, \tag{30}$$

$$E_0^{(3)} = 0. \tag{31}$$

The exact ground-state energy for this oscillator (up to five decimals) is $E_0 = 0.667\,75\lambda^{1/3}$. The present result up to third order is $0.670\,60\lambda^{1/3}$. However, the author feels if more and more orders are included then the exact result will be obtained. Further, the perturbation series in this context is also convergent. To show this we express the ground-state energy as the sum of different orders given by

$$E_0 = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} + \cdots + E_0^{(k)} + \cdots, \tag{32}$$

where $E_0^{(k)}$ is the Kth-order correction to the ground state. Using the argument of Halliday and Suranyi[2] we find each term of the perturbation series has an extra factor $\langle K|H_{2N}|r\rangle/E_0^{(0)} - E_r^{(0)}$, where $r = K \pm 2, K \pm 4$, and $E_r^{(0)}$ is the zeroth-order energy of the rth level of the unperturbed Hamiltonian $H(W)$. For $K \to \infty$ this ratio is either $\frac{2}{3}$ or $\frac{1}{6}$. Furthermore, considering the less convergent situation $r = K \pm 2$, we have the series

$$f(N)\,[1 - \tfrac{2}{3} + (\tfrac{2}{3})^2 + \cdots] + \text{finite terms},$$

where $f(N)$ is a finite quantity and N is finite but large. The sum of the infinite terms in the bracket is $\frac{3}{5}$. Thus the series is a convergent one for any value of the coupling constant. The main advantage of this method is that the energy level expression for the oscillator can be calculated analytically. For weak coupling constants $\lambda << 1$, one needs to calculate a few order perturbation corrections. In the present paper numerical results are calculated using a CASIO-fx-82 scientific calculator.

ACKNOWLEDGMENT

I am thankful to the referee who suggested improvements to all facets of the work discussed in this manuscript. Suggestions of this kind are always welcome.

[1]C. M. Bender and T. T. Wu, Phys. Rev. Lett. **21**, 406 (1968); Phys. Rev. **184**, 1231 (1969); B. Simon, Ann. Phys. (NY) **58**, 76 (1969).

[2]I. G. Halliday and P. Suranyi, Phys. Rev. D **21**, 1529 (1980), and references cited therein.

[3]K. Banerjee and S. P. Bhatnagar, Phys. Rev. D **18** (12), 4767 (1978).

[4]B. Rath, Rad. Eff. **80**, 183 (1984).

[5]G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1968), p. 23.

Bounds on the number of bound states for potentials with critical decay at infinity

Joachim Stubbe
Département de Physique Théorique, Université de Genève, 24, quai Ernest-Ansermet, CH-1211 Genève, Switzerland

(Received 28 June 1989; accepted for publication 15 November 1989)

Let V be a potential whose negative part V_- decays like $c|x|^{-2}$ at infinity. If c is not too large, then the Schrödinger operator $H_V = -\Delta + V$ on $\mathbb{R}^n$, $n \geqslant 3$, has only a finite number of bound states although the associated classical phase space volume is infinite. Optimal conditions are derived for the absence of bound states and a family of bounds on the total number of bound states for the operator H_V of the form $N_0(V) \leqslant c_{\gamma,n} \int_{\mathbb{R}^n} (V + c/|x|^2)_-^{\gamma} |x - x_0|^{2\gamma - n} d^n x$. The basic tool used to obtain results is a family of sharp Sobolev inequalities for the operator $A(c) = -\Delta - c/|x|^2$.

I. INTRODUCTION. MAIN RESULTS

A basic problem in the study of Schrödinger operators is to obtain conditions on the potential V in order to guarantee that this potential has at most N bound states. Let $N_0(V)$ be the number of bound states of the Schrödinger operator $H_V = -\Delta + V$ on $\mathbb{R}^n$. It is well known that under suitable conditions on V the number of bound states is asymptotically determined by the classical phase space volume associated with V in the sense

$$\lim_{\lambda \to \infty} \frac{N_0(\lambda V)}{\int_{\mathbb{R}^n} |\lambda V_-|^{n/2} d^n x} = (2\pi)^{-n} B_n. \tag{1}$$

Here V_- denotes the negative part of V and B_n is the volume of the n-dimensional unit ball. For a proof of (1), we refer to Ref. 1.

Formula (1) suggests that we look for a bound of the form

$$N_0(V) \leqslant L_{0,n} \int_{\mathbb{R}^n} |V_-|^{n/2} d^n x, \tag{2}$$

provided the integral on the right-hand side exists, i.e., $V_- \in L^{n/2}$. Here we should remark that in view of the inequality $N_0(V) \leqslant N_0(-V_-)$ it is always sufficient to obtain a bound in terms of the negative part of the potential.[2]

Estimates of the type (2) have been found by Lieb,[3] Cwickel,[4] and Rosenbljum,[5] and an optimal condition that guarantees $N_0(V) = 0$ has been derived by Glaser *et al.*[6] The latter paper also contains a large family of nonexistence conditions involving different powers of the potentials under the integral, which lead, e.g., to the bound of Courant–Hilbert[7] or Bargmann.[8] Such bounds on $N_0(V)$ have been obtained later by Glaser, Grosse, and Martin[9] and more recently by Blanchard, Stubbe, and Rezende,[10] improving a method proposed by Li and Yau.[11]

The validity of the asymptotic formula (1) as well as the inequality (2) indicate that the finiteness (resp. infiniteness) of $N_0(V)$ is determined by the decay rate of the potential at infinity. In fact, we have $N_0(V) < \infty$ if $V(x) \geqslant -c/|x|^{2+\varepsilon}$, while $N_0(V)$ is infinite if $V(x) \leqslant -c/|x|^{2-\varepsilon}$ for some positive numbers c and ε.[12] The borderline case is given by the condition

$$\lim_{|x| \to \infty} V_-(x)|x|^2 = c. \tag{3}$$

In this case, inequality (2) is inconclusive since the right-hand side is infinite. However, if $n \geqslant 3$ and c satisfies the condition

$$0 \leqslant c < ((n-2)/2)^2, \tag{4}$$

then $N_0(V)$ is finite, as we shall also see from the results of this paper, provided V satisfies certain conditions to be specified later. The number of bound states of H_V is infinite if the reversed inequality holds. The aim of the present paper is to present very simple conditions on potentials V having the critical decay given in (3) such that $N_0(V)$ in finite, and to give good estimates on $N_0(V)$ in terms of an integral norm. The basic tool for the derivation of our results is a family of sharp Sobolev inequalties for the operator

$$A(c) = -\Delta - c/|x|^2 \quad \text{on } \mathbb{R}^n, \quad 0 \leqslant c < ((n-2)/2)^2. \tag{5}$$

Theorem 1: Let $n \geqslant 3$. For each non-negative $c < ((n-2)/2)^2$, there exists a constant $K_{n,2n/(n-2)}(c)$ such that, for any $f \in H^1(\mathbb{R}^n)$, the following inequality holds:

$$\|f\|^2_{2n/(n-2)} \leqslant K^2_{n,2n/(n-2)}(c)(\|\nabla f\|_2^2 - c\|(1/|x|)f\|_2^2). \tag{6}$$

The optimal constant $K_{n,2n/(n-2)}(c)$ is given by

$$K_{n,2n/(n-2)}(c) = [\pi n(n-2)]^{-1/2}(\Gamma(n)/\Gamma(n/2))^{1/n} \times (1 - 4c/(n-2)^2)^{-(n-1)/2n}. \tag{7}$$

The first result on H_V gives an optimal condition for the absence of bound states as in the work of Glaser *et al.*[6]

Theorem 2: Let V be a potential satisfying the conditions (3) and (4). If V satisfies the inequality

$$K^n_{n,2n(n-2)}(c) \int_{\mathbb{R}^n} \left(V(x) + \frac{c}{|x|^2}\right)_-^{n/2} d^n x < 1, \tag{8}$$

then $N_0(V) = 0$.

Our second result on H_V is the following inequality on the number of its bound states.

Theorem 3: Let V be a potential satisfying the conditions (3) and (4), then we have

$$N_0(V) \leqslant e^{(n-2)/2} K^n_{n,2n/(n-2)}(c) \times \int_{\mathbb{R}^n} \left(V(x) + \frac{c}{|x|^2}\right)_-^{n/2} d^n x, \tag{9}$$

 0022-2488/90/051177-04$03.00

with $K_{n,2n/(n-2)}(c)$ the constant obtained in Theorem 1.

In Sec. II the proofs of these theorems will be presented. Since we believe that the techniques should be more or less well known we shall only sketch the main lines of the proofs. There is a useful generalization of the Sobolev-type inequality (6) which will imply generalizations of Theorems 2 and 3. In order to avoid confusion we shall state these results separately in Sec. III. Finally, in Sec. IV, we shall discuss our results.

II. PROOFS OF THEOREMS 1–3

Proof of Theorem 1: First of all, we observe that the operator $A(c)$ satisfies the elementary inequality

$$\begin{aligned}(f, A(c)f)_2 &\equiv \|\nabla f\|_2^2 - c\|(1/|x|)f\|_2^2 \\ &> (1 - 4c/(n-2)^2)\|\nabla f\|_2^2, \end{aligned} \tag{10}$$

which follows from the "local uncertainty principle lemma" (see, e.g., Ref. 13). Therefore we may apply the technique of Lieb, who gave a simple rearrangement inequality proof of (6) in the case $c = 0$.[14] We have to prove the existence of a maximizing function of the functional

$$S_c(f) = \|f\|_{2n/(n-2)}^2 / (f, A(c)f)_2, \tag{11}$$

and we define $K_{n,2n/(n-2)}(c)$ by

$$K_{n,2n/(n-2)}^2(c) = \sup\{S_c(f) \mid f \in H^1(\mathbb{R}^n),\ f \neq 0\}. \tag{12}$$

As proved in Ref. 14 we may restrict ourselves to functions that are spherically symmetric and decreasing in $|x|$. The coordinate change

$$F([(n-2)/2]u) = e^{[(n-2)/2]u} f(e^u), \quad u \in \mathbb{R}, \tag{13}$$

will give us

$$S_c(f) = (nB_n)^{-2/n}((n-2)/2)^{-2(n-1)/n} \times \frac{\|F\|_{2n/(n-2)}^2}{\|F'\|_2^2 + (1 - 4c/(n-2)^2)\|F\|_2^2}, \tag{14}$$

where the norms for F are defined by integrals over the real axis. The functional of F on the right-hand side of (14) can be easily maximized. In particular, rescaling F by $F_\sigma(x) = F(x/\sigma)$, with $\sigma = \sigma_c = (1 - 4c/(n-2)^2)^{1/2}$, we arrive at the same functional for F as in Ref. 14, which is independent of c:

$$S_c(f) = (nB_n)^{-2/n}(n - 2/2)^{-2(n-1)/n}(1 - 4c/(n-2)^2)^{-(n-1)/n} \times \|F\|_{2n/(n-2)}^2 / (\|F'\|_2^2 + \|F\|_2^2). \tag{14'}$$

Using Theorem 4.2 of Ref. 14 we get

$$K_{n,2n/(n-2)}^2(c) = K_{n,2n/(n-2)}^2(0)\,(1 - 4c/(n-2)^2)^{-(n-1)/n}, \tag{15}$$

where $K_{n,2n/(n-2)}(0)$ is the constant obtained in Ref. 14 for the case $c = 0$. As a maximizing function for $S_c(f)$ we may choose

$$\begin{aligned} f(|x|) &= |x|^{-[(n-2)/2](1-(1-4c/(n-2)^2)^{1/2})} \\ &\quad \times (1 + |x|^{2(1-4c/(n-2)^2)^{1/2}})^{-(n-2)/2} \\ &= |x|^{-[(n-2)/2](1-\sigma_c)}(1 + |x|^{2\sigma_c})^{-(n-2)/2}. \end{aligned} \tag{16}$$

Remark 2.1: Note that if $c < 0$ in (6) we cannot apply the rearrangement technique and we can show that there is no maximizer for $S_c(f)$. However, $S_c(f)$ is still bounded and the sharp constant is given by $K_{n,2n/(n-2)}(0)$. On the other hand, if we restrict our attention to spherically symmetric functions f, then we can easily prove the existence of a maximizer given by (16). Note that the maximizing f is not decreasing in $|x|$ and $f(0) = 0$. Again the sharp constant is given by formula (15) [resp. formula (7)].

Proof of Theorems 2 and 3: The basic idea is to relate the number of bound states $N_0(V)$ to the eigenvalues of the following boundary value problem on a bounded domain D in $\mathbb{R}^n$:

$$\begin{aligned} &A(c)f(x) = \lambda q(x)f(x) \quad \text{on} \quad D \subset \mathbb{R}^n, \\ &f|_{\partial D} = 0, \end{aligned} \tag{17}$$

where q is a positive function such that $q \in L^{n/2}(\mathbb{R}^n)$ and $A(c)$ is the operator defined in (5).

The eigenvalue problem (17) has an infinite sequence of eigenvalues

$$0 < \lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_k \to \infty. \tag{18}$$

By a standard reduction argument of Birman and Schwinger (see, e.g., Refs. 2 and 11) it can be shown that the number of eigenvalues λ_k that are less than 1 is equal to $N_0(V)$.

To prove Theorem 2 we choose $q(x) = (V(x) + c/|x|^2)_-$ and apply the Sobolev-type inequality of Theorem 1 to the first normalized eigenfunction f_1 of (17). More precisely, we have

$$\begin{aligned} 1 &= \int_D f_1^2\, q\, d^n x \\ &\leqslant \left(\int_D f_1^{(2n/(n-2))}\, d^n x\right)^{(n-2)/n} \left(\int_D q^{n/2}\, d^n x\right)^{2/n} \\ &\leqslant K_{n,2n/(n-2)}^2(c)(f_1, A(c)f_1)_2 \left(\int_D q^{n/2}\, d^n x\right)^{2/n} \\ &= \lambda_1 K_{n,2n/(n-2)}^2(c) \left(\int_D \left(V(x) + \frac{c}{|x|^2}\right)_-^{n/2} d^n x\right)^{2/n}, \end{aligned} \tag{19}$$

which implies the nonexistence condition stated in Theorem 2.

Theorem 3 will be proved by the method of Blanchard, Stubbe, and Rezende.[10] The idea is to study the heat kernel

$$H_0(x,y,t)=\sum_{i=1}^{\infty} e^{-\lambda_i t} f_i(x) f_i(y) \tag{20a}$$

and the "weighted" heat kernel

$$H_{-1}(x,y,t)=\sum_{i=1}^{\infty} (2\lambda_i)^{-1} e^{-\lambda_i t} f_i(x) f_i(y) \tag{20b}$$

associated to the operator $A(c)$ in $L^2(D, q\,d^n x)$, where $(f_i)_{i\in\mathbb{N}}$ is the set of normalized eigenfunctions of (17). Starting from

$$\begin{aligned} h_{-1}(t) &\equiv \sum_{i=1}^{\infty} (2\lambda_i)^{-1} e^{-2\lambda_i t} \\ &= \int_D \int_D H_{-1}(x,y,t) H_0(x,y,t) \\ &\quad \times q(x) q(y) d^n x\, d^n y, \end{aligned} \tag{21}$$

we arrive, after some simple steps involving Hölder's inequality and the Sobolev-type inequaltiy of Theorem 1, at the following differential inequality for $h_{-1}(t)$:

$$\begin{aligned} h_{-1}(t) &\leqslant \tfrac{1}{2} K^2_{n,2n/(n-2)}(c) \\ &\quad \times \left(\int_D q^{n/2}(x) d^n x\right)^{2/n} \left(-\frac{dh_{-1}(t)}{dt}\right)^{(n-2)/n}. \end{aligned} \tag{22}$$

Indeed, we write

$$\begin{aligned} h_{-1}(t) &= \int_D \int_D H_{-1}(x,y,t) [H_0(x,y,t) q(y)^{(n+2)/4}]^{2/n} \\ &\quad \times [H_0(x,y,t)^2 q(y)]^{(n-2)/2n} d^n y\, q(x) d^n x \end{aligned}$$

and apply Hölder's inequality with exponents $2n/(n-2)$, $n/2$, and $2n/(n-2)$ for the y integration and then with exponents 2, n, and $2n/(n-2)$ for the x integration to obtain

$$\begin{aligned} h_{-1}(t) &\leqslant \left[\int_D q(x) \left(\int_D |H_{-1}(x,y,t)|^{2n/(n-2)} d^n y\right)^{(n-2)/n} d^n x\right]^{1/2} \\ &\quad \times \left[\int_D q(x) \left(\int_D H_0(x,y,t) q(y)^{(n+2)/4} d^n y\right)^2 d^n x\right]^{1/n} \\ &\quad \times \left[\int_D q(x) \left(\int_D H_0(x,y,t)^2 q(y) d^n y\right) d^n x\right]^{(n-2)/2n}. \end{aligned}$$

From (21), it is easy to see that the last term equals $(-dh_{-1}/dt)^{(n-2)/2n}$. As in Ref. 11, we can estimate the second term of the above expression by $(\int_D q(x)^{n/2}\,dx)^{1/n}$ using the fact that it is decreasing in t, hence bounded by its value at $t=0$, which can be evaluated by using the completeness of the eigenfunctions. To estimate the first term we apply the Sobolev inequality of Theorem 1 and the identity

$$\begin{aligned} \frac{1}{2} h_{-1}(t) &= \int_D q(x) \int_D |\nabla_y H_{-1}(x,y,t)|^2 \\ &\quad - c \left|\frac{H_{-1}(x,y,t)}{|y|}\right|^2 d^n y\, d^n x, \end{aligned}$$

which gives the bound

$$K_{n,2n/(n-2)}(c) \cdot (\tfrac{1}{2} h_{-1}(t))^{1/2}.$$

Combining these estimates we obtain inequality (22). Integrating and optimizing with respect to t yields

$$\lambda_k^{n/2} \int_D q^{n/2}\, d^n x \geqslant k\, K^{-n}_{n,2n/(n-2)}(c) e^{(2-n)/2}, \tag{23}$$

which implies Theorem 3 by the Birman–Schwinger principle since

$$N_0(V) \leqslant N_0(-(V + c/|x|^2)_- - c/|x|^2).$$

III. GENERALIZATIONS OF THE MAIN RESULTS

Generalizations of Theorems 2 and 3 are consequences of the following generalized Sobolev-type inequality for the operator $A(c)$.

Theorem 1': Let $n \geqslant 3$ and $(2-n)/2 \leqslant b \leqslant 1$. Define

$$p = \frac{2n}{2b+n-2}, \quad \gamma = \frac{p}{p-2}, \quad \sigma_c = \left(1 - \frac{4c}{(n-2)^2}\right)^{1/2}.$$

(a) If $0 \leqslant c < ((n-2)/2)^2$ and $0 \leqslant b \leqslant 1$, then there exists a constant $K_{n,p}(c)$ such that, for any $f \in H^1(\mathbb{R}^n)$, the following inequality holds:

$$\| |x|^{-b} f \|_p^2 \leqslant K^2_{n,p}(c) (\|\nabla f\|_2^2 - c \|(1/|x|) f\|_2^2).$$

(b) If either $(2-n)/2 \leqslant b < 0$ and $c < [(n-2)/2]^2$ or $c < 0$, then there exists a constant $K_{n,p}(c)$ such that

$$\| |x|^{-b} f \|_p^2 \leqslant K^2_{n,p}(c) (\|\nabla f\|_2^2 - c \|(1/|x|) f\|_2^2)$$

holds for any spherically symmetric function $f \in H^1(\mathbb{R}^n)$.

In any case the optimal constant $K_{n,p}(c)$ is given by

$$\begin{aligned} K_{n,p}(c) &= K_{n,p}(0)\, \sigma_c^{-1/4 - 1/2p} \\ &= \omega_{n-1}^{-1/2\gamma} (n-2)^{1/2\gamma - 1} \\ &\quad \times \left(\frac{\gamma - 1}{\gamma}\right)^{(\gamma-1)/2\gamma} \left(\frac{\Gamma(2\gamma)}{\Gamma(\gamma+1)\Gamma(\gamma)}\right)^{1/2\gamma} \\ &\quad \times \sigma_c^{(1/2)(1/2\gamma - 1)}, \end{aligned} \tag{24}$$

where $\omega_{n-1} = 2\pi^{n/2}/\Gamma(n/2)$ is the area of the unit sphere in $\mathbb{R}^n$. If $(2-n)/2 < b < 1$, in each case (a) and (b) there is a "maximizing" function of the form

$$f(|x|) = |x|^{-[(n-2)/2](1-\sigma_c)}\,(1+|x|^{[(n-2)/\gamma]\sigma_c})^{1-\gamma}. \tag{25}$$

Now, the generalizations of Theorems 2 and 3 can be stated as follows.

Theorem 2': (a) Let $x_0 \in \mathbb{R}^n$ and $\gamma \geqslant n/2$. Let V be a potential satisfying the decay conditions (3) and (4). If

$$K^{2\gamma}_{n,p}(c) \int_{\mathbb{R}^n} \left(V(x) + \frac{c}{|x|^2}\right)_-^{\gamma} |x-x_0|^{2\gamma-n}\, d^n x < 1, \tag{26a}$$

then H_V has no bound state.

(b) Let $1 \leqslant \gamma < n/2$ and V be a spherically symmetric potential satisfying the decay conditions (3) and (4). If

$$K^{2\gamma}_{n,p}(c) \int_{\mathbb{R}^n} \left(V(x) + \frac{c}{|x|^2}\right)_-^{\gamma} |x|^{2\gamma-n}\, d^n x < 1, \tag{26b}$$

then $N_0(V) = 0$.

Theorem 3': (a) Let γ and V satisfy the conditions of Theorem 2' (a). Then

$$N_0(V) \leqslant K^{2\gamma}_{n,p}(c) e^{\gamma-1} \times \int_{\mathbb{R}^n} \left(V(x) + \frac{c}{|x|^2}\right)_-^{\gamma} |x-x_0|^{2\gamma-n}\, d^n x. \tag{27a}$$

(b) Let γ and V satisfy the conditions of Theorem 2'(b). Then the number of spherically symmetric bound states $N^s_0(V)$ satisfies the estimate

$$N^s_0(V) \leqslant K^{2\gamma}_{n,p}(c) e^{\gamma-1} \int_{\mathbb{R}^n} \left(V(x) + \frac{c}{|x|^2}\right)_-^{\gamma} |x|^{2\gamma-n}\, d^n x. \tag{27b}$$

IV. DISCUSSION

First of all we want to show that for any given p the nonexistence condition of Theorem 2' (resp. Theorem 2) is optimal in the following sense: The constant $K^2_{n,p}(c)$ cannot be replaced by a smaller number. Indeed, let us choose V as

$$V(x) = -c|x|^{-2} - |x|^{-2+[(n-2)/(\gamma-1)]\sigma_c} \times (1+|x|^{[(n-2)/(\gamma-1)]\sigma_c})^{-2}, \tag{28}$$

If $1 < \gamma < \infty$ (i.e., $\infty > p > 2$). In the case $\gamma = n/2$ or, equivalently, $p = 2n/(n-2)$, this reads as

$$V(x) = -c|x|^{-2} - |x|^{-2(1-\sigma_c)}(1+|x|^{2\sigma_c})^{-2}. \tag{28'}$$

It can be easily verified that the optimizing function for the Sobolev inequality given by (25) is a solution of the Schrödinger equation for H_V with eigenvalue zero. Furthermore the chain of inequalities in (19) (resp. in its generalized form) reduces to equalities for this choice of f_1 and V which proves our claim. For the values $p = 2$ and $p = \infty$ this follows by a simple limiting argument.

From (28) we see that the optimal potential always has in inverse square singularity at the orgin. The reduced potential $q(x) = (V(x) + c/|x|^2)_-$ is singular at the origin with a power weaker than the inverse square potential since $0 < \sigma_c < 1$, if $0 < c < ((n-2)/2)^2$. Now $q(x)$ is regular at the origin if $c = 0$. At infinity $q(x)$ behaves like $|x|^{-2(1+\sigma_c)}$.

The bound on $N_0(V)$ presented in Theorem 3 (resp. Theorem 3') is not optimal and the factor $e^{(n-2)/2}$ (resp. $e^{\gamma-1}$) should not be present. However, in the important case $n = 3$ it provides us with very good results. The case $\gamma = 1$ also yields an optimal bound on $N_0(V)$, which is a generalization of Bargmann's bound.[8,10,12]

Theorems 2 and 3 have some nice consequences. Results in this direction can also be found, e.g., in Ref. 12. Let $V \in L^{n/2}_{\mathrm{loc}}(\mathbb{R}^n)$, $n \geqslant 3$, and

$$V(x) \geqslant -c|x|^{-2}, \quad \begin{matrix} \text{if } |x| > R \text{ for some } R, \\ 0 < c < ((n-2)/2)^2. \end{matrix} \tag{29}$$

Then $N_0(V) < \infty$. In particular, if (29) holds for all x, then $N_0(V) = 0$. On the other hand, if $c > ((n-2)/2)^2$, then $N_0(V)$ is generally not bounded as shown in Ref. 15.

Finally, it should be stressed that the results of this paper also apply to potentials whose negative part behaves like $|x|^{-2}$ at the origin.

ACKNOWLEDGMENTS

I am very grateful to H. Grosse, A. Martin, and L. Vazquez for the stimulating discussions at CERN.

This work was supported by the Swiss National Science Foundation.

[1] A. Martin, Helv. Phys. Acta **45**, 140 (1972).
[2] B. Simon, in *Studies in Mathematical Physics*, edited by E. H. Lieb, B. Simon, and A. S. Wightman (Princeton U.P., Princeton, NJ, 1976), pp. 305–326.
[3] E. H. Lieb, Bull. Am. Math. Soc. **82**, 751 (1976); Proc. Symp. Pure Math. **36**, 241 (1980).
[4] M. Cwickel, Ann. Math. **106**, 93 (1977).
[5] G. V. Rosenbljum, Dokl. Akad. Nauk SSR **202**, 1012 (1972).
[6] V. Glasser, H. Grosse, A. Martin, and W. Thirring, in Ref. 2, pp. 169–194.
[7] R. Courant and D. Hilbert, *Methods of Mathematical Physics I* (Interscience, New York, 1953), p. 446.
[8] V. Bargmann, Proc. Natl. Acad. Sci. USA **39**, 961 (1952).
[9] V. Glaser, M. Grosse, and A. Martin, Commun. Math. Phys. **59**, 197 (1978).
[10] Ph. Blanchard, J. Stubbe, and J. Rezende, Lett. Math. Phys. **14**, 215 (1987).
[11] P. Li and S.-T. Yau, Commun. Math. Phys. **88**, 309 (1983).
[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV* (Academic, New York, 1978), p. 87.
[13] M. Reed and B. Simon, *Methods of Modern Mathematical Physics II* (Academic, New York, 1975), p. 169.
[14] E. H. Lieb, Ann. Math. **118**, 349 (1983).
[15] W. Kirsch and B. Simon, Ann. Phys. (NY) **183**, 122 (1988).

On obtaining the scattering potential and its moments in the partial wave Born approximation

H. A. Mavromatis and A. M. Al-Jalal
Physics Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

(Received 7 July 1989; accepted for publication 29 November 1989)

Mathematical techniques are used to invert the the partial wave Born approximation. Two expressions are obtained for the scattering potential. Additionally simple expressions are given for various moments of this scattering potential. These results are illustrated by some analytic examples. Finally a few numerical calculations are performed using both "Born approximation" and "experimental" phase shifts in the relevant expressions.

I. FORMAL EXPRESSIONS FOR $V_l(r)$ AND ITS MOMENTS

To solve the inverse problem in quantum mechanical scattering theory one must use the Gel'fand–Levitan Marchenko equation.[1] Alternatively it is possible to invert one of the simplest approximations in quantum mechanical scattering theory, the partial wave Born approximation,[2] which relates a particular integral of the scattering potential $V_l(r)$ to the phase shift $\delta_l(k)$ that this potential produces, namely,

$$-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\approx\int_0^\infty j_l^2(kr)\,V_l(r)r^2\,dr. \tag{1}$$

[Various authors use $\sin(\delta_l(k))$ or $\tan(\delta_l(k))$[2] instead of $\delta_l(k)$.]

This is known to be a fairly good approximation for channels where the potential is "weak" in some sense, i.e., the phase shifts produced are small.

Here E is the energy of the mass m scattered particle, $k=\sqrt{2mE/\hbar^2}$, and $V_l(r)$ is the scattering potential in the angular momentum l channel. The function $j_l(kr)$ is the usual spherical Bessel function.[3]

In the literature[4] a set of functions $g_l(kr)$ was obtained such that

$$\int_0^\infty j_l^2(kr)g_l(kr')dk=\frac{r'^2}{r^2}\{\delta(r-r')+\delta(r+r')-(-1)^l2\delta(r')\}. \tag{2}$$

[Reference 4 overlooks the $\delta(r+r')$ and $\delta(r')$ terms in expression (2). Though the former term does not affect subsequent results, the latter does.] The first three g_l functions and the corresponding j_l's are listed in Table I for the convenience of the reader.

Premultiplying Eq. (1) by $g_l(kr')$ and then integrating over k one obtains with the help of Eq. (2) that

$$\int_0^\infty\left(-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right)g_l(kr')dk=r'^2V_l(r')-2(-1)^lr'^2\delta(r')\int_0^\infty V_l(r)dr$$

or

$$V_l(r')=\frac{1}{r'^2}\int_0^\infty\left(-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right)g_l(kr')dk+2(-1)^l\delta(r')\int_0^\infty V_l(r)dr. \tag{3}$$

The second term on the rhs of Eq. (3) appears to make this inversion approach impractical since if one wishes to evaluate the potential from the phase shifts in this approximation the "unknown" potential also appears on the rhs of this expression.

This difficulty does not arise however if one wishes to obtain "moments" of the potential. Thus

$$\int_0^\infty r'^nV_l(r')dr'=\int_0^\infty r'^{n-2}\int_0^\infty\left(-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right)g_l(kr')dk\,dr', \tag{4}$$

TABLE I. Spherical Bessel functions $j_l(\rho)$, the $g_l(\rho)$ functions, and $\int[g_l(\rho)/\rho^2]d\rho$, for $l=0$, 1, and 2.

l	0	1	2
$j_l(\rho)$	$\frac{\sin(\rho)}{\rho}$	$\frac{\sin(\rho)}{\rho^2}-\frac{\cos(\rho)}{\rho}$	$\left(\frac{3}{\rho^3}-\frac{1}{\rho}\right)\sin(\rho)-\frac{3}{\rho^2}\cos(\rho)$
$g_l(\rho)$	$-\frac{8\rho^2}{\pi}\cos(2\rho)$	$\frac{8}{\pi}\left\{(\rho^2-2)\cos(2\rho)+\left(\frac{1}{\rho}-2\rho\right)\sin(2\rho)\right\}$	$-\frac{8}{\pi}\left\{\left(\rho^2-18+\frac{36}{\rho^2}\right)\cos(2\rho)+\left(-6\rho+\frac{33}{\rho}-\frac{18}{\rho^3}\right)\sin(2\rho)\right\}$
$\int\frac{g_l(\rho)}{\rho^2}d\rho$	$-\frac{4}{\pi}\sin(2\rho)$	$\frac{4}{\pi}\left\{\left(1-\frac{1}{\rho^2}\right)\sin(2\rho)+\frac{2}{\rho}\cos(2\rho)\right\}$	$-\frac{4}{\pi}\left\{\left(1-\frac{15}{\rho^2}+\frac{9}{\rho^4}\right)\sin(2\rho)+\left(\frac{6}{\rho}-\frac{18}{\rho^3}\right)\cos(2\rho)\right\}$

 0022-2488/90/051181-08$03.00

where $n>0$, since the integral $\int_0^\infty r'^n\delta(r')dr'$ vanishes.

A second approach to inverting Eq. (1) involves using a formula that we have derived, and can be easily verified for $l=0$, 1, 2, and 3, namely,

$$\left(\frac{d}{d\rho^2}\right)^l \frac{d}{d\rho}\left(\frac{d}{d\rho^2}\right)^l (\rho^{l+1}j_l(\rho))^2 = \sin(2\rho), \tag{5}$$

where

$$\frac{d}{d\rho^2} \equiv \frac{1}{2\rho}\frac{d}{d\rho}.$$

A derivation of Eq. (5) for general l is given in Appendix B.

Using this expression and Eq. (1) one obtains

$$\left(\frac{d}{dk^2}\right)^l \frac{d}{dk}\left(\frac{d}{dk^2}\right)^l \left(k^{2l+2}\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\right)$$
$$= \int_0^\infty \sin(2kr)V_l(r)r^{2l+1}dr, \tag{6}$$

which is just the Fourier sine transform[5] of $(1/2)V_l(r)r^{2l+1}$.

Multiplying both sides of Eq. (6) by $(4/\pi)\sin(2kr')$ and integrating over k one then obtains

$$V_l(r') = \frac{4}{\pi r'^{2l+1}}\int_0^\infty \left(\frac{d}{dk^2}\right)^l \frac{d}{dk}\left(\frac{d}{dk^2}\right)^l \times\left(k^{2l+2}\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\right)\sin(2kr')dk, \tag{7}$$

since

$$\frac{4}{\pi}\int_0^\infty \sin(2kr)\sin(2kr')dk = \delta(r-r')+\delta(r+r').$$

Equation (7) appears a more useful inversion expression than Eq. (3) because there is no unknown term on the rhs of Eq. (7). Unlike Eq. (3) however, this equation requires knowledge of derivatives of the phase shifts.

Thus

$$V_0(r) = \frac{4}{\pi r}\int_0^\infty \left(\frac{d}{dk}\left(k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right\}\right)\right)\sin(2kr)dk, \tag{8a}$$

$$V_1(r) = \frac{4}{\pi r^3}\int_0^\infty \left(\left(\frac{d}{dk^2}\right)\frac{d}{dk}\left(\frac{d}{dk^2}\right)\left(k^4\left\{-\frac{\hbar^2}{2m}\frac{\delta_1(k)}{k}\right\}\right)\right) \times\sin(2kr)dk, \tag{8b}$$

and so on for larger l.

Expressions (3) and (7) are in fact related as is shown in Appendix A. A new expression for $g_l(\rho)$ that results from examining this relationship between Eq. (3) and Eq. (7) is also given in this appendix.

Examing Eq. (1) for large k, in which limit this approximation is most appropriate, one obtains using the expression $\lim_{\rho\to\infty} j_l(\rho)\to\sin(\rho-l\pi/2)/\rho$ that as $k\to\infty$

$$\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\to\frac{1}{k^2}\int_0^\infty \sin^2(kr-l\pi/2)V_l(r)dr$$
$$= \frac{1}{2k^2}\int_0^\infty (1+(-1)^{l+1}\cos(2kr))V_l(r)dr.$$

Thus

$$\lim_{k\to\infty} k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\} = \frac{1}{2}\int_0^\infty V_l(r)dr, \tag{9}$$

since the integral $\int_0^\infty \cos(2kr)V_l(r)dr$ goes to zero as k tends to infinity due to the rapid oscillations of $\cos(2kr)$ in this limit.

Thus the troublesome terms on the rhs of Eq. (3) can be replaced by known quantities:

$$V_l(r') = \frac{1}{r'^2}\int_0^\infty \left(-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right)g_l(kr')dk + 4(-1)^l\delta(r') \times\lim_{k'\to\infty} k'^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k')}{k'}\right\}.$$

The functions $g_l(kr)$ have the property

$$\int_0^\infty \frac{g_l(kr)}{k^2r^2}dk = 4(-1)^{l+1}\delta(r),$$

a result which can easily be verified for $l=0$, 1, and 2 and which is discussed in Appendix C. Using this expression one can also write Eq. (3) in the form

$$V_l(r') = \frac{1}{r'^2}\int_0^\infty \left(-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k} - \frac{1}{k^2} \times\lim_{k'\to\infty} k'^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k')}{k'}\right\}\right)g_l(kr')dk. \tag{10}$$

One thus has two useful expressions, Eq. (7) or Eq. (10) for inverting the partial wave Born approximation.

Equation (4) involves double integrals for the evaluation of moments of the potential $V_l(r)$. An analogous double integral expression may be obtained starting from Eq. (7), namely,

$$\int_0^\infty r'^n V_l(r')dr' = \frac{4}{\pi}\int_0^\infty\int_0^\infty \frac{\sin(2kr')}{r'^{2l+1-n}}dr'\left(\left(\frac{d}{dk^2}\right)^l \times\frac{d}{dk}\left(\frac{d}{dk^2}\right)^l\left(k^{2l+2}\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\right)\right)dk. \tag{11}$$

Simpler expressions for some moments of the scattering potential can also be obtained as follows. From the Born approximation (1) one has obtained above an explicit expression for the integral of the potential over r since from Eq. (9)

$$\int_0^\infty V_l(r)dr = \lim_{k\to\infty} 2k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}. \tag{12}$$

If, on the other hand, one wishes to evaluate certain higher moments of V_l, for example the volume integral of V_l, $\int_0^\infty r^2V_l(r)dr$, etc., one may have recourse to Eq. (6).

In the limit $k\to 0$, $\sin(2kr)\to 2kr$ and one has

$$\int_0^\infty r^{2l+2}V_l(r)dr = \lim_{k\to 0}\frac{1}{2k}\left(\frac{d}{dk^2}\right)^l \frac{d}{dk}\left(\frac{d}{dk^2}\right)^l \times\left(k^{2l+2}\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\right). \tag{13}$$

More generally, for $n = 0,1,2,\ldots$,

$$\int_0^\infty r^{2(l+1+n)} V_l(r)\,dr = \lim_{k\to 0} \frac{(-1)^n}{2k}\left(\frac{1}{2}\frac{d}{dk}\right)^{2n}\left(\frac{d}{dk^2}\right)^l \frac{d}{dk}\left(\frac{d}{dk^2}\right)^l \times\left(k^{2l+2}\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\right). \tag{14}$$

Expressions for arbitrary moments follow from Eq. (11). Thus, for instance, substituting $n = 2l$ in Eq. (11) yields:

$$\int_0^\infty r^{2l} V_l(r')\,dr' = 2\int_0^\infty \left(\left(\frac{d}{dk^2}\right)^l \frac{d}{dk}\left(\frac{d}{dk^2}\right)^l \times\left(k^{2l+2}\left\{-\frac{\hbar^2}{2m}\frac{\delta_l(k)}{k}\right\}\right)\right)dk. \tag{15}$$

II. ANALYTICAL EXAMPLES

Some analytical examples that illustrate Eqs. (8a), (8b), (12), (13), (14), and (15) follow.

A. Example 1

Given

$$V_0(r) = V_0, \quad r<a,$$
$$= 0, \quad r>a.$$

From Eq. (1)

$$-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k} \approx V_0\left\{\frac{a}{2k^2} - \frac{\sin(2ka)}{4k^3}\right\},$$

$$k^2\left(-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right) \approx V_0\left(\frac{a}{2} - \frac{\sin(2ka)}{4k}\right),$$

$$\frac{d}{dk}k^2\left(-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right) \approx \frac{V_0}{4}\left(\frac{\sin(2ka)}{k^2} - \frac{2a\cos(2ka)}{k}\right).$$

Substituting this expression into Eq. (8a) and integrating over k,

$$V_0(r) \approx \frac{1}{r}\frac{4}{\pi}\frac{V_0}{4}\int_0^\infty \left(\frac{\sin(2ka)}{k^2} - \frac{2a\cos(2ka)}{k}\right) \times\sin(2kr)\,dk$$
$$= \frac{V_0}{r\pi}\left\{\int_0^\infty \frac{\sin(2ka)\sin(2kr)}{k^2}\,dk - a\int_0^\infty \frac{\sin(2k(r-a)) + \sin(2k(r+a))}{k}\,dk\right\}$$
$$= (V_0/r\pi)((\pi/2)2a - \pi a) = 0, \quad r>a$$
$$= (V_0/r\pi)((\pi/2)2r) = V_0, \quad r<a.$$

On the other hand, from Eq. (12)

$$\int_0^\infty V_0(r)\,dr = \lim_{k\to\infty} 2k^2\left(-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right) \approx \lim_{k\to\infty} 2k^2 V_0\left(\frac{a}{2k^2} - \frac{\sin(2ka)}{4k^3}\right) = V_0 a,$$

while from Eq. (13)

$$\int_0^\infty r^2 V_0(r)\,dr = \lim_{k\to 0}\frac{1}{2k}\frac{d}{dk}k^2\left(-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right)$$
$$\approx \lim_{k\to 0}\frac{1}{2k}\frac{V_0}{4}\left(\frac{\sin(2ka)}{k^2} - \frac{2a\cos(2ka)}{k}\right)$$
$$= \lim_{k\to 0}\frac{V_0}{8k}\left(\frac{2ka}{k^2} - \frac{8k^3a^3}{6k^2} - \frac{2a}{k} + 4k^2\frac{a^3}{k} + \cdots\right)$$
$$= \frac{V_0 a^3}{3}.$$

Finally from Eq. (14) with $n = 1$

$$\int_0^\infty r^4 V_0(r)\,dr$$
$$= \lim_{k\to 0} -\frac{1}{2k}\left(\frac{1}{2}\frac{d}{dk}\right)^2 \frac{d}{dk}k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right\}$$
$$\approx \lim_{k\to 0} -\frac{V_0}{32k}\left(\frac{d}{dk}\right)^2\left(\frac{\sin(2ka)}{k^2} - \frac{2a\cos(2ka)}{k}\right)$$
$$= \lim_{k\to 0}\frac{V_0}{16k^2}\left\{\frac{\sin(2ka)}{k}\left(6a^2 - \frac{3}{k^2}\right) - 2a\cos(2ka)\left(2a^2 - \frac{3}{k^2}\right)\right\}$$
$$= \lim_{k\to 0}(V_0/16k^2)\{-8k^2a^5 - \tfrac{4}{3}k^2a^5 + 8k^2a^5 + 4k^2a^5 + \cdots\}$$
$$= V_0 a^5/5.$$

B. Example 2

Given $V_0(r) = V_0 e^{-\alpha r^2}$.

From Eq. (1)

$$-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k} \approx \frac{V_0}{4k^2}\sqrt{\frac{\pi}{\alpha}}(1 - e^{-k^2/\alpha}),$$

$$\frac{d}{dk}k^2\left(-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right) \approx \frac{V_0}{2\alpha}\sqrt{\frac{\pi}{\alpha}}ke^{k^2/\alpha}.$$

Substituting this expression in Eq. (8a):

$$V_0(r) \approx \frac{4}{\pi r}\int_0^\infty \frac{V_0}{2\alpha}\sqrt{(\pi/\alpha)}e^{-k^2/\alpha}\sin(2kr)\,dk$$
$$= V_0 e^{-\alpha r^2}.$$

On the other hand, from Eq. (12)

$$\int_0^\infty V_0(r)dr = \lim_{k\to\infty} 2k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right\}$$
$$\approx \lim_{k\to\infty} 2k^2(V_0/4k^2)\sqrt{(\pi/\alpha)}(1-e^{-k^2/\alpha})$$
$$= (V_0/2)\sqrt{\pi/\alpha},$$

while from Eq. (13)

$$\int_0^\infty r^2V_0(r)dr = \lim_{k\to 0}\frac{1}{2k}\frac{d}{dk}k^2\left(-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right)$$
$$\approx \lim_{k\to 0}\frac{1}{2k}\frac{V_0}{2\alpha}\sqrt{\frac{\pi}{\alpha}}ke^{-k^2/\alpha}$$
$$= (V_0/4\alpha)\sqrt{\pi/\alpha}.$$

Finally from Eq. (14) with $n=1$

$$\int_0^\infty r^4V_0(r)dr$$
$$= \lim_{k\to 0} -\frac{1}{2k}\left(\frac{1}{2}\frac{d}{dk}\right)^2\frac{d}{dk}k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_0(k)}{k}\right\}$$
$$\approx \lim_{k\to 0}\frac{V_0}{16k\alpha}\sqrt{\frac{\pi}{\alpha}}\left(\frac{d}{dk}\right)^2 ke^{-k^2/\alpha}$$
$$= \lim_{k\to 0}\frac{V_0}{8\alpha^2}\sqrt{\frac{\pi}{\alpha}}\left\{3-\frac{2k^2}{\alpha}\right\}e^{-k^2/\alpha}$$
$$= (3V_0/8\alpha^2)\sqrt{\pi/\alpha}.$$

C. Example 3

Given $V_1(r) = V_1r^2e^{-\alpha r^2}$.

From Eq. (1)

$$-\frac{\hbar^2}{2m}\frac{\delta_1(k)}{k} \approx \frac{V_1}{4}\sqrt{\frac{\pi}{\alpha}}\frac{1}{k^4}\left\{1+\frac{k^2}{2\alpha} - e^{-k^2/\alpha}\left(1+\frac{3k^2}{2\alpha}+\frac{k^4}{\alpha^2}\right)\right\},$$

$$\frac{d}{dk^2}\frac{d}{dk}\frac{d}{dk^2}k^4\left(-\frac{\hbar^2}{2m}\frac{\delta_1(k)}{k}\right)$$
$$\approx \frac{V_1}{4}\sqrt{\frac{\pi}{\alpha}}e^{-k^2/\alpha}\left(\frac{15k}{2\alpha^3}-\frac{10k^3}{\alpha^4}+\frac{2k^5}{\alpha^5}\right).$$

Substituting into Eq. (8b)

$$V_1(r) \approx \frac{4}{\pi r^3}\int_0^\infty \frac{V_1}{4}\sqrt{\frac{\pi}{\alpha}}e^{-k^2/\alpha}\left(\frac{15k}{2\alpha^3}-\frac{10k^3}{\alpha^4}+\frac{2k^5}{\alpha^5}\right) \times\sin(2kr)dk$$
$$= V_1r^2e^{-\alpha r^2}.$$

On the other hand, from Eq. (12)

$$\int_0^\infty V_1(r)dr = \lim_{k\to\infty} 2k^2\left\{-\frac{\hbar^2}{2m}\frac{\delta_1(k)}{k}\right\}$$
$$\approx \lim_{k\to\infty}\frac{V_1}{2}\sqrt{\frac{\pi}{\alpha}}\frac{1}{k^2}\left\{1+\frac{k^2}{2\alpha} - e^{-k^2/\alpha}\left(1+\frac{3k^2}{2\alpha}+\frac{k^4}{\alpha^2}\right)\right\}$$
$$= (V_1/4\alpha)\sqrt{\pi/\alpha},$$

while from Eq. (15)

$$\int_0^\infty r^2V_1(r)dr$$
$$\approx 2\int_0^\infty \frac{V_1}{4}\sqrt{\frac{\pi}{\alpha}}e^{-k^2/\alpha}\left(\frac{15k}{2\alpha^3}-\frac{10k^3}{\alpha^4}+\frac{2k^5}{\alpha^5}\right)dk$$
$$= (3V_1/8\alpha^2)\sqrt{\pi/\alpha},$$

and from Eq. (13)

$$\int_0^\infty r^4V_1(r)dr$$
$$= \lim_{k\to 0}\frac{1}{2k}\frac{d}{dk^2}\frac{d}{dk}\frac{d}{dk^2}\left(k^4\left\{-\frac{\hbar^2}{2m}\frac{\delta_1(k)}{k}\right\}\right)$$
$$\approx \lim_{k\to 0}\frac{V_1}{8k}\sqrt{\frac{\pi}{\alpha}}e^{-k^2/\alpha}\left(\frac{15k}{2\alpha^3}-\frac{10k^3}{\alpha^4}+\frac{2k^5}{\alpha^5}\right)$$
$$= (15V_1/16\alpha^3)\sqrt{\pi/\alpha},$$

The expressions (8a), (8b), (12), (13), (14), and (15) all give exact results in the above three examples because the Born approximation phase shifts and their analytic derivatives were used for each potential, and the integrals were evaluated with k varying from 0 to ∞ while the limits were evaluated in a mathematically exact fashion from these analytic phase shifts. In practice, however, in an actual scattering problem, one extracts from the empirical cross section experimental rather than Born approximation phase shifts, which at most extend up to a certain maximum energy (or k) beyond which point the nonrelativistic formalism (hence also the Born approximation) breaks down. For proton–proton scattering this is about 350 MeV. The integrals (8a), (8b), and (15) must thus be truncated at some finite k. An additional error is invariably introduced when one takes derivatives of the experimental phase shifts.

The most interesting quantities to obtain are of course the potentials themselves since one can then evaluate any moment of the potential one is interested in. To do this one has to evaluate the integrals (8a) and (8b) despite the above caveats. Further discussion of these integrals is given in Sec. III. On the other hand, to evaluate Eq. (12) in a practical case requires that one estimate what limit $k\,\delta_l(k)$ tends to as k goes to infinity by looking at what happens to $k\,\delta_l(k)$ for large (but nonrelativistic) k's. One expects that one can estimate this limit rather well because the Born approximation improves at high energies. We illustrate this in Fig. 1 where we plot the "experimental" and Born $k\,\delta_0(k)$ for a nucleon of mass ≈ 1 amu scattered off two square well potentials of range 2 fm and depth 5 and 40 MeV, respectively (the "experimental" phase shifts are calculated by matching wave functions, obtained from the Schrödinger equation, and their derivatives at the edge of the square well potential). One sees from this figure that already at $k = 4$ fm both the Born and the "experimental" phase shifts are converging to the same limit $k\delta(k) \approx 0.24$ fm^{-1} for the $V = -5$ MeV potential and to $k\delta(k) \approx 1.9$ fm^{-1} for the $V = -40$ MeV potential. These correspond to values -10.1 and -79.9 MeV fm for $\int_0^\infty V_0(r)dr$, which are very close to the values of the corresponding integrals, namely -10.0 and -80.0 MeV fm, respectively. At the k approaches zero limit, which is

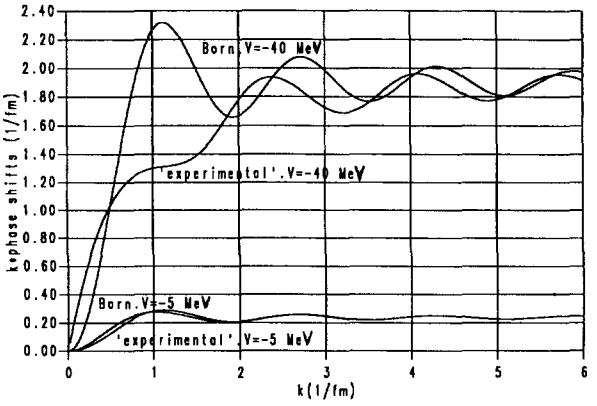


FIG. 1. $k\delta_0(k)$ calculated, as a function of k, from the Born approximation and the "experimental" phase shifts for a particle of mass 1 amu scattered from two different square well potentials of range 2 fm and depths 5 and 40 MeV, respectively.

needed in Eqs. (13) and (14), however, one does not expect that the Born approximation will be very good. Hence one expects different results using the experimental and the Born approximation expressions. That this is in fact the case is illustrated in Fig. 2 where for the above two potentials we plot $(1/2k)(d/dk)(k^2\{-(\hbar^2/2m)(\delta_0(k)/k)\})$. Here even for the weak $V=-5$ MeV potential the "experimental" value is $\approx 50\%$ greater in magnitude than the Born value for this quantity. Thus of the above set of expressions for reconstructing the potential the least reliable appear to be Eqs. (13) and (14).

III. SOME NUMERICAL RESULTS

In this section we show the results of some numerical calculations of $V_l(r)$ using Eq. (7) or equivalently Eq. (10). In particular, we consider scattering of a nucleon from a square well potential of depth V and range a. The first question that arises in this context is up to what $k=k_{\max}$ one must integrate these expressions. Analytically, if Born approximation phase shifts, Eq. (1), are inserted in Eqs. (7) or (10), the potential used in Eq. (1) to obtain these phase shifts will be recovered, provided the integration is performed over all k. Thus if the Born approximation phase shifts are used in Eqs. (7) or (10), the deviation from the square well shape of the resulting potential will reflect on the goodness of our calculations as a function of the cutoff $k_{\max}$.

Figures 3 and 4 show the resulting potential for $l=0$ when these integrals are carried out in steps of 0.001 fm^{-1} and k runs from 0.0001 fm^{-1} to $k_{\max}=2$ and 4 fm^{-1}, respectively. Comparison of these two figures shows that performing the integrals up to $k_{\max}=4$ fm^{-1} gives quite good results while integrating up to $k_{\max}=2$ fm^{-1} gives results which are reasonably good. Therefore in all our calculations the integrals for $V_l(r)$ are evaluated up to $k_{\max}=4$ fm^{-1}. For a nucleon, $k=4\,\text{fm}^{-1}$ corresponds to 333 MeV in kinetic energy which is about a third of the rest energy of the nucleon (931 MeV). This justifies our using nonrelativistic expressions, as for example the Born approximation itself and in obtaining exact phase shifts for the square well potential.

If, instead of Born approximation phase shifts, exact or so to speak "experimental" phase shifts, calculated by matching the exact wave function and its derivative at the edge of the square well potential, i.e., at $r=a$, are used, one does not expect to obtain the exact square shape either analytically or numerically from Eqs. (7) or (10). Rather the deviation from the "experimental" square shape, for a particular l tells us how good the Born approximation is for a particular l as a function of the depth and range of the potential, bearing in mind that some error is also introduced due to the fact that the integrals over k are truncated.

The "experimental" phase shifts produced by a nucleon scattered from three square well potentials of the same width, $a=2$ fm, and different depths, namely $V_0=-5$ MeV, -10 MeV, and -40 MeV, are shown in Figs. 5, 6, and 7, respectively.

The results of evaluating $V_l(r)$ using Eq. (10) for these three sets of phase shifts are shown in Figs. 8, 9, and 10. As expected, it can be seen from these three figures that the Born approximation is better for shallower potentials and

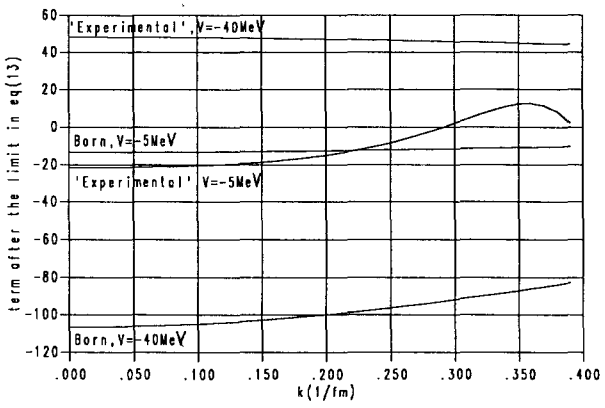


FIG. 2. The term after the limit in Eq. (13) $[-(1/2k)(\hbar^2/2m)(d/dk)(k\delta_0(k))]$ is calculated as a function of k by using the Born approximation and the "experimental" phase shifts for a particle of mass 1 amu scattered from two different square well potentials of range 2 fm and depths 5 and 40 MeV, respectively.

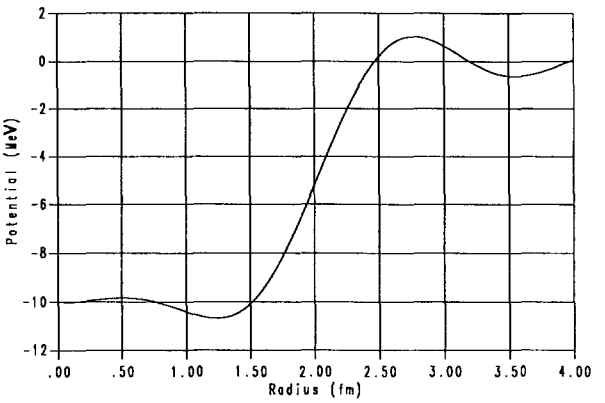


FIG. 3. Potential calculated from Eq. (10) using Born approximation $l=0$ phase shifts calculated for a nucleon scattered off a square well potential of depth 10 MeV and width 2 fm. The integral is performed from k equals 0.01 to 2 in steps of 0.001 fm^{-1}.

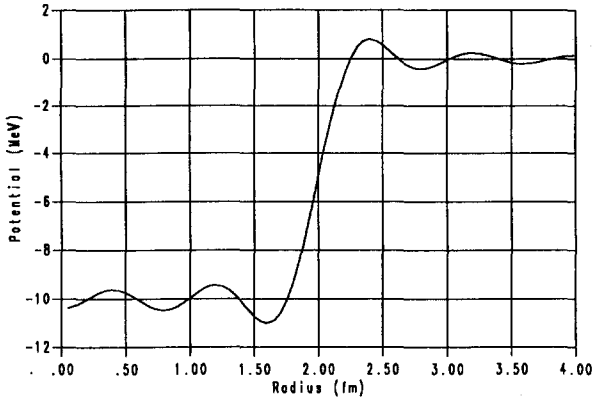


FIG. 4. Potential calculated from Eq. (10) using Born approximation $l = 0$ phase shifts calculated for a nucleon scattered from a square well potential of depth 10 MeV and width 2 fm. The integral is performed from k equals 0.01 to 4 in steps of 0.001 fm^{-1}.

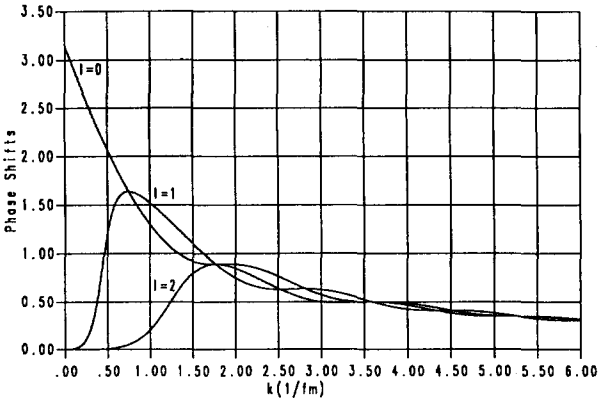


FIG. 7. "Experimental" phase shifts produced by a nucleon scattered from a square well potential of width 2 fm and depth 40 MeV.

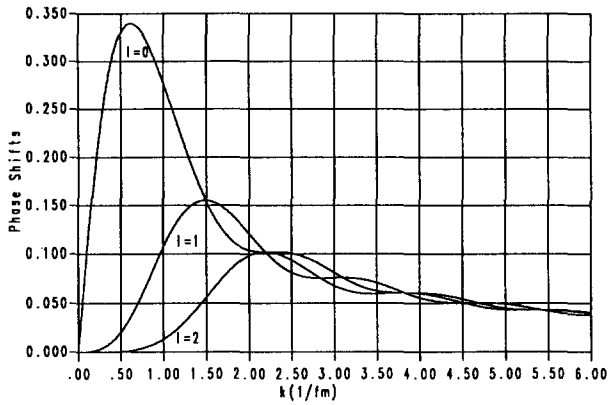


FIG. 5. "Experimental" phase shifts produced by a nucleon scattered from a square well potential of width 2 fm and depth 5 MeV.

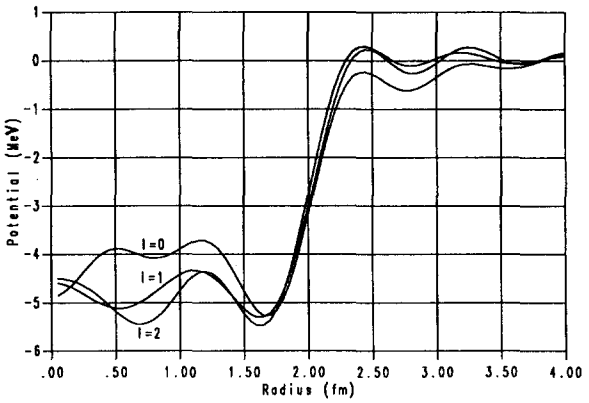


FIG. 8. Potentials calculated from Eq. (10) using "experimental" $l = 0$, 1, and 2 phase shifts calculated for a nucleon scattered off a square well potential of depth 5 MeV and width 2 fm. The integrals are performed from k equals 0.01 to 4 in steps of 0.001 fm^{-1}.

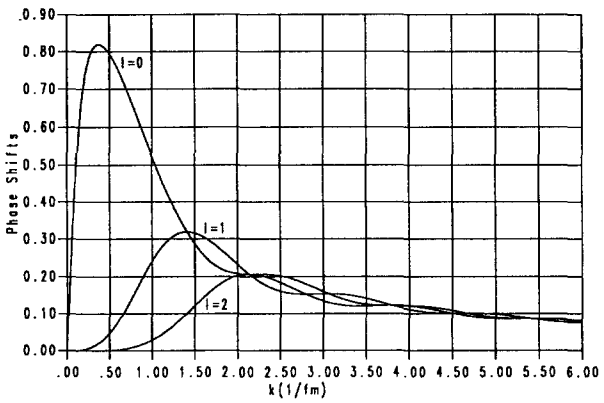


FIG. 6. "Experimental" phase shifts produced by a nucleon scattered from a square well potential of width 2 fm and depth 10 MeV.

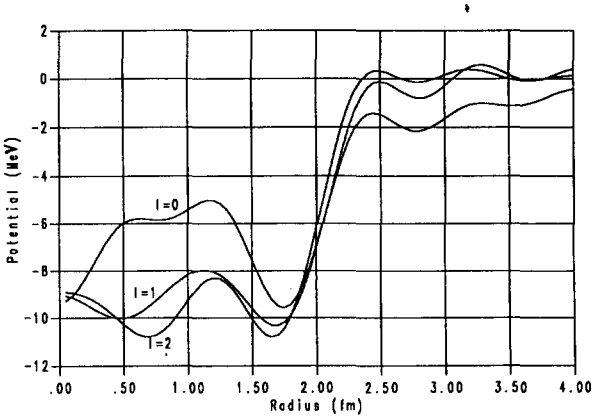


FIG. 9. Potentials calculated from Eq. (10) using "experimental" $l = 0$, 1, and 2 phase shifts calculated for a nucleon scattered off a square well potential of depth 10 MeV and width 2 fm. The integrals are performed from k equals 0.01 to 4 in steps of 0.001 fm^{-1}.

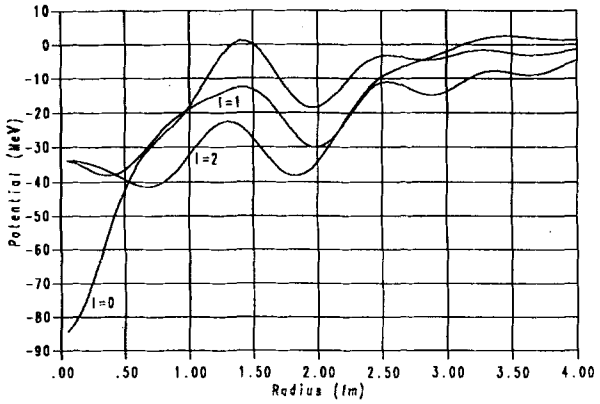


FIG. 10. Potentials calculated from Eq. (10) using "experimental" $l = 0, 1$, and 2 phase shifts calculated for a nucleon scattered off a square well potential of depth 40 MeV and width 2 fm. The integrals are performed from k equals 0.01 to 4 in steps of 0.001 fm^{-1}.

higher angular momenta (smaller phase shifts). Thus even if a given potential produces $l = 0$ phase shifts, which make the Born approximation extremely inaccurate for this channel, one gets increasingly more accurate results for the potential as one goes to higher l channels. Also if the potential is a function of l (and possibly additionally of j), inverting the Born approximation may give a reasonable result for the potential in higher l channels even if it is not applicable say for $l = 0$.

Though the integrands in expressions (7) and (10) are different, we have verified in detail for the case $l = 0$ that they generate essentially the same potential $V(r)$ when we integrate out to $k_{\max} = 4$ fm^{-1}. The above techniques may easily be extended to two-body scattering by using the corresponding phase shifts, reduced masses, and relative energies.

APPENDIX A

Integrating Eq. (7) successively by parts one obtains

$$V_l(r') = \frac{(-1)^{l+1}}{r'^{2l+1}} \int_0^\infty \left\{ -\frac{\hbar^2}{2m} \frac{\delta_l(k)}{k} \right\} \left\{ k^{2l+2} \left(\frac{\partial}{\partial k} \frac{1}{2k} \right)^l \times \frac{\partial}{\partial k} \left(\frac{\partial}{\partial k} \frac{1}{2k} \right)^l \frac{4}{\pi} \sin(2kr') \right\} dk + \text{surface terms.} \tag{A1}$$

Comparing (A1) to Eq. (3) one obtains

$$g_l(\rho) = -\frac{4}{\pi} \rho^{2l+2} \left(\frac{d}{d\rho} \frac{1}{2\rho} \right)^l \frac{d}{d\rho} \left(\frac{d}{d\rho} \frac{1}{2\rho} \right)^l \sin(2\rho), \tag{A2}$$

while the surface terms in (A1) equal $2(-1)^l \delta(r') \int_0^\infty V_l(r) dr$.

The expression (A2) provides an alternative way to obtain $g_l(\rho)$ to those given in Ref. 4.

From (A2) it is clear that (for large r) the leading term of (A2) is $(8(-1)^{l+1}/\pi)\rho^2 \cos(2\rho)$.

APPENDIX B

From Ref. 6 ($k!$ is missing in this expression as given by Ref. 6)

$$j_l^2(\rho) = \frac{\pi}{2\rho} J^2_{l+1/2}(\rho) = \frac{\pi}{2\rho} \sum_{k=0}^\infty \frac{(-1)^k (\rho/2)^{2l+1+2k} \Gamma(2l+2k+2)}{\Gamma(2l+k+2)\Gamma(l+k+\frac{3}{2})^2 k!}. \tag{B1}$$

Using the duplication formula[7]

$$j_l^2(\rho) = 2^{2l} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{2l+2k} (l+k)!^2}{(2l+k+1)!(2l+2k+1)!k!}, \tag{B2}$$

$$\rho^{2l+2} j_l^2(\rho) = 2^{2l} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{4l+2k+2} (l+k)!^2}{(2l+k+1)!(2l+2k+1)!k!},$$

$$\frac{d}{2\rho\, d\rho} \rho^{2l+2} j_l^2(\rho) = 2^{2l} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{4l+2k+2-2} (2l+k+1)(l+k)!^2}{(2l+k+1)!(2l+2k+1)!k!},$$

$$\left(\frac{d}{2\rho\, d\rho} \right)^l \rho^{2l+2} j_l^2(\rho) = 2^{2l} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{2l+2k+2} (l+k)!^2}{(l+k+1)!(2l+2k+1)!k!}, \tag{B3}$$

$$\frac{d}{d\rho} \left(\frac{d}{2\rho\, d\rho} \right)^l \rho^{2l+2} j_l^2(\rho) = 2^{2l+1} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{2l+2k+1} (l+k)!}{(2l+2k+1)!k!}, \tag{B4}$$

$$\frac{d}{2\rho\, d\rho} \frac{d}{d\rho} \left(\frac{d}{2\rho\, d\rho} \right)^l \rho^{2l+2} j_l^2(\rho) = 2^{2l+1-1} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{2l+2k+1-2} (l+k)!(2l+2k+1)}{(2l+2k+1)!k!},$$

$$\left(\frac{d}{2\rho\, d\rho} \right)^l \frac{d}{d\rho} \left(\frac{d}{2\rho\, d\rho} \right)^l \rho^{2l+2} j_l^2(\rho) = 2^{l+1} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{2k+1} (l+k)!}{(2l+2k)(2l+2k-2)\cdots(2k+2)(2k+1)!k!}$$

$$= 2^{l+1} \sum_{k=0}^\infty \frac{(-1)^k 2^{2k} \rho^{2k+1} (l+k)!}{(l+k)!2^l (2k+1)!}, \tag{B5}$$

$$\left(\frac{d}{2\rho\, d\rho} \right)^l \frac{d}{d\rho} \left(\frac{d}{2\rho\, d\rho} \right)^l \rho^{2l+2} j_l^2(\rho) = \sum_{k=0}^\infty \frac{(-1)^k (2\rho)^{2k+1}}{(2k+1)!} = \sin(2\rho). \tag{B6}$$

APPENDIX C

The equation

$$\int_0^\infty \frac{g_l(kr)}{k^2r^2}\,dr = 4(-1)^{l+1}\delta(k) \tag{C1}$$

can be easily verified for $l=0$, 1, and 2 using expression (A2).

Thus substituting Eq. (A2) into Eq. (C1) one obtains

$$-\frac{4}{\pi}\int_0^\infty k^{2l}r^{2l}\left(\frac{\partial}{\partial r}\frac{1}{2r}\right)^l\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}\frac{1}{2r}\right)^l \sin(2kr)dr = 4(-1)^{l+1}\delta(k), \tag{C1'}$$

or

$$\frac{(-1)^l}{\pi 2^{2l}k^{2l+1}}\int_0^\infty r^{2l}\left(\frac{\partial}{\partial r}\frac{1}{r}\right)^l\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}\frac{1}{r}\right)^l \sin(2kr)dr = \delta(k). \tag{C2}$$

For $l=0$ the lhs of Eq. (C2) reduces to

$$\frac{1}{\pi k}\int_0^\infty \frac{\partial}{\partial r}\sin(2kr)dr = \frac{1}{\pi}\lim_{2r\to\infty}\frac{\sin(2kr)}{k},$$

a standard δ function representation.[8]

For $l=1$ the lhs of Eq. (C2) becomes

$$\begin{aligned}
&\frac{-1}{4\pi k^3}\int_0^\infty r^2\frac{\partial}{\partial r}\left\{\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}\frac{\sin(2kr)}{r}\right)\right\}dr \\
&= \frac{-1}{4\pi k^3}\left\{r\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}\frac{\sin(2kr)}{r}\right) - 2\frac{\partial}{\partial r}\frac{\sin(2kr)}{r}\right\}\Bigg|_0^\infty \\
&= \frac{-1}{4\pi k^3}\left\{\left(\frac{4}{r^2}-4k^2\right)\sin(2kr) - \frac{8k\cos(2kr)}{r}\right\}\Bigg|_0^\infty \\
&= \frac{1}{\pi}\lim_{2r\to\infty}\frac{\sin(2kr)}{k}.
\end{aligned}$$

Generally, the leading term of $g_l(kr) = [8(-1)^{l+1}/\pi]\,k^2r^2\cos(2kr)$, when substituted into Eq. (C1), yields the delta function while other terms cancel among themselves. Thus

$$\frac{8(-1)^{l+1}}{\pi}\int_0^\infty \cos(2kr)dr = 4(-1)^{l+1}\delta(k),$$

while

$$\int_0^\infty\left(\frac{g_l(kr)}{k^2r^2} - \frac{8(-1)^{l+1}}{\pi}\cos(2kr)\right)dr = 0.$$

The indefinite integrals corresponding to Eq. (C1) for $l=0$, 1, and 2 are given in Table I.

[1] F. Calogero and A. Degasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982).
[2] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), 2nd ed., p. 244.
[3] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1980), p. 966.
[4] H. A. Mavromatis and K. Schilcher, J. Math. Phys. 9, 1627 (1968).
[5] A. Erde'lyi *et al.*, *Tables of Integral Transforms* (Bateman Manuscript Project) (McGraw-Hill, New York, 1954), Vol. 1, p. 63.
[6] I. S. Gradshteyn and I. M. Ryzhik, Ref. 3, p. 960.
[7] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), p. 256.
[8] E. Merzbacher, Ref. 2, p. 84.

A path integral for a nonrelativistic spinor particle

Takashi Miura
Physics Department, Faculty of Engineering, Kanagawa University, Yokohama 221, Japan

(Received 25 October 1988; accepted for publication 10 January 1990)

By formulating a path integral in the SU(2) fiber bundle space, a nonrelativistic spinor particle moving in external electromagnetic fields is studied. This formulation is a direct extension of Schulman's [Phys. Rev. **176**, 1588 (1968)] continuous treatment for spin; it unifies all spin states to a spherical top model. This theory enjoys a definite classical theory with spin–orbital interactions and the corresponding semiclassical approximation for the path integral. This method is also applicable to a particle moving in a more general Yang–Mills field.

I. INTRODUCTION

As is well known, the amplitude for the arrival of a spinor particle moving in external electromagnetic fields is given by $\exp(iS)$, with S the classical action for a scalar particle; the orientation of the spin rotates at each point in its path at an angular velocity equal to the field strength at that point.[1] However, this solution remains incomplete unless the fields are constant in space because it is expressed as an infinite folding of the hypercomplex numbers. These are most easily shown by formulating the path integral solution for the Pauli equation in the form of the product integral and discussing its semiclassical approximation.[2,3]

On the other hand, spin was originally introduced in relation to the motion of a rigid rotator; however, it soon became clear that a rigid rotator must propagate all spins, unlike a spinor, whose spin value always remains fixed, i.e., $j=\frac{1}{2}$. When Bopp and Haag[4] showed the possibility of describing a spin by using the representation of the SU(2) group, a rigid rotator (a spherical top) model was again revived as a carrier of spin. Schulman[5] studied a path integral in the SU(2) [or SO(3)] group space as a continuous model for spin motion and obtained an exact propagator for the free motion. This interesting study suggested that the propagator of a spinor particle may be treated as a wave packet in the total space of the SU(2) fiber bundle whose fiber is the SU(2) group space and whose base space is the usual physical space or space-time. In this case the final result may be obtained by projecting it to the subspace of the fixed spin value $j=\frac{1}{2}$. However, without interactions, this projection may reduce the result for spin rotation only to an identity operation.[5]

This paper is aimed at studying a complete treatment of an interacting spinor in this multidimensional fiber bundle formulation. DeWitt[6] first formulated this bundle space as an arena of dynamics; since then it has been widely studied in the Kaluza–Klein theory[7] for gauge interactions. An important property of the Kaluza–Klein theory is the unification of all j states of the corresponding Yang–Mills theory. It is a result of this unification that the former has its specific classical theory with gauge interactions,[8] unlike the latter, whose path integral will take the form of the product integral.

In formulating a path integral in the bundle space, the boundary condition of paths in such a compact space as the SU(2) is a key problem to be solved; it has not yet been solved in its full generality of compact manifolds. As to the SU(2) fiber space, Schulman[5] already solved it for the free motion by reducing it to the periodicity around the one-dimensional maximal torus. The full periodicity about all the three Euler angles is to be considered for our interacting case, which can be done by referring to the result that Prokhorov[9] derived for n-dimensional spherical coordinates in a flat space. Another problem to be solved is to formulate the path integral in the physical space by taking the projected effect of the motion in the fibers into consideration. This problem can be solved by using the convenient path integral formulation originated by DeWitt-Morette[10]: Its convenience is due to the fact that it enables us to use formulas studied in stochastic integrals such as the Cameron–Martin formula. Application of the Cameron–Martin formula to our problem enables us to formulate a specific classical theory with spin–orbital interactions, as well as the usual electromagnetic interactions for a charged scalar particle.

The construction of this paper is as follows: In Sec. II we derive our multidimensional formulation from the usual Pauli one in such a way as to make their relationship clear. In Sec. III the path integral in the SU(2) fiber space is formulated by referring to the Prokhorov method applied to our problem, as given in Appendix A. We project the result to $j=\frac{1}{2}$ subspace in Sec. IV and apply the Cameron–Martin formula given in Appendix B to obtain the final result. Conclusions are given in Sec. V.

II. DERIVATION OF THE MULTIDIMENSIONAL FORMULATION

We begin with the usual Pauli equation for a nonrelativistic charged spinor particle moving in an external electromagnetic field. The Hamiltonian is then given as

$$\mathscr{H}=\frac{1}{2m}\left(p_k-\frac{e}{c}A_k\right)^2+eV+\frac{e\hbar}{2mc}(\sigma_k\cdot H_k) \tag{2.1}$$

and the corresponding Schrödinger equation is

$$i\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2m}\left(\partial_k+\frac{ie}{\hbar c}A_k\right)^2\psi + eV\psi+\frac{e\hbar}{2mc}(\sigma_k\cdot H_k)\psi, \tag{2.2}$$

 0022-2488/90/051189-09$03.00

where σ_k is the Pauli matrix and $H_k = (\mathrm{rot}\, A)_k$ is the external magentic field.

In order to obtain a solution of Eq. (2.2) via formally applied path integration, we use a product integral[2] of the type

$$\int \mathscr{D}(x)\exp\left\{\frac{i}{\hbar}\int\left[\frac{m}{2}\left(\frac{dx_k}{dt}\right)^2+\frac{e}{c}A_k\frac{dx_k}{dt}-eV - \frac{e\hbar}{2mc}(\sigma_k(t)\cdot H_k(x))\right]dt\right\}. \tag{2.3}$$

The sum (2.3) over all paths $x_k(t)$ of the hypercomplex amplitude cannot easily be evaluated for an arbitrary magnetic field (except the spatially constant one) because of the time-ordered operation of $\sigma_k(t)$.

Therefore, we reduce the Pauli formalism (2.2) to the continuous treatment for spin by introducing the unitary irreducible two-dimensional representation $D_{1/2}(g)$ of the group SU(2) in the C^2 spinor space.

First, we define the generator of the right translation of SU(2) as

$$V^{\alpha}_{(\beta)}(g) = \frac{\partial R(h,g)^{\alpha}}{\partial h^{\beta}} = \frac{\partial (gh)^{\alpha}}{\partial h^{\beta}}\bigg|_{h=e}. \tag{2.4}$$

Then $\hat{V}_k = V^{\alpha}_k(\partial/\partial g^{\alpha})$ satisfies

$$\hat{V}_k(g)D_{1/2}(g) = D_{1/2}(g)i\sigma_k/2, \tag{2.5}$$

where α labels the group coordinates and k runs over (x, y, z), which is prearranged to coincide with the label (β) of the group algebra at $h = e$. The proof[7] of Eq. (2.5) is roughly as follows: Consider the right translation of $g\in$SU(2) by a group element $e^{\theta^k Q_k}$ near the identity $h = e$. Then by the definition of V_k above, $V^{\alpha}_k(g)\theta^k = g^{\alpha}\theta^k Q_k$ and defining $D_{1/2}(g) = \langle p|g|q\rangle$ with $(i\sigma_k/2) = \langle q|Q_k|r\rangle$, we obtain $\hat{V}_k(g)D_{1/2}(g) = \langle p|gQ_k|r\rangle$.

If we apply $D_{1/2}$ to the lhs of Eq. (2.2) and define the probability amplitude $\Psi(g,x)$ for the initial state as

$$\Psi(g,x) = D_{1/2}(g)\psi(x,t), \tag{2.6}$$

Eq. (2.2) reduces to

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\partial_k+\frac{ie}{\hbar c}A_k\right)^2\Psi + eV\Psi - \frac{i\hbar e}{mc}H^k\hat{V}_k\Psi, \tag{2.7}$$

where $\hat{V}_k$ is proportional to the angular momentum operator $\hat{p}_k$ written with the group coordinate g, i.e.,

$$\hat{p}_k = -i\hbar\hat{V}_k. \tag{2.8}$$

The concrete representation of $\hat{V}_k$ can be obtained[11] directly from definition (2.4) by using the Euler angles (φ,θ,ψ):

$$\hat{V}_x = \cos\psi\frac{\partial}{\partial\theta}+\frac{\sin\psi}{\sin\theta}\frac{\partial}{\partial\varphi}-\cot\theta\sin\psi\frac{\partial}{\partial\psi},$$
$$\hat{V}_y = -\sin\psi\frac{\partial}{\partial\theta}+\frac{\cos\psi}{\sin\theta}\frac{\partial}{\partial\varphi}-\cot\theta\cos\psi\frac{\partial}{\partial\psi},$$
$$\hat{V}_z = \frac{\partial}{\partial\psi}. \tag{2.9}$$

Since we are analyzing Eq. (2.7) with path integration, we introduce the kinetic energy term in the SU(2) group space with the Lagrange multiplier method as

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\partial_k+\frac{ie}{\hbar c}A_k\right)^2\Psi + eV\Psi - \frac{2\hbar^2}{I}(\Delta_3+\lambda)\Psi - \frac{i\hbar e}{mc}H^k\hat{V}_k\Psi, \tag{2.10}$$

where I is the multiplier and takes the value of a quarter of the moment of inertia corresponding to Δ_3. Then a special solution of Eq. (2.10) that satisfies the constraint

$$(\Delta_3+\lambda)\Psi = 0 \tag{2.11}$$

solves Eq. (2.7).

As is well known, Δ_3 in Eq. (2.11) is given with the Euler angles as

$$\Delta_3 = \frac{\partial^2}{\partial\theta^2}+\cot\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\varphi^2}+\frac{\partial^2}{\partial\psi^2} - 2\cos\theta\frac{\partial^2}{\partial\varphi\,\partial\psi}\right) \tag{2.12}$$

and

$$\lambda = j(j+1), \tag{2.13}$$

where j is one of the values of $(0,\frac{1}{2},1,\frac{3}{2},..)$, i.e., for $j=\frac{1}{2}$, as in the case of (2.6), $\lambda=\frac{3}{4}$.

If we solve Eq. (2.10) in the form of the propagator, we will obtain a solution which involves not only the $j=\frac{1}{2}$ representation, but also all j-value representations. Then the constraint (2.11) with $\lambda=\frac{3}{4}$ means only projecting it to the subspace of $j=\frac{1}{2}$, i.e., Fourier transforming with $D^{*}_{1/2}(g)$ to the angular momentum representation corresponding to $j=\frac{1}{2}$. More strictly, this Fourier coefficient times $D_{1/2}(g)$ satisfies both Eqs. (2.10) and (2.11) and the former only solves Eq. (2.2), as is easily shown by tracing the above reduction backward.

For later convenience, here we introduce the group coordinates $(\vartheta_1,\vartheta_2,\vartheta_3)$ defined as

$$\vartheta_1 = \theta/z,\quad \vartheta_2 = (\psi+\varphi)/2,\quad \vartheta_3 = (\psi-\varphi)/2 \tag{2.14}$$

and rewrite the equations thus far formulated as follows.

For Ψ given by (2.6), we introduce a term $\exp(i2\hbar\lambda t/I)$ and set

$$\Psi(g,x,t) = D_{1/2}(g)e^{i2\hbar\lambda t/I}\psi(x,t). \tag{2.15}$$

Then Eq. (2.10) becomes

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\partial_k+\frac{ie}{\hbar c}A_k\right)^2\Psi + eV\Psi - \frac{\hbar^2}{2I}\Delta_{\vartheta}\Psi - \frac{i\hbar e}{mc}H^k\hat{V}_k\Psi, \tag{2.16}$$

with

$$(\Delta_{\vartheta}+3)\Psi = 0. \tag{2.17}$$

The Laplacian Δ_{ϑ} written with ϑ is given as

$$\Delta_{\vartheta} = \frac{\partial^2}{\partial\vartheta_1^2}+2\cot 2\vartheta_1\frac{\partial}{\partial\vartheta_1} + \frac{1}{\cos^2\vartheta_1}\frac{\partial^2}{\partial\vartheta_2^2}+\frac{1}{\sin^2\vartheta_1}\frac{\partial^2}{\partial\vartheta_3^2} \tag{2.18}$$

and $\hat{V}_k$ is given with ϑ_i as

$$\hat{V}_k = V^i_k\frac{\partial}{\partial\vartheta_i} = \frac{1}{2}g^i_k\frac{\partial}{\partial\vartheta_i}, \tag{2.19}$$

with

$$g^i_k = \begin{bmatrix} \cos(\vartheta_2+\vartheta_3) & \sin(\vartheta_2+\vartheta_3)\tan\vartheta_1 & -\sin(\vartheta_2+\vartheta_3)\cot\vartheta_1 \\ -\sin(\vartheta_2+\vartheta_3) & \cos(\vartheta_2+\vartheta_3)\tan\vartheta_1 & -\cos(\vartheta_2+\vartheta_3)\cot\vartheta_1 \\ 0 & 1 & 1 \end{bmatrix}. \tag{2.20}$$

III. PATH INTEGRAL IN THE SU(2) GROUP SPACE

The equation of motion (2.16) is the Schrödinger type formulated in the total space of the SU(2) fiber bundle. If we solve (2.16) with the initial condition (2.15) and project the solution to the subspace of $j=\frac{1}{2}$, we obtain the amplitude of a Pauli spinor that moves in an external electromagnetic field; it is also observed as the Pauli spinor of the same type. Therefore, we follow this procedure as our basic formalism for our problem in spite of the fact that it is not quite equivalent to the Pauli formalism unless we apply the projection [i.e., use the constraint (2.17)] to all of the unobserved intermediate states.

To obtain the solution of Eq. (2.16) via path integration, we must first obtain the classical Lagrangian corresponding to it, which can easily be done by rewriting (2.16) with the Hermitian momenta conjugate to ϑ_i (for notational simplicity we sometimes use the subscript, where, to be exact, one must use the superscript, as in ϑ^i), i.e.,

$$\hat{p}_1 = -i\hbar\frac{\partial}{\partial\vartheta_1} - i\hbar\cot 2\vartheta_1, \quad \hat{p}_2 = -i\hbar\frac{\partial}{\partial\vartheta_2},$$
$$\hat{p}_3 = -i\hbar\frac{\partial}{\partial\vartheta_3}. \tag{3.1}$$

For Eq. (2.16) we obtain

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{\mathscr{H}}\Psi, \tag{3.2}$$

where the Hamiltonian $\hat{\mathscr{H}}$ is given as

$$\hat{\mathscr{H}} = \frac{1}{2m}\left(\hat{p}_k - \frac{e}{c}A_k\right)^2 + eV + \frac{1}{2I}\left[\left(\hat{p}_1^2 - \frac{\hbar^2}{4}\right) + \frac{(\hat{p}_2^2-\hbar^2/4)}{\cos^2\vartheta_1} + \frac{(\hat{p}_3^2-\hbar^2/4)}{\sin^2\vartheta_1} - \frac{3}{4}\hbar^2\right] + \frac{e}{2mc}H^k[g^i_k\hat{p}_i]_t \tag{3.3}$$

with

$$[g^i_k\hat{p}_i]_t = \tfrac{1}{2}[\hat{p}_i g^i_k + g^i_k\hat{p}_i]. \tag{3.4}$$

From the classical Hamiltonian that corresponds to (3.3), we can easily obtain the Lagrangian as

$$L = L_0(x,t) + L_1(x,\vartheta,t), \tag{3.5}$$

where

$$L_0 = \frac{m}{2}\left(\frac{dxk}{dt}\right)^2 + \frac{e}{c}A_k\frac{dx_k}{dt} - eV, \tag{3.6}$$

$$L_1 = \frac{I}{2}g_{ij}\left(\frac{d\vartheta^i}{dt} - \frac{e}{2mc}H^k g^i_k\right)\left(\frac{d\vartheta^j}{dt} - \frac{e}{2mc}H^l g^j_l\right), \tag{3.7}$$

where g^j_l is the transverse of $g_k{}^i$ and

$$g_{ij} = (g^{ij})^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos^2\vartheta_1 & 0 \\ 0 & 0 & \sin^2\vartheta_1 \end{bmatrix}. \tag{3.8}$$

The formal path integral solution for Eq. (2.16) may then be written as

$$\int\int \mathscr{D}[x_k(t)]\mathscr{D}[\vartheta^i(t)]\exp\left\{\frac{i}{\hbar}\int L\,dt\right\} = \int \mathscr{D}(x_k)\left[\exp\left\{\frac{i}{\hbar}\int L_0\,dt\right\}\int\mathscr{D}(\vartheta) \times\exp\left\{\frac{i}{\hbar}\int L_1(x,\vartheta,t)dt\right\}\right]. \tag{3.9}$$

Therefore, we are the first to study the path integral in the group space

$$\int \mathscr{D}(\vartheta)\exp\left\{\frac{i}{\hbar}\int L_1(x,\vartheta,t)dt\right\}, \tag{3.10}$$

in which x is to be regarded as a parameter.

In order to formulate a path integral in a compact manifold as (3.10), the periodicity property of the boundary condition is an essential constituent; we study it here for the SU(2) group space. We are able to solve this problem in almost the same way as Prokhorov[9] did for the spherical coordinates in flat space. The detailed formulation of this problem is given in Appendix A. The main results are as follows.

We first discuss the periodicity of the kinetic energy part of the problem by beginning with Schulman's infinitesimal propagator for a free top[5] and defining the geodesic distance Γ using our variables:

$$\cos\Gamma = \cos\vartheta_1\cos\vartheta'_1\cos(\vartheta_2-\vartheta'_2) + \sin\vartheta_1\sin\vartheta'_1\cos(\vartheta_3-\vartheta'_3). \tag{3.11}$$

Then the corresponding action and the propagator itself have the same symmetry (or periodicity) as Γ, which can be understood as the invariance under the following four substitutions:

$$\vartheta_1,\vartheta_2,\vartheta_3 \to \vartheta_1+2\pi k, \quad \vartheta_2+2\pi l, \quad \vartheta_3+2\pi m,$$
$$\vartheta_1 \to -\vartheta_1+\pi, \quad \text{with} \quad \vartheta_2\to\vartheta_2+\pi,$$
$$\vartheta_1 \to -\vartheta_1, \quad \text{with} \quad \vartheta_3\to\vartheta_3+\pi,$$
$$\vartheta_1\to\vartheta_1+\pi, \quad \text{with} \quad \vartheta_2\to\vartheta_2+\pi, \quad \vartheta_3\to\vartheta_3+\pi, \tag{3.12}$$

where k, l, and m are integers.

Following Prokhorov,[9] we can directly write the prolongation function $Q_{SU(2)}$ for the full periodicity of ϑ by using the invariance under the substitutions (3.12) as

$$Q_{SU(2)}(\vartheta;\vartheta') = Q(\vartheta_1-\vartheta'_1)Q(\vartheta_2-\vartheta'_2)Q(\vartheta_3-\vartheta'_3) + Q(\vartheta_1+\vartheta'_1+\pi)Q(\vartheta_2-\vartheta'_2+\pi)Q(\vartheta_3-\vartheta'_3) + Q(\vartheta_1+\vartheta'_1)Q(\vartheta_2-\vartheta'_2)Q(\vartheta_3-\vartheta'_3+\pi) + Q(\vartheta_1-\vartheta'_1+\pi)Q(\vartheta_2-\vartheta'_2+\pi)Q(\vartheta_3-\vartheta'_3+\pi), \tag{3.13}$$

where

$$Q(\vartheta_i - \vartheta_i') = \sum_{k=-\infty}^{\infty} \delta(\vartheta_i - \vartheta_i' + 2\pi k). \tag{3.14}$$

Although we are not studying a free top, but instead a top with its potentials proportional to $\hat{V}_k|(\vartheta)$, we can fortunately use $Q_{SU(2)}$ as given above with no modifications because the $\hat{V}_k|(\vartheta)$ also do not change under the substitutions (3.12); this can be shown by a straightforward calculation.

Then the final results are written in the form of the phase space path integral as

$$\Psi(\vartheta) = \int_{-\infty}^{\infty} U(\vartheta,\vartheta',t)\sqrt{g'}\, d\vartheta' \times \int_{SU(2)} Q_{SU(2)}(\vartheta',\vartheta'')\Psi_0(\vartheta'')d\vartheta'', \tag{3.15}$$

where Ψ_0 is the initial value defined only on the finite region of the compact SU(2) space and

$$U(\vartheta,\vartheta',t) = g^{-1/4}g'^{-1/4} \int \prod_{\tau=0}^{t} \frac{d^3\vartheta^i d^3p_i}{(2\pi\hbar)^3} \times \exp\left\{\frac{i}{\hbar}\int [p_i\vartheta^i - H_{\text{eff}}(\vartheta,p)]dt\right\}. \tag{3.16}$$

The effective Hamiltonian H_{eff} is given as

$$H_{\text{eff}} = \frac{1}{2I}\left\{\left(p_1^2 - \frac{\hbar^2}{4}\right) + \frac{(p_2^2 - \hbar^2/4)}{\cos^2\vartheta_1} + \frac{(p_3^2 - \hbar^2/4)}{\sin^2\vartheta_1} - \frac{3}{4}\hbar^2\right\} - \frac{e}{2mc}H^k(x)g_k^i(\vartheta)p_i + \Delta H_{\text{eff}}, \tag{3.17}$$

where ΔH_{eff} is the effective potential originating from the operator ordering. As we can easily see from the Hamiltonian operator (3.3), ΔH_{eff} becomes zero if we take the Hermitian ordering given by (3.4) because there is no ordering problem in the kinetic part of our dymanics. (However, if we use the standard orderings, as Prokhorov[9] did, ΔH_{eff} becomes more complicated, i.e.,

$$\Delta H_{\text{eff}} = (ie\hbar/2mc)\cot 2\vartheta_1 \times [H_x\cos(\vartheta_2+\vartheta_3) - H_y\sin(\vartheta_2+\vartheta_3)].)$$

Therefore, we proceed from (3.16) to an equivalent, but more convenient formulation for analyzing our problem, i.e., the phase space path integral with the Gaussian integrator without a limiting procedure,[10] since it corresponds to the Hermitian ordering (3.3) if (3.17) is used as the effective classical Hamiltonian in this formulation. In Appendix B, there is a brief summary of this path integral, along with its relation to the corresponding Lagrangian formulation[12]: We apply it here to our problem and obtain the WKB approximation for formula (3.16) with (3.17) as

$$U_0(\vartheta,\vartheta';t) = w(\bar{\vartheta},\bar{p})\exp[iS(\bar{\vartheta},\bar{p})/\hbar] = w^L(\bar{\vartheta})\exp[iS^L(\bar{\vartheta})/\hbar], \tag{3.18}$$

where

$$w(\bar{\vartheta},\bar{p}) = w^L(\bar{\vartheta}) = (2\pi i\hbar/I)^{-3/2} \times g^{-1/4}[\det M_{\alpha\beta}(t,t')]^{1/2}g'^{-1/4} \tag{3.19}$$

and $\bar{\vartheta}$ is taken along the external paths from $\vartheta' = \vartheta(t')$ to $\vartheta = \vartheta(t)$. The superscript L indicates that quantities are obtained via the Lagrangian formulation. The VanVleck–Morette determinant $\det M_{\alpha\beta}$ is defined as

$$\det M_{\alpha\beta} = I^{-3}\det\frac{\partial^2\bar{S}^L}{\partial\vartheta^\alpha\partial\vartheta'^\beta} \tag{3.20}$$

and the classical action $\bar{S}^L = S^L(\bar{\vartheta})$ is obtained by integrating (3.7) along the extremal paths.

IV. PROJECTION TO THE SUBSPACE OF $j=\frac{1}{2}$ AND THE RESULTING WKB SOLUTION

In order to obtain the final result we have to eliminate the group coordinates by projecting to the subspace of $j=\frac{1}{2}$ and to approximate the path integral (3.9) further with respect to the space variable x_k. This projection of Fourier transforming with $D^*_{1/2}(\vartheta)$ to an angular momentum representation may be regarded as a dual procedure of (3.15) since the initial state $\Psi_0(\vartheta'')$ is equal to $D_{1/2}(\vartheta)$; therefore we treat them at the same time.

Before dealing with solution (3.18), we discuss this problem in a rather simple case, i.e., the infinitesimal WKB propagator given by

$$U_0(\epsilon) = (2\pi i\hbar\epsilon)^{-3/2}\exp[iS^L(\bar{\vartheta})/\hbar], \tag{4.1}$$

where

$$S^L = (I/2\epsilon)g_{ij}[\Delta\vartheta_i + (e/2mc)H^k g_k^i\epsilon] \times [\Delta\vartheta^j + (e/2mc)H^l g^j{}_l\epsilon], \tag{4.2}$$

with

$$\Delta\vartheta = \vartheta - \vartheta'.$$

Then we may integrate (3.15) in the form

$$\Psi(\vartheta) = \int\int\int_{-\infty}^{\infty} U_0(\epsilon)D_{1/2}(\vartheta')\sqrt{g'}\, d\vartheta_1'\, d\vartheta_2'\, d\vartheta_3' \tag{4.3}$$

since the prolongation operation $\int_{SU(2)} d\vartheta'' Q_{SU(2)}(\vartheta',\vartheta'')$ in (3.15) only prolongs $\Psi_0 = D_{1/2}(\vartheta'')$ from the compact SU(2) space to the infinite R^3. The evaluation of the integral (4.3) gives

$$\Psi(\vartheta) = D_{1/2}(\vartheta'_{\text{ex}}), \tag{4.4}$$

where

$$\vartheta'^i_{\text{ex}} = \vartheta^i + (e/2mc)H^k(x)g_k^i(\vartheta)\epsilon, \tag{4.5}$$

which is given by the condition

$$\left.\frac{\partial S^L(\bar{\vartheta})}{\partial\vartheta'}\right|_{\vartheta'=\vartheta'_{\text{ex}}} = 0, \tag{4.6}$$

making the classical action $\bar{S}^L(\vartheta'_{\text{ex}})$ also zero.

Then projecting $\Psi(\vartheta)$ by $D^*_{1/2}(\vartheta)$ we obtain

$$U_0(x,\epsilon)\Big|_{j=1/2} = \int_{SU(2)} D^*_{1/2}(\vartheta)D_{1/2}(\vartheta'_{\text{ex}}) \times\sqrt{g}\, d\vartheta_1\, d\vartheta_2\, d\vartheta_3.$$

Expanding $D_{1/2}(\vartheta'_{ex})$ around ϑ by using (4.5) we obtain the result

$$U_0(x,\epsilon)\Big|_{j=1/2} = 1 + \int_{SU(2)} \sqrt{g}\, d^3\vartheta\, D''_{1/2}(\vartheta) \times \frac{e\epsilon}{2mc} H^k g^j_k \frac{\partial}{\partial \vartheta_i} D_{1/2}(\vartheta) = 1 + \frac{ie}{2mc} H^k \sigma_k \epsilon = \exp\left[\frac{ie}{2mc} H^k \sigma_k \epsilon\right], \tag{4.7}$$

which leads us back to the Pauli-type theory. In order to confirm that our formulation is equivalent to the usual Pauli-type formulation as far as infinitesimal propagation is concerned, we need only show that the higher-order correction to the WKB solution (4.7) vanishes identically. In discussing the higher-order correction, we may put $H^k_{(x)} \equiv 0$ because the potential term only appears in the WKB solution since we use Hermitian ordering. Then the exact infinitesimal propagator can be written as

$$U(\epsilon) = \left(\frac{2\pi i\hbar\epsilon}{I}\right)^{-3/2} \exp\left\{\frac{i}{\hbar}\frac{I}{2\epsilon}\Gamma^2 + \frac{i\hbar\epsilon}{I}\right\},$$

where Γ is the infinitesimal geodesic distance between ϑ and ϑ' (Appendix A).

Then we obtain the following expression in place of (4.3):

$$\Psi(\vartheta) = \int_{-\infty}^{\infty}\int\int U(\epsilon) D_{1/2}(\vartheta')\sqrt{g'}\, d^3\vartheta'.$$

This can easily be integrated by introducing the normal coordinates and yields

$$\begin{aligned}\Psi(\vartheta) &= e^{i\hbar\epsilon/I}\left[D_{1/2}(\vartheta) + (i\hbar\epsilon/2I)\Delta_\vartheta D_{1/2}(\vartheta) - (i\hbar\epsilon/6I)R_G D_{1/2}(\vartheta)\right] \\ &= e^{i\hbar\epsilon/I} D_{1/2}(\vartheta)\left[1 + i\hbar\epsilon/2I\right] \\ &= e^{i3\hbar\epsilon/2I} D_{1/2}(\vartheta),\end{aligned}$$

where $R_G = 6$ is the scalar curvature of the group space. Here we take the initial phase factor $e^{i2\hbar\lambda t/I} = e^{i3\hbar t/2I}$ in (2.15) into consideration because it cannot be neglected in the higher-order discussion. Then (4.7) becomes

$$U_0(x,\epsilon)\Big|_{j=1/2} = \int_{SU(2)} D^*_{1/2}(\vartheta) D_{1/2}(\vartheta)\sqrt{g}\, d^3\vartheta \times e^{-i3\hbar(t-t')/2I} e^{i3\hbar\epsilon/2I} = 1,$$

which guarantees the vanishing of the higher-order correction for (4.7).

Now we return to solution (3.18) with (3.19). For $\Psi(\vartheta)$ we obtain, from (3.15),

$$\Psi(\vartheta) = \int_{-\infty}^{\infty}\int\int w^L(\vartheta)\exp\left[\frac{iS^L(\bar{\vartheta})}{\hbar}\right] D_{1/2}(\vartheta')\sqrt{g'}\, d^3\vartheta', \tag{4.8}$$

where the classical action is given by

$$S^L(\bar{\vartheta}) = \frac{I}{2}\int_{t'}^{t} g_{ij}\left(\frac{d\bar{\vartheta}^i}{dt} + \frac{e}{2mc} H^k g^i_k\right)\left(\frac{d\bar{\vartheta}^j}{dt} + \frac{e}{2mc} H^l g^j_l\right). \tag{4.9}$$

By using the identity

$$I^3[\det M_{\alpha\beta}]\left[\det \frac{\partial^2 S^L}{\partial\vartheta'^i \partial\vartheta'^j}\right]^{-1} = \left(\frac{g}{g'}\right)^{1/2}, \tag{4.10}$$

we obtain, via the stationary phase evaluation of (4.8),

$$\Psi(\vartheta) = \sum D_{1/2}(\vartheta'_{ex})\exp\left[\frac{iS^L(\bar{\vartheta}'_{ex})}{\hbar}\right], \tag{4.11}$$

where Σ means the sum over all extremal points ϑ'_{ex} of ϑ'. Then projecting with $D^*_{1/2}(\vartheta)$ we obtain

$$\begin{aligned}U_0(x,t)\Big|_{j=1/2} &= \int_{SU(2)} \sqrt{g}\, d^3\vartheta \sum_{\vartheta'_{ex}} D^*_{1/2}(\vartheta) \times \exp\left\{\frac{i}{\hbar}\bar{S}^L(\vartheta,\vartheta'_{ex})\right\} D_{1/2}(\vartheta'_{ex}) \\ &= \int_{-\infty}^{\infty}\int\int \sqrt{g}\, d^3\vartheta D^*_{1/2}(\vartheta) \times \exp\left\{\frac{i}{\hbar}\bar{S}^L(\vartheta,\vartheta'_{ex})\right\} D_{1/2}(\vartheta'_{ex})\end{aligned} \tag{4.12}$$

because the distribution of the extremal points ϑ'_{ex} for a fixed ϑ reflects the periodicity of the SU(2) space, i.e., (4.6) also satisfies the invariance under the substitutions (3.12).

Therefore, the result that is obtained via the stationary phase approximation applied to (4.12) is written as

$$\begin{aligned}U_0(x,t)|_{j=1/2} &= \sum_{\vartheta_{ex}} \left|\det \frac{\partial^2 \bar{S}^L(\vartheta,\vartheta'_{ex})}{\partial\vartheta^\alpha \partial\vartheta^\beta}\right|^{-1/2}_{\vartheta=\vartheta_{ex}} \times (2\pi\hbar i)^{3/2} g^{1/2} D^*_{1/2}(\vartheta_{ex}) \times \exp\left\{\frac{i}{\hbar}\bar{S}^L(\vartheta_{ex},\vartheta'_{ex})\right\} D_{1/2}(\vartheta'_{ex}) \\ &= \left(\frac{2\pi i\hbar}{I}\right)^{3/2} \sum g'^{1/4} (\det M_{\alpha\beta})^{-1/2} g^{1/4} \times D^*_{1/2}(\vartheta_{ex})\exp\left\{\frac{i\bar{S}^L(\vartheta_{ex},\vartheta'_{ex})}{\hbar}\right\} D_{1/2}(\vartheta'_{ex}),\end{aligned} \tag{4.13}$$

where Σ means the sum over the relative winding numbers between ϑ_{ex} and ϑ'_{ex}, which can be expressed almost in the same manner as the case of $Q_{SU(2)}$, with ϑ_{ex} (or ϑ'_{ex}) in (4.13) being replaced by the corresponding substituted periodic ones in (3.12). It is worth mentioning here that ϑ'_{ex} and ϑ_{ex} are determined by the conditions that the initial and the final classical spin angular momenta vanish, i.e.,

$$p'_2 = \frac{\partial\bar{S}^L}{\partial\vartheta'}\Big|_{\vartheta'_{ex}} = 0, \quad p_f = \frac{\partial\bar{S}^L}{\partial\vartheta}\Big|_{\vartheta_{ex}} = 0. \tag{4.14}$$

Now we insert (4.13) into formula (3.9) and path integrate it with respect to the space trajectories $x_k(t)$. Then as we briefly show in Appendix B, (3.9) can be expressed in the following form of the Feynman–Kac formula:

$$K(x,x') = \int \mathcal{D}(x_k)\exp\left\{\frac{i}{\hbar}\int L_0(x,t)dt\right\} U_0(x,t)|_{j=1/2}$$
$$= \exp\left(\frac{iS_0(\bar{x}_k)}{\hbar}\right)\int U_0(x,t)|_{j=1/2}\, dw_0^L + \text{ho}, \tag{4.15}$$

where $S_0(\bar{x}_k)$ is the classical action for a scalar particle; w_0^L is the corresponding integrator whose integrand $\bar{U}_0(x,t)$ is given by (4.13); and ho means the higher-order corrections. Rewriting the integrand (4.13) as

$$U_0(x,t)|_{j=1/2} = \sum w^L \exp[iS_{\text{ex}}^L/\hbar] \tag{4.16}$$

and expanding $S_{\text{ex}}^L(x,\bar{\vartheta}(x)) \equiv \bar{S}^L(\vartheta_{\text{ex}},\vartheta'_{\text{ex}})$ around the classical trajectories $\bar{x}$, we obtain, for (4.15),

$$K(x,x') = \sum w^L(\bar{x})\exp\left\{\frac{i}{\hbar}[S_0(\bar{x}) + S'_{\text{ex}}(\bar{x})]\right\}$$
$$\times\int dw_0^L \exp\left\{\frac{i}{\hbar}\int \nabla_k\nabla_l\bar{L}_1 x_l x_k\, dt\right\} + \text{ho}, \tag{4.17}$$

where $\bar{L}_1 = dS_{\text{ex}}^L/dt$ is given by

$$\bar{L}_1 = \frac{I}{2}g_{ij}(\bar{\vartheta})\left[\frac{d\bar{\vartheta}^i}{dt} + \frac{e}{2mc}H^k g_k^i\right]\left[\frac{d\bar{\vartheta}^j}{dt} + \frac{e}{2mc}H^l g^j{}_l\right], \tag{4.18}$$

with $\bar{\vartheta}^i$ (and both of its end points) depending on x via the Euler equations about ϑ derived from L_1 [and (4.14)].

Then the integral (4.17) can be evaluated by using the Cameron–Martin formula as

$$K_{\text{WKB}}(x,x') = \sum_{\tilde{\vartheta}} w^L(\tilde{x})\left(\frac{2\pi i\hbar}{m}\right)^{-3/2}(\det M_{kl}(\tilde{x}))^{1/2}$$
$$\times\exp\left\{\frac{i}{\hbar}[S_0(\tilde{x}) + S_{\text{ex}}^L(\tilde{x})\right\}, \tag{4.19}$$

where the VanVleck–Morette determinant $M_{kl}(\tilde{x})$ corresponds to the classical action $S_0(\tilde{x}) + S_{\text{ex}}^L(\tilde{x})$ and the classical trajectories $\tilde{x}$ satisfy the equations of motion:

$$m\frac{d^2\tilde{x}_k}{dt^2} - \frac{e}{c}\left[E + \frac{d\tilde{x}}{dt}\times H\right]_k + \nabla_k\bar{L}_1 = 0, \tag{4.20}$$

where (E,H) is the external electromagnetic field.

V. CONCLUSION

The SU(2) fiber bundle treatment for a spinor particle with a half-spin studied in this paper is equivalent to the usual Pauli theory as far as the initial and final states are concerned, but it constructs a wave packet involving all spins for the nonobserved intermediate states. As a result, the treatment becomes an equivalent alternative to the Pauli theory in the case of infinitesimal propagation, as proved in Sec. IV. This equivalence is guaranteed by the method we applied to derive the one from the other and which may be applicable to a similar derivation[7] of a more general gauge theory of the Kaluza–Klein type from the corresponding Yang–Mills type. Our formulation has enabled us to formulate not only a definite path integral, but also a definite classical theory with spin–orbital interactions (and thus the corresponding semiclassical approximation to the path integral) for a spinor particle. These features will also be common to the formulation of the path integral in the Kaluza–Klein space for a particle interacting with external gauge fields, in which one introduces a new dimensional constant such as $\sqrt{I/m}$ in our theory.[8] This extension of our path integral to a more general Kaluza–Klein space, as well as its relativistic formulation, will be studied in another work.

ACKNOWLEDGMENT

I would like to thank C. DeWitt-Morette for her stimulating conversations and discussions during the completion of this work.

APPENDIX A

In this Appendix, we formulate the phase space path integral in the compact SU(2) manifold and at the same time obtain the prolongation function[9] $Q_{\text{SU}(2)}$, which reflects the periodicity of this group space and makes it possible to prolong the domain of integration from the finite region of the compact SU(2) to the infinite R^3. We begin with the exact propagator for a free top studied by Schulman,[5] i.e.,

$$U(\epsilon) = \sum_n \left(\frac{I}{2\pi i\hbar\epsilon}\right)^{3/2}\frac{\Gamma_n}{\sin\Gamma}\exp\left[\frac{i\hbar R_G\epsilon}{12I} + \frac{iI\Gamma_n^2}{2\hbar\epsilon}\right], \tag{A1}$$

where ϵ is the infinitesimal time lapse and $\Gamma(\Gamma_n = \Gamma + 4\pi n$, where n are integers) is defined as

$$\cos\Gamma = \cos\vartheta_1\cos\vartheta'_1\cos(\vartheta_2 - \vartheta'_2)$$
$$+ \sin\vartheta_1\sin\vartheta'_1\cos(\vartheta_3 - \vartheta'_3), \tag{A2}$$

with ϑ_i given with the Euler angles (φ,θ,ψ) as

$$\vartheta_1 = \theta/2,\ \vartheta_2 = (\psi+\vartheta)/2,\ \vartheta_3 = (\psi-\vartheta)/2. \tag{A3}$$

Then the metric of the SU(2) mainfold is given by using the parameters

$$ds^2 = (d\theta^2 + d\varphi^2 + d\psi^2 + 2\cos\theta\, d\varphi\, d\psi)/4$$
$$= d\vartheta_1^2 + ws^2\,\vartheta_1\, d\vartheta_2^2 + \sin^2\vartheta_1\, d\vartheta_3^2. \tag{A4}$$

For any contributing small Γ_n, we can approximate the factors in (A1) as

$$\exp[(iI/2\hbar\epsilon)\Gamma_n^2] = \exp[(iI/2\hbar\epsilon)](2 - 2\cos\Gamma$$
$$+ (1/12)\Gamma_n^4) + O(\epsilon^2),$$
$$\Gamma_n/\sin\Gamma = 1 + \tfrac{1}{6}\Gamma_n^2 + O(\epsilon^2) = e^{i\hbar\epsilon/2I} + O(\epsilon^2),$$
$$\exp[(iI/24\hbar\epsilon)\Gamma_n^4] = 1 + (iI/24\hbar\epsilon)\Gamma_n^4 + O(\epsilon^2)$$
$$= \exp[(iI/2\hbar\epsilon)\Gamma_n^2] + O(\epsilon^2). \tag{A5}$$

The scalar curvature of the group space given by the metric (A4) is

$$R_G = 4\times(3/2) = 6. \tag{A6}$$

Then we may rewrite (A1) in the form

$$U(\epsilon) = \left(\frac{I}{2\pi i\hbar\epsilon}\right)^{3/2}\exp\left(\frac{3i\hbar\epsilon}{8I} + \frac{iI(2-2\cos\Gamma)}{2\hbar\epsilon}\right). \tag{A7}$$

This formula for the infinitesimal propagator is appropriate for applying the Prokhorov[9] results (with the neces-

sary modifications for our problem) because it retains the periodicity of Γ (Γ_n) within the required approximation and makes it possible to write the phase factor of (A7) with our parameters, i.e.,

$$2-2\cos\Gamma = 2-2\cos\vartheta_1\cos\vartheta_1'\cos(\vartheta_1-\vartheta_2') - 2\sin\vartheta_1\sin\vartheta_1'\cos(\vartheta_3-\vartheta_3'). \quad \text{(A8)}$$

The exponential that involves the cosine as in (A7) was expanded by Prokhorov after Langguth and Inomata[13] in a series of Bessel functions J_m as

$$\exp[-iz\cos(\vartheta-\vartheta')] = \sum_{-\infty}^{\infty} J_m(z)\cos\left[im(\vartheta-\vartheta')+\frac{3\pi}{2}\right].$$

Prokhorov used the asymptotic formula for J_m in the limit of $\epsilon\to 0$, i.e., $z\to\infty$:

$$J_m(z) = \sqrt{2/\pi z}\cos[z-(m+\tfrac{1}{2})\pi+(m^2-\tfrac{1}{4})/2z] + O(z^{-3/2}).$$

Then by using the Poisson summation formula

$$\sum_{-\infty}^{\infty} f(m\hbar)e^{im(\vartheta-\vartheta'')} = \int\int_{-\infty}^{\infty}\frac{dp\,d\vartheta'}{\hbar}f(p)e^{(i/\hbar)p(\vartheta-\vartheta')}Q(\vartheta'-\vartheta''),$$

with

$$Q(\vartheta'-\vartheta'') = \sum_{-\infty}^{\infty}\delta(\vartheta'-\vartheta''+2\pi m), \quad \text{(A9)}$$

Prokhorov obtained the following important expression:

$$\exp[-iz\cos(\vartheta-\vartheta'')] = \sqrt{\frac{2\pi i}{z}}\int\int_{-\infty}^{\infty}\frac{dp\,d\vartheta'}{2\pi\hbar}\exp\left(\frac{i}{\hbar}p(\vartheta-\vartheta')\right)\times\left\{\exp\left[-iz-\frac{i}{\hbar}\frac{p^2-\hbar^2/4}{2\hbar z}\right]Q(\vartheta'-\vartheta'') + \exp\left[iz+\frac{i}{\hbar}\frac{p^2-\hbar^2/4}{2\hbar z}-\frac{i\pi}{2}\right]\times Q(\vartheta'-\vartheta''+\pi)\right\}+O(z^{-3/2}). \quad \text{(A10)}$$

The application of formula (A10) to our propagator (A7) is straightforward for the variables ϑ_2 and ϑ_3 with $z_2'' = I\cos\vartheta_1\cos\vartheta_1''/\hbar\epsilon$ and $z_3'' = I\sin\vartheta_1\sin\vartheta_1''/\hbar\epsilon$.

In this case we obtain four terms, each of which involves the factor $\exp[\pm i(z_2''\mp z_3'')]$. For example, the factor $\exp[i(z_2''-z_3'')]$ appears in company with the factor

$$\frac{1}{\sqrt{z_2''z_3''}}\exp\left[\frac{i}{\hbar}\frac{p_2^2-\hbar^2/4}{2\hbar z_2''}+\frac{i\pi}{2}-\frac{i}{\hbar}\frac{p_3^2-\hbar^2/4}{2\hbar z_3''}\right]\times Q(\vartheta_2'-\vartheta_2''+\pi)Q(\vartheta_3'-\vartheta_3''). \quad \text{(A11)}$$

As to the variable ϑ_1, we apply (A10) to the factors $\exp[\pm i(z_2''\mp z_3'')]$ and obtain, e.g.,

$$\exp[i(z_2''-z_3'')] = \exp\left[-\left(\frac{iI}{\hbar\epsilon}\right)\cos(\vartheta_1'+\vartheta_2''+\pi)\right] = \sqrt{\frac{2\pi i\hbar\epsilon}{I}}\int\int_{-\infty}^{\infty}\frac{dp_1\,d\vartheta_1'}{(2\pi\hbar)}\exp\left(\frac{i}{\hbar}p_1(\vartheta_1-\vartheta_1')\right)\left\{\exp\left[-\frac{iI}{\hbar\epsilon}-\frac{i\epsilon}{\hbar}\frac{p_1^2-\hbar^2/4}{2I}\right]Q(\vartheta_1'+\vartheta_1''+\pi) + \exp\left[\frac{iI}{\hbar\epsilon}+\frac{i\epsilon}{\hbar}\frac{p_1^2-\hbar^2/4}{2I}+\frac{i\pi}{2}\right]Q(\vartheta_1'+\vartheta_1'')\right\}. \quad \text{(A12)}$$

Since $\exp(-iI/\hbar\epsilon)$ cancels the corresponding factor in (A7) and, making the integration variables with respect to ϑ_1'' uniform in notation to ϑ_2' and ϑ_3', $Q(\vartheta_1'+\vartheta_1''+\pi)$ changes z'' as

$$z_2''(\vartheta_1'') = z_2(-\vartheta_1'-\pi) = -z_2(-\vartheta_1') = -z_2(\vartheta_1'),$$
$$z_3'' = z_3(-\vartheta_1'-\pi) = -z_3(-\vartheta_1') = z_3(\vartheta_1'), \quad \text{(A13)}$$

and

$$\exp(i\pi/2)/\sqrt{z_2''z_3''} = 1/\sqrt{z_2'z_3'} = (g\cdot g')^{-1/4},$$

with

$$g = \det g_{ij} = \cos^2\vartheta_1\sin^2\vartheta_1$$

the term that involves (A11) becomes

$$\int\int\int_{-\infty}^{\infty}\frac{dp_1\,d\vartheta_1'\,dp_2\,d\vartheta_2'\,dp_3\,d\vartheta_3'}{(2\pi\hbar)^3(g\cdot g')^{1/4}}\exp\left(\frac{i3\hbar\epsilon}{8I}\right)\exp\left(\frac{i}{\hbar}[p_1(\vartheta_1-\vartheta_1')+p_2(\vartheta_2-\vartheta_2')+p_3(\vartheta_3-\vartheta_3')]\right)\times\exp\left\{-\frac{i\epsilon}{2I\hbar}\left[\left(p_1^2-\frac{\hbar^2}{4}\right)+\frac{(p_2^2-\hbar^2/4)}{\cos\vartheta_1\cos\vartheta_1'}+\frac{(p_3^2-\hbar^2/4)}{\sin\vartheta_1\sin\vartheta_1'}\right]\right\}Q(\vartheta_1'+\vartheta_1''+\pi)Q(\vartheta_2'-\vartheta_2''+\pi)Q(\vartheta_3'-\vartheta_3''). \quad \text{(A14)}$$

Of course, (A12) also involves a term proportional to $\exp[iI/\hbar\epsilon]$, which gives a factor $\exp(i2I/\hbar\epsilon)$ in (A7); however, this term vanishes if we introduce a small absorbing factor in I as $I(1+\delta)$, with $\delta>0$, as did Langguth and Inomata.[13]

Therefore, all that is needed to obtain the required expression for the infinitesimal propagator (A7) is only to collect terms such as (A14) for all $\exp[\pm i(z_2''\mp z_3'')]$, which yields just (A14); only the prolongation function is replaced by

$$Q_{SU(2)} = Q(\vartheta_1' - \vartheta_1'')Q(\vartheta_2' - \vartheta_2'')Q(\vartheta_3' - \vartheta_3'') + Q(\vartheta_1' + \vartheta_1'' + \pi)Q(\vartheta_2' - \vartheta_2'' + \pi)Q(\vartheta_3' - \vartheta_3'') + Q(\vartheta_1' + \vartheta_1'')Q(\vartheta_2' - \vartheta_2'')Q(\vartheta_3' - \vartheta_3'' + \pi) + Q(\vartheta_1' - \vartheta_1'' + \pi)Q(\vartheta_2' - \vartheta_2'' + \pi)Q(\vartheta_3' - \vartheta_3'' + \pi), \tag{A15}$$

which coincides with the propagation function obtained via the symmetry discussion given in Sec. II.

Then substituting (A14) with $Q_{SU(2)}$ into the formula for ψ, i.e.,

$$\psi(\vartheta,\epsilon) = \int_{SU(2)} U(\epsilon)\psi(\vartheta'',t=0)\sqrt{g''}\,d\vartheta'' \tag{A16}$$

we obtain

$$\psi(\vartheta,\epsilon) = \int_{-\infty}^{\infty}\int \frac{d^3p\,d^3\vartheta'}{(2\pi\hbar)^3}\left(\frac{g'}{g}\right)^{1/4} \times\exp\left(\frac{i}{\hbar}\left[p_i\Delta\vartheta^i - \epsilon H_{\text{eff}}\right]\right)\Psi_0(\vartheta'), \tag{A17}$$

with

$$\Psi_0 = \int_{SU(2)} Q_{SU(2)}(\vartheta',\vartheta'')\psi_0(\vartheta'')d\vartheta'', \tag{A18}$$

where H_{eff} is the classical Hamiltonian with effective potentials.

From (A18) we know that $Q_{SU(2)}$ is the kernel of an operator that allows us to define the function ψ_0 outside the region $0<\vartheta_1\leqslant\pi/2$, $0\leqslant\vartheta_2,\ \vartheta_3<2\pi$; Prokhorov[9] named it a prolongation function, i.e., $Q_{SU(2)}$ is the kernel of the identity operator, with the integration over the compact region which constitutes the direct extension of the delta function $\Sigma\delta(\vartheta' - \vartheta'' + 2\pi n)$ on the maximal torus.

Since there are no ordering problems as long as we treat only a free top, i.e., only the kinetic part of our potential problem in the SU(2) group space, the effective Hamiltonian in (A17) can be written in any ordering, e.g., we can set

$$H_{\text{eff}}(\vartheta,p) = \frac{1}{2I}\left[\left(p_1^2 - \frac{\hbar^2}{4}\right) + \frac{(p_2^2 - \hbar^2/4)}{\cos^2\vartheta_1} + \frac{(p_3^2 - \hbar^2/4)}{\sin^2\vartheta_1} - \frac{3}{4}\pi\right], \tag{A19}$$

which corresponds to the standard ordering of the operator in the quantum Hamiltonian in the SU(2) manifold

$$\hat{H} = \frac{1}{2I}\left[\left(\hat{p}_1^2 - \frac{\hbar^2}{4}\right) + \frac{(\hat{p}_2^2 - \hbar^2/4)}{\cos^2\vartheta_1} + \frac{(\hat{p}_3^2 - \hbar^2/4)}{\sin^2\vartheta_1} - \frac{3}{4}\hbar^2\right] \tag{A20}$$

written with the Hermitian momenta

$$\hat{p}_i = -i\hbar g^{-1/4}\partial_i\, g^{1/4}. \tag{A21}$$

Using (A19) in (A17), we can show by straightforward evaluation of integration that ψ in (A17) satisfies the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi. \tag{A22}$$

Of course, another effective Hamiltonian $H_{\text{eff}}(\vartheta,p,\Delta\vartheta)$ corresponding to any ordering will produce the same result (A22).

Then the path integral becomes[9]

$$\psi(\vartheta,t) = \int_{-\infty}^{\infty}\int \sqrt{g'}\,d\vartheta'\,U(\vartheta,\vartheta';t)\Psi_0(\vartheta'), \tag{A23}$$

with

$$U(\vartheta,\vartheta';t) = g^{-1/4}g'^{-1/4}\int_{-\infty}^{\infty}\int\prod_0^t \frac{d^3p\,d^3\vartheta}{(2\pi\hbar)^3} \times\exp\left\{\frac{i}{\hbar}\left[p_i\frac{d\vartheta^i}{dt} - H_{\text{eff}}\right]dt\right\}. \tag{A24}$$

APPENDIX B

Here we briefly summarize the phase space path integral integral without the limiting procedure with the Gaussian integrator, which was developed in Ref. 10.

In this formulation, the propagator for the transition $A\to B$ is given by the formula

$$K(B,A) = \exp\left[\frac{i}{\hbar}S(\bar{q},\bar{p})\right] \times\int_Z \exp\left[\frac{i}{\hbar}\Sigma\right]dw(z), \tag{B1}$$

where the following notation has been used.

The action has been expanded around the classical path $(\bar{q},\bar{p})$ from $A(\bar{q}(t_a),t=t_a)$ to $B(\bar{q}(t_b),t=t_b)$:

$$S(q,p) = S(\bar{q},\bar{p}) + \tfrac{1}{2}S''(\bar{q},\bar{p})zz + \Sigma(\bar{q},\bar{p};z), \tag{B2}$$

with

$$s = (x\cdot y)\in Z, \tag{B3}$$

where Z is the space of vector fields $z=(x,y)$ along $(\bar{q},\bar{p})$ such that $x(t_a)=x(t_b)=0$. (Of course, other boundary conditions are also possible, but we restrict ourselves here only to this condition, which is of necessity in the present work.)

The Gaussian integrator w on Z is defined by

$$\int_Z\begin{bmatrix}x^\alpha(t)\\ y_a(t)\end{bmatrix}[x^\eta(s),y_\eta(s)]dw = i\begin{bmatrix}G^{\alpha\eta}(r,s) & G^\eta_\alpha(r,s)\\ G^\alpha_\eta(r,s) & G_{\alpha\eta}(r,s)\end{bmatrix}, \tag{B4}$$

with the corresponding boundary conditions on G, i.e., $G^{\alpha\eta}(t_a,s) = G^{\alpha\eta}(r,t_b) = G^\alpha_\eta(r,t_b) = G^\eta_\alpha(t_a,s) = 0$ and no restrictions on others.

The covariance $G^{\alpha\eta}$ is given as the elementary kernel of the Jacobi equation. The Jacobi equation (the second variation of the action) for a system with a potential $k_0 H^i(q)p^i$ in curved phase space is given as

$$\begin{bmatrix} -(1/m)R^{\delta}_{\alpha\beta\xi}g^{\beta\gamma}\bar{p}_\gamma\bar{p}_\delta - k_0 H^i_{,\alpha\xi}\bar{p}_i & -\delta^\xi_\alpha\nabla_t \\ \delta^\alpha_\xi\nabla_t - k_0 H^i_{,\xi}\delta^\alpha_i & -(1/m)g^{\alpha\xi} \end{bmatrix} \times \begin{bmatrix} x^\xi \\ y_\xi \end{bmatrix} = 0 \tag{B5}$$

and $G^{\alpha\eta}$ satisfies

$$-m^2\nabla_r^2 G^{\alpha\eta}(r,s) - g^{\alpha\rho}g^{\beta\gamma}R^{\delta}_{\rho\beta\xi}\bar{p}_\gamma\bar{p}_\delta G^{\xi\eta}(r,s) + k_0 H^i_{,\beta\xi}g^{\beta\alpha}p_i G^{\xi\eta}(r,s) ,$$
$$-mk_0 H^i_{,\xi}\delta^\alpha_i\nabla_t G^{\xi\eta}(r,s) = mg^{\alpha\eta}\delta_s(r) , \tag{B6}$$

with the boundary condition given above.

Other G's are determined with $G^{\alpha\eta}$ as

$$G^\eta_\alpha(r,s) = mg_{\alpha\xi}\nabla r G^\xi(r,s),$$
$$G^\alpha_\eta(r,s) = mg_{\eta\xi}\nabla_s G^{\alpha\xi}(r,s),$$

and

$$G_{\alpha\eta}(r,s) = m^2 g_{\alpha\xi}\, g_{\eta\delta}\nabla_r\nabla_s G^{\xi\delta}(r,s) - mg_{\alpha\eta}\,\delta_s(r) . \tag{B7}$$

From (B6) we see that $G^{\alpha\eta}$ may be written with the Jacobi fields as

$$G^{\alpha\eta}(r,s) = \theta(s-t)\,J(t,t_a)M(t_a,t_b)\,J(t_b,s) - \theta(t-s)\,J(t,t_b)M(t_b,t_a)\,J(t_a,s) , \tag{B8}$$

where θ is the step function and J and M are given with $J = \det J^{\alpha\beta}$ as

$$J^{\alpha\beta}(t,t_a) = -J^{\beta\alpha}(t_a,t) = m\left(\frac{\partial q^\alpha}{\partial p_{\beta\alpha}}\right) \tag{B9}$$

and

$$M_{\alpha\beta}(t_a,t_b) = \frac{1}{m}\left(\frac{\partial p_{\beta b}}{\partial q^\alpha_a}\right) = \frac{1}{m}\left(\frac{\partial^2\bar{S}}{\partial q^\alpha_a\,\partial q^\beta_b}\right).$$

When we discuss only the WKB approximation solution, we obtain, from (B1),

$$K_0(B,A) = \exp\left[\frac{i}{\hbar}S(\bar{q},\bar{p})\right]\int_Z dw = w(z)\,e^{(i/\hbar)\bar{S}} . \tag{B10}$$

The solution (B10) can also be obtained for the same system developed in Ref. 12:

$$K_0 = \exp\left[\frac{i}{\hbar}S^L(\bar{q})\right]\int_X dw^L = w^L(X)\,e^{(i/\hbar)\bar{S}^L} , \tag{B11}$$

where the superscript L indicates that the quantities are obtained in the Lagrangian formulation.

Since we can obtain

$$S(\bar{q},\bar{p}) = S^L(\bar{q}) , \tag{B12}$$

we find, for the normalization $w(Z)$ and $w^L(X)$,

$$w(Z) = w(X) = (2\pi i\hbar/m)^{-n/2}g_b^{-1/4} \times (\det M_{\alpha\beta}(t_b,t_a))^{1/2}g_a^{-1/4} \tag{B13}$$

by inversing the covariance (B8), where n is the dimension of the configuration space and g is the determinant of its metric tensor.

If we are to construct the path measure dw for a free motion in the curved phase space and regard the potential $k_0 H^i(q)p_i$ as the integrand of the Feynman–Kac formula, (B10) becomes

$$K_0 = \exp\left[\frac{i}{\hbar}S_F(\bar{q},\bar{p})\right]\int_{Z'}\exp\left(\frac{1}{\hbar}k_0\int H^i p_i\,dt\right)dw_F . \tag{B14}$$

We must use the Cameron–Martin formula[10] for evaluating (B14), which leads us back to (B10) with (B6).

[1] R. P. Feynman, Phys. Rev. **84**, 108 (1951), Appendix D.
[2] C. Morette, Phys. Rev. **81**, 848 (1951).
[3] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley–Interscience, New York, 1981).
[4] F. Bopp and R. Haag, Z. Naturforsch. **52**, 644 (1950).
[5] L. S. Schulman, Phys. Rev. **176**, 1558 (1968).
[6] B. S. DeWitt, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964); *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[7] D. S. Toms, in *An Introduction to Kaluza–Klein Theories*, edited by H. Lee (World Scientific, Singapore, 1983).
[8] A. Salam and J. Strathdee, Ann. Phys. **141**, 316 (1982).
[9] L. V. Prokhorov, Sov. J. Nucl. Phys. **39** (2), 312 (1984).
[10] C. DeWitt-Morette, Ann. Phys. **97**, 367 (1976); C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rev. **50**, 255 (1979).
[11] N. J. Vilenkin, Am. Math. Soc. Transl. Math. Monograph (1968).
[12] B. Nelson and B. Sheeks, Commun. Math. Phys. **84**, 515 (1982); J. Math. Phys. **22**, 1944 (1981).
[13] W. Langguth and A. Inomata, J. Math. Phys. **20**, 499 (1979).

Conformal symmetries and fixed points in space-time

G. S. Hall
Department of Mathematical Sciences, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB9 2TY, Scotland, United Kingdom

(Received 29 March 1989; accepted for publication 6 December 1989)

A general study is made of conformal vector fields on a four-dimensional Lorentz manifold with particular emphasis being laid on the structure of the zeros (critical points) of such vector fields. The implications for general relativity are investigated and a discussion of conformal vector fields in generalized plane wave space-times is given. An attempt is made to clarify the well-known theorem of Bilyalov and Defrise-Carter.

I. INTRODUCTION

There has been much recent interest in the study of conformal symmetries in general relativity. However, very few general theorems seem to be known about them. It is true that a good deal of information regarding conformal symmetries in space-time is stored in the theorems of Bilyalov[1] and Defrise-Carter[2] (for the case of a positive definite metric, see Refs. 3,4). These theorems, although extremely useful, employ several implicit and crucial assumptions. Although these assumptions are presumably part of the spirit of the theorem it is felt that some discussion and clarification is needed. The essential aim of these theorems is to construct a certain function (a conformal factor—see Sec. II). This function, unfortunately, may behave badly near the zeros of conformal vector fields and, in any case, is only locally defined. In the next section the theorem is stated and some discussion and examples are given to demonstrate the problems involved. This provides a useful motivation for the main part of the paper (Sec. III) which is to give a discussion of conformal symmetries in general relativity, with special emphasis laid on their fixed point structure, and to point out the similarities and differences in this fixed point structure between (proper) conformal, homothetic, and isometric symmetries. The (generalized) plane wave space-times play an important role in the theorem in Ref. 2 and Sec. IV contains a discussion of their conformal symmetries.

Let M be a space-time manifold with metric g of Lorentz signature. A (not identically zero) global vector field ξ on M is called a *conformal vector field* if in each chart of M the following relation holds:

$$\xi_{a;b}=\phi g_{ab}+F_{ab} \quad (\Leftrightarrow \underset{\xi}{\mathcal{L}} g=2\phi g). \tag{1}$$

Here, $F_{ab}(=-F_{ba})$ is the *conformal bivector*, ϕ is a real-valued function on M, $\mathcal{L}$ denotes the Lie derivative, and a semicolon denotes the covariant derivative arising from g in the usual way. The manifold M and all structures on M will be supposed smooth. If one uses a comma to denote a partial derivative one then finds (see, e.g., Ref. 5—but note some difference in conventions)

$$F_{ab;c}=R_{abcd}\xi^d-2\phi_{,[a}g_{b]c}, \tag{2}$$

$$\phi_{;ab}=-\tfrac{1}{2}L_{ab;c}\xi^c-\phi L_{ab}+R_{c(a}F_{b)}{}^c, \tag{3}$$

where R_{abcd} are the curvature tensor components, $R_{ab}\equiv R^c_{acb}$, the Ricci tensor components $R\equiv R_{ab}g^{ab}$, and $L_{ab}\equiv R_{ab}-\frac{1}{6}Rg_{ab}$. The vector field ξ is called *homothetic* if $\phi=\text{const}\neq 0$ on M and *Killing* if $\phi\equiv 0$ on M. Otherwise it is called *proper conformal*. The condition (3) is equivalent to a statement about the Lie derivative of the Ricci tensor which, with the usual abuse of notation, is

$$\underset{\xi}{\mathcal{L}} R_{ab}=-2\phi_{;ab}-(\phi_{;cd}g^{cd})g_{ab}. \tag{4}$$

For reference later in the paper the idea of a generalized plane wave is required. A space-time M is called a *generalized pp wave* if it is nonflat (in the sense that the curvature tensor does not vanish on a nonempty open subset of M) and if it admits a global, covariantly constant, nowhere zero, null bivector V.[6] Given the former nonflat condition the latter condition is equivalent to M admitting a global, covariantly constant, nowhere zero, null vector field l and, either, M having a Weyl tensor C of Petrov type N or 0 at each $p\in M$, or, M having a Ricci tensor that is either zero or of Segre type $\{(211)\}$ with zero eigenvalue at each $p\in M$. The null vector field necessarily spans the repeated principal null direction of V and of C at points where $C\neq 0$ and the unique null eigendirection of the Ricci tensor at points where the Ricci tensor is nonzero, the latter taking the form $R_{ab}=\rho l_a l_b$ in any chart, where ρ is a real-valued function. Using $*$ to denote the duality operator, the complex, self-dual Weyl tensor $\overset{+}{C}\equiv C+iC^*$ takes the component form $\overset{+}{C}_{abcd}=\beta\overset{+}{V}_{ab}\overset{+}{V}_{cd}$ in any chart where $\overset{+}{V}\equiv V+iV^*$ and β is a complex-valued function. A generalized *pp*-wave is called a *generalized plane wave* if, using standard v,u,x,y coordinates of the former,[6] the function β depends only on u. The Bianchi identity then shows that ρ depends only on u. The usual energy conditions then imply that ρ is a non-negative function on M and that such plane waves may be realized as null Einstein–Maxwell fields. However the energy conditions will not be enforced in what is to follow. In the vacuum case the standard definitions given in Ref. 6 are recovered. (The definition given here of a generalized *pp* wave is stricter than the definition of a nonvacuum *pp* wave given in Ref. 7 but the definition of the generalized plane wave is essentially the same.)

 0022-2488/90/051198-10$03.00

II. CONFORMAL SYMMETRIES

The general idea in Refs. 1 and 2 is, starting with the existence of a group (or local group) of conformal symmetries on a space-time (M,g) represented by a finite-dimensional Lie algebra of conformal vector fields on M, to investigate the possibility of replacing the metric g by a metric $g'=e^{2\sigma}g$ conformally related to g, where σ is a real-valued function on M, so that the above conformal symmetries with respect to g become either homothetic symmetries or isometries with respect to g' (i.e., the above Lie algebra of conformal vector fields with respect to g becomes a Lie algebra of homothetic or Killing vector fields with respect to g'). The combination of the results in Refs. 1 and 2 essentially says that if M is not conformally flat then under certain assumptions (described in the next paragraph) the transition from conformal to homothetic symmetries (in the sense described above) can always be made and if, further, M is not conformally related to a generalized plane wave, the transition from conformal symmetries to isometries can always be made.

Although these results are useful in describing the generic local behavior of Lie algebras of conformal vector fields they have limitations when applied more generally. These limitations arise because of the implicit assumptions made in Refs. 1 and 2 and were, doubtless, well understood from the beginning. The remainder of this section will be used to discuss and clarify these assumptions and to exhibit a space-time with a type of singular behavior in its conformal group of transformations and to which the results in Ref. 2 must be applied with care.

In Ref. 1 it is implicitly assumed that the orbits of the conformal transformations (or local transformations) are everywhere of the same dimension and nature (timelike, spacelike or null) and this, of course, need not be the case. In Ref. 2 a certain type of real-valued function on M is employed called a *proper conformal scalar* (see also Ref. 4). Now one may write M as the disjoint union $M=U\cup V$, where U is the open subset of M at each point of which some proper conformal scalar is nonzero and V is the subset of M consisting of all those points at which all proper conformal scalars vanish. In Ref. 2 it is implicitly assumed that either $U=\phi$ or $V=\phi$ which again need not be the case. In fact, even more is assumed in Ref. 2 and this will be returned to later.

To see what can go wrong consider the following metric defined on the manifold $\mathbb{R}^4$ with coordinates labeled (v, u,x,y):

$$ds^2=e^{ux}\,du\,dv+dx^2+dy^2. \tag{5}$$

This metric was first given in Ref. 8 and discussed in more detail in Ref. 9. It is everywhere of Petrov type III and admits exactly two independent Killing vector fields $\eta_1\equiv\partial/\partial v$ and $\eta_2\equiv\partial/\partial y$, the former null and the latter spacelike and covariantly constant, together with the homothetic vector field ξ with components in the above global chart given by $\xi^a\equiv(3v,-u,x,y)$. The global vector fields η_1, η_2, and ξ can be shown to span the total Lie algebra of *conformal* vector fields on M, there being no proper conformal ones. The Lie algebra bracket relations are $[\eta_1,\xi]=3\eta_1$, $[\eta_2,\xi]=\eta_2$, and $[\eta_1,\eta_2]=0$. The vector fields, being complete, give rise to a three-dimensional Lie group of homothetic (bijective) diffeomorphisms $M\to M$. The homothetic vector field ξ has a unique zero at the origin of the above coordinates. The orbits of the above group can now be described. The null hypersurface $u=0$ through the origin breaks up into a two-dimensional null orbit $u=x=0$ and two three-dimensional null orbits $u=0$, $x>0$ and $u=0$, $x<0$. All other orbits are three-dimensional and timelike. The origin is, of course, not distinguished by its being a zero of a homothetic vector field since an appropriate combination of η_1, η_2, and ξ will yield a homothetic vector field that vanishes at any given point of the two-dimensional null orbit $u=x=0$. Denoting the metric in Eq. (5) by g, it is easily seen that no conformally related metric $e^{2\sigma}g$ could admit all the above conformal vector fields as Killing vector fields on M because, considering ξ, the function σ would be required to satisfy the relation $\mathcal{L}_\xi\sigma+c\equiv\sigma_{,a}\xi^a+c=0$, where c is the homothetic constant arising from ξ and g. Since $c\neq 0$ there can be no solution of this equation in any neighborhood of the origin. [More generally, if M admits a proper conformal or homothetic vector field ξ with a zero at p such that the corresponding function ϕ in Eq. (1) satisfies $\phi(p)\neq 0$—and this is always true if ξ is homothetic—then it is impossible to find a metric in any neighborhood W of p conformally related in W to the original metric on M with respect to which ξ is a Killing vector on W.]

The problem with the metric in Eq. (5) lies in the two-dimensional null orbit $u=x=0$. Strictly speaking, this metric does not contradict the result as stated in Ref. 1 but it is clear that it cannot be handled by the proof given there because of the assumption (mentioned above) made about the orbits. Concerning the result in Ref. 2 it can be shown that there is no proper conformal scalar in the space-time represented by the metric (5) which is nonzero at any point of the orbit $u=x=0$ but that this is not true for the complement of this orbit. Hence, in an earlier notation, neither U nor V is empty for the space-time of Eq. (5). However the conclusions of Ref. 2 hold in some neighborhood of any point not in this orbit. [It should be remarked here that the energy-momentum tensor constructed from the metric (5) has Segre type {31} or its degeneracy at each $p\in M$ and so fails the usual classical energy conditions everywhere.[9] However, metrics with properties similar to (5) can be constructed which do satisfy the energy conditions.]

It was remarked earlier that the assumption in Ref. 2 was stronger than the assumption $M=U$ or $M=V$. This is because even if $M=U$, a proper conformal scalar which is nonzero at $p\in M$ may lead to an appropriate function σ only over some neighborhood of p. Of course if a global, nowhere zero proper conformal scalar exists then the results in Ref. 2 show that a global solution σ to the above problem exists. However, if no such global proper conformal scalar exists, only local solutions for σ may be available. For example, let M' be Minkowski space with the usual global coordinates x,y,z,t and metric η and let ξ be the standard homothetic vector field on M' with components

(x,y,z,t). Clearly no function σ exists on any neighborhood of the origin O (the unique zero of ξ) such that ξ is a Killing vector field with respect to $e^{2\sigma}\eta$ (by the remark in the previous paragraph) but one can ask if such a global function σ exists on $M'\setminus\{O\}$. The answer is yes, a choice of σ being $\sigma = -\frac{1}{2}\log(x^2+y^2+z^2+t^2)$. [Although Minkowski space is excluded in Ref. 1 and 2, this example still usefully makes the point.] Consider now the metric g in Eq. (5) on the manifold M ($\approx\mathbb{R}^4$) and let N denote the two-dimensional submanifold given by $u=x=0$ at each point of which some global homothetic vector field on M vanishes. Although one can construct a function σ on some neighborhood of any $p\in M\setminus N$ such that the vector fields η_1, η_2, and ξ are Killing vector fields of the metric $e^{2\sigma}g$ on this neighborhood (e.g., $\sigma = ux + \log|u|$ works at all points where $u\neq 0$) there is no global function σ on $M\setminus N$ with this property. To see this note that if such a function σ existed, the fact that η_1 and η_2 are Killing vector fields with respect to both g and $e^{2\sigma}g$ implies that $\sigma = \sigma(u,x)$. The conversion of ξ into a Killing vector field with respect to $e^{2\sigma}g$ then implies that $x\sigma_{,x} - u\sigma_{,u} + 1 = 0$ and so it is sufficient to show that there is no global function σ on $\mathbb{R}^2\setminus\{(0,0)\}$ satisfying this partial differential equation. This can be established using the following argument due to John Pulham. Suppose such a σ exists on $\mathbb{R}^2\setminus\{(0,0)\}$ and let k be the vector field on $\mathbb{R}^2\setminus\{(0,0)\}$ whose components are $(x,-u)$. The integral curves of k from some initial point (x_0,u_0) are given by $t\to(x_0e^t,u_0e^{-t})$ and along each such curve ux is constant. Suppose x_0, $u_0>0$ so the integral curve from (x_0,u_0) intersects the diagonal set of $\mathbb{R}^2\setminus\{(0,0)\}$ in the point $(\sqrt{x_0u_0},\sqrt{x_0u_0})$ with parameter value $t = \log\sqrt{u_0/x_0}$. But along each such integral curve, the partial differential equation for σ implies $d\sigma/dt = -1$ so that $\sigma(t) = -t + \sigma(0)$. Hence σ satisfies, for $x,u>0$,

$$\sigma(\sqrt{xu},\sqrt{xu}) = -\log\sqrt{u/x} + \sigma(x,u).$$

The two operations of firstly fixing $u>0$ and letting $x\to 0$ and secondly fixing $x>0$ and letting $u\to 0$ together with the assumed well-behaved nature of σ at the points $(x,0)$ and $(0,u)$ $(u,x>0)$ then lead to a contradiction by considering the behavior of $\sigma(v,v)$ $(v>0)$ as $v\to 0$.

The following is an example where such local functions σ can be joined together smoothly to give a *global* smooth function with the required properties. Suppose a space-time (M,g) admits a Lie algebra $\mathscr{A}$ of conformal vector fields such that each $p\in M$ admits an open, connected coordinate neighborhood U together with a real-valued function σ on U such that the restriction to U of each member of $\mathscr{A}$ is Killing with respect to the metric $e^{2\sigma}g$ on U. Suppose further that the members of $\mathscr{A}$ evaluated at any $p\in M$ span the tangent space T_pM to M at p and finally that M is simply connected. Then there exists a global real-valued function $\hat{\sigma}$ on M such that each member of $\mathscr{A}$ is Killing with respect to the metric $e^{2\hat{\sigma}}g$ on M. To see this one uses a technical device employed in a different context in Ref. 10 and so only a sketch proof is given here. First, it is clear that each of the local functions σ is determined up to an arbitrary additive constant on the appropriate open set U since for each $\xi\in\mathscr{A}$ satisfying Eq. (1) one has $\sigma^1{}_{,a}\xi^a + \phi = 0 = \sigma^2{}_{,a}\xi^a + \phi$ on U for any such functions σ^1 and σ^2. Let U_α be a labeling of all open, connected coordinate neighborhoods admitting such local solutions σ with $\alpha\in S$ for some labeling set S and for a fixed α let $U_\alpha^\beta(\beta\in\mathbb{R})$ represent U_α together with a choice of function σ on U_α (represented by β). A set $\tilde{M}$ is then constructed from the disjoint union of all such U_α^β by identifying two points $a\in U_\alpha^\beta$ and $a'\in U_{\alpha'}^{\beta'}$ if $\pi_\alpha(a) = \pi_{\alpha'}(a')$ (where π_α: $U_\alpha^\beta\to U_\alpha$ is the obvious projection map) and if the functions σ associated with U_α and $U_{\alpha'}$ and labeled by β and β' agree at $\pi_\alpha(a)$. It then follows (cf. Ref. 10) that a natural Hausdorff four-dimensional manifold structure arises on $\tilde{M}$ determined by that of M, that the (obviously defined) projection π: $\tilde{M}\to M$ is locally a diffeomorphism and that (M,g) and $(\tilde{M},\pi^*g)$ are locally isometric. Also the Lie algebra $\mathscr{A}$ naturally carries over to a Lie algebra $\tilde{\mathscr{A}}$ on $\tilde{M}$ and by its construction $\tilde{M}$ admits a global function $\tilde{\sigma}$ such that the members of $\tilde{\mathscr{A}}$ are Killing vector fields on $\tilde{M}$ with respect to the metric $e^{2\tilde{\sigma}}(\pi^*g)$. The final step in the argument is made by choosing a particular U_α in the original collection of open, connected coordinate neighborhoods in M and a particular choice of σ on U_α thus defining a certain U_α^β. This U_α^β uniquely determines a connected subset $\tilde{U}_\alpha^\beta$ of $\tilde{M}$. Let $\tilde{\tilde{M}}$ be the component of $\tilde{M}$ containing $\tilde{U}_\alpha^\beta$. Then $\tilde{\tilde{M}}$ is a Hausdorff, connected, open four-dimensional submanifold of $\tilde{M}$ and $\pi(\tilde{\tilde{M}}) = M$. It follows that $\tilde{\tilde{M}}$ is a covering space of M and since M is simply connected M and $\tilde{\tilde{M}}$ are diffeomorphic and hence (M,g) and $(\tilde{\tilde{M}},\tilde{\tilde{g}})$ are isometric, where $\tilde{\tilde{g}}$ is the restriction of π^*g to $\tilde{\tilde{M}}$. The restriction $\tilde{\tilde{\sigma}}$ of the function $\tilde{\sigma}$ to $\tilde{\tilde{M}}$ when projected down to M yields the desired function $\hat{\sigma}$ and the proof is complete.

Clearly the above proof made no use of the four-dimensionality of M or the signature of g and so holds quite generally.

The discussion of this section shows that for a more-detailed study of the theorems in Refs. 1 and 2 further information is required concerning the zeros of conformal vector fields on space-time. A global vector field on M gives rise to a local group of local diffeomorphisms of M and a zero p of such a vector field is a fixed point of those local diffeomorphisms which act on p. Now there are certain geometrical features regarding such zeros that are independent of the metric (or at least of its conformal class) and hence remain unchanged under a conformal rescaling of the metric. Since these features can differ significantly in the cases when the vector field is proper conformal, homothetic, or Killing they provide restrictions on those cases where one of these types of vector field can be converted into another by means of a conformal change in the metric. They will be discussed in some detail in the next section.

The statement of the theorem in Ref. 2 shows that a special role is played by the generalized plane waves. A discussion of conformal symmetries in these space-times will be given in Sec. IV.

III. FIXED POINT STRUCTURE

The study of *affine* (and, in particular, isometric and homothetic) transformations of a space-time M is simplified to a certain extent because of a type of linearity that

such transformations enjoy. This linearity can be observed in several different ways. Let ξ be a (not identically zero) global affine vector field on M. First, ξ is uniquely determined on M (since M is connected) by the specification of ξ and its first covariant derivative at any point of M. Second, the natural lift (Ref. 11, Vol. 1) of ξ to the frame bundle $L(M)$ of M yields a vector field on $L(M)$ that is nowhere zero. Third, the one-parameter local group of local affine diffeomorphisms of M arising in the usual way from ξ, because of their geodesic (and affine parameter) preserving property, commute in a natural way with the exponential map (Ref. 11, Vol. 1). This fact is particularly useful in the geometrical interpretation of such transformations[9,12] and in the study of their fixed points.[9] Proper conformal vector fields are, however, more complicated. In particular, one must specify ξ and its first and second covariant derivatives [or, equivalently, ξ, F, ϕ, and $\phi_{,a}$ in the notation of Eqs. (1) and (2)] at a point of M to uniquely determine ξ on M.[5,10] The following points can now be made with regard to the zeros of Killing, homothetic, and proper conformal vector fields first raised in Sec. I (the zeros of the more general types of affine vector fields being discussed elsewhere).

A. Local linearization

Using a technique well known from the theory of linear differential equations[13–15] one can study the behavior of the integral curves of "linear" vector fields in the vicinity of a zero p of such a vector field ξ. One requires knowledge of the algebraic structure of the matrix $\xi^a{}_{,b}$ evaluated at p and the procedure is applicable to affine vector fields because of the third consequence of linearity mentioned at the beginning of this section. This results from the fact that the local diffeomorphisms arising from ξ and which act on some neighborhood of p are mirrored in the tangent space T_pM by the exponential map and controlled there by a set of linear maps derived from the matrix $A = \xi^a{}_{,b}(p)$. More precisely, if ψ is the exponential diffeomorphism from some open neighborhood U' of $0 \in T_pM$ to some open neighborhood U of p and χ_t represents the one parameter local group of local diffeomorphisms arising from the affine vector field ξ (for appropriate t) then (Ref. 11, Vol. 1)

$$\psi \circ \chi_{t^*} = \chi_t \circ \psi, \quad \chi_{t^*} = \exp(tA) \tag{6}$$

holds wherever it makes sense. In the resulting exponential coordinate system in U the integral curves of ξ satisfy the linear differential equation $(d/dt)(\mathbf{x}) = A\mathbf{x}$, where $A = \xi^a{}_{,b}(p)$ evaluated in this coordinate system. This follows by considering the corresponding integral curves that arise in T_pM under the map ψ. A consequence of this result is that in this coordinate system the components ξ^a of ξ are *linear* functions of the coordinates.

If ξ is not an affine vector field there is a more general theorem (the Sternberg linearization theorem—see, e.g., Ref. 16) which says that if ξ is any smooth vector field on any smooth n-dimensional manifold such that $\xi(p) = 0$ $(p \in M)$ and such that the eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ of the matrix $\xi^a{}_{,b}(p)$ counted properly (i.e., with their algebraic multiplicities as roots of the associated characteristic equation) have the property that

$$\lambda_i \neq m_1\lambda_1 + \cdots + m_n\lambda_n \quad (1 \leqslant i \leqslant n), \tag{7}$$

whenever the m_i are non-negative integers satisfying $2 \leqslant m_1 + \cdots + m_n$ then there exists a coordinate system about p in which the components ξ^a are linear functions of the coordinates. Thus, in such a coordinate system, the integral curves of ξ satisfy a linear differential equation similar to that given in the previous paragraph.

Now suppose that ξ is a conformal vector field on a space-time M with a zero at p. The algebraic structure (Segre type) of the matrix $A = \xi^a{}_{,b}(p)$ is the same as that of the matrix $F^a_b(p)$ [Eq. (1)] but with eigenvalues augmented by $\phi(p)$. Thus it is important to know the possible algebraic structures of the bivector F at p, i.e., to study the eigenvector–eigenvalue problem $F^a_b k^b = \lambda k^a$ at p for $k \in T_pM$ (or its complexification) and $\lambda \in \mathbb{C}$. One can easily show that if $F \neq 0$ at p, the allowed Segre types for F at p are $\{(11)z\bar{z}\}$, $\{11(11)\}$, $\{(31)\}$, or $\{11z\bar{z}\}$ where, in every case, the repeated eigenvalues (enclosed in round brackets) are zero and that these types correspond, respectively, to the cases where F is spacelike, timelike, null or nonsimple at p. The corresponding eigenvectors and eigenvalues are easily computed in each case and it should be noted that only null (real or complex) eigenvectors may have a corresponding nonzero eigenvalue. The various possibilities for the algebraic structure of the matrix $A = \xi^a{}_{,b}(p)$ in terms of the algebraic type for F at p and the value of $\phi(p)$ is listed in a way convenient for later use in Table I. It will be assumed that $\phi(p) \leqslant 0$ which can always be arranged by changing the sign of ξ if necessary and it is convenient to distinguish between the cases where $\phi(p) < 0$ and where $\phi(p) = 0$. It then follows that the condition (7) is always satisfied only for the cases $A1$, $B1$, and $C1$ in this table and so in these cases the corresponding conformal vector field can be linearized in the above sense in some coordinate neighborhood of p. The condition Eq. (7) is also satisfied for types $D1$ and $D3$ provided the ratio $\phi(p)\beta^{-1}$ is irrational but it is never satisfied for types $A2$, $B2$, $C2$, $D2$, $D4$, or E. Later, an example of a conformal vector field will be given which has a zero at p but which cannot be linearized in any coordinate neighborhood of p.

B. Zero sets and the Weyl and Ricci tensors

Let ξ be a global vector field on M with a zero at p and suppose one can construct a coordinate neighborhood U of p in which the components ξ^a are linear functions of the coordinates. It follows that the set of all zeros (the *zero set*) of ξ in U can be given the structure of a regular submanifold (a submanifold N of M is *regular* if its natural manifold topology coincides with the topology it inherits from the manifold topology of M) of U and hence of M of dimension $4 - \text{rank } A$ [where $A = \xi^a{}_{,b} = \xi^a{}_{,b}(p)$ and in the sense that if rank $A = 4$ the zero p is topologically isolated according to the inverse function theorem]. These remarks apply to conformal vector fields on M which are linearizable on some coordinate neighborhood of p. Even without

TABLE I. In this table F is the conformal bivector at the zero, p, of ξ $A = \xi^a{}_{,b}(p)$,and α and β are nonzero real numbers except in the timelike cases of D and E when $\alpha = 0$. In cases $A1$,$B1$,$C1$,and $D1$ all the eigenvalues of A have a negative real part. In case $D2$ three eigenvalues of A have negative real part and one has zero real part whilst in $D3$ [where it is assumed $\phi(p) < 0$] and $D4$ three eigenvalues of A have negative real part and one has positive real part. The eigenvalues in the case $D2$ are $2\phi(p)$, 0, $\phi(p) \pm i\alpha$ and in the case $D4$ are $3\phi(p)$, $-\phi(p)$, $\phi(p) \pm i\alpha$.

Nature of F	Eigenvalues of A	Restrictions	Type label
Zero	All equal to $\phi(p)$	$\phi(p) < 0$	$A1$
Zero	All equal to zero	$\phi(p) = 0$	$A2$
Null	All equal to $\phi(p)$	$\phi(p) < 0$	$B1$
Null	All equal to zero	$\phi(p) = 0$	$B2$
Spacelike	$\phi(p),\phi(p),\phi(p) \pm i\alpha$	$\phi(p) < 0$	$C1$
Spacelike	$0,0,\pm i\alpha$	$\phi(p) = 0$	$C2$
F timelike ($\alpha = 0$) or nonsimple	$\phi(p) \pm \beta,\phi(p) \pm i\alpha$	$\lvert\beta\rvert < -\phi(p)$	$D1$
		$\lvert\beta\rvert = -\phi(p)$	$D2$
		$-2\phi(p) \neq \lvert\beta\rvert > -\phi(p)$	$D3$
		$\lvert\beta\rvert = -2\phi(p)$	$D4$
F timelike ($\alpha = 0$) or nonsimple	$\pm\beta, \pm i\alpha$	$\phi(p) = 0$	E

the linearizability condition if A has type $A1$, $B1$, $C1$, $D1$, $D3$, or $D4$ then $A = \xi^a{}_{,b}(p)$ is nonsingular and the inverse function theorem guarantees that the zero p is isolated.

If ξ is a global affine vector field (and, in particular, if ξ is homothetic or Killing) on M with a zero at p the exponential map reveals the existence of a coordinate domain U in which ξ is linearizable. In this case, however, one has the extra consequence that if p is not isolated the submanifold of zeros of ξ in U is *totally geodesic* (Ref. 11, Vol. 2).

Now let ξ be a conformal vector field on M with a zero at p. Of interest later is the subset of points of M which lie on an integral curve of ξ (other than the trivial curve at p) which gets arbitrarily close to p in the sense that they lie on some nontrivial integral curve c: $I \to M$ of ξ for some open interval I of $\mathbb{R}$ such that there exists t_0 with either $t_0 \in \mathbb{R}$ or $t_0 = \pm\infty$ and $c(t) \to p$ as $t \to t_0$.

It is often important to have information regarding certain algebraic features of a space-time at a zero p of a conformal vector field ξ on M. In this respect it is useful to note that, since $\xi(p) = 0$, the condition $\mathcal{L}_\xi C^a{}_{bcd} = 0$ and Eq. (1) give the following algebraic relation at p:

$$C^e{}_{bcd}F^a{}_e - C^a{}_{ecd}F^e{}_b - C^a{}_{bed}F^e{}_c - C^a{}_{bce}F^e{}_d = 2\phi(p)C^a{}_{bcd}. \tag{8}$$

Equation (8) can be used to compute the Petrov type(s) at p if $\phi(p)$ and the algebraic type of the bivector F at p are known. Another result arises by considering the local diffeomorphisms χ_t arising from ξ and which act on some open neighborhood of p. These local diffeomorphisms have p as a fixed point and the associated linear maps $\chi_{t*}: T_pM \to T_pM$ are homothetic if $\phi(p) \neq 0$ and isometric if $\phi(p) = 0$ in an obvious way. They also provide useful information on the Weyl tensor at p as will be seen in the next section. Next, a contraction of Eq. (1) with $\xi^a\xi^b$ shows that along any integral curve of ξ with natural parameter t, $\xi^a\xi_a$ behaves like $A\exp(2\int\phi\,dt)$ for some constant A. It follows that ξ is of the same type (timelike, spacelike, or null) at each point of c. Also, a consideration of the eigenbivector–eigenvalue problem for the Weyl tensor shows that the Petrov type is the same at each point of c and that if γ is any Weyl eigenvalue (Petrov scalar) which is nonzero at some point of c, γ behaves like $B\exp(-2\int\phi\,dt)$ along c for some constant B (as a consequence of the relation $\mathcal{L}_\xi\gamma = -2\phi\gamma$). It follows that $\gamma\xi^a\xi_a$ is constant along c. The pleasant behavior of the Weyl tensor described above arises because of the conformal nature of the symmetry. However, as Eq. (4) suggests, there is no analogous behavior for the Ricci tensor (unless ξ is *special conformal*, i.e., $\phi_{a;b} = 0$ on M—see, e.g., Ref. 17). In the case that ξ is *homothetic* (including Killing), Eq. (4) yields $\mathcal{L}_\xi R_{ab} = 0$ and an equation analogous to Eq. (8) arises for the Ricci tensor at the zero p of ξ

$$R_{ac}F^c{}_b + R_{bc}F^c{}_a = -2\phi(p)R_{ab}, \tag{9}$$

and which is useful for computing the algebraic (Segre) type of the Ricci tensor[18] at p when the algebraic type of F at p is known. In this case the Ricci eigenvalues and Segre type behave in an analogous way to that described above for Petrov scalars and Petrov type along integral curves of ξ.[9]

C. Fixed point theorems

In this section the results relating to the conditions at, and in the neighborhood of, a zero p of a conformal vector field on M will be given. Some are well known but they are all collected together for easy reference. Throughout, $A \equiv \xi^a{}_{,b}(p)$.

Theorem 1: Let ξ be a Killing vector field on M with a zero at p. Then A is of type $B2$,$C2$, or E and the following hold.

(i) If A is of type $B2$ (respectively, $C2$) the Petrov type at p is N or 0 (respectively, D or 0) and the Ricci tensor has Segre type $\{(21)1\}$, $\{(31)\}$ or 0 [respectively, $\{1,1(11)\}$, $\{z\bar{z}(11)\}$, $\{2(11)\}$, or 0] or some degeneracy of these types at p. There exists an open neighborhood U of p such that the zero set of ξ in U is a two-dimensional, totally geodesic, regular null (respectively timelike) submanifold of U. No integral curve of ξ gets arbitrarily close to p.

(ii) If A is of type E the Petrov type at p is D or 0 and the Ricci tensor has Segre type $\{(1,1)11\}$ (if A is simple timelike) and $\{(1,1)(11)\}$ (if A is nonsimple) or some degeneracy of these types. If A is simple timelike there exists an open neighborhood U of p such that the zero set of ξ in U is a two-dimensional, totally geodesic spacelike regular submanifold of U whereas if A is nonsimple the zero p is (topologically) isolated. Finally there exists an open neighborhood U of p such that exactly four distinct maximal integral curves of ξ in U get arbitrarily close to p and each of these curves is a null geodesic.

Proof: The results describing Petrov types are well known.[6] Those describing Segre types are also known[7,19,20] and can be derived directly from Eq. (9) by noting that this equation implies that the blade of F (if F is simple) and the dual pair of blades of F (if F is nonsimple) are eigenspaces of the Ricci tensor. The results describing the zero sets of ξ follow by using the exponential map and Eq. (6)[9] and noting that fixed points of the χ_t correspond to fixed points of χ_{t*} under the exponential map ψ and hence to eigenvectors of F at p with zero eigenvalue. Their totally geodesic nature follows from Sec. III B and the fact that, when two-dimensional, they are of the same nature (timelike, spacelike or null) at each of their points follows from their having that nature at p and the fact that they are *autoparallel* submanifolds [since they are totally geodesic and the connection on M is symmetric (Ref. 11, Vol. 2)]. The remainder of the proof follows from considerations of the exponential mapping, the algebraic forms for F at p and the fact that the linearizability of ξ in normal coordinates implies that $\xi^a{}_{,b} = A^a{}_b = F^a{}_b(p)$ in this coordinate system from which one may derive the integral curves of ξ. Note that $\xi^a\xi_a$ is constant along an integral curve of ξ, from Killing's equation, and so any integral curve of ξ which gets arbitrarily close to p must be null. These occur only when A is type E and the integral curves of ξ that get arbitrarily close to p arise in an obvious way from the two independent null eigendirections of F at p and are thus null geodesics.

Theorem 2: Let ξ be a homothetic vector field on M with a zero at p. Then A is of type $A1$, $B1$, $C1$, or $D1$–$D4$ and the following hold.

(i) If A has type $A1$, $B1$, $C1$, or $D1$ then p is an isolated zero of ξ and there exists a neighborhood U of p such that any point of U lies on an integral curve of ξ which gets arbitrarily close to p (that is, p is an *asymptotically stable* zero of ξ.[13] M is flat in a neighborhood of p.

(ii) If A is of type $D3$ or $D4$, p is an isolated (*hyperbolic*[13]) zero of ξ and there exists an open neighborhood U of p a three-dimensional null submanifold N of U containing p and a null geodesic δ in U such that $N\cap\delta = \{p\}$ and such that any point in $N\cup\delta$ lies on an integral curve of ξ that gets arbitrarily close to p. If A is of type $D3$, or of type $D4$ with F nonsimple at p the Weyl and Ricci (and hence Riemann tensor) are zero at p. If A is of type $D4$ with F timelike at p the Weyl tensor is of Petrov type III or 0 and the Ricci tensor is of type $\{(31)\}$ with zero eigenvalue or zero at p.

(iii) If A is of type $D2$ there exists an open neighborhood U of p such that the zero set of ξ in U is a null geodesic in U through p. For each p' in this zero set there exists a three-dimensional null submanifold N' in U whose intersection with the zero set of ξ in U is precisely $\{p'\}$ and such that each point of N' lies on an integral curve of ξ that lies in N' and gets arbitrarily close to p'. The Petrov type at p is N or 0 (and is necessarily 0 if F is nonsimple at p) and the Ricci tensor at p has type $\{(211)\}$ with zero eigenvalue or 0. If the Riemann tensor does not vanish in a neighborhood of p then some neighborhood of p is isometric to a (generalized) plane wave space-time.

Proof: Parts of the proof can be found in Refs. 9, 21 and 8). The first part of (i) follows because all the eigenvalues of A have negative real part (see Table I). The final statement in (i) can be established by using the result $\mathscr{L}_\xi R^a{}_{bcd} = 0$ and the local diffeomorphisms χ_t arising from ξ to investigate the components $R^a{}_{bcd}$ of the curvature tensor in a normal coordinate neighborhood V of p arising from the exponential map ψ (see, e.g., a similar method used for computing the Ricci tensor components in Sec. III of Ref. 9 and cf. 8, 21). This argument, variations of which will be used elsewhere in this paper, starts by assuming without loss of generality that the above coordinate neighborhood V coincides with the neighborhood U in the statement of part (i) of the theorem. One then chooses a basis of T_pM adapted in an obvious way to the algebraic type of A to establish coordinates in V under ψ. The linear differential equations satisfied by the integral curves of ξ then enable the corresponding map χ_t to be calculated and so χ_{t*} can be calculated for the curvature tensor and the corresponding Lie derivative set equal to zero. The consequence is that any curvature component in this coordinate system would, if nonzero at some $q\in V$ become arbitrarily large along the integral curve of ξ through q toward p. The result then follows. For part (ii) the algebraic form of A in cases $D3$ and $D4$ (Table I) allows the choice of a real null tetrad $\{l,n,r,s\}$ at p (whose only nonvanishing inner products are $l^an_a = r^ar_a = s^as_a = 1$) such that l,r, and s are eigenvectors of A whose eigenvalues have negative real part and n is an eigenvector of A whose eigenvalue has positive real part. Using the ordered basis n,l,r,s of T_pM to construct normal coordinates u,v,x,y in some open neighborhood U of p one concludes that those points in U satisfying $u = 0$ and those satisfying $x=y=v=0$ constitute, respectively, a three-dimensional and a one-dimensional submanifold of U such that any point in either of them lies on an integral curve of ξ which gets arbitrarily close to p. That the latter of these submanifolds is part of a null geodesic in M is clear. That the former is null follows by first using the method above for computing the curvature components to show the metric tensor components in the above normal coordinates satisfy $g_{22} = g_{23} = g_{24} = 0$ and hence $g^{11} = 0$ on $u = 0$ and then noting that $u_{,a}u_{,b}g^{ab}=0$ on $u = 0$. Thus the existence of the submanifolds N and δ in the statement of the theorem is guaranteed. Concerning the remarks about Petrov and Segre types in Theorem 2 it follows from the homothetic nature of the action of the maps χ_{t*} on T_pM that *all the eigenvalues of the Weyl and Ricci tensors are zero at p* (Ref. 9) and hence that the Petrov type at p is III, N, or 0

and the Segre type of the Ricci tensor at p is $\{(211)\}$, $\{(31)\}$ (both with eigenvalue zero) or 0. The exact algebraic nature of A supplies further distinctions between these remaining possibilities either by the method given in Ref. 9 (in Ref. 9, p. 678, line 20 there is a misprint; the factor "2" should be "4") or by substituting the appropriate canonical forms for the above Petrov and Segre types and for the bivector F at p (from Table I) into Eqs. (8) or (9). This latter calculation can be simplified by noting that if the Weyl tensor C is nonzero at p (and hence in some open neighborhood W of p) and if k is a Debever–Penrose vector field for C in W, the equation $\mathcal{L}_\xi C^a{}_{bcd} = 0$ implies that $\mathcal{L}_\xi k \propto k$ in W and so k is a null eigenvector of F at p from (1). The same conclusion is reached if the Ricci tensor is nonzero at p and if k is its (unique up to scaling because of the above restriction on the Segre types) null eigenvector field. One then substitutes into (8) or (9) convenient canonical forms for F and for the appropriate types of $C^a{}_{bcd}$ (see e.g., Ref. 7) or R_{ab} (Ref. 7 and 18). The remainder of the calculation is straightforward and reveals also that if the Weyl and Ricci tensors are both nonzero at p then the repeated principal null direction of the former and the null eigendirection of the latter agree at p. The proof of those parts of (iii) not covered by the above discussion (and excluding the last sentence) follow since if A is of type $D2$ it has three eigenvalues with negative real part whilst the other is zero. Given that the curvature tensor does not vanish in a neighborhood of p this is the only case when the zero p is not isolated and leads to the conclusion regarding (generalized) plane waves.[21,9]

Further, remarks in Sec. III B show that the Ricci eigenvalues and Petrov scalars vanish along any non-null integral curve of ξ which gets arbitrarily close to p and so the Petrov and Segre types described in the above results are the only ones possible on the null submanifold N and N'. The components of ξ in the normal coordinates used in the above proof can be easily calculated.[9] An example when A is of type $D4$ with F timelike at p is the metric Eq. (5). Examples of these types can be constructed in Minkowski space by taking appropriate linear combinations of the homothetic vector field with components $\xi'^a = (x,y,z,t)$ in the usual global coordinates with Killing vector fields which vanish at the origin and noting that ξ' has zero homothetic bivector.

In the case when ξ is a proper conformal vector field on M such that $\xi(p) = 0$ the situation is more complicated. Since any Killing or homothetic vector field on M with zero at p can be converted into a proper conformal vector field with a zero at p by an appropriate conformal rescaling of the metric g on M all the features described in the above theorems are possible for ξ except (possibly) those which describe nonconformally invariant concepts (in this case, statements about the Ricci and Riemann tensors and the totally geodesic property). Rather than formally state a theorem analogous to 1 and 2 for proper conformal vector fields which would be somewhat verbose and repetitive (and, at present, incomplete), a brief discussion of features similar to those described in the previous theorems will be given.

Suppose ξ is a proper conformal vector field satisfying $\xi(p) = 0$. If $\phi(p) \neq 0$ [and so $\phi(p) < 0$ by an earlier convention] then by results earlier in this section and in Sec. III B one see that all Petrov scalars vanish (i.e., the Petrov type is III, N, or 0) at p and along all non-null integral curves of ξ which get arbitrarily close to p. The relation between the Petrov type at p and the type of the corresponding matrix A is just as in the homothetic case (Theorem 2). If, on the other hand, $\phi(p) = 0$ and the associated conformal bivector F is nonzero at p the corresponding isometric action on T_pM shows that the Petrov type at p is either D,N, or 0 just as for Killing vector fields. The relation between the Petrov type and the type of A is as in Theorem 1. If ϕ and F both vanish at p (i.e., A is of type $A2$) then it can be shown[22] that the Weyl tensor vanishes at p. An example of this latter type will be given later.

If ξ is a proper conformal vector field such that $\xi(p) = 0$ and A is nonsingular at p (i.e., A is of type $A1$, $B1$, $C1$, $D1$, $D3$, or $D4$), then the zero p is isolated (Sec. III B). If the matrix A is of type $A1$, $B1$, or $C1$, then the eigenvalue condition (7) holds and so by the linearization theorem there exists a coordinate neighborhood U of p in which the components ξ^a are linear functions of the coordinates. Thus $\xi^a{}_{,b}$ is a constant matrix on U equal to A when both are evaluated in these coordinates. For the cases $A1$, $B1$, and $C1$ all the eigenvalues of A have negative real part and the theory of linear differential equations then shows that p is an asymptotically stable zero of ξ so that there exists a neighborhood V of p such that every point of V lies on an integral curve of ξ which approaches p arbitrarily closely as $t \to \infty$. In the case $D1$, although the condition Eq. (7) fails, the same conclusions can be drawn from a theorem in nonlinear differential equations Ref. 13 chap. 9). In the other cases Eq. (7) also fails and so ξ may not be linearizable in a neighborhood of p. However, if any of these latter cases (excluding type $A2$) obtain *but ξ is linearizable in some neighborhood U of p* then the structure of the integral curves of ξ near p are similar to those obtained above in the Killing (for $\phi(p) = 0$) and homothetic (for $\phi(p) \neq 0$) cases. For example, if $\phi(p) \neq 0$ (i.e., $\phi(p) < 0$) and A has type $D2$ with F timelike one chooses an open coordinate neighborhood U of p in which ξ is linearizable and $\phi < 0$ and null vectors l and n at p such that $l_a n^a = 1$ and, at p, $F_{ab} = 2\phi(p) l_{[a} n_{b]}$. One may then make a *linear* transformation of the coordinates in U so that, at p, $l_a = (1,0,0,0)$, $n_a = (0,1,0,0)$ and $A^a_b = \phi(p)\mathrm{diag}(0,2,1,1)$. In U, the integral curves of ξ satisfy $dx^a/dt = A^a{}_b x^b$ and if one labels the coordinates (u,v,x,y) this leads to $u = c_1$, $v = c_2 \exp[2\phi(p)t]$, $x = c_3 \exp[\phi(p)t]$, $y = c_4 \exp[\phi(p)t]$ for constant c_1–c_4. The components of ξ are $\xi^a = (0,2v,x,y)$ and thus there is a one-dimensional submanifold in U consisting of zeros of ξ given by $v = x = y = 0$ which can be shown to be null because $\phi < 0$ in U. The above expression for the integral-curves of ξ can be used to calculate the behavior of the metric components in U. One uses them to calculate $\chi_t{}^* g$ and

then uses the fact that $\chi_{t*}g = \exp(-2\int\phi\,dt)g$. In order to avoid divergences as a zero of ξ is approached (i.e., as $t\to\infty$) one must take $g_{22}=g_{23}=g_{24}=0$ in U (and so $g^{11}=0$ in U). It follows that each of the three-dimensional submanifolds $u=\text{const}$ in U is null and is such that any point on, say, the submanifold $u=u_0$ lies on an integral curve of ξ which lies in this submanifold and gets arbitrarily close to the zero of ξ with coordinates $(u_0,0,0,0)$. The Petrov type is N or 0 in some neighborhood of p. Similar arguments hold if F is nonsimple only now the Weyl tensor necessarily vanishes at p. Still in the linearizable case if A is of type $D3$ or $D4$ similar arguments show that one obtains a null three-dimensional and a null one-dimensional submanifold in U (which intersect in the isolated zero p) such that any point on either of these submanifolds lies on an integral curve of ξ in that submanifold which gets arbitrarily close to p. The Petrov type at p is 0 in the $D3$ case and 0 or III in the $D4$ case. However, if ξ is linearizable and $\phi(p)=0$, only where A is of type E can there exist integral curves of ξ which get arbitrarily close to p and one may choose U such that these curves are four in number and null. If F is also simple the zero set of ξ is a two-dimensional spacelike submanifold of U for some open neighborhood U of p whereas if F is nonsimple, p is an isolated zero of ξ. For types $B2$ and $C2$ there is an open neighborhood U of p such that the zero set of ξ in U is a two-dimensional submanifold of U and U may be chosen such that this submanifold is timelike for type $C2$.

If ξ is a proper conformal vector field with $\xi(p)=0$ and A is of type $A1$, $B1$, $C1$, or $D1$ with $\phi(p)\beta^{-1}$ irrational in the $D1$ case then ξ is necessarily linearizable and p is an isolated, asymptotically stable zero of ξ. In this case, similar methods to those used in the previous paragraph but applied to the Weyl tensor $C^a{}_{bcd}$ together with the condition $\mathcal{L}_\xi C^a{}_{bcd}=0$ show that the Weyl tensor vanishes in some neighborhood of p [cf. the last part of Theorem 2(i)].

If, however, ξ cannot be linearized in some coordinate neighborhood of p the situation seems more complicated. For example, consider Minkowski space with the usual global coordinates x,y,z,t and construct the vector field ξ given by

$$\xi^a=\tfrac{1}{4}(\eta_{cd}x^cx^d)k^a-\tfrac{1}{2}(\eta_{cd}x^ck^d)x^a, \tag{10}$$

where η is the Minkowski metric tensor, $k=(1,0,0,0)$ and $x^a=(x,y,z,t)$. Then ξ satisfies Eq. (1) with $\phi=-(1/2)(\eta_{ab}k^ax^b)$ and $F_{ab}=k_{[a}x_{b]}$. The set of zeros of ξ are those points whose coordinates x^a satisfy $\eta_{ab}x^ax^b=\eta_{ab}x^ak^b=0$ (a "null cone" in three-dimensional Minkowski space). Thus ξ has a zero at the origin (of type $A2$) but the set of zeros cannot be given the structure of a regular submanifold of M in any neighborhood of the origin. Hence ξ cannot be linearized in any neighborhood of the origin. (Note that if the origin is removed from Minkowski space, the two remaining components of the set of zeros of ξ can now be given regular submanifold structures in M but neither is totally geodesic.)

IV. CONFORMAL SYMMETRIES IN GENERALIZED *pp* AND PLANE WAVES

In this section, a general geometrical discussion is given of conformal symmetries in generalized *pp* and plane waves as defined in Sec. I. Some of these results are implicitly contained in Ref. 23 as a result of direct calculation.

Let M be a space-time whose Weyl tensor is nowhere zero and which admits a global nowhere zero, covariantly constant null vector field l. [The results of this section still hold (with a few obvious modifications) if the Weyl tensor is nonzero over an open dense subset of M.] If $p\in M$ assume that coordinates are chosen in some neighborhood U of p such that u is one of them and that $l_a=u_{,a}$ in U. The Ricci identity then gives $R_{abcd}l^d=0$ and it is easily shown using the Bel criteria that the Weyl tensor is algebraically special with l a repeated principal null direction at each $p\in M$. Now suppose that ξ is a proper conformal vector field on M satisfying Eqs. (1)–(3). The equation $\mathcal{L}_\xi C^a{}_{bcd}=0$ implies that $\mathcal{L}_\xi l\propto l$ and so it follows from (1) that

$$F_{ab}l^b=vl_a, \tag{11}$$

for some function v on M. Define also the function χ on M by $\chi=\xi^al_a$. By taking the covariant derivative of Eq. (11) and using Eq. (2) one finds

$$-\phi_{,a}l_c+(\phi_{,b}l^b)g_{ac}=v_{,c}l_a, \tag{12}$$

which from elementary rank considerations shows that $\phi_{,b}l^b=0$ and then that $\phi=\phi(u)$ and $v=v(u)$ in U (in fact (12) shows that $\phi+v$ is constant). A contraction of (1) with l^a then reveals that $\chi=\chi(u)$ and $\chi'=\phi-v=2\phi+e(e\in\mathbb{R})$ where $'$ denotes d/du. One also sees that if $\chi\equiv 0$ (so that, locally, ξ lies in the hypersurfaces $u=\text{const}$) then $l^a\xi_{a;b}=0$ and substituting from Eq. (1), covariantly differentiating the result and using Eq. (2) shows that $\phi'\equiv 0$ (and $v'\equiv 0$). Hence for such space-times, conformal vector fields which locally lie in the hypersurfaces $u=\text{const}$ are homothetic or Killing.

Now suppose as well that M is a vacuum space-time and hence, a (vacuum) *pp* wave space-time.[6] It is a consequence of Brinkmann's theorem (a corollary of which says that if Ω is a positive function on M and g and Ωg are vacuum metrics on M then either Ω is constant or g and Ωg are *pp* wave metrics) that *the only vacuum space-times admitting proper conformal vector fields are the pp* waves. This latter result can also be deduced from (3) since, then, either ϕ is constant or $\phi_{,a}$ is a (necessarily null) covariantly constant vector field on M. If a vacuum *pp* wave admits a proper conformal vector field then the above discussion and Eq. (3) show that ϕ is a *linear* function of u. It follows that a vacuum *pp* wave space-time M can admit at most one independent proper conformal vector field in the sense that if ξ_1 and ξ_2 are distinct proper conformal vector fields on M then some linear combination $a\xi_1+b\xi_2(a,b,\in\mathbb{R})$ of them is homothetic or Killing. Also for vacuum space-times v is a linear and x a quadratic function of u.

Returning to the general case described in the second paragraph of this section suppose now that ξ is a Killing vector field on M with Killing bivector F. Then Eq. (11)

shows that l is an eigenvector of F and, since now $\phi = 0$ and also $\phi + v$ constant, v is constant and χ is a linear function of u. If also $\chi \equiv \xi^a l_a \equiv 0$, then $v = 0$. These results are helpful in the geometrical interpretation of the local diffeomorphisms ϕ_t arising from ξ using parallely propagated tetrads along the integral curves of ξ since they restrict the algebraic type of F (which is necessarily the same along each integral curve of ξ).[9,12]

Now let M above be a vacuum plane wave. Then M admits exactly five independent global Killing vector fields which lie in the $u = \text{const}$ null hypersurfaces and at most one other independent global Killing vector field not confined to these hypersurfaces (i.e., its associated function χ is not identically zero). That there is at most one of the latter type follows from isotropy considerations and the fact that the Petrov type is N. The general form of the metric for M is well known[6,7] as is the fact that M admits a homothetic vector field lying in the above null hypersurfaces whose zeros constitute a null geodesic in M[2,9] (see Sec. III). Thus the Lie algebra of conformal vector fields is at least six-dimensional and possibly seven-dimensional, the extra conformal vector field being either Killing (the homogeneous case) or proper conformal. In neither case can the extra conformal vector field lie in the hypersurfaces $u = \text{const}$ as follows from remarks above. (The only way a homothetic vector field can exist that does not lie in these hypersurfaces is by taking appropriate linear combinations in the homogeneous case.) That the maximum dimension of this algebra is seven can easily be established. In fact from the above remarks and those regarding independent proper conformal vector fields earlier it suffices only to show that M cannot admit a Killing vector field ξ_1 and a proper conformal vector field ξ_2 neither of which lie in the $u = \text{const}$ hypersurfaces. If such a situation exists one simply changes g to the conformally related metric $g' \equiv e^{2\sigma} g$ with $\sigma = -\log(u+d)$ $(d \in \mathbb{R})$ on some appropriate open submanifold of M. The metric g' is easily shown to be a vacuum plane wave metric and the constant d may be chosen so that ξ_1 and ξ_2 are (independent) proper conformal vector fields with respect to g' in contradiction to an above result. Three other brief remarks can be added here. First, for each $p \in M$ there exists a homothetic vector field whose zeros constitute a null geodesic through p, a geometrical discussion of which was given in Sec. III. It is a consequence of the nowhere vanishing of the Weyl tensor that these types of homothetic vector fields are the only ones possessed by M which have a zero. Second, considerations of isotropy and Petrov type show that a Killing vector on M not lying in the $u = \text{const}$ hypersurfaces has no zeros. Finally, the work above shows that a vacuum plane wave, apart from the six-dimensional Lie algebra of homothetic and Killing vector fields which are everywhere orthogonal to l may (but need not) have either an extra Killing or proper conformal vector field (but not both) which is not everywhere orthogonal to l. In fact these two possibilities are conformally related in the following sense. A conformal change of metric $g \to g' = e^{2\sigma} g$, $\sigma = -\log(u+d)$ $(d \in \mathbb{R})$ will preserve the vacuum and plane wave conditions on some open submanifold of M and any such extra Killing vector field is now a proper conformal vector field with respect to g'. Conversely, given an extra proper conformal vector field ξ on M the associated quadratic function x factorises over $\mathbf{R}$ (Ref. 23) and so one has $\phi = au + b$, $\chi = a(u - c_1)(u - c_2)$ $(a,b,c_1,c_2 \in \mathbf{R})$. It follows that one can choose d in the above conformal transformation such that either ξ is Killing or else homothetic (and either possibility leads to the existence of an extra Killing vector field in the above sense) with respect to g'.

Finally let M above be a generalized plane wave whose Weyl tensor is nowhere zero. The Ricci tensor takes the general coordinate form $R_{ab} = \rho(u) l_a l_b$. Under a conformal change of metric $g \to g' = e^{2\sigma} g$, $\sigma = \sigma(u)$, the metric g' is a vacuum (and hence a vacuum plane wave) metric if and only if $2\sigma'' - 2\sigma'^2 = \rho(u)$. This is readily converted into two first order differential equations and a solution $\sigma(u)$, at least for some open interval of values of u about any initial value u_0 of u, is assured by Picard's theorem. *Hence any such generalized plane wave is locally conformally related to a vacuum plane wave.* Also, as pointed out above, any vacuum plane wave can be locally conformally rescaled $g \to g' = e^{2\sigma} g$ so that its Lie algebra of conformal vector fields contains only homothetic or Killing vector fields with respect to g'. It follows that if g is a generalized plane wave metric on M whose Weyl tensor is nowhere zero, the associated Lie algebra of conformal vector fields has dimension at most seven and for each $p \in M$ there is a neighborhood U of p and a function of $\sigma'(u)$ such that the metric $e^{2\sigma} g$ on U (also a generalized plane wave) has a Lie algebra of conformal vector fields on U consisting entirely of homothetic and Killing vector fields (cf. Ref. 2).

ACKNOWLEDGMENTS

The author wishes to thank John Pulham, John Steele, and Jose da Costa for many useful discussions.

[1] R. F. Bilyalov, Sov. Phys. **8**, 878 (1964).
[2] L. Defrise-Carter, Commun. Math. Phys. **40**, 273 (1975).
[3] K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1955).
[4] T. Siguri and S. Ueno, Tensor, N. S. **24**, 253 (1972).
[5] R. P. Geroch, Commun. Math. Phys. **13**, 180 (1969).
[6] J. Ehlers and W. Kundt, in *Gravitation: An Introduction to Current Research*, edited L. Witten (Wiley, New York, 1962).
[7] D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (V.E.B. Deutscher Verlag der Wissenschaften, Berlin, 1980).
[8] J. K. Beem, Lett. Math. Phys. **2**, 317 (1978).
[9] G. S. Hall, Gen. Rel. Grav. **20**, 671 (1988).
[10] G. S. Hall, Class. Quant. Grav. **6**, 157 (1989).
[11] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, Vol. 1, 1963 and Vol. 2, 1969).
[12] G. S. Hall, "Symmetries in General Relativity" in *Relativity Today*, edited Z. Perjes (World Scientific, Singapore, 1988).
[13] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra* (Academic, New York, 1974).
[14] R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications* (Addison-Wesley, Reading, MA, 1983).
[15] V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer, New York, 1983).
[16] E. Nelson, *Topics in Dynamics*, 1. Flows (Princeton U.P. and University of Tokyo Press, Princeton, 1969).
[17] G. S. Hall, to appear in Gen. Relat. Gravit. (1990).

[18] G. S. Hall, J. Phys. A **9**, 541 (1976).
[19] R. F. Crade and G. S. Hall, Phys. Lett. A **75** 17 (1979).
[20] M. A. H. MacCallum, in *Essays on General Relativity,* edited by F. Tipler (Academic, New York, 1980).
[21] D. Alexeevski, Ann. Glob. Anal Geom. **3**, 59 (1985).
[22] G. S. Hall and J. D. Steele, to appear in Gen. Relat. Gravit. (1990).
[23] H. Salazar, A. Garcia, and J. F. Plebanski, J. Math. Phys. **24**, 2191 (1983).

Future null infinity of Robertson–Walker space-times

Osvaldo M. Moreschi[a)]
International Centre for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy

(Received 6 September 1988; accepted for publication 22 November 1989)

The future null infinity for all noncontracting Robertson–Walker space-times is studied systematically. A theorem is proved that establishes the expected relation between the nature of $\mathscr{I}^+$ and the appearance or absence of cosmic event horizons.

I. INTRODUCTION

In the standard model of cosmology it is assumed that, on the largest scale, the universe can be reasonably represented by a Robertson–Walker (RW) space-time.

When we observe the universe we usually obtain information encoded in the electromagnetic radiation that arrives from the particular observed object. As we try to study more distant objects we are forced to direct our attention to earlier times. Thus when we are studying the largest scale of our universe we are, for all practical purposes, at future null infinity of the observed region.

It is then necessary in the study of the largest scale of the universe to have a clear picture of future null infinity of the Robertson–Walker space-times.

In this paper we make a systematic study of future null infinity for noncontracting RW space-times.

Some particular examples have been extensively discussed in the literature.[1–3] However, here we develop the techniques that allow the study of all noncontracting models. In any case, we review the Minkowski and de Sitter examples in Sec. II and the dust Friedmann models in Sec. III.

In Sec. IV the general method is described, and it is applied to characteristic asymptotic behavior.

Section V contains some closing remarks.

In the remainder of this introductory section we will mainly present the notation to be used.

Let us consider noncontracting Robertson–Walker models. Their line element can be given by

$$ds^2 = dt^2 - A(t)^2\, dL_K^2, \tag{1.1}$$

where dL_K can be expressed in several equivalent ways:

$$dL_K^2 = d\hat{r}^2/(1-K\hat{r}^2) + \hat{r}^2\, d\Sigma^2, \tag{1.2}$$

$$dL_K^2 = (d\tilde{r}^2 + \tilde{r}^2\, d\Sigma^2)/(1+K\tilde{r}^2/4)^2, \tag{1.3}$$

$$dL_K^2 = d\chi^2 + \mathscr{f}_K^2(\chi)d\Sigma^2. \tag{1.4}$$

Here $d\Sigma^2$ is the line element of the unit sphere, which can also be expressed in a variety of ways, for example,

$$d\Sigma^2 \equiv d\theta^2 + \sin(\theta)^2\, d\phi^2 \tag{1.5}$$

$$= (4\, d\zeta\, d\bar{\zeta})/(1+\zeta\bar{\zeta})^2 = (d\zeta\, d\bar{\zeta})/P_0^2, \tag{1.6}$$

with

$$P_0 \equiv (1+\zeta\bar{\zeta})/2. \tag{1.7}$$

Finally the function $\mathscr{f}_K$ is defined by

$$\mathscr{f}_K(\chi) = \begin{cases} \sinh(\chi), & \text{for } K=-1, \quad 0\leqslant\chi<\infty, \\ \chi, & \text{for } K=0, \quad 0\leqslant\chi<\infty, \\ \sin(\chi), & \text{for } K=1, \quad 0\leqslant\chi\leqslant\pi. \end{cases} \tag{1.8}$$

The range of the coordinate t is associated with the behavior of the function $A(t)$. When there is an initial singularity followed by a continuing expansion one takes $0<t<\infty$.

By a noncontracting RW space we mean that the scalar $A(t)$ must satisfy

$$\frac{\partial A}{\partial t} \geqslant 0. \tag{1.9}$$

In the study of future null infinity of asymptotically flat space-times, the use of null coordinates and/or null tetrads adapted to scri has proved useful. It is also useful in our case to introduce a null coordinate u by the equation

$$du = \frac{dt}{A(t)} - d\chi. \tag{1.10}$$

Using this relation to replace dt in the line element one obtains

$$ds^2 = A^2(du^2 + 2\, du\, d\chi - \mathscr{f}_K^2\, d\Sigma^2). \tag{1.11}$$

The range of the coordinate u is determined by the asymptotic behavior of the scalar $A(t)$; that is, for example, if

$$\lim_{t\to\infty} \int_{t_0}^{t} \frac{dt'}{A(t')} = \infty, \tag{1.12}$$

then the function u is unbounded from above and its range in the cases $K=-1$ and $K=0$ is $-\infty<u<\infty$. Instead, if the limit is finite then the coordinate u is bounded from above (let us say $u<u_0$); this can be thought of as the manifestation of the appearance of cosmic event horizons.

We can also define the coordinate r by the expression

$$r \equiv \int_{t_0}^{t(u,\chi)} A(t')dt' - r_0(u); \tag{1.13}$$

which satisfies

$$\frac{\partial r}{\partial \chi} \equiv \frac{\partial r}{\partial \chi}\bigg|_u = \frac{\partial r}{\partial t}\frac{\partial t}{\partial \chi} = A\frac{\partial t}{\partial \chi} = A^2, \tag{1.14}$$

so that

$$dr = (A^2 - \dot{r}_0)du + A^2\, d\chi, \tag{1.15}$$

where we are using the notation $\dot{r}_0 \equiv dr_0/du$.

Then using r instead of the coordinate χ, the Robertson–Walker line element is given by

$$ds^2 = (2\dot{r}_0 - A^2)du^2 + 2\, du\, dr - A^2\mathscr{f}_k^2\, d\Sigma^2. \tag{1.16}$$

Note that the line element is spherically symmetric; thus we do not change the angular coordinates. The relation between the original nonangular coordinates and the new ones can be expressed by the following differential equations:

a) Present address: FAMAF, Laprida 854, 5000 Córdoba, Argentina.

 0022-2488/90/051208-09$03.00

$$d\chi = (\dot{r}_0/A^2 - 1)du + (1/A^2)dr, \tag{1.17}$$

$$dt = (\dot{r}_0/A)du + (1/A)dr. \tag{1.18}$$

Let us now introduce a null tetrad adapted to this new coordinate system. We can deduce from the line element that $u = \text{const}$ are null hypersurfaces, so we take the null vector ℓ to be

$$\ell = du. \tag{1.19}$$

Then we have

$$g^{ab}\ell_a\ell_b = 0, \tag{1.20}$$

$$\ell^a\nabla_a\ell^b = 0, \tag{1.21}$$

and

$$\ell^a = g^{ab}\ell_b = \left(\frac{\partial}{\partial r}\right)^a; \tag{1.22}$$

that is, r is an affine parameter of the future directed null geodesics in the null hypersurface $u = \text{const}$.

Complex null vectors m^a and $\bar{m}^a$ are taken such that

$$m^a = \frac{1}{A\mathcal{J}_k} m_0^a, \quad \text{with } m_0^a \equiv \sqrt{2}\, P_0 \frac{\partial}{\partial\zeta}. \tag{1.23}$$

The last null vector n is defined by

$$n^a = \left(\frac{\partial}{\partial u} + U\frac{\partial}{\partial r}\right)^a, \quad \text{with } U \equiv \frac{1}{2}(A^2 - 2\dot{r}_0). \tag{1.24}$$

This null tetrad satisfies the usual contraction relations

$$n^a\ell_a = -m^a\bar{m}_a = 1, \tag{1.25}$$

while all other contractions are zero.

Although now the expansion parameter A is a function of u and r, it is convenient to retain the original functional dependence of $A(t)$. Therefore we will use a prime to denote derivatives with respect to t, that is,

$$A' \equiv \frac{\partial A(t)}{\partial t}, \tag{1.26}$$

and evaluate coordinate derivatives of A in terms of A'. They are

$$\frac{\partial A}{\partial u} = A'\frac{\dot{r}_0}{A} \tag{1.27}$$

and

$$\frac{\partial A}{\partial r} = \frac{A'}{A}. \tag{1.28}$$

We obtain similar expressions evaluating coordinate derivatives of A':

$$\frac{\partial A'}{\partial u} = \frac{A''}{A}\dot{r}_0 \tag{1.29}$$

and

$$\frac{\partial A'}{\partial r} = \frac{A''}{A}. \tag{1.30}$$

Using the fact that

$$\frac{\partial \mathcal{J}_K}{\partial\chi} = \sqrt{1 - K\mathcal{J}_K^2}, \tag{1.31}$$

the coordinate derivatives of $\mathcal{J}_K$ can be expressed by

$$\frac{\partial \mathcal{J}_K}{\partial u} = \frac{\partial \mathcal{J}_K}{\partial\chi}\frac{\partial\chi}{\partial u} = \sqrt{1 - K\mathcal{J}_K^2}\left(\frac{\dot{r}_0}{A^2} - 1\right) \tag{1.32}$$

and

$$\frac{\partial \mathcal{J}_K}{\partial r} = \frac{\partial \mathcal{J}_K}{\partial\chi}\frac{\partial\chi}{\partial r} = \sqrt{1 - K\mathcal{J}_K^2}\,\frac{1}{A^2}. \tag{1.33}$$

In terms of the Geroch–Held–Penrose (GHP) notation[4] the only components of the curvature tensor different from zero are

$$\Phi_{22} = (K + A'^2 - AA'')/4, \tag{1.34}$$

$$\Phi_{11} = (K + A'^2 - AA'')/4A^2, \tag{1.35a}$$

$$\Lambda = (K + A'^2 + AA'')/4A^2, \tag{1.35b}$$

$$\Phi_{00} = (K + A'^2 - AA'')/A^4. \tag{1.36}$$

At this stage it is natural to try to follow some of the techniques that were used in the definitions of future asymptotically flat space-times.[5] These techniques include the use of a conformal factor Ω, which will bring infinity to a finite distance in the conformally related manifold $\tilde{M}$ with metric

$$\tilde{g}_{ab} = \Omega^2 g_{ab}. \tag{1.37}$$

We have constructed our null tetrad and coordinate system out of a family of null hypersurface that define a null congruence reaching the asymptotic region under study. Then, since the function r is an affine parameter along these null geodesics, we know that taking Ω proportional to $1/r$ will bring future null infinity to a finite affine distance.

We will study the consequences of taking $\Omega = 1/r$ in Secs. III–V. But let us first review two of the most celebrated isotropic cosmological models.

II. TWO EXCEPTIONAL CASES: MINKOWSKI AND de SITTER

Among the RW family, two space-times deserve special treatment: they are Minkowski and de Sitter space. These are the only[6] metrics that can be expressed in more than one of the forms

$$ds^2 = dt^2 - A(t)^2\, dL_K^2; \tag{1.1}$$

that is, Minkowski can be represented by a line element of this form with $K = 0$ and $K = -1$, while de Sitter can be represented by the three possible values of K.

A. Minkowski space

In Minkowski space we have

$$0 = (K + A'^2 - AA'')/4A^2 \tag{2.1}$$

and

$$0 = (K + A'^2 + AA'')/4A^2; \tag{2.2}$$

therefore

$$A' = \sqrt{-K}, \tag{2.3}$$

with solutions

$$A = t, \quad \text{for } K = -1, \tag{2.4}$$

and

$$A = \text{const} = 1, \quad \text{for } K = 0; \tag{2.5}$$

where, without loss of generality, we have chosen the arbitrary constants to be 1. So we can express $A(t)$ by

$$A(t) = -Kt + (1+K). \tag{2.6}$$

Minkowski space is particularly special among the RW modes since it is flat; and obviously asymptotically flat. When $K = -1$, one refers to it as the Milne model.[7]

It has been customary to represent the null infinity of Minkowski space by its conformal map into the Einstein universe, which is obtained from the $K = 0$ form of the metric through the transformation

$$t' = \arctan(t+\chi) + \arctan(t-\chi), \tag{2.7}$$

$$\chi' = \arctan(t+\chi) - \arctan(t-\chi), \tag{2.8}$$

with the coordinate range $-\pi < t' + \chi' < \pi$ and $-\pi < t' - \chi' < \pi$.

The Minkowski line element can then be expressed by

$$\begin{aligned} ds^2 &= dt^2 - d\chi^2 - \chi^2\, d\Sigma^2 \\ &= \left[\frac{\sec((t'+\chi')/2)\sec((t'-\chi')/2)}{2}\right]^2 \\ &\quad \times (dt'^2 - d\chi^2 - \sin^2\chi'\, d\Sigma^2), \end{aligned} \tag{2.9}$$

where, from the last line, one can see the conformal relation between the Minkowski metric and that of the Einstein space-time. Null infinity agrees with the region where the conformal factor is zero, and, as is well known, it is formed by two null hypersurfaces in the Einstein space corresponding to future and past null infinity, respectively.

Alternatively one could also study future null infinity of this space without making any reference to the Einstein universe.

Here we end our short characterization of Minkowski space, and go on to consider other nontrivial cases.

B. de Sitter space

If we only require the trace-free part of the Ricci tensor to be zero we obtain

$$0 = K + A'^2 - AA', \tag{2.10}$$

and, due to the contracted Bianchi identities,

$$\Lambda = (K + A'^2 + AA'')/4A^2 = \mathbb{C}/2, \tag{2.11}$$

where $\mathbb{C}$ is a constant. From these equations we deduce

$$(K + A'^2)/A^2 = \mathbb{C} \tag{2.12}$$

and

$$AA''/A^2 = \mathbb{C}. \tag{2.13}$$

The solutions of which, for $\mathbb{C} > 0$, are

$$t = \mathbb{C}^{-1/2} \ln(A\mathbb{C}^{1/2} + (A^2\mathbb{C} - K)^{1/2}), \tag{2.14}$$

or, in terms of $A(t)$,

$$A(t) = \begin{cases} (1/\sqrt{\mathbb{C}})\sinh(t\mathbb{C}^{1/2}), & \text{for } K = -1, \\ (1/2\sqrt{\mathbb{C}})e^{(t\mathbb{C}^{1/2})}, & \text{for } K = 0, \\ (1/\sqrt{\mathbb{C}})\cosh(t\mathbb{C}^{1/2}), & \text{for } K = 1, \end{cases} \tag{2.15}$$

which can also be expressed in one line as

$$A(t) = (1/\sqrt{\mathbb{C}})(e^{(t\mathbb{C}^{1/2})} + Ke^{(-t\mathbb{C}^{1/2})}/2). \tag{2.16}$$

Defining the function

$$t' = \arctan(\sinh(t\mathbb{C}^{1/2})), \tag{2.17}$$

the de Sitter metric can be expressed by

$$ds^2 = A(t)^2(dt'^2 - d\chi^2 - \sin(\chi)^2\, d\Sigma^2). \tag{2.18}$$

We see that this space is conformal to a portion of the static Einstein universe with the range of the time coordinate given by $-\pi/2 < t' < \pi/2$. One then takes the conformal factor Ω to be given by A^{-1}. By doing so one finds that $\mathscr{I}^+$ turns out to be a spacelike hypersurface in the static Einstein universe.

As we have observed these two exceptional examples are easily related to the Einstein space. This suggests we should refer every RW model to its conformal image in the Einstein universe since "it is a kind of maximal universal conformally flat spacetime."[3]

Instead of carrying out this program, in Sec. IV we will try to follow closely the techniques used in the study of asymptotically flat space-times. Note that de Sitter space is particularly easily related to the Einstein universe because it can be expressed as a line element with $K = 1$.

III. IS THE FRIEDMAN MODEL ASYMPTOTICALLY FLAT?

One of the key features of the definition of GeFAF space-times[5] is the flatness condition on the Riemann tensor, which in the regular case is

$$R_{abc}{}^d = \Omega\hat{R}_{abc}{}^d + \delta R_{abc}{}^d, \tag{3.1}$$

where $\hat{R}_{abc}{}^d$ is a regular tensor at future null infinity and $\delta R_{abc}{}^d$ goes to zero faster than Ω.

In our case, for the Robertson–Walker space-times, we expect to have a generalization of this behavior that will look like

$$R_{abc}{}^d = \mathcal{A}(\Omega)\hat{R}_{abc}{}^d + \delta R_{abc}{}^d; \tag{3.2}$$

where now $\mathcal{A}(\Omega)$ is some function of Ω that might diverge for $\Omega \to 0$.

It is amusing in any case to compare Eq. (3.1) with Eqs. (1.34)–(1.36). Observing equations (3.81), (3.83), (3.85), and (3.86) of Ref. 5, we see that this comparison will imply the following relations:

$$\Phi_{22} = (K + A'^2 - AA'')/4 = \omega\hat{\Phi}_{22} + \delta\Phi_{22}, \tag{3.3}$$

$$\Phi_{11} = (K + A'^2 - AA'')/4A^2 = \omega^3\hat{\Phi}_{11} + \delta\Phi_{11}, \tag{3.4}$$

$$\Phi_{00} = (K + A'^2 - AA'')/A^4 = \omega^5\hat{\Phi}_{00} + \delta\Phi_{00}, \tag{3.5}$$

$$\Lambda = (K + A'^2 + AA'')/4A^2 = \omega^3\hat{\Lambda} + \delta\Lambda, \tag{3.6}$$

where we are using ω in order to differentiate from the true conformal factor Ω. It is observed then that this comparison establishes the following proportional expressions:

$$(K + A'^2 - AA'') \propto \omega, \tag{3.7}$$

$$1/A \propto \omega, \tag{3.8}$$

$$(K + A'^2 + AA'') \propto \omega. \tag{3.9}$$

Therefore one should have

$$(K + A'^2 - AA'') \propto 1/A \tag{3.10}$$

and

$$(K + A'^2 + AA'') \propto 1/A; \tag{3.11}$$

which implies

$$K + A'^2 = Q/A \tag{3.12}$$

and

$$AA'' = Q'/A, \tag{3.13}$$

for some constants Q and Q'. Computing the derivative of the first expression, we obtain

$$(K + A'^2)' = 2A'A'' = -(Q/A^2)A'; \tag{3.14}$$

so the second equation is a consequence of the first and, in particular, we have that $Q' = -Q/2$.

The first equation is nothing other than the Friedmann equation. It is then somewhat surprising that by playing around with the idea of RW space-times that look future asymptotically flat, we do not find a trivial or extremely complicated new model, but instead the dust Friedmann model.

It can easily be seen that in the Friedmann model and the curvature components are given by

$$\Phi_{22} = (K + A'^2 - AA'')/4 = 3Q/8A, \tag{3.15}$$

$$\Phi_{11} = (K + A'^2 - AA'')/4A^2 = 3Q/8A^3, \tag{3.16}$$

$$\Phi_{00} = (K + A'^2 - AA'')/A^4 = 3Q/2A^5, \tag{3.17}$$

$$\Lambda = (K + A'^2 + AA'')/4A^2 = Q/8A^3. \tag{3.18}$$

We will now relate these expressions with the family of conformal factors Ω proportional to the inverse of the affine null distances, that is,

$$\Omega \propto 1/r. \tag{3.19}$$

It will be useful to note that

$$dA = \frac{\partial A}{\partial u}\,du + \frac{\partial A}{\partial r}\,dr = A'\frac{\dot{r}_0}{A}\,du + \frac{A'}{A}\,dr$$

$$= \frac{A'}{A}\,d(r_0 + r), \tag{3.20}$$

since then we can express

$$r + r_0 = \int \frac{A\,dA}{A'} = \int \frac{A\,dA}{\sqrt{Q/A - K}} = \begin{cases} \sqrt{\dfrac{Q}{A} + 1}\left(\dfrac{A^2}{2} - \dfrac{3QA}{4}\right) + \dfrac{3Q^2}{8}\ln\left(\dfrac{\sqrt{Q/A + 1} + 1}{\sqrt{Q/A + 1} - 1}\right), & \text{for } K = -1, \\ \dfrac{2A^{5/2}}{5\sqrt{Q}}, & \text{for } K = 0. \end{cases} \tag{3.21}$$

It is observed from the last expression that in the limit $r \to \infty$ one also has $A \to \infty$; and the leading order of this equation gives

$$r \approx A^2/2, \quad \text{for } K = -1, \tag{3.22}$$

and

$$r \approx 2A^{5/2}/5\sqrt{Q}, \quad \text{for } K = 0. \tag{3.23}$$

Let us next study the two cases corresponding to the two possible values of the constant K.

A. Case $K = -1$

In this case we can take the conformal factor Ω to be

$$\Omega = 1/A^2; \tag{3.24}$$

that is, Ω is asymptotically given in terms of the affine distance r by

$$\Omega \approx 1/2r. \tag{3.25}$$

Let us see what the conformal metric looks like at future null infinity. From Eq. (1.16) we obtain

$$d\tilde{s}^2 = \Omega^2\,ds^2 = A^{-4}(2\dot{r}_0 - A^2)du^2 + 2A^{-4}\,du\,dr - A^{-2}\ell_k^2\,d\Sigma^2. \tag{3.26}$$

The first two terms of this equation present no problem; however, we should study in detail the asymptotic behavior of the third term. Using Eq. (1.14) asymptotically one obtains

$$\frac{\partial r}{\partial \chi} \equiv \left.\frac{\partial r}{\partial \chi}\right|_u = A^2 \approx 2r; \tag{3.27}$$

from which it is deduced that, for very large r,

$$r \approx e^{2\chi}. \tag{3.28}$$

Then since, for very large χ, one has that $\ell_k^2 \approx e^{2\chi}/4$, it is deduced that asymptotically

$$A^{-2}\ell_k^2 \approx r/4A^2 \approx \tfrac{1}{8}. \tag{3.29}$$

So the conformal metric at future null infinity is given by

$$d\tilde{s}^2|_{\Omega = 0} = -du\,d\Omega - \tfrac{1}{8}\,d\Sigma^2, \tag{3.30}$$

which is a nondegenerate regular metric. This means that the present choice of Ω does the job of bringing infinity to a finite distance and provides us with a well behaved conformal metric at future null infinity in this case.

It is then reasonable to ask under what conditions on $A(t)$ will the choice of Ω as the inverse of the affine distance along null geodesics have this property. We will study this in a later section.

The Riemann tensor can now be expressed in terms of Ω, obtaining

$$\Phi_{22} = (3Q/8)\Omega^{1/2}, \tag{3.31}$$

$$\Phi_{11} = (3Q/8)\Omega^{3/2}, \tag{3.32a}$$

$$\Lambda = (Q/8)\Omega^{3/2}, \tag{3.32b}$$

$$\Phi_{00} = (3Q/2)\Omega^{5/2}. \tag{3.33}$$

Equation (3.2) implies asymptotic behavior of the form

$$\Phi_{22} = \mathcal{K}(\Omega)\hat{\Phi}_{22} + \delta\Phi_{22}, \tag{3.34}$$

$$\Phi_{11} = \mathcal{K}(\Omega)\Omega^2\hat{\Phi}_{11} + \delta\Phi_{11}, \tag{3.35a}$$

$$\Lambda = \mathcal{K}(\Omega)\Omega^2\hat{\Lambda} + \delta\Lambda, \tag{3.35b}$$

$$\Phi_{00} = \mathcal{K}(\Omega)\Omega^4\hat{\Phi}_{00} + \delta\Phi_{00}. \tag{3.36}$$

It can be seen from the last equation that

$$\mathcal{K}(\Omega) = \Omega^{5/2 - 8/2} = \Omega^{-3/2}. \tag{3.37}$$

This means that the Riemann tensor diverges as $\Omega^{-3/2}$ as one approaches null infinity. And to avoid any confusion, it is convenient to remark here that the Friedmann $K=-1$ model is not asymptotically flat, and the scalar ω used above does not coincide with the conformal factor Ω used in the construction of the conformal metric.

Let us now consider the other expanding Friedmann model.

B. Case $K=0$

We now try to repeat the same construction as before by taking the conformal factor Ω proportional to the inverse of the affine distance; more precisely we take

$$\Omega = 1/A^{5/2}. \tag{3.38}$$

The conformal metric is given in this case by

$$d\tilde{s}^2 = \Omega^2\, ds^2 = A^{-5}(2\dot{r}_0 - A^2)d\mu^2 + 2A^{-5}\, d\mu\, dr - A^{-3}\mathcal{J}_k^2\, d\Sigma^2. \tag{3.39}$$

Studying the asymptotic behavior of $A^{-3}\mathcal{J}_k^2$ in a similar fashion, we obtain

$$\frac{\partial r}{\partial \chi} \equiv \left.\frac{\partial r}{\partial \chi}\right|_{\mu} = A^2 \approx \left(\frac{5Q^{1/2}r}{2}\right)^{4/5}; \tag{3.40}$$

which means that asymptotically we have

$$\chi \approx (2/5Q^{1/2})^{4/5} 5 r^{1/5}. \tag{3.41}$$

Then for large r one obtains

$$A^{-3}\mathcal{J}_k^2 = A^{-3}\chi^2 \approx (5Q^{1/2}r/2)^{-6/5}(2/(5Q^{1/2}))^{8/5}5^2 r^{2/5} \approx (5Q^{1/2}/2)^{-14/5}5^2 r^{-4/5}; \tag{3.42}$$

so the conformal metric obtained in this way turns out to be degenerate at future null infinity.

Note that traditionally the Friedmann models have been studied through their conformal representation in the Einstein universe, where this bad behavior of the conformal metric is absent. This reinforces the initiative of relating every RW model conformally to the Einstein universe. However, we should also note that in this last example the bad behavior of the metric at future null infinity is associated with the choice of the conformal factor as the inverse of the affine distance, since it produces complications with the asymptotic behavior of the angular part of the metric; and thus it is clear that these complications will disappear if we choose the conformal factor as the inverse of the luminosity distance, although probably at the expense of introducing other complications.

It is therefore clear that we need a systematic study of future null infinity in the RW models. We do this in the next section.

IV. THE GENERAL CASE

We are therefore confronted with the question of how to choose an appropriate conformal factor that will make the conformal metric regular (meaning at least continuous) at future null infinity. Of course, this is associated with the behavior of the function $A(t)$, which is the only nontrivial input in the metric.

Let us be more precise in our discussion. We will define the manifold $\mathscr{I}^+$ to be future null infinity of a noncontracting C^∞ Robertson–Walker space-time (M, g_{ab}) if there exists a manifold $\tilde{M}$ with boundary $\mathscr{I}^+$, metric $\tilde{g}_{ab}$, and a function Ω on $\tilde{M}$ such that a neighborhood of $\mathscr{I}^+$ in $\tilde{M}$ is diffeomorphic to a neighborhood of $\mathscr{I}^+$ in the manifold $M \cup \mathscr{I}^+$ and (a) on M, Ω is C^∞, $\Omega > 0$, and $\tilde{g}_{ab} = \Omega^2 g_{ab}$; and (b) at $\mathscr{I}^+$, $\Omega = 0$, Ω is C^0; at every point of $\mathscr{I}^+$ there end future directed null geodesics of $\tilde{M}$, and $\tilde{g}_{ab}$ is nondegenerate.

Note that we are implicitly requiring the conformal metric to be C^1 at scri. Also it should be observed that nothing is said about the differentiability properties of Ω at scri, that is, we only require it to be continuous.

At this point it is important to recall that if ν is an affine parameter along the null geodesics contained in the null hypersurfaces $u = \text{const}$, but with respect to the conformal metric $\tilde{g}_{ab}$, then it can be related to r by the equation

$$\frac{\partial \nu}{\partial r} = -\Omega^2. \tag{4.1}$$

Since ν is a natural coordinate of $\tilde{M}$, it is then clear how the choice of Ω determines the differentiable structure of the conformal manifold.

Because of the present symmetries, the scalar Ω is taken to be a function depending only on u and r; so one can define ν to be

$$\nu = -\int_{r_1}^{r} \Omega(u,r')^2\, dr' + \nu_0(u), \tag{4.2}$$

with $\nu = 0$ at scri. In this way we will have

$$d\nu = -\Omega^2\, dr + \left(\dot{\nu}_0 - \int_{r_1}^{r} 2\Omega\dot{\Omega}\, dr'\right) du, \tag{4.3}$$

where as before a dot means partial derivative with respect to the coordinate u. For example,

$$\dot{\Omega} = \left.\frac{\partial \Omega}{\partial u}\right|_{r=\text{const}}. \tag{4.4}$$

Let us note that the existence of the coordinate ν in the conformal manifold $\tilde{M}$ requires that

$$\lim_{r\to\infty} \int_{r_1}^{r} \Omega(u,r')^2\, dr' \text{ must exist.} \tag{4.5}$$

The conformal metric can then be expressed by

$$\begin{aligned} d\tilde{s}^2 &= \Omega^2\, ds^2 \\ &= \Omega^2(2\dot{r}_0 - A^2)du^2 + 2\Omega^2\, du\, dr - \Omega^2 A^2 \mathcal{J}_k^2\, d\Sigma^2 \\ &= \left(\Omega^2(2\dot{r}_0 - A^2) - 2\int_{r_1}^{r} 2\Omega\dot{\Omega}\, dr' + 2\dot{\nu}_0\right)du^2 \\ &\quad - 2\, du\, d\nu - \Omega^2 A^2 \mathcal{J}_k^2\, d\Sigma^2. \end{aligned} \tag{4.6}$$

It is crucial to notice that this expression for the conformal metric is an invariant one, that is, it has been geometrically defined; and that du, $d\nu$, and $d\Sigma^2$ have a clear invariant meaning. Therefore the three terms appearing in the above expression are geometrically well defined and it makes sense to refer to the asymptotic behavior of each of them. From the last term, in particular, we observe that the algebraic condition that will make this term regular at $\mathscr{I}^+$ is $\Omega^2 A^2 \mathcal{J}_K^2 \propto \text{const}$.

One may ask, "Is it possible then that by taking Ω to be the inverse of the luminosity distance r_L the conformal metric is regular at $\mathscr{I}^+$?"

The luminosity distance is a scalar that satisfies

$$\frac{\partial r_L}{\partial r} = -\rho r_L. \tag{4.7}$$

The natural choice for r_L is

$$r_L = A\mathcal{f}_K. \tag{4.8}$$

So by taking the conformal factor Ω to be the inverse of the luminosity distance we obtain

$$\Omega = 1/r_L = 1/A\mathcal{f}_K \tag{4.9}$$

and

$$d\tilde{s}^2 = \left(\frac{1}{\mathcal{f}_K^2}\left(\frac{2\dot{r}_0}{A^2} - 1\right) - 2\int_{r_1}^{r} 2\Omega\dot{\Omega}\, dr' + 2\dot{\omega}_0\right) du^2 - 2\, du\, d\omega - d\Sigma^2. \tag{4.10}$$

It is then clear that for this metric to be regular at scri we should have the expression

$$\mathscr{W} \equiv \int_{r_1}^{r} \Omega\dot{\Omega}\, dr' \tag{4.11}$$

regular at scri.

The last expression can be further transformed by noting that

$$\begin{aligned}\dot{\Omega} &= \frac{1}{A\mathcal{f}_K^2}\sqrt{1 - K\mathcal{f}_K^2} - \frac{\dot{r}_0}{A\mathcal{f}_K}\left(\frac{\sqrt{1 - K\mathcal{f}_K^2} + A'\mathcal{f}_K}{A^2\mathcal{f}_K}\right) \\ &= \frac{1}{A\mathcal{f}_K^2}\sqrt{1 - K\mathcal{f}_K^2} + \frac{\dot{r}_0}{A\mathcal{f}_K}\rho \\ &= \Omega\frac{\sqrt{1 - K\mathcal{f}_K^2}}{\mathcal{f}_K} + \Omega\dot{r}_0\rho \\ &= \Omega\left(\frac{\sqrt{1 - K\mathcal{f}_K^2}}{\mathcal{f}_K} + \dot{r}_0\rho\right),\end{aligned} \tag{4.12}$$

which implies

$$\mathscr{W} \equiv \int_{r_1}^{r} \Omega^2\left(\frac{\sqrt{1 - K\mathcal{f}_K^2}}{\mathcal{f}_K} + \dot{r}_0\rho\right) dr'. \tag{4.13}$$

Then using the relation

$$\frac{\partial\Omega}{\partial r} = -\frac{1}{r_L^2}\frac{\partial r_L}{\partial r} = \frac{1}{r_L}\rho = \Omega\rho, \tag{4.14}$$

we can express $\mathscr{W}$ by

$$\begin{aligned}\mathscr{W} &= \int_{r_1}^{r}\left(\Omega^2\frac{\sqrt{1 - K\mathcal{f}_K^2}}{\mathcal{f}_K} + \dot{r}_0\Omega\frac{\partial\Omega}{\partial r}\right) dr' \\ &= \int_{r_1}^{r}\left(\Omega^2\frac{\sqrt{1 - K\mathcal{f}_K^2}}{\mathcal{f}_K}\right) dr' + \dot{r}_0\frac{\Omega^2}{2}\Bigg|_{r_1}^{r}.\end{aligned} \tag{4.15}$$

This means that we only need to consider the first term in the asymptotic region in order to determine the regularity of the conformal metric at $\mathscr{I}^+$.

At this stage it is clear that we need a detailed study of the asymptotic behavior of the functions A and χ as r goes to infinity. We have explicitly mentioned that we consider $A' \geqslant 0$. Next let us observe that the optical scalar ρ is given by

$$\rho = -(\sqrt{1 - K\mathcal{f}_K^2} + A'\mathcal{f}_K)/A^2\mathcal{f}_K, \tag{4.16}$$

and therefore satisfies $\rho < 0$; which implies that

$$\frac{\partial r_L}{\partial r} > 0. \tag{4.17}$$

But furthermore, by taking the conformal factor $\Omega = 1/r_L$, we are implicitly assuming that in the limit $r \to \infty$, we should have $\Omega \to 0$, or equivalently $r_L \to \infty$. Can we have $\lim_{r\to\infty} r_L < \infty$? We will see later that, under the present conditions, $r_L \to \infty$ as one approaches $\mathscr{I}^+$.

In order to obtain information on the asymptotic behavior of the coordinate χ along the null geodesics contained in the null hypersurfaces $u = \text{const}$, it is convenient to study the equation

$$du = \frac{dt}{A(t)} - d\chi, \tag{1.10}$$

since, when $u = \text{const}$, we have

$$\chi = \int_{t_0}^{t}\frac{dt'}{A(t')}. \tag{4.18}$$

Defining χ_∞ by

$$\chi_\infty = \int_{t_0}^{\infty}\frac{dt'}{A(t')}, \tag{4.19}$$

we want to know whether χ_∞ is bounded (finite) or unbounded (infinite). We can then classify the RW spacetimes in two classes such that class B corresponds to χ_∞ bounded and class U corresponds to χ_∞ unbounded. Case B coincides with the appearance for each observer of an event horizon; since, for example, an observer traveling along the geodesic $\chi = \text{const} > \chi_\infty$, $\theta = \text{const}$, and $\varphi = \text{const}$ will never be able to get information from the events with $t > t_0$ and $\chi < \chi_\infty$. Case U refers to those models where there is no event horizon.

Case B: Since in this case $\mathcal{f}_K(\chi_\infty)$ is finite we deduce that a sufficient condition for $\mathscr{W}$ to be regular at $\mathscr{I}^+$ is that

$$\lim_{r\to\infty}\int_{r_1}^{r}\Omega^2\, dr' \text{ exists;} \tag{4.20}$$

which agrees with Eq. (4.5), and it is also equivalent in this case to the condition that

$$\int_{r_1}^{r}\frac{1}{A^2}\, dr' \text{ exists for } r \to \infty. \tag{4.21}$$

Case U: Since χ_∞ is unbounded, we disregard the possibility $K = 1$, and we consider the cases $K = -1$ and $K = 0$ separately.

Case U, $K = -1$: We now also require that the expression

$$\int_{r_1}^{r}\Omega^2\, dr' = \int_{r_1}^{r}\frac{1}{A^2\mathcal{f}_K^2}\, dr' \text{ exists for } r \to \infty, \tag{4.22}$$

which also coincides with Eq. (4.5).

Case U, $K = 0$: In this situation it is required that

$$\int_{r_1}^{r}\frac{\Omega^2}{\mathcal{f}_K}\, dr' = \int_{r_1}^{r}\frac{1}{A^2\mathcal{f}_K^3}\, dr' \text{ exists for } r \to \infty, \tag{4.23}$$

which is weaker than the condition for the existence of the function ω.

It is clear in all cases considered that $r_L \to \infty$ as $r \to \infty$, since at least A or f_K is unbounded in this limit, and both satisfy $A > 0$ and $f_K > 0$.

We have that the condition of regularity of $\mathscr{W}$ in the cases considered above is satisfied if the function ω is well defined. Let us now observe that in all cases ω is well defined.

In case B we have that the integral (4.19) exists; while for ω to be well defined we should have that

$$\int_{r_1}^{r} \frac{1}{A^2 f_K^2}\, dr' \text{ exists for } r \to \infty; \tag{4.24}$$

which in this case is equivalent to the condition that

$$\int_{r_1}^{r} \frac{1}{A^2}\, dr' \text{ exists for } r \to \infty. \tag{4.25}$$

This integral is easily shown to be equivalent to (4.19) since

$$\int_{r_1}^{\infty} \frac{1}{A^2}\, dr' = \int_{t_1}^{\infty} \frac{1}{A^2} \frac{\partial r'}{\partial t}\bigg|_u dt = \int_{t_1}^{\infty} \frac{1}{A}\, dt. \tag{4.26}$$

Therefore ω is well defined in case B.

In case U one has

$$\int_{t_1}^{\infty} \frac{1}{A}\, dt = \infty, \tag{4.27}$$

and one would like to have

$$\int_{r_1}^{\infty} \frac{1}{A^2 f_K^2}\, dr' < \infty. \tag{4.28}$$

But let us note that since K can only have the values -1 and 0, one has that

$$f(\chi) \geqslant \chi; \tag{4.29}$$

from which one concludes that

$$\begin{aligned} &\lim_{r\to\infty} \int_{r_1}^{r} \frac{1}{A^2 f_K^2}\, dr' \\ &\quad = \lim_{t\to\infty} \int_{t_1}^{t} \frac{1}{A^2 f_K^2} \frac{\partial r}{\partial t'}\bigg|_u dt' \\ &\quad = \lim_{t\to\infty} \int_{t_1}^{t} \frac{1}{A f_K^2}\, dt' = \lim_{\chi\to\infty} \int_{\chi_1}^{\chi} \frac{1}{A f_K^2} \frac{\partial t'}{\partial \chi'}\bigg|_u d\chi' \\ &\quad = \lim_{\chi\to\infty} \int_{\chi_1}^{\chi} \frac{1}{f_K^2}\, d\chi' \leqslant \lim_{\chi\to\infty} \int_{\chi_1}^{\chi} \frac{1}{\chi'^2}\, d\chi' < \infty. \end{aligned} \tag{4.30}$$

Therefore we see that also in case U the function ω is well defined.

Then since the regularity conditions of $\mathscr{W}$ are satisfied if ω exists, we conclude that the choice

$$\Omega = 1/A f_K \tag{4.31}$$

provides us with a construction of $\mathscr{I}^+$ for all noncontracting RW models.

It now remains to be seen when $d\Omega$ is regular at $\mathscr{I}^+$, and what type of hypersurface $\mathscr{I}^+$ is, namely, whether it is timelike, null, or spacelike.

In terms of the coordinates $(u,\omega,\zeta,\bar{\zeta})$ of the conformal manifold $\tilde{M}$ the differential of Ω is given by

$$\begin{aligned} d\Omega = &\left[\frac{\sqrt{1 - K f_K^2}}{A f_K^2} + \Omega_\omega \left(2\mathscr{W} - \dot{\nu}_0 - \frac{\dot{r}_0}{A^2 f_K^2} \right) \right] \\ &\times du + \Omega_\omega\, d\omega, \end{aligned} \tag{4.32}$$

where

$$\Omega_\omega \equiv (A' f_k + \sqrt{1 - K f_K^2})/A. \tag{4.33}$$

Then $d\Omega$ will be regular at $\mathscr{I}^+$ if and only if Ω_ω is regular at $\mathscr{I}^+$; furthermore, at $\mathscr{I}^+$ we have

$$d\Omega|_{\mathscr{I}^+} = \Omega_\omega|_{\mathscr{I}^+}\, d\omega. \tag{4.34}$$

If $d\Omega$ is well behaved at $\mathscr{I}^+$ the character of the hypersurface $\mathscr{I}^+$ can be studied from the expression

$$\tilde{g}(d\Omega, d\Omega) = (A'/A)^2 - (\sqrt{1 - K f_K^2}/A f_K)^2, \tag{4.35}$$

where one can see that since it is the difference of two positive terms, in principle it could be positive, negative, or zero.

In order to have a concrete picture of the different possible behavior, let us consider the following examples:

(a) constant,

$$\begin{aligned} &A = 1, \\ &\frac{A'}{A} = 0, \end{aligned} \tag{4.36}$$

(b) very slow,

$$\begin{aligned} &A = \frac{t}{t+1}, \\ &\frac{A'}{A} = \frac{1}{t(t+1)}, \end{aligned} \tag{4.37}$$

(c) slow,

$$\begin{aligned} &A = \ln t, \\ &\frac{A'}{A} = \frac{1}{t \ln (t)}, \end{aligned} \tag{4.38}$$

(d) power law,

$$\begin{aligned} &A = t^n, \quad n > 0, \\ &\frac{A'}{A} = \frac{n}{t}, \end{aligned} \tag{4.39}$$

(e) fast,

$$\begin{aligned} &A = \cosh t, \\ &\frac{A'}{t} = \tanh t. \end{aligned} \tag{4.40}$$

The first three cases belong to class U. The examples (d) with $n \leqslant 1$ are also of class U, while, when $n > 1$, they belong to class B. Finally the last example is of class B. So we see that $A = t$ is a boundary case among our examples which divide them into classes U and B.

Table I shows the value of Ω_ω and $\tilde{g}(d\Omega, d\Omega)$ at $\mathscr{I}^+$.

It is observed that although we are forced to take the conformal factor as the inverse of the luminosity distance, in order to obtain a regular metric at scri, this conformal factor is not suitable for the study of the nature of scri, since we have seen that $d\Omega$ is sometimes zero or not defined at scri.

However, we can choose the coordinate ω such that

TABLE I. Values of $\Omega_{,}$ and $\tilde{g}(d\Omega,d\Omega)$ at scri.

Example	Class	K	$\Omega_{,}\vert_{\mathscr{I}^+}$	$\tilde{g}(d\Omega,d\Omega)\vert_{\mathscr{I}^+}$	Remarks
(a)	U	-1	∞	$\cdots$	
		0	1	0	Minkowski space
(b)	U	-1	∞	$\cdots$	
		0	1	0	
(c)	U	-1	∞	$\cdots$	
		0	1	0	
(d) $n<1$	U	-1	∞	$\cdots$	
		0	0	$\cdots$	$A=t^{2/3}$ is a Friedmann model
(d) $n=1$	U	-1	2	0	Minkowski space
		0	1	0	
		-1	0	$\cdots$	
(d) $n>1$	B	0	0	$\cdots$	
		1	0	$\cdots$	
		-1	$\sinh\chi_\infty$	1	
(e)	B	0	χ_∞	1	
		1	$\sin\chi_\infty$	1	de Sitter space

$\omega=0$ at $\mathscr{I}^+$ by appropriately defining $\omega_0(\alpha)$; and by doing so we can use $d\omega$ for the study of the nature of $\mathscr{I}^+$, since obviously ω is a regular function at scri. In fact, one finds that

$$g(d\omega,d\omega)=1/f_K^2; \tag{4.41}$$

so scri is nontimelike, and, furthermore, in case U it is null and in case B spacelike.

We have just proved the following thereom.

Theorem: The future null infinity of noncontracting Robertson–Walker models is null or spacelike according to the absence or presence of cosmic event horizons, respectively.

This completes the statements appearing in the literature[3] which claimed, based on intuitive arguments, that when $\mathscr{I}^+$ is null one expects no event horizons, and when $\mathscr{I}^+$ is spacelike each observer will be assigned an event horizon. It is important to emphasize that although the same intuitive arguments were used for the nature of past null infinity, which normally coincides with the initial cosmic singularity, the analog theorem is not true. For example, the $K=-1$ de Sitter model would violate it, since past null infinity of de Sitter space (which is spacelike) does not coincide with the initial null cone cosmic singularity of the $K=-1$ model (which does not possess particle horizons).

It was asked previously under what circumstances would the choice of the conformal factor as the inverse of the affine distance provide us with a construction of $\mathscr{I}^+$. To see this let us note that

$$\frac{\partial r_L}{\partial r}=\Omega_{,\omega}. \tag{4.42}$$

Thus, since we know that the choice of the inverse of the luminosity distance as the conformal factor provides us with a construction of $\mathscr{I}^+$, we are sure that the affine distance will do the job if

$$0<\Omega_{,\omega}\vert_{\mathscr{I}^+}<\infty. \tag{4.43}$$

It is clear then from Table I why in the case $K=0$ of the Friedman model it is inappropriate to take $\Omega=1/r$.

V. FINAL COMMENTS

We will proceed here with a quick recapitulation of the topics we have covered, and take the opportunity to insert some comments.

In order to put our work into perspective we have reviewed in Secs. II and III what could be considered the most significant Robertson–Walker models, namely, Minkowski space, de Sitter space, and the open dust Friedmann models. We have approached the first two spaces through their standard conformal representations in the static Einstein universe.

In Sec. III we have observed that a naive comparison of the asymptotic behavior of a regular asymptotically flat space-time with the Riemann tensor of a RW model leads us to the dust Friedmann models. Actually, as we have seen, these models are not asymptotically flat. This might seem a little curious, since, for example, in the $K=-1$ dust Friedmann model the scalar $A(t)$ asymptotically approaches the functional form $A\approx t$ for large t. And, since Minkowski space can be represented as a RW model with $K=-1$ and $A=t$, one would be tempted to conclude that the $K=-1$ dust Friedmann model is asymptotically flat at future null infinity. Instead we have seen that this is not the case; in fact, the curvature tensor of this model diverges at future null infinity. One might, however, introduce the notion of timelike infinity and argue that the $K=-1$ dust Friedmann model is asymptotically flat in that region; we, instead, for the moment concentrate our attention on future null infinity.

We have also studied the use of the conformal factor as the inverse of the affine distance and found the following: while in the case $K=-1$ for the dust Friedmann model it provides us with a well behaved metric at future null infinity, in the case $K=0$ it produces a degenerate metric at scri.

From the contents of Sec. IV we have proved that the choice of the conformal factor as the inverse of the luminosity distance permits the construction of $\mathscr{I}^+$ for all noncontracting RW models. We have also seen that although this choice is necessary if one wants to obtain a regular metric at scri, one cannot always use the gradient of this conformal factor for the study of the nature of $\mathscr{I}^+$, since in some cases $d\Omega$ is zero or not defined at scri. In any case it was proved that in the noncontracting RW models $\mathscr{I}^+$ is not timelike, and more precisely the theorem of the last section relates its nature to the appearance or absence of cosmic event horizons.

ACKNOWLEDGMENTS

The author is grateful to G. DeToma for having checked equations appearing in Secs. I and II, and to P. Cuddeford for suggesting several improvements of the manuscript.

This work was carried out during the author's stay at the International Centre for Theoretical Physics, to which he is particularly indebted for its hospitality and financial support.

[3] S. W. Hawkins and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., London, 1973).

[2]R. Penrose, in *Battelle Rencontres*, edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).
[3]R. Penrose, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).
[4]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[5]O. M. Moreschi, Class. Quantum Grav. **4**, 1063 (1987).
[6]R. J. Torrence and W. E. Couch, Gen. Relativ. Gravit. **18**, 585 (1986).
[7]W. Rindler, *Essential Relativity: Special, General, and Cosmological* (Springer, Berlin, 1969).

Lagrange multipliers and Gauss–Bonnet-type invariants in Riemann–Cartan space

R. de Azeredo Campos
Departamento de Física, Universidade de Brasília, 70.910 Brasília DF, Brazil

(Received 17 May 1988; accepted for publication 13 December 1989)

A Lagrange multiplier approach is applied to Gauss–Bonnet-type invariants in Riemann–Cartan geometry. Lanczos-type identities are then derived. The use of the variational procedure in the context of gravitational theories with a Gauss–Bonnet invariant limit is discussed.

I. INTRODUCTION

Recently it has been pointed out that the universal Gauss–Bonnet (GB) form seems to be of crucial importance as a correction to supergravity Lagrangians in higher space-time dimensions.[1] By contrast, in four dimensions the role of the GB combination is at first rather restricted. It is a topological invariant and its addition to an action that describes a gravitational theory does not modify the resulting field equations.[2] In spite of this, one can suggest quadratic curvature combinations that contain as a limiting case the four-dimensional GB invariant. Such quadratic curvature terms produce field equations that, only in the GB invariant limit, reduce to an identity. For instance, one can propose a gravitational model in four-dimensional Riemann–Cartan space U_4 in which a particular choice of parameters and a condition on the torsion tensor yield Einstein-type equations plus the Lanczos identity.

In the variational derivation of field equations for theories of gravitation the Lagrange multiplier technique considerably abbreviates the calculations. Modified variational principles, which include Lagrange multipliers, have been used in the literature[3–6] in the contexts of U_4 and of Riemann space V_4.

In this paper a particular form of variational procedure is applied to some GB-type invariance and also to curvature combinations related to GB-type invariants in U_4. The variations of the topological invariants lead to Lanczos-type identities, which are verified with the aid of well-known identities for the curvature tensor.

In Sec. II, we present our notation and the form of variational method adopted here. In Sec. III, Lanczos-type identities are derived. The relation between a GB-type invariant proposed in the literature[7] and the second identity for the curvature tensor[8] is clarified. Section IV deals with an application of the variational approach to the construction of alternative gravitational models in U_4 with a GB invariant limit.

II. PRELIMINARIES

For the curvature and Ricci tensors we follow the definitions (as in Ref. 8)

$$R_{\alpha\beta\gamma}{}^{\delta} \equiv 2\partial_{[\alpha}\Gamma_{\beta]\gamma}{}^{\delta} + 2\Gamma_{[\alpha}{}^{\delta}{}_{|\rho}\Gamma_{|\beta]}{}^{\rho}{}_{\gamma}, \tag{1}$$

$$R_{\alpha\beta} \equiv R_{\lambda\alpha\beta}{}^{\lambda}. \tag{2}$$

The metricity condition $\nabla_\sigma g_{\alpha\beta} = 0$ leads to the connection of U_4,

$$\Gamma_{\alpha\beta}{}^{\gamma} \equiv \{{}^{\gamma}_{\alpha\beta}\} + S_{\alpha\beta}{}^{\gamma} - S_{\beta}{}^{\gamma}{}_{\alpha} + S^{\gamma}{}_{\cdot\alpha\beta}, \tag{3}$$

where $S_{\alpha\beta}{}^{\gamma} \equiv \Gamma^{\gamma}_{[\alpha\beta]}$ is the torsion tensor, with the square brackets denoting antisymmetrization.

The modified variational principle in Riemann space V_4 consists of the introduction of a Lagrange multiplier $\Lambda_\alpha{}^{\cdot\beta\gamma}$, which imposes a constraint to be included in the action. The variations are then carried out with respect to the metric $g_{\alpha\beta}$, the symmetric affine connection $\{{}^{\gamma}_{\alpha\beta}\}$ and $\Lambda_\alpha{}^{\cdot\gamma\beta}$ independently.[3,4] This procedure is extendible to the U_4 space.[5] A particular form of this extension will be followed here.

Let us first consider the general action

$$I = \int \{\mathbb{L} + \mathbb{L}_M + \sqrt{-g}\Lambda_\alpha{}^{\cdot\beta\gamma}(\Gamma^{\alpha}_{(\beta\gamma)} - \{{}^{\alpha}_{\beta\gamma}\} - 2S^{\alpha}{}_{\cdot(\beta\gamma)})\} d^4x, \tag{4}$$

where $\mathbb{L} = \mathbb{L}(g_{\mu\nu}, \Gamma^{\lambda}_{(\mu\nu)}, \partial_\sigma\Gamma^{\lambda}_{(\mu\nu)}, S_{\mu\nu}{}^{\lambda}, \partial_\sigma S_{\mu\nu}{}^{\lambda})$ is a pure gravitational Lagrangian, $\Gamma^{\lambda}_{(\mu\nu)}$ is the symmetric part of the connection and $\mathbb{L}_M = \mathbb{L}_M(g_{\mu\nu}, \Gamma^{\lambda}_{(\mu\nu)}, S^{\lambda}_{\mu\nu}, \phi_A, \partial_\mu\phi_A)$ is a matter Lagrangian, with the matter field ϕ_A. By varying independently $g_{\mu\nu}$, $\Gamma^{\lambda}_{(\mu\nu)}$, $S_{\mu\nu}{}^{\lambda}$, and $\Lambda^{\cdot\mu\nu}_{\lambda}$, we get

$$\frac{\delta\mathbb{L}}{\delta g_{\mu\nu}} + \sqrt{-g}M^{\mu\nu} = -\frac{\delta\mathbb{L}_M}{\delta g_{\mu\nu}}, \tag{5}$$

$$\frac{\delta\mathbb{L}}{\delta\Gamma^{\lambda}_{\mu\nu}} + \sqrt{-g}\Lambda^{\cdot\mu\nu}_{\lambda} = -\frac{\delta\mathbb{L}_M}{\delta\Gamma^{\lambda}_{(\mu\nu)}}, \tag{6}$$

$$\frac{\delta\mathbb{L}}{\delta S_{\mu\nu}{}^{\lambda}} - 2\sqrt{-g}\Lambda^{[\mu\nu]}{}_{\cdot\lambda} = -\frac{\delta\mathbb{L}_M}{\delta S_{\mu\nu}{}^{\lambda}}, \tag{7}$$

$$\Gamma^{\lambda}_{(\mu\nu)} = \{{}^{\lambda}_{\mu\nu}\} + 2S^{\lambda}_{(\mu\nu)}. \tag{8}$$

Equation (8) is equivalent to (3), or to the assumption of the U_4 postulate. The symmetric tensor $M^{\mu\nu}$ is obtained by the variation of the term that contains $\Lambda^{\cdot\beta\gamma}_{\alpha}$ in (4), and is given by

$$M^{\mu\nu} \equiv \tfrac{1}{2}\overset{*}{\nabla}_\sigma \Lambda^{\{\nu\sigma\mu\}}, \tag{9}$$

where $\overset{*}{\nabla}_\sigma \equiv \nabla_\sigma + 2S_{\sigma\lambda}{}^{\lambda}$, ∇_σ is the covariant derivative associated with $\Gamma_{\mu\nu}{}^{\lambda}$, and the abbreviation $\Lambda^{\{\nu\sigma\mu\}} \equiv \Lambda^{\nu\sigma\mu} - \Lambda^{\sigma\mu\nu} + \Lambda^{\mu\nu\sigma}$ is used. From Eq. (6) one obtains the Lagrange multiplier for a given $\mathbb{L}$. By substituting this expression into (5) and (7), two groups of field equations result, involving the basic fields $g_{\alpha\beta}$ and $S_{\alpha\beta}{}^{\cdot\gamma}$. Thus the variational derivation via Lagrange multipliers reconstitutes the field equations

 0022-2488/90/051217-04$03.00

that can be obtained alternatively from the conventional Hilbert principle.

III. DERIVATION OF LANCZOS-TYPE IDENTITIES

We first recall that

$$\int \sqrt{-g}\,\varepsilon_{\alpha\beta\gamma\delta}\varepsilon^{\mu\nu\lambda\rho}R^{\alpha\beta}_{\cdot\cdot\mu\nu}R^{\gamma\delta}_{\cdot\cdot\lambda\rho}\,d^4x$$

is an invariant in V_4, the usual Gauss–Bonnet form, where the conventions $\varepsilon^{0123}=+1/\sqrt{-g}$, $\varepsilon_{0123}=\sqrt{-g}$ are adopted for the permutation tensor $\varepsilon^{\mu\nu\lambda\rho}$. The identity related to this invariant[9] can be extended to U_4 and rewritten as[10,11]

$$\delta\int\sqrt{-g}\,(R_{\mu\nu\lambda\rho}R^{\lambda\rho\mu\nu}-4R_{\mu\nu}R^{\nu\mu}+R^2)\,d^4x\equiv 0. \tag{10}$$

In U_4, the properties $R_{\mu\nu\lambda\rho}=-R_{\mu\nu\lambda\rho}=-R_{\mu\nu\rho\lambda}$ hold, but the symmetry relations $R_{\mu\nu\lambda\rho}=R_{\lambda\rho\mu\nu}$ and $R_{\mu\nu}=R_{\nu\mu}$ are valid only in V_4 where $S_{\alpha\beta\gamma}=0$ globally. Therefore, in V_4, Eq. (10) reads

$$\delta\int\sqrt{-g}\,(R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho}-4R_{\mu\nu}R^{\mu\nu}+R^2)d^4x\equiv 0, \tag{10'}$$

an identity frequently misquoted as being the GB identity in U_4. The true form is (10).

Let us take $\mathbb{L}_1\equiv\sqrt{-g}\,(R_{\alpha\beta\gamma\delta}R^{\gamma\delta\alpha\beta}-4R_{\alpha\beta}R^{\beta\alpha}+R^2)$ as the Lagrangian $\mathbb{L}$ in (4), with $\mathbb{L}_M=0$. The variational expressions (5)–(8) must give then the corresponding Lanczos identity. From (6) we get

$$\begin{aligned}\Lambda_\sigma{}^{\mu\nu}=2\Big[&\overset{*}{\nabla}{}_\sigma(g^{\mu\nu}R)-\overset{*}{\nabla}{}^{(\mu}(\delta^{\nu)}_\sigma R)-4\overset{*}{\nabla}_\sigma R^{(\mu\nu)}\\&+4\overset{*}{\nabla}_\rho(\delta^{(\mu}_\sigma R^{\nu)\rho})-2\overset{*}{\nabla}_\rho R^{(\mu\cdot\ \nu)\rho}{}_\sigma\Big]\\&-4\big[S_{\lambda\rho}{}^{(\mu}R^{\nu)\cdot\ \lambda\rho}{}_\sigma-4S_{\sigma\rho}{}^{(\mu}R^{\nu)\rho}+RS^{(\mu\nu)}_{\cdot\ \sigma}\big].\end{aligned} \tag{11}$$

From Eq. (5) it follows that

$$M^{\mu\nu}+N^{\mu\nu}=0. \tag{12}$$

with $M^{\mu\nu}$ defined at (9) and

$$\begin{aligned}N^{\mu\nu}\equiv&\tfrac{1}{2}g^{\mu\nu}(R^2-4R_{\lambda\rho}R^{\rho\lambda}+R_{\lambda\tau\rho\sigma}R^{\rho\sigma\lambda\tau}\\&-2(RR^{(\mu\nu)}-4R_\rho{}^{(\mu}R^{\nu)\rho}-R^{\lambda\sigma\cdot\ (\mu}_{\ \ \rho}R^{\nu\rho)}_{\cdot\cdot\lambda\sigma}.\end{aligned} \tag{13}$$

Using the contracted Bianchi identities

$$\nabla_\tau R_{\mu\nu\lambda}{}^{\cdot\cdot\tau}-2\nabla_{[\mu}R_{\nu]\lambda}=4S_{\tau[\mu\rho}R_{\nu]\rho\lambda\cdot\cdot}{}^{\tau}+2S_{\mu\nu}{}^{\rho}R_{\rho\lambda}, \tag{14}$$

$$\nabla^\mu R=2\nabla_\sigma R^{\mu\sigma}+4S_\sigma{}^{\mu\rho}R_{\rho\cdot\ \cdot}{}^{\sigma}+2S_{\sigma\lambda\rho}R^{\mu\rho\lambda\sigma}, \tag{15}$$

and also

$$\nabla_{[\alpha}\nabla_{\beta\ ]}R^{\mu\nu}=\tfrac{1}{2}R_{\alpha\beta\rho}{}^{\mu}R^{\rho\nu}+\tfrac{1}{2}R_{\alpha\beta\rho}{}^{\nu}R^{\mu\rho}-S_{\alpha\beta\cdot}{}^{\rho}\nabla_\rho R^{\mu\nu}, \tag{16}$$

we verify that (12) is precisely the Lanczos identity in U_4, as expected.

We consider now the invariant $\int\mathbb{L}_2\,d^4x$, with

$$\mathbb{L}_2=\sqrt{-g}\,\varepsilon^{\mu\nu\lambda\rho}R^{\alpha\beta}_{\cdot\cdot\mu\nu}R_{\alpha\beta\lambda\rho}. \tag{17}$$

The Lagrange multiplier in this case is

$$\begin{aligned}\Lambda^{\xi\kappa\tau}=4\big[&\nabla_\sigma R^{\sigma(\kappa}_{\cdot\cdot\lambda\rho}\varepsilon^{\tau)\xi\lambda\rho}+R^{\mu\sigma}_{\cdot\cdot\lambda\rho}S_{\sigma\rho\cdot}{}^{(\kappa}\varepsilon^{\tau)\xi\lambda\rho}\\&+2R^{\sigma(\kappa}_{\cdot\cdot\lambda\rho}\varepsilon^{\tau)\xi\lambda\rho}S_{\sigma\mu\cdot}{}^{\mu}\big].\end{aligned} \tag{18}$$

The substitution of (18) into (5) and (7), again with $\mathbb{L}_M=0$, leads to expressions that are identically satisfied. This can be checked using the Bianchi identity. Thus the Lanczos-type formula associated with the Gauss–Bonnet-type invariant (17) turns out to be related to the usual Bianchi identity in U_4.

Let us apply the technique to an identity presented in Ref. 7. One can form, with the Lagrangian

$$\mathbb{L}_3=\sqrt{-g}\,\varepsilon^{\mu\nu\lambda\rho}(R_{\mu\nu\lambda\rho}+4S_{\mu\sigma\nu}S^{\cdot\cdot\sigma}_{\lambda\rho}+4S_{\mu\sigma\cdot}{}^{\sigma}S_{\lambda\rho\nu}), \tag{19}$$

the invariant $\int\mathbb{L}_3\,d^4x$. From L_3 we derive the multiplier

$$\Lambda^{\alpha\beta\gamma}=-2S_{\sigma\lambda\cdot}{}^{(\beta}\varepsilon^{\gamma)\sigma\lambda\alpha}. \tag{20}$$

Equations (20) and (7) lead to an identity trivially verified. From (20) and (5) a relation similar to (12) results, in this case with

$$M^{\mu\nu}=2\overset{*}{\nabla}{}_\sigma S_{\alpha\beta\cdot}{}^{(\mu}\varepsilon^{\nu)\alpha\beta\sigma} \tag{21}$$

and

$$\begin{aligned}N^{\mu\nu}=&\varepsilon^{\alpha\beta\lambda(\mu}R_{\alpha\beta\lambda\cdot}{}^{\nu)}+4\varepsilon^{\alpha\lambda\rho(\mu}S_{\alpha\sigma\cdot}{}^{\nu)}S_{\lambda\rho\cdot}{}^{\sigma}\\&+4S_{\alpha\sigma\cdot}{}^{\sigma}\varepsilon^{\alpha\lambda\rho(\mu}S_{\lambda\rho\cdot}{}^{\nu)}.\end{aligned} \tag{22}$$

The second identity for the curvature tensor

$$R_{[\alpha\beta\lambda\ ]\cdot}{}^{\mu}=2\nabla_{[\alpha}S_{\beta\lambda\ ]\cdot}{}^{\mu}-4S_{[\alpha\beta\cdot}{}^{\rho}S_{\lambda\ ]\rho\cdot}{}^{\mu} \tag{23}$$

is then required to verify that $(M^{\mu\nu}+N^{\mu\nu})$ vanishes identically. Thus one can say that (23) is the Lanczos form of the GB-type identity discussed in Ref. 7. This is made clear by the use of the Lagrange multiplier procedure.

IV. APPLICATION TO GRAVITATIONAL MODELS

Several classes of $R+R^2$ gravitational theories have been studied recently.[12–17] The main motivation is the search for renormalizable divergences in the S matrix in the quantum version of these theories. The particular class of models we shall consider here contains as a limit the Einstein Lagrangian plus the Gauss–Bonnet invariant in the torsionless case. We also find that torsion is a dynamical variable for various choices of the parameters in the Lagrangian. The existence of the Gauss–Bonnet invariant limit assures the restitution of Einstein gravity when the torsion tensor vanishes.

Our Lagrangian is

$$L=\alpha\sqrt{-g}\,R+\mathbb{L}_Q, \tag{24}$$

$$\begin{aligned}\mathbb{L}_Q=\sqrt{-g}\,\big[&R^2+\beta R^{\mu\nu}R_{\mu\nu}+\gamma R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu}\\&-(\beta+4)R^{\mu\nu}R_{\mu\nu}+(1-\gamma)R^{\mu\nu\alpha\beta}R_{\alpha\beta\mu\nu}\\&+\xi R^{\alpha\beta\mu\nu}R_{\alpha[\beta\mu\nu]}+\eta S^{\alpha\beta\mu}S_{\beta\mu\alpha}\\&+\theta S_{\alpha\mu}{}^{\mu}S^{\alpha\nu}{}_{\nu}+\lambda S^{\alpha\beta\mu}S_{\alpha\beta\mu}\big].\end{aligned} \tag{25}$$

The quadratic combination $\mathbb{L}_Q$ generates the Lanczos

identity in V_4 when the torsion tensor vanishes. In this limit we have $\mathbb{L}_Q = \sqrt{-g}(R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\alpha\beta\mu\nu}R^{\alpha\beta\mu\nu}$. We note that

$$R_{\alpha[\beta\mu\nu]} \equiv 3^{-1}(R_{\alpha\beta\mu\nu} + R_{\alpha\mu\nu\beta} + R_{\alpha\nu\beta\mu}) = 0$$

in V_4. In (25), $\beta, \gamma, \xi, \eta, \theta$, and λ are free parameters. If we take all these parameters to be equal to zero the action $\int d^4x$ becomes the Gauss–Bonnet invariant in U_4. The choice $\alpha = \frac{1}{2}K$, where K is the Einstein gravitational constant, is needed to obtain the Einstein limit when $S_{\alpha\beta\gamma} = 0$. In this case, (24) becomes the most general Lagrangian in V_4 with at most quadratic terms in the curvature which reproduces Einstein equations in vacuum. (Parity-violating terms such as $\varepsilon^{\mu\nu\lambda\rho}R^{\alpha\beta}{}_{\mu\nu}R_{\alpha\beta\lambda\rho}$ or $\varepsilon^{\alpha\beta\mu\nu}S_{\mu\nu\beta}S_{\alpha\lambda}{}^{\lambda}$ were rejected. The cosmological term $\sqrt{-g}\Lambda$ also was not taken into account.)

The use of the variational procedure discussed in Sec. II leads to two groups of field equations, which are (in vacuum), respectively,

$$\alpha G^{(\mu\nu)} - M^{\mu\nu} - N^{\mu\nu} = 0, \tag{26}$$

$$\alpha(S^{\mu\nu}{}_{\sigma} - 2S_{\sigma}{}^{[\mu\nu]} + 4\delta_{\sigma}{}^{[\mu}S^{\nu]\lambda}{}_{\lambda}) + E_{\sigma}{}^{\mu\nu} + F_{\sigma}{}^{\mu\nu} + H_{\sigma}{}^{\mu\nu} + J_{\sigma}{}^{\mu\nu} = 0, \tag{27}$$

where

$$E_{\sigma}{}^{\mu\nu} \equiv \Big[\beta \overset{*}{\nabla}_{\lambda}(R^{\lambda[\nu}\delta_{\sigma}{}^{\mu]} - \delta_{\sigma}{}^{[\mu}R^{\nu]\lambda} - \delta_{\sigma}R^{[\mu\nu]}) + R^{[\lambda\nu]}(S_{\sigma\lambda}{}^{\mu} + S^{\mu}{}_{\lambda\sigma}) - R^{[\lambda\mu]}(S_{\sigma\lambda}{}^{\nu} + S^{\nu}{}_{\lambda\sigma})\Big] + 2R_{[\sigma\lambda]}S^{\lambda[\mu\nu]}, \tag{28}$$

$$F_{\sigma}{}^{\mu\nu} \equiv -2\gamma\Big[2\overset{*}{\nabla}_{\lambda}(R^{\lambda[\mu\nu]}{}_{\sigma} + R_{\sigma}{}^{[\mu\nu]\lambda}) + \overset{*}{\nabla}_{\lambda}(R_{\sigma}{}^{\lambda\mu\nu} + R^{\mu\nu\lambda}{}_{\sigma}) + 2S_{\lambda\rho}{}^{[\mu}(R^{\nu]}{}_{\sigma}{}^{\lambda\rho} - R^{[\lambda|\rho|\nu]}{}_{\sigma}) - S_{\lambda\rho\sigma}(R^{\mu\nu\lambda\rho} + R^{\lambda\rho\mu\nu})\Big], \tag{29}$$

$$H_{\sigma}{}^{\mu\nu} \equiv \tfrac{2}{3}\xi\Big[-\overset{*}{\nabla}_{\lambda}(R^{\mu\nu\lambda}{}_{\sigma} + 2R^{\lambda[\mu\nu]}{}_{\sigma} + S_{\lambda\rho}{}^{[\nu}(2R^{[\lambda|\rho|\mu]}{}_{\sigma} - 2R^{|\rho|\mu]\lambda}{}_{\sigma} + R^{\mu]\lambda}{}_{\sigma}{}^{\rho}) - S_{\lambda\rho\sigma}(R^{\lambda\rho\mu\nu} + R^{\lambda[\mu\nu]\rho})\Big], \tag{30}$$

$$J_{\sigma}{}^{\mu\nu} \equiv \eta S_{\sigma}{}^{[\mu\nu]} + \Theta\delta_{\sigma}{}^{[\nu}S^{\mu]\rho}{}_{\rho} + \lambda S^{\mu\nu}{}_{\sigma}. \tag{31}$$

The tensor $M^{\mu\nu}$ in (26) is given by (9), with

$$\Lambda_{\sigma}{}^{\mu\nu} = A_{\sigma}{}^{\mu\nu} + B_{\sigma}{}^{\mu\nu} + C_{\sigma}{}^{\mu\nu} + D_{\sigma}{}^{\mu\nu}, \tag{32}$$

$$A_{\sigma}{}^{\mu\nu} \equiv -2\alpha(g^{\mu\nu}S_{\lambda\sigma}{}^{\lambda} - \delta_{\sigma}^{(\mu}S_{\lambda}{}^{\nu)\lambda} + S_{\sigma}{}^{(\mu\nu)}), \tag{33}$$

$$B_{\sigma}{}^{\mu\nu} \equiv -4\big[(g^{\mu\nu}R - 4R^{(\mu\nu)})S_{\lambda\sigma}{}^{\lambda} + R\delta_{\sigma}^{(\nu}S^{\mu)\rho}{}_{\rho} + RS_{\sigma}{}^{(\mu\nu)} + 4S_{\lambda\sigma}{}^{(\nu}R^{\mu)\lambda}\big] - 2\big[\nabla^{(\nu}(\delta_{\sigma}{}^{\mu)}R) - \nabla_{\sigma}(g^{\mu\nu}R) + 2\beta(R^{[\lambda\nu]}S_{\sigma\lambda}{}^{\mu} + R^{[\lambda\mu]}S_{\sigma\lambda}{}^{\nu})\big], \tag{34}$$

$$C_{\sigma}{}^{\mu\nu} \equiv \beta\big[2(R^{\lambda(\mu}\delta_{\sigma}^{\nu)} - \delta_{\sigma}^{(\nu}R^{\mu)\lambda})S_{\lambda\rho}{}^{\rho} + \nabla_{\lambda}R^{\lambda(\mu}\delta_{\sigma}^{\nu)} - \nabla_{\lambda}(\delta_{\sigma}^{(\nu}R^{\mu)\lambda})\big] - 4(2\delta_{\sigma}^{(\mu}R^{\nu)\lambda}S_{\lambda\rho}{}^{\rho} + \nabla_{\lambda}(\delta_{\sigma}^{(\mu}R^{\nu)\lambda}) - \nabla_{\sigma}R^{(\mu\nu)}) + 2\gamma\big[2S_{\rho\lambda}{}^{\rho}(R^{\lambda(\mu\nu)}{}_{\sigma} - R_{\sigma}{}^{(\mu\nu)\lambda}) + S_{\lambda\rho}{}^{(\mu}(R^{|\lambda|\rho|\nu)}{}_{\sigma} - R^{\nu)\lambda\rho}{}_{\sigma}) + \nabla_{\lambda}(R^{(\nu|\lambda|\mu)}{}_{\sigma} - R^{\lambda(\mu\nu)}{}_{\sigma})\big] + 2(S_{\lambda\rho}{}^{(\mu}R^{\nu)\lambda\rho}{}_{\sigma} + 2S_{\lambda\rho}{}^{\lambda}R^{(\mu|\rho|\nu)}{}_{\sigma} - \nabla_{\lambda}R^{(\nu|\lambda|\mu)}{}_{\sigma}), \tag{35}$$

$$D_{\sigma}{}^{\mu\nu} \equiv -\tfrac{4}{3}\xi\big[S_{\rho\lambda}{}^{\rho}(R^{(\mu\nu)}{}_{\sigma} - R_{\sigma}{}^{(\mu\nu)\lambda}) + S_{\lambda\rho}{}^{(\mu}(R^{|\lambda|\rho|\nu)}{}_{\sigma} - R^{|\lambda|\nu)\rho}{}_{\sigma} + \tfrac{1}{2}\nabla_{\lambda}(R_{\sigma}{}^{(\mu\nu)\lambda} - R^{\lambda(\mu\nu)}{}_{\sigma})\big]. \tag{36}$$

The $G^{(\mu\nu)}$ in (26) is the Einstein tensor and, for $N^{\mu\nu}$, we have

$$N^{\mu\nu} = -2\{RR^{(\mu\nu)} + g^{\lambda(\nu}R^{\mu)\rho}[\beta R_{\lambda\rho} - (\beta+4)R_{\rho\lambda}] + (\gamma + \tfrac{\xi}{3})R^{(\nu}{}_{\lambda\tau\rho}R^{\mu)\lambda\tau\rho} + (1-\gamma)R^{\lambda\tau(\mu}{}_{\rho}R^{\nu)\rho}{}_{\lambda\tau} + (\xi/3)[R^{(\mu}{}_{\lambda\tau\rho}R^{\nu\tau\rho\lambda} + R_{\lambda\tau}{}^{(\mu}\rho R^{|\lambda|\nu)\rho\tau}]\} + \tfrac{1}{2}g^{\mu\nu}[R^2 + \beta R_{\lambda\rho}R^{\lambda\rho} - (\beta+4)R_{\lambda\rho}R^{\rho\lambda} + (\gamma + \xi/3)R_{\lambda\tau\rho\sigma}R^{\lambda\tau\rho\sigma} + (1-\gamma)R_{\lambda\rho\tau\sigma}R^{\rho\sigma\lambda\tau} + \tfrac{2}{3}\xi R_{\lambda\tau\rho\sigma}R^{\lambda\rho\sigma\tau}] - \Theta(S^{\mu\sigma}{}_{\sigma}S^{\nu\rho}{}_{\rho} - \tfrac{1}{2}g^{\mu\nu}S_{\alpha\sigma}{}^{\sigma}S^{\alpha\rho}{}_{\rho}) - \lambda(2S^{\mu\alpha\beta}S^{\nu}{}_{\alpha\beta} - S^{\alpha\beta\mu}S_{\alpha\beta}{}^{\nu} - \tfrac{1}{2}S_{\alpha\beta\gamma}S^{\alpha\beta\gamma}g^{\mu\nu}) - \eta(S^{\mu\alpha\beta}S_{\beta}{}^{\nu}{}_{\alpha} + \tfrac{1}{2}g^{\mu\nu}S_{\alpha\beta\gamma}S^{\gamma\beta\alpha}). \tag{37}$$

In the torsionless limit, (27) vanishes identically and (26) corresponds to the Einstein equations plus the Lanczos identity in V_4. Thus in this case (and for $\alpha = \frac{1}{2}$ K) we regain general relativity. The vanishing of (27) can be checked with the help of Eqs. (14) and (15).

The existence of torsion waves in the general case for the class of models based on the Lagrangian (24) can be verified in the weak torsion approximation defined by the assumptions (i) $g_{\alpha\beta} = \eta_{\alpha\beta}$, where $\eta_{\alpha\beta}$ is the Minkowski metric; and (ii) terms in which the torsion and/or its derivations appear quadratically are neglected.

In this approximation, (26) becomes an identity and (27) reduces to an equation for the propagating torsion, which will be given explicitly in the following cases.

(a) The nonvanishing components of the torsion are the vector components of the $(\frac{1}{2}, \frac{1}{2})$ representation $V_{\alpha} \equiv S_{\alpha\mu}{}^{\mu}$. For this case, $S_{\alpha\beta\gamma} = \frac{2}{3}\eta_{\gamma[\beta}V_{\alpha]}$ and (27) becomes

$$\Box V^{\mu} - \tfrac{3}{4}\left((3\theta + 2\lambda - 10\alpha/4\beta + 7\gamma + 2\xi)\right)V^{\mu} = 0. \tag{38}$$

(b) The torsion is determined by the axial vector of the $(\frac{1}{2}, \frac{1}{2})$ representation $A_{\alpha} \equiv \frac{1}{6}\varepsilon_{\lambda\mu\nu\alpha}S^{\lambda\mu\nu}$, so that $S_{\alpha\beta\gamma} = \varepsilon_{\alpha\beta\gamma\delta}A^{\delta}$. This gives (with the additional condition $\partial_{\mu}A^{\mu} = 0$)

$$\Box A_{\sigma} + \tfrac{9}{2}\left((\alpha - \lambda - \eta)/(3\beta + 12\gamma + 2\xi)\right)A_{\sigma} = 0. \tag{39}$$

Proca fields like those in Eqs. (38) and (39) have been found in some different classes of gravitational models and show the existence of mass torsion waves. We observe that a particular form of (26) and (27) was discussed in Ref. 11. We also remark that the main interest in the class of models presented here lies in the existence not only of torsion propagation but also of an Einstein limit accomplished via the Lanczos identity. This implies that the theory based on (24) satisfies the classical gravitational tests when the effect of torsion is negligible. We have also found that for axial torsion the field equations (26) and (27) with the weak torsion ansatz admit the same *pp*-wave solutions as in general relativity. More details about this class of models and also about the construction of other types of theories with the Gauss–Bonnet invariant limit are under investigation.

[1] B. Zwiebach, Phys. Lett. B **156**, 315 (1985); D. G. Boulware and S. Deser, Phys. Rev. Lett. **55**, 2656 (1985); S. Deser and A. N. Redlich, Phys. Lett. B **176**, 350 (1986); S. Ferrara and M. Villasante, J. Math. Phys. **30**, 104 (1989).
[2] D. Lovelock, J. Math. Phys. **12**, 498 (1971).
[3] J. R. Ray, J. Math. Phys. **19**, 100 (1977).
[4] J. L. Safko and F. Elston, J. Math. Phys. **17**, 1531 (1976).
[5] W. K. Atkins, W. M. Baker, and W. R. Davis, Phys. Lett. A **61**, 363 (1977).
[6] F. W. Hehl, E. A. Lord, and L. L. Smalley, Gen. Relativ. Gravit. **13**, 1037 (1981).
[7] H. T. Nieh and M. L. Yan, J. Math. Phys. **23**, 373 (1982).
[8] J. A. Schouten, *Ricci Calculus* (Springer, Berlin, 1954).
[9] R. Bach, Math. Z. **9**, 110 (1921); C. Lanczos, Ann. Math. **39**, 842 (1938).
[10] K. Hayashi and T. Shirafuji, Prog. Theor. Phys. **65**, 525 (1981).
[11] R. de Azeredo Campos and P. S. Letelier, Prog. Theor. Phys. **75**, 1359 (1986).
[12] F. W. Hehl, Y. Ne'eman, J. Nitsch, and P. von Heyde, Phys. Lett. B **78**, 102 (1978).
[13] W. Adamowicz, Gen. Relativ. Gravit. **12**, 677 (1979).
[14] E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. D **21**, 3269 (1980).
[15] K. Hayashi and T. Shirafuji, Prog. Theor. Phys. **64**, 866 (1980).
[16] Y. Z. Zhang, Phys. Rev. D **28**, 1866 (1983).
[17] F. Müller-Hoissen and J. Nitsch, Gen. Relativ. Gravit. **17**, 747 (1985).

Stochastic calculus and Schrödinger semigroups admitting a ground state

Dražen Pantić
University of Belgrade, Center for Multidisciplinary Studies, Slobodana Penezića Krcuna 35/IV, 1100 Belgrade, Yugoslavia

(Received 9 August 1989; accepted for publication 15 November 1989)

Presented are some new results concerning distorted Brownian motion, the Ito process associated with the Schrödinger operator with ground state. These results are obtained by means of Malliavin stochastic calculus. Under the condition of continuously differentiable drift coefficient with bounded partial derivatives, it is proved that the image of a bounded function after action of the expectation operator of distorted Brownian motion is twice continuously differentiable. This approach enabled us to prove that distorted Brownian motion has twice continuously differentiable transition densities. With the additional assumption that the drift is bounded, it is proved that the corresponding initial value problem has an infinitely differentiable fundamental solution.

I. INTRODUCTION

In this paper, we shall investigate regularity properties of the distribution of distorted Brownian motion. Distorted Brownian motion is a diffusion process that arises naturally in the formulation of the Schrödinger dynamics in terms of a ground state.[1–4]

We shall, following Ref. 4, assume that the Schrödinger operator $H = -\Delta/2 + V$ has a strictly positive ground state ρ_0 such that $\rho_0 \in L_{1,\mathrm{loc}}$, $V\rho_0 \in L_{1,\mathrm{loc}}$, and, for $f \in \mathbb{C}_0^\infty$,

$$\tfrac{1}{2}(\Delta f|\rho_0)_{L_2} = (f|V\rho_0)_{L_2}.$$

If $\nabla\rho_0 \in L_{1,\mathrm{loc}}$ we can define a positive self-adjoint operator $\tilde{H}$ in $L_2(m)$, $(m(dx) = \rho_0(x)^2\,dx)$, as Dirichlet forms, so that for $f \in \mathbb{C}_0^2$, $\tilde{H}f = -(\Delta f/2 + b\cdot\nabla f)$, where $b = \nabla \ln \rho_0$.[1]

For a potential V in Simon's class $\mathbb{K}_d$ (Ref. 3), let us define a Schrödinger semigroup $\{e^{-tH}\}$ in L_2 as

$$e^{-tH}f(x) = \mathbb{E}_x\left\{f(B_t)\exp\left(-\int_0^t V(B_s)ds\right)\right\},$$

where $\{B_t, t \geqslant 0\}$ is standard d-dimensional Brownian motion starting from zero. If U is a unitary map $U: L_2 \to L_2(m)$ defined as $Uf = f/\rho_0$, then the semigroups $\{e^{-t\tilde{H}}\}$ and $\{e^{-tH}\}$ are related,[3] as

$$e^{-t\tilde{H}} = U\cdot e^{-tH}\cdot U^{-1}. \tag{1}$$

The symmetric diffusion process $\{X_t, t \geqslant 0\}$ on $\mathbb{R}^d$ could be associated to the operator $\tilde{H}$, such that for a bounded, measurable function f, $(f \in \mathbb{M}_b)$, belonging to $L_2(m)$, $e^{-t\tilde{H}}f(x) = \mathbb{E}_x\{f(X_t)\}$, where $\mathbb{E}_x$ is expectation w.r.t. the measure generated by $\{X_t, t \geqslant 0)$ such that $X_0 = x$.

As for the symmetry we have, for $f, g \in \mathbb{M}_b \cap L_2(m)$

$$(f|\mathbb{E}\{g(X_t)\})_{L_2(m)} = (\mathbb{E}\{f(X_t)\}|g)_{L_2(m)}. \tag{2}$$

Further, if $b \in \mathbb{R}^d \otimes L_2(m)$, then $\{X_t\}$ is the solution of a stochastic differential equation

$$X_t - X_0 = \int_0^t b(X_s)ds + B_t,$$

where Brownian motion $\{B_t\}$ starts from zero and $P\{X_0 \in dx\} = m(dx)$.[1]

Following Ref. 5, the process $\{X_t\}$ is called the distorted Brownian motion.

Let us mark with (A) the condition that $b = \nabla \ln \rho_0$ has continuous and bounded first-order partial derivatives.

Then there exists,[6] a unique solution $\{X_{t,x}^{r,u}\}$ of the equation

$$X_{t,x}^{r,u} - x = \int_0^t b(X_{s,x}^{r,u})ds + B_t + r\int_0^t u_s\,ds, \tag{3}$$

where $r > 0$, $x \in \mathbb{R}^d$ and $\{u_t, t \geqslant 0\}$ is a process in $\mathbb{R}^d$, adapted to the natural filtration of $\{B_t, t \geqslant 0\}$, such that for some $\delta > 0$ and all t

$$\mathbb{E}\left\{\exp\left(\delta\int_0^t \|u_s\|^2\,ds\right)\right\} < \infty. \tag{4}$$

We shall denote $X_{t,x} = X_{t,x}^{0,0}$. It is clear that for $f \in M_b \cap L_2(m)$ $e^{-t\tilde{H}}f(x) = \mathbb{E}\{f(X_{t,x})\} = \mathbb{E}_x\{f(X_t)\}$.

Assuming condition (A), the following facts have been proven in Ref. 7.

(D1) For each $t > 0$ there exists the derivative of $X_{t,x}$ with respect to the initial condition x. The matrix $DX_{t,x} = [\partial X_{t,x}^{(i)}/\partial x_j]_{i,j=1}^d$ is nonsingular and $\sup\{\|DX_{t,x}\|: x \in \mathbb{R}^d\} \leqslant e^{Ct}$. (All matrix norms are to be understood in Hilbert–Schmidt sense.)

(D2) There exists a directional derivative in the direction u, $D_u X_{t,x} = \lim_{r\to 0}(X_{t,x}^{r,u} - X_{t,x})/r$, where u is as in Ref. 5.

For $\mathbf{u} = [u_{(1)},\ldots,u_{(d)}]$, the Bismut matrix $\mathbb{D}_{\mathbf{u}}X_{t,x} = [D_{u_{(1)}}X_{t,x},\ldots,D_{u_{(d)}}X_{t,x}]$, (Ref. 8) is

$$\mathbb{D}_u X_{t,x} = DX_{t,x}\int_0^t (DX_{s,x})^{-1}\cdot\mathbf{u}_s\,ds. \tag{5}$$

(D3) For $f \in \mathbb{M}_b \cap L_2(m)$ and $t > 0$, $e^{-t\tilde{H}}f$ is a continuously differentiable function, and

$$\nabla_x \mathbb{E}_x\{f(X_t)\} = \mathbb{E}_x\left\{f(X_t)\cdot\int_0^t DX_s\,dB_s\cdot t^{-1}\right\}. \tag{6}$$

Immediate consequences of the previous assertions are the following two lemmas.

Lemma 1: Let $\{f_s, 0 \leqslant s \leqslant t\}$ be a family of bounded functions, such that

 0022-2488/90/051221-05$03.00

$$\int_0^t \|f_s\|_\infty s^{-1/2}\,ds < \infty.$$

Then mapping $x \to \mathbb{E}_x\{\int_0^t f_s(X_s)ds\}$ is differentiable and

$$\nabla_x \mathbb{E}_x\left\{\int_0^t f_s(X_s)ds\right\} = \mathbb{E}_x\left\{\int_0^t f_s(X_s)\left(\int_0^s DX_v\, dB_v\right)s^{-1}\,ds\right\}. \tag{7}$$

Proof: We shall prove the Lemma for $d=1$. Multidimensional extension is straightforward.

Let us fix $s\in(0,t)$. Then (D3) implies that

$$\frac{\partial}{\partial x}\mathbb{E}_x\{f_s(X_s)\} = \mathbb{E}_x\left\{f_s(X_s)\int_0^s DX_v\, dB_v s^{-1}\right\}.$$

As

$$\left|\left|\mathbb{E}_x\left\{f_s(X_s)\int_0^s DX_v\, dB_v s^{-1}\right\}\right|\right|_\infty$$
$$\leqslant s^{-1}\|f_s\|_\infty \sup_x \mathbb{E}_x\left\{\left|\left|\int_0^s DX_v\, dB_v\right|\right|^2\right\}^{1/2}$$
$$= s^{-1}\|f_s\|_\infty \sup_x \mathbb{E}_x\left\{\int_0^s \|DX_v\|^2 dv\right\}^{1/2}$$
$$\leqslant s^{-1}\|f_s\|_\infty \left\{\int_0^s e^{2Cv}dv\right\}^{1/2}$$
$$\leqslant s^{-1/2}\|f_s\|_\infty e^{Ct}$$

we can conclude that integral

$$\mathbb{E}_x\left\{\int_0^t f_s(X_s)\left(\int_0^s DX_v\, dB_v\right)s^{-1}\,ds\right\}$$

converges uniformly w.r.t. x, so relation (7) holds. Q.E.D.

Lemma 2: Let f be a bounded function with bounded and continuous first-order partial derivatives. Then the functional $f(X_{t,x})$ admits the following representation:

$$f(X_{t,x}) = \mathbb{E}_x\{f(X_{t,x})\} + \int_0^t \mathbb{E}_{X_{s,x}}\{f'(X_{t-s})dX_{t-s}\}dB_s. \tag{8}$$

Proof: Since $f(X_{t,x})$ is a square integrable functional on $\mathcal{F}_t$ by Clark's representation,[9] we have

$$f(X_{t,x}) = \mathbb{E}\{f(X_{t,x})\} + \int_0^t H_{t,x}(s)dB_s$$

for a square integrable, adapted process $\{H_{t,x}(s),0\leqslant s\leqslant t\}$. Then for another square integrable, adapted process $\{u_s,0\leqslant s\leqslant t\}$ as in (4), using integration by parts formula, Ref. 8, and (D2), one gets

$$\mathbb{E}\left\{\int_0^t H_{t,x}(s)\cdot u_s\,ds\right\}$$
$$= \mathbb{E}\left\{f(X_{t,x})\int_0^t u_s\,dB_s\right\}$$
$$= \mathbb{E}\{D_u(f(X_{t,x}))\}$$
$$= \mathbb{E}\left\{f'(X_{t,x})DX_{t,x}\int_0^t (DX_{s,x}|^{-1}\cdot u_s\,ds\right\}$$
$$= \mathbb{E}\left\{\int_0^t \mathbb{E}\{f'(X_{t,x})DX_{t,x}(DX_{s,x})^{-1}\Big|\mathcal{F}_s\,ds\right\}.$$

Since u is arbitrary,

$$H_{t,x}(s) = \mathbb{E}\{f'CX_{t,x})DX_{t,x}(DX_{s,x})^{-1}|\mathcal{F}_s\}.$$

Now, as

$$DX_{t,x} - DX_{s,x} = \int_s^t b'(X_{u,x})\cdot DX_{u,x}\,du \text{ (Ref. 7)},$$

using Markov property, we can conclude that

$$\mathbb{E}\{f'(X_{t,x})DX_{t,x}(DX_{s,x})^{-1}|\mathcal{F}_s\} = \mathbb{E}_{X_{s,x}}\{f'(X_{t-s})DX_{t-s}\},$$

so the Lemma is proved. Q.E.D.

II. REGULARITY PROPERTIES OF THE SEMIGROUP $\{e^{-t\tilde{H}}\}$

Now, we shall extend the concept of stochastic calculus from Ref. 7 and prove the following theorem.

Theorem 1: Assume function b satisfies condition (A) and let $\{X_{t,x}\}$ be the family of solutions of Eq. (3). Then for $f\in\mathbb{C}_b^1$,

$$\mathbb{E}\{\nabla f(X_{t,x})\} = \mathbb{E}\left\{f(X_{t,x})\cdot\left(\int_0^t DX_{s,x}\,dB_s\cdot t^{-1} - \int_0^t b'(X_{s,x})dB_s\right)\right\}. \tag{9}$$

Proof: In Ref. 7 it is proved that $\{DX_{t,x}\}$ satisfies the following Cauchy equation:

$$DX_{t,x} - 1 = \int_0^t b'(X_{u,x})\cdot DX_{u,x}\,du.$$

Then, it is elementary that $(DX_{t,x})^{-1}$ exists and that

$$1 - (DX_{t,x})^{-1} = \int_0^t (DX_{s,x})^{-1}\cdot b'(X_{s,x})ds,$$

so multiplying by $DX_{t,x}$ on both sides of the above equality one gets

$$DX_{t,x} - 1 = DX_{t,x}\int_0^t DX_{s,x}^{-1}\cdot b'(X_{s,x})ds.$$

Defining $\tilde{u}_\cdot = b'(X_{\cdot,x})$, by (D3) one has, $D_{\tilde{u}}X_{t,x} = DX_{t,x} - 1$. Also for $\hat{u}_\cdot = DX_{\cdot,x}$, it is $DX_{t,x} = t^{-1}\cdot D_{\hat{u}}X_{t,x}$, hence

$$1 = t^{-1}\cdot D_{\hat{u}}X_{t,x} - D_{\tilde{u}}X_{t,x}.$$

For a function $f\in\mathbb{C}_b^1$ we have

$$f'(X_{t,x}) = t^{-1}\cdot D_{\tilde{u}}(f(X_{t,x})) - D_{\hat{u}}(f(X_{t,x})),$$

and by integration of the parts formula (Ref. 8) the conclusion of the theorem holds. Q.E.D.

Using Lemmas 1 and 2 and Theorem 1 we are able to prove the following theorem.

Theorem 2: If the function b has continuous and bounded first-order partial derivatives then, for a bounded measurable function f and $t>0$ the mapping $x\to\mathbb{E}\{f(X_{t,x})\}$ is twice continuously differentiable with bounded derivatives.

Proof: Without the loss of generality we can carry the proof out for $d=1$.

In (D1) it is proved that $\mathbb{E}\{f(X_{t,x})\}$ is once continuously differentiable for $f\in\mathbb{M}_b$, while the boundeness of the first derivative is obvious from relation (6). By the semigroup property $\mathbb{E}\{f(X_{t+s,x})\} = \mathbb{E}\{\mathbb{E}_{X_{s,x}}\{f(X_t)\}\}$, so it is sufficient to prove the theorem for $f\in\mathbb{C}_b^1$.

Theorem 1 and (D1) imply that, for $f\in C_b^1$

$$\frac{\partial}{\partial x}\mathbb{E}\{f(X_{t,x})\}=\mathbb{E}\{f'(X_{t,x})\}+\mathbb{E}\left\{f(X_{t,x})\int_0^t b'(X_{s,x})dB_s\right\}.$$

Now, using representation for $f(X_{t,x})$ as in Lemma 2 we have

$$\frac{\partial}{\partial x}\mathbb{E}\{f(X_{t,x})\}=\mathbb{E}\{\hat{f}'(X_{t,x})\}+\mathbb{E}\left\{\int_0^t \mathbb{E}_{X_{s,x}}\{f'(X_{t-s})DX_{t-s}\}b'(X_{s,x})ds\right\}.$$

The first term $\mathbb{E}\{f'(X_{t,x})\}$ is differentiable w.r.t. x by (D1). Lemma 1 implies that the second term is differentiable as the family of functions $f_s(x)=b'(x)\mathbb{E}_x\{f'(X_{t-s})DX_{t-s}\}$ obviously satisfies conditions from Lemma 1.

So, we have proved that $\mathbb{E}\{f(X_{t,x})\}$ is twice differentiable, and using Markov property as in Lemma 2 in the end we get

$$\frac{\partial^2}{\partial x^2}\mathbb{E}\{f(X_{t,x})\}=\mathbb{E}\left\{f'(X_{t,x})\int_0^t DX_{s,x}\,dB_s\cdot t^{-1}\right\}+\mathbb{E}\left\{f'(X_{t,x})DX_{t,x}\int_0^t b'(X_{s,x})(DX_{s,x})^{-1}\times\left(\int_0^s DX_{u,x}dB_u\right)s^{-1}\,ds\right\}.$$

The right side of the above formula is a continuous and bounded function in x so the theorem is proved. Q.E.D.

Using Theorem 1 and the so-called Malliavin lemma (see Ref. 8) one can conclude that for $t>0$ and $x\in\mathbb{R}^d$ the distribution of the random variable $X_{t,x}$ is absolutely continuous with respect to the Lebesgue measure, and consequently, w.r.t. the measure m.

That fact is accurately stated in the following theorem.

Theorem 3: Let everything be as in Theorem 2. Then there exists a function $\rho\colon \mathbb{R}_+\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}_+$ such that for $f\in\mathbb{M}_b\cap L_2(m)$, $t>0$, $x\in\mathbb{R}^d$

$$e^{-t\bar{H}}f(x)=\mathbb{E}\{f(X_{t,x})\}=\int_{\mathbb{R}^d}f(y)\cdot\rho_t(x,y)\cdot\rho_0(y)^2\,dy.$$

Moreover, there exist constant K depending only on dimension d such that

$$\sup\{\rho_t(x,y)\colon y\in\mathbb{R}^d\}\leqslant K\cdot\max(t^{-d/2},1)\cdot\rho_0(x)^{-2}.$$

Function $\rho_t(x,y)$ is once continuously differentiable w.r.t. variable t, and twice continuously differentiable w.r.t. spatial variables x and y.

Proof: Let $\{P_t(x,B)\colon t>0, x\in\mathbb{R}, B\in\mathcal{B}(\mathbb{R})\}$ be a family of transition probabilities of the process $\{X_t\}$. Then relation (8) from Theorem 2 could be expressed as

$$\int_{\mathbb{R}^d}\nabla f(y)\cdot P_t(x,dy)=\int_{\mathbb{R}^d}f(y)\cdot r_t(x,y)\cdot P_t(x,dy),$$

where

$$r_t(x,y)=\mathbb{E}\left\{\int_0^t DX_{s,x}\,dB_s\cdot t^{-1}-\int_0^t b'(X_{s,x})dB_s\,|X_{t,x}=y\right\}. \tag{10}$$

Now, for $q\geqslant 2$, it is

$$\left(\int_{\mathbb{R}^d}\|r_t(x,y)\|^q\cdot P_t(x,dy)\right)^{1/q}=\mathbb{E}\left\{\left|\left|\mathbb{E}\left\{\int_0^t DX_{s,x}\,dB_s\cdot t^{-1}-\int_0^t b'(X_{s,x})\,dx_s|X_{t,x}\right\}\right|\right|^q\right\}^{1/q}$$
$$\leqslant\mathbb{E}\left\{\left|\left|\int_0^t DX_{s,x}\,dB_s\cdot t^{-1}-\int_0^t b'(X_{s,x})dB_s\right|\right|^q\right\}^{1/q}$$
$$\leqslant\mathbb{E}\left\{\left|\left|\int_0^t DX_{s,x}\,dB_s\cdot t^{-1}\right|\right|^q\right\}^{1/q}+\mathbb{E}\left\{\left|\left|\int_0^t b'(X_{s,x})dB_s\right|\right|^q\right\}^{1/q}.$$

Now, by Burkholder's inequality [Ref. 10, p. 286], we have

$$\mathbb{E}\left\{\left|\left|\int_0^t DX_{s,x}\,dB_s\right|\right|^q\right\}\leqslant C(q)t^{q/2-1}\cdot\mathbb{E}\left\{\int_0^t\|DX_{s,x}\|^q\,ds\right\}\leqslant C(q)t^{q/2}e^{Ctq},$$

and

$$\mathbb{E}\left\{\left|\left|\int_0^t b'(X_{s,x})dB_s\right|\right|^q\right\}\leqslant C(q)t^{q/2-1}\cdot\mathbb{E}\left\{\int_0^t\|b'(X_{s,x})\|^q\,ds\right\}\leqslant C(q)\cdot t^{q/2}\cdot C,$$

so

$$\left(\int_{\mathbb{R}^d}\|r_t(x,y)\|^q\cdot P_t(x,dy)\right)^{1/q}\leqslant Kt^{-1/2}e^{Ct}. \tag{11}$$

We can always chose q to be greater than d and using Theorem (1.28) from Ref. 10 conclude that there exists a density w.r.t. Lebesgue measure $q_t(x,y)$ such that

$$P_t(x,dy)=q_t(x,y)dy$$

and

$$\sup\{q_t(x,y)\colon x,y\in\mathbb{R}^d\}\leqslant\tilde{C}t^{-d/2}e^{dCt}.$$

For $t\leqslant 1$ we have $\sup\{q_t(x,y)\colon x,y\in\mathbb{R}^d\}\leqslant Kt^{-d/2}$ and for $s=1+u$, by the Chapman–Kolmogorov theorem,

$$q_s(x,y)=\int_{\mathbb{R}^d} q_1(x,z)q_u(z,y)\,dz$$

$$\leqslant K\int_{\mathbb{R}^d} q_u(z,y)\,dz=K.$$

So, we have $\sup\{q_t(x,y):\ x,y\in\mathbb{R}^d\}\leqslant K\max(t^{-d/2},1)$.

Since assumption (A) implies that $\rho_0>0$, we can define density w.r.t. measure m as $\rho_t(x,y)=q_t(x,y)\rho_0(y)^{-2}$. Then we have, for $f\in\mathbb{M}_b\cap L_2(m)$, $t>0$, $x\in\mathbb{R}^d$,

$$e^{-t\tilde{H}}f(x)=\mathbb{E}\{f(X_{t,x})\}$$

$$=\int_{\mathbb{R}^d} f(y)\cdot\rho_t(x,y)\cdot\rho_0(y)^2\,dy.$$

The symmetry of distorted Brownian motion, relation (2), implies that $\rho_t(x,y)=\rho_t(y,x)$, and consequently

$$\sup\{\rho_t(x,y):\ y\in\mathbb{R}^d\}\rho_0(x)^2\leqslant K\max(t^{-d/2},1).$$

For fixed $x\in\mathbb{R}^d$ and $s>0$, $\rho_s(x,\cdot)$ is a bounded function on $\mathbb{R}^d$. The Chapman–Kolomogrov formula

$$\rho_{t+s}(x,y)=\int_{\mathbb{R}^d}\rho_s(x,z)\rho_t(z,y)\rho_0(z)^2\,dz$$

$$=\int_{\mathbb{R}^d}\rho_s(x,z)\rho_t(y,z)\rho_0(z)^2\,dz$$

$$=\mathbb{E}\{\rho_s(x,X_{t,y})\}\qquad(12)$$

and Theorem 2 imply that mapping $y\to\rho_{t+s}(x,y)$ is twice continuously differentiable. Symmetry of x and y in $\rho_{t+s}(x,y)$ yields that, for fixed y, mapping $x\to\rho_{t+s}(x,y)$ is also twice continuously differentiable.

The differentiability of map $t\to\rho_t(x,y)$ is now easily established. By the Ito formula

$$\rho_s(x,X_{t,y})-\rho_s(x,y)$$

$$=\int_0^t\nabla_y\rho_s(x,X_{v,y})\,dB_v+\int_0^t\left[\frac{1}{2}\Delta_y\rho_s(x,X_{v,y})\right.$$

$$\left.+b(X_{v,y})\cdot\nabla_y\rho_s(x,X_{v,y})\right]dv,$$

and taking into account expectation, it is obtained that

$$\rho_{s+t}(x,y)-\rho_s(x,y)=\int_0^t\mathbb{E}\left\{\frac{1}{2}\Delta_y\rho_s(x,X_{v,y})\right.$$

$$\left.+b(X_{v,y})\cdot\nabla_y\rho_s(x,X_{v,y})\right\}dv,$$

hence the proof of the theorem is completed. Q.E.D.

In addition to the assumption that b has continuous and bounded partial derivatives of the first order, let us now suppose that function b is also bounded.

On the space $\mathbb{C}_b$ of bounded, continuous functions, with the usual supremum norm, we can define C_0 semigroup of operators $\{P_t,t\geqslant 0\}$, as $P_tf(x)=\mathbb{E}_x\{f(X_t)\}$. Let A be a generator of semigroup $\{P_t,t\geqslant 0\}$. Then using the Ito formula one can easily see that $\mathbb{C}_b^2\subset\mathrm{Dom}(A)$ and, for $f\in\mathbb{C}_b^2$, $Af=\Delta f/2+b\cdot\nabla f$, while for $f\in\mathbb{C}_0^2$ $Af=-\tilde{H}f$. Also, Theorem 2 implies that for $t>0$, $P_t:\mathbb{C}_b\to\mathbb{C}_b^2\subset\mathrm{Dom}(A)$, hence mapping $t\to P_tf$ is strongly differentiable for each $f\in\mathbb{C}_b$, i.e., $\{P_t,t\geqslant 0\}$ is a differentiable C_0 semigroup.[11] Then, for $t>0$ and $n\in\mathbb{N}$, $P_t:\mathbb{C}_b\to\mathrm{Dom}(A^n)$ and $A^n\circ P_t$ is a bounded operator on $\mathbb{C}_b$.[11] From Theorem 2 we know that, for $t>0$ and a bounded function f, $\mathbb{E}_x\{f(X_t)\}\in\mathbb{C}_b^2\subset\mathbb{C}_b$ hence

$$\mathbb{E}_x\{f(X_{t+s})\}=P_s(\mathbb{E}\{f(X_t)\})\in\bigcap_{n=1}^{\infty}\mathrm{Dom}(A^n).$$

Theorems 2 and 3 imply that for each fixed $v>0$ and $x\in\mathbb{R}^d$ (respectively, $y\in\mathbb{R}^d$), $\rho_t(x,\cdot)$ [respectively, $\rho_t(\cdot,y)$] is in $\mathbb{C}_b^2$, hence

$$A_x\rho_v(x,y)=\Delta_x\rho_v(x,y)/2+b(x)\cdot\nabla_x\rho_v(x,y).$$

Then using relation (12) one has

$$A_x\rho_{t+s}(x,y)=A_x\mathbb{E}_x\{\rho_s(X_t,y)\}$$

$$=\mathbb{E}_x\{A_x\rho_s(X_t,y)\}$$

$$=\int_{\mathbb{R}^d}\rho_t(x,z)\cdot A_z\rho_s(z,y)\rho_0(z)^2\,dz$$

$$=\int_{\mathbb{R}^d}A_z\rho_t(x,z)\cdot\rho_s(z,y)\rho_0(z)^2\,dz$$

$$=\mathbb{E}_y\{A_y\rho_t(x,X_s)\}=A_y\rho_{t+s}(x,y),$$

so

$$A_x\rho_t(x,y)=A_y\rho_{t+s}(x,y).$$

Now, as

$$\rho_t(x,\cdot)\in\bigcap_{n=1}^{\infty}\mathrm{Dom}(A_y^n)$$

and

$$\rho_t(\cdot,y)\in\bigcap_{n=1}^{\infty}\mathrm{Dom}(A_x^n)$$

and $\mathrm{Dom}(A)\in\mathbb{C}^2$ we have that $\rho_t\subset\mathbb{C}^\infty(\mathbb{R}^d\times\mathbb{R}^d)$. Having in mind that $(\partial/\partial t)\rho_t=A\rho_t$, we have that $\rho\in\mathbb{C}^\infty(\mathbb{R}_+\times\mathbb{R}^d\times\mathbb{R}^d)$.

We shall use the above considerations to prove the following theorem.

Theorem 4: Let $b=\nabla\ln\rho_0$ be a bounded function with bounded and continuous first-order derivatives, and $\{X_{t,x},t\geqslant 0\}$ distorted Brownian motion starting from $x\in\mathbb{R}^d$. Then the initial value problem $(\partial/\partial t)u(t,x)=\frac{1}{2}\Delta_xu(t,x)+b(x)\cdot\nabla_x(t,x)$, $u(0,x)=f(x)\in\mathbb{C}_b$ has a unique solution of the form

$$u(t,x)=\int_{\mathbb{R}^d}f(y)\cdot\rho_t(x,y)\cdot\rho_0(y)^2\,dy$$

where ρ, distorted Brownian motion's transition density w.r.t. measure $\rho_0(y)^2\,dy$, could be chosen such that ρ belongs to $\mathbb{C}^\infty(\mathbb{R}_+\times\mathbb{R}^d\times\mathbb{R}^d)$.

Proof: The only unproven part is uniqueness of solutions. Let $u\in\mathbb{C}^{1,2}([0,\infty)\times\mathbb{R}^d)$ such that

$$\frac{\partial}{\partial t}u(t,x)=\frac{1}{2}\Delta_xu(t,x)+b(x)\cdot\nabla_xu(t,x),$$

and $u(0,x)=0$. For $0\leqslant t\leqslant T$ let us define $h(t,x)=u(T-t,x)$. Then, for $t\leqslant T$,

$$\frac{\partial}{\partial t}h(t,x)+\frac{1}{2}\Delta_xh(t,x)+b(x)\cdot\nabla_xh(t,x)=0.$$

By the Ito formula we have

$$h(t,X_t)-h(s,X_s)=\int_s^t\nabla h(v,X_v)\,dB_v$$

$$+\int_s^t \left[\frac{\partial}{\partial t} h(v,X_v) + \frac{1}{2}\Delta_x h(v,X_v) + b(X_v)\cdot\nabla_x h(v,X_v)\right] dv,$$

hence, $\mathbb{E}_x\{h(t,X_t)\} = \mathbb{E}_x\{h(s,X_s)\}$. Taking $s=0$ and $t=T$ one gets

$$0 = \mathbb{E}_x\{u(0,X_T)\} = \mathbb{E}_x\{u(T,X_0)\} = u(T,x). \quad \text{Q.E.D.}$$

[1]S. Albeverio, R. Hoegh-Krohn, and L. Streit, "Energy forms, Hamiltonians, and distorted Brownian motion," J. Math. Phys. **18**, 907 (1977).

[2]S. Albeverio, F. Gesztesy, W. Karwowsky, and L. Streit, "On the connection between Schrödinger and Dirichlet forms," J. Math. Phys. **26**, 2546 (1985).

[3]B. Simon, "Schrödinger semigroups," Bull. Am. Math. Soc. **7**, 447 (1982).

[4]L. Streit, "Energy forms: Schrödinger theory, Processes," Phys. Rep. **77**, 361 (1981).

[5]H. Ezawa, J. R. Klauder, and L. A. Shepp, "A path space picture for Feynman–Kac functional," Ann. Phys. **88**, 588 (1974).

[6]Ju. B. Blagoveschenskii and M. I. Freidlin, "Some properties of diffusion processes depending on the parameter," Sov. Math. **2**, 633 (1961).

[7]D. Pantić, "Stochastic calculus on distorted Brownian motion," J. Math. Phys. **29**, 207 (1988).

[8]M. Zakai, "The Malliavin calculus," Acta Appl. Math. **3**, 175 (1985).

[9]A. Ju. Veretennikov, "Probabilistic approach to hypoellipticity," Usp. Mat. Nauk. **38**, 113 (1983).

[10]S. Kusuouka and D. Stroock, "Applications of the Malliavin calculus, Part I," in *Proceedings of the Taniguchi International Symposium on Stochastic Analysis*, Katata and Kyoto, 1982, edited by K. Ito (North-Holland, Amsterdam, 1984), pp. 271–306.

[11]A. Pazy, *Semigroups of Liner Operators and Application to Partial Differential Equations* (Springer, New York, 1983).

Deformations of complex structures and conformal field theories on Riemann surfaces

E. Guadagnini
Dipartimento di Fisica dell'Universitá di Pisa, Pisa, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Pisa, Italy

M. Martellini
Dipartimento di Fisica dell'Universitá di Milano, Milan, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, Pavia, Italy

M. Mintchev
Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Pisa, Italy and Dipartimento di Fisica dell'Universitá di Pisa, Pisa, Italy

(Received 13 December 1988; accepted for publication 29 November 1989)

The role of complex structure deformations in representing the Virasoro algebra associated to any conformal field theory on a genus $p \geqslant 1$ compact Riemann surface is investigated.

I. INTRODUCTION

In the early days of conformal field theory it was already clear that the Virasoro algebra[1] plays an essential role in describing the symmetry content of conformal invariant models defined on the plane.[2] In the last few years there has been considerable progress in the study of conformal field theories on a general Riemann surface. This development, which was stimulated by string theories[3] and the study of critical phenomena,[4,7] naturally poses the problem of generalizing the Virasoro algebra to Riemann surfaces of nontrivial genus. Recently there have been a few attempts[8] at solving this problem in abstract form. In the present paper, we will be mainly concerned with the approach of Eguchi and Ooguri.[9] It is worth mentioning that in the formalism of Ref. 9, the Ward identities reflect the conformal invariance, also taking into account the deformations of the complex structure on the underlying genus $p \geqslant 1$ compact Riemann surface Σ_p. By means of these generalized (with respect to the familiar $p = 0$) Ward identities,[4] Eguchi and Ooguri show that the Euclidean correlation functions of the system give rise to a representation $\mathscr{I}$ of two commuting (conventionally called left and right) Virasoro algebras. Each of these algebras is originally defined on any coordinate patch of Σ_p, but may be consistently glued together to form a global algebraic structure on Σ_p. It turns out that this structure is universal (p independent) and consequently, the commutation relations are those from the $p = 0$ case. All information about the nontrivial genus is carried by the representation $\mathscr{I}$. In order to make a comparison, we recall that for $p = 0$ the space $\mathscr{I}$ has the structure[4]

$$\mathscr{I} = \oplus_{\alpha,\beta>0} \mathscr{I}_L(h_\alpha, c) \otimes \mathscr{I}_R(\bar{h}_\beta, c), \tag{1.1}$$

where $\mathscr{I}_{L(R)}(h,c)$ is a highest weight representation of the left (right) Virasoro algebra characterized by central charge c and highest weight h. Moreover, in conformal field theory $h \geqslant 0$ and the representation $\mathscr{I}_{\text{vac}} \equiv \mathscr{I}_L(0,c) \otimes \mathscr{I}_R(0,c)$, which we call the vacuum sector, appears exactly once in the direct sum (1.1).

In the present paper we further pursue the investigation of Ref. 9 by extracting from the conformal Ward identities information about the structure of the representation $\mathscr{I}$ for $p \geqslant 1$. We show that some new phenomena related to the presence of complex structure deformations occur. In particular, by deforming the complex structure on Σ_p, one obtains in general unitarily inequivalent field theories. The reason is that the energy-momentum tensor T_{zz} develops a nontrivial vacuum expectation value $\langle T_{zz} \rangle$ that *depends* on the complex structure. By means of the conformal Ward identities we prove that $\langle T_{zz} \rangle$ cannot be absorbed in a redefinition of T_{zz}, maintaining conformal invariance. As a consequence of this phenomenon, field theories that for $p = 0$ give rise to a representation of the Virasoro algebra of the type given by Eq. (1.1) in general do not preserve this property when quantized on Σ_p with $p \geqslant 1$. Indeed, we show that for $p \geqslant 1$ one has the following alternative. One can have either nondegenerate vacuum or standard Hermiticity properties of the Virasoro generators on $\mathscr{I}_{\text{vac}}$, but not both. Therefore, in both cases the vacuum sector $\mathscr{I}_{\text{vac}}$ is not a unitary highest weight representation. We also prove that, different from the $p = 0$ case, for $p \geqslant 1$ some of the descendent states in $\mathscr{I}_{\text{vac}}$ couple to the vacuum.

Finally, we discuss some sufficient conditions for the vanishing of $\langle T_{zz} \rangle$. These conditions are expressed in terms of the partition function and in this sense are not universal, but depend on the model. We also show that as a consequence of modular invariance, $\langle T_{zz} \rangle$ vanishes on the orbifold points of the moduli space $\mathscr{M}_p$ associated to Σ_p.

Our paper is organized as follows. In Sec. II we derive the general form of $\langle T_{zz} \rangle$ as a solution of the conformal Ward identities and clarify its meaning. Following Ref. 9, we introduce in Sec. III the Virasoro algebra and study the impact of $\langle T_{zz} \rangle$ on its representation. Section IV contains some examples on the torus. A comparison to the finite temperature theory on the cylinder is also given. Section V is devoted to our conclusions and we comment on string unitarity.

II. THE ENERGY-MOMENTUM TENSOR ON A RIEMANN SURFACE

In this section we consider a field theory with the classical action

$$S_{\text{cl}} = \int_{\Sigma_p} d^2x \sqrt{g} \mathscr{L}(u_a, \partial_\mu u_a), \tag{2.1}$$

 0022-2488/90/051226-08$03.00

where u_a are two-dimensional matter fields and Σ_p is a compact Riemann surface of genus p equipped with the metric $g_{\mu\nu}$. It is worth stressing that $S_{\rm cl}$ need not be a Gaussian action. We assume further that (2.1) is invariant under Weyl rescalings and both diffeomorphisms and local Lorentz transformations on Σ_p. This assumption implies (see, e.g., Ref. 10) the following functional dependence:

$$S_{\rm cl} = S_{\rm cl}(m;u_a,\partial_\mu u_a). \tag{2.2}$$

Here m is a point in the moduli space

$$m \in \mathcal{M}_p \equiv \mathrm{Met}(\Sigma_p)/\mathrm{Diff}(\Sigma_p) \otimes \mathcal{W}, \tag{2.3}$$

where $\mathrm{Met}(\Sigma_p)$ is the set of all metrics on Σ_p, while $\mathrm{Diff}(\Sigma_p)$ and $\mathcal{W}$ denote the groups of diffeomorphisms and Weyl rescalings, respectively. As is well known, $\mathcal{M}_p$ is finite dimensional. For the real dimension $\dim \mathcal{M}_p \equiv 2M_p$ one has $M_0 = 0$, $M_1 = 1$, and $M_p = 3p-3$ for $p \geqslant 2$. It can be shown that $\mathcal{M}_p$ has a natural complex structure. For the relative complex coordinates we shall use the notation $\{y_j : j = 1,\ldots,M_p\}$.

In performing the quantization, $\mathrm{Diff}(\Sigma_p)$ and $\mathcal{W}$ are in general potentially anomalous. Assuming that the Lorentz symmetry is not anomalous, by means of local counterterms all anomalies can be shifted to $\mathcal{W}$. In what follows we shall assume that this shift has been performed. Consequently, for the quantum action S_q one has

$$S_q = S_q(y,\bar{y};\sigma(x);u_a,\partial_\mu u_a), \tag{2.4}$$

where $\sigma(x)$ stands for the Liouville mode. The associated partition function will be denoted by

$$Z(y,\bar{y};\sigma(x)). \tag{2.5}$$

The next step toward the conformal Ward identities is the construction of the energy-momentum tensor $T_{\mu\nu}(x)$ and the conformal fields $\phi_k(x)$. In general $T_{\mu\nu}$ and ϕ_k are composite operators built up in terms of the elementary fields $u_a(x)$. Finally, the conformal Ward identities are expressed in terms of the Euclidean correlation functions

$$\langle T_{\mu\nu}(x_1)\cdots\phi_k(x_k)\cdots\rangle. \tag{2.6}$$

For convenience we switch from now on to complex isothermal coordinates, i.e., Σ_p is covered by an atlas $\mathcal{A} = \{U_\alpha\}$ of local coordinates such that on each coordinate patch one has $g_{zz} = g_{\bar{z}\bar{z}} = 0$ and $g_{z\bar{z}} = g_{\bar{z}z}$. In these coordinates the conformal Ward identities with one energy-momentum insertion read as[9]

$$\begin{aligned}\langle T_{zz}\phi_1(z_1,\bar{z}_1)\cdots\phi_n(z_n,\bar{z}_n)\rangle &= \sum_{j=1}^{M_p} h_{zz,}{}^{j}(z)\int d^2w\sqrt{g}g^{w\bar{w}}\eta^{w}{}_{\bar{w},j}(w,\bar{w})\langle T_{ww}\phi_1(z_1,\bar{z}_1)\cdots\phi_n(z_n,\bar{z}_n)\rangle - \sum_{k=1}^{n}\left[h_k\nabla_{z_k}G^{z_k}{}_{zz}(z_k,z) + G^{z_k}{}_{zz}(z_k,z)\nabla^{(s_k)}_{z_k}\right]\\ &\times\langle\phi_1(z_1,\bar{z}_1)\cdots\phi_n(z_n,\bar{z}_n)\rangle - \frac{c}{48\pi}\int d^2w\sqrt{g}G^{w}{}_{zz}(w,z)\partial_w R(w,\bar{w})\langle\phi_1(z_1,\bar{z}_1)\cdots\phi_n(z_n,\bar{z}_n)\rangle,\end{aligned} \tag{2.7}$$

where h_k is the conformal weight of ϕ_k defined in terms of the dimension d_k and the spin s_k by $h_k = \frac{1}{2}(d_k + s_k)$. The covariant derivative $\nabla^{(s)}_z$ is given by

$$\nabla^{(s)}_z = \partial_z + is\omega_z,$$

where ω_z is the spin connection. The function $G^{w}{}_{zz}$ is the Green kernel for the operator ∇^w and satisfies

$$\nabla^w G^{w}{}_{zz}(w,z) = \frac{1}{\sqrt{g}}\delta(z-w) - \sum_{j=1}^{M_p} g^{w\bar{w}}\eta^{w}{}_{\bar{w},j}(w,\bar{w})h_{zz,}{}^{j}(z), \tag{2.8}$$

where $\{h_{zz,}{}^{j}\}$ is a basis of holomorphic quadratic differentials on Σ_p, while $\{\eta^{z}{}_{\bar{z},j}\}$ are the Beltrami differentials dual to $h_{zz,}{}^{j}$, i.e.,

$$\int d^2z\sqrt{g}g^{z\bar{z}}\eta^{z}{}_{\bar{z},i}(z,\bar{z})h_{zz,}{}^{j}(z) = \delta^j_i. \tag{2.9}$$

Finally, R in (2.7) is the scalar curvature and the numerical factor c is the central charge of the associated Virasoro algebra. A similar identity holds for $T_{\bar{z}\bar{z}}$.

Let us now concentrate on the vacuum expectation value $\langle T_{zz}\rangle$. From (2.7) one obtains

$$\langle T_{zz}\rangle - \sum_{j=1}^{M_p} h_{zz,}{}^{j}\int d^2w\sqrt{g}g^{w\bar{w}}\eta^{w}{}_{\bar{w},j}\langle T_{ww}\rangle = -\frac{c}{48\pi}\int d^2w\sqrt{g}G^{w}{}_{zz}(w,z)\partial_w R, \tag{2.10}$$

which represents an integral equation for $\langle T_{zz}\rangle$. For solving (2.10) it is convenient to introduce the field[9]

$$t_{zz}(z,\bar{z}) \equiv \partial_z\Gamma^{z}{}_{zz} - \tfrac{1}{2}(\Gamma^{z}{}_{zz})^2, \tag{2.11}$$

which behaves as a Schwarzian differential[11] under holomorphic transformations. A simple computation shows that

$$\partial_z R = -2g^{z\bar{z}}\partial_{\bar{z}}t_{zz}. \tag{2.12}$$

Inserting (2.12) in the rhs of (2.10) and integrating by parts, with the help of (2.8) one obtains

$$\langle T_{zz}\rangle + \frac{c}{24\pi}t_{zz} - \sum_{j=1}^{M_p} h_{zz,}{}^{j}\int d^2w\sqrt{g}g^{w\bar{w}}\eta^{w}{}_{\bar{w},j}\times\left(\langle T_{ww}\rangle + \frac{c}{24\pi}t_{ww}\right) = 0. \tag{2.13}$$

By means of (2.9) one can easily verify that the general solution of (2.13) is

$$\langle T_{zz}\rangle = \sum_{j=1}^{M_p} h_{zz,}{}^{j}\mathcal{Y}_j(y,\bar{y}) - \frac{c}{24\pi}t_{zz}(z,\bar{z}), \tag{2.14}$$

where $\mathcal{Y}_j$ defines a vector field on $\mathcal{M}_p$ which is still to be determined. Equation (2.14) deserves some comments.

Under the holomorphic change of coordinates $z \to f(z)$, t_{zz} transforms as

$$t_{zz}(dz)^2 \mapsto t_{zz}(dz)^2 - \{f,z\}(dz)^2, \tag{2.15}$$

where $\{f,z\}$ is the Schwarzian derivative[11]

$$\{f,z\} = \frac{f'''(z)}{f'(z)} - \frac{3}{2}\left[\frac{f''(z)}{f'(z)}\right]^2. \tag{2.16}$$

Therefore, for $c \neq 0$, T_{zz} is not a tensor field, but as expected on general grounds, generates a Schwarzian under general holomorphic reparametrizations.

In the case $c \neq 0$, the rhs of (2.14) depends in general on $\bar{z}$, which makes impossible the introduction of the Virasoro algebra by means of a Laurent expansion of T_{zz}. From Eq. (2.12) one obtains

$$\partial_{\bar{z}}\, t_{zz} = -\tfrac{1}{2} g_{z\bar{z}}\, \partial_z R, \tag{2.17}$$

which implies that T_{zz} is a meromorphic function if and only if the metric is of constant curvature. From now on we assume that this is the case. For constant curvature Eq. (2.10) states that the projection of $\langle T_{zz}\rangle$, orthogonal to the quadratic differentials, vanishes. Therefore, one has

$$\langle T_{zz}\rangle = \sum_{j=1}^{M_p} h_{zz,}{}^{j}(z)\, V_j(y,\bar{y}), \tag{2.18}$$

where V_j is the vector field on $\mathcal{M}_p$ one obtains from $\mathcal{V}_j$ by specifying that the metric has constant curvature. For the slice of constant curvature metrics, the absence in the rhs of (2.18) of a Schwarzian differential is consistent over the whole surface. In fact, in this case the atlas $\mathcal{A}$ can be chosen so that the transition functions between overlapping patches are isometries; it is also known that $\{f,z\}$ vanishes if f is an isometry. The existence of an atlas $\mathcal{A}$ with the above properties is crucial for defining the theory globally on Σ_p because for $c \neq 0$ only the isometries are unitarily implementable and do not modify physics. On the contrary, general holomorphic coordinate transformations connect physically inequivalent theories and can be used, for example, to derive[5] the central charge as a kind of Casimir effect.

In the case $p = 0$ one obtains from (2.18) that $\langle T_{zz}\rangle = 0$, which is extensively used in deriving the familiar properties of the Virasoro algebra on the sphere.

In the rhs of Eq. (2.18) one recognizes the general expression of the classical energy-momentum tensor used in Ref. 8. All dynamical information is contained in V_j. In order to clarify the physical meaning of (2.18) for $p \geqslant 1$, it is useful to compute V_j explicitly. We recall first that for an infinitesimal deformation δy^j of the complex structure one has[10,12]

$$\delta g_{\bar{z}\bar{z}} = 2 \sum_{j=1}^{M_p} g_{\bar{z}z}\, \eta^{z}{}_{\bar{z},j}\, \delta y^j. \tag{2.19}$$

Therefore, under deformations,

$$\delta(\langle \phi_1 \cdots \phi_n\rangle Z) = \sum_{j=1}^{M_p} \delta y^j \int d^2w \sqrt{g} g^{w\bar{w}} \eta^{w}{}_{\bar{w},j} \langle T_{ww}\phi_1 \cdots \phi_n\rangle Z, \tag{2.20}$$

where Z is the partition function (2.5). Equivalently,

$$\frac{\partial}{\partial y^j}(\langle \phi_1 \cdots \phi_n\rangle Z) = \int d^2w \sqrt{g} g^{w\bar{w}} \eta^{w}{}_{\bar{w},j} \langle T_{ww}\phi_1 \cdots \phi_n\rangle Z, \tag{2.21}$$

which in view of (2.9) implies

$$Z\langle T_{zz}\phi_1 \cdots \phi_n\rangle = \sum_{j=1}^{M_p} h_{zz,}{}^{j} \frac{\partial}{\partial y^j}(Z\langle \phi_1 \cdots \phi_n\rangle). \tag{2.22}$$

For $n = 0$ one obtains

$$Z\langle T_{zz}\rangle = \sum_{j=1}^{M_p} h_{zz,}{}^{j} \frac{\partial}{\partial y^j} Z. \tag{2.23}$$

Comparing (2.18) and (2.23) we deduce that

$$V_j(y,\bar{y}) = \frac{\partial}{\partial y^j} \log Z(y,\bar{y}), \tag{2.24}$$

provided that $Z(y,\bar{y}) \neq 0$. Here and in what follows we suppress the σ dependence of Z [see Eq. (2.5)] for notational simplicity and because it is not essential for our purposes. In the points $m \in \mathcal{M}_p$ where Z vanishes, Eq. (2.23) carries no information about $\langle T_{zz}\rangle$. In such points $\langle T_{zz}\rangle$ requires a separate investigation—see Sec. IV for an example of this kind.

Let us now concentrate on Eqs. (2.18) and (2.24). The net result is that T_{zz} develops in general a nontrivial vacuum expectation value $\langle T_{zz}\rangle$, which physically can be interpreted as a sort of Casimir effect due to handles. We stress that in our case $\langle T_{zz}\rangle$ is not related to the central charge and in this sense is not the effect considered in Ref. 5. The phenomenon we have described occurs even when the central charge vanishes. From the point of view of field theory, condensates like $\langle T_{zz}\rangle$ usually parametrize unitarily inequivalent field representations. In our case the field theories corresponding to different complex structures on Σ_p are in general unitarily inequivalent. Expression (2.24) for V_j teaches us that it is exactly $\langle T_{zz}\rangle$ that reflects the fact that a deformation of the complex structure leads in general to an inequivalent two-dimensional field theory. Indeed, if one associates with the deformation (2.19) the "charge" operator[13]

$$Q_j \equiv \int d^2w \sqrt{g} g^{w\bar{w}} \eta^{w}{}_{\bar{w},j}\, T_{ww}, \tag{2.25}$$

from (2.18) one obtains

$$\langle Q_j\rangle = V_j(y,\bar{y}). \tag{2.26}$$

From the above discussion it should be evident that the identity

$$\langle T_{zz}\rangle = \sum_{j=1}^{M_p} h_{zz,}{}^{j} \frac{\partial}{\partial y^j} \log Z(y,\bar{y}) \tag{2.27}$$

is an intrisic property of the system and that there exists no field redefinition

$$T_{zz} \mapsto T'_{zz} = T_{zz} - C_{zz} \tag{2.28}$$

compatible with conformal invariance and such that

$$\langle T'_{zz}\rangle = 0. \tag{2.29}$$

Since this point is crucial for what follows, we now give an independent proof of the above statement (considering for simplicity the case $c = 0$). Before giving the argument, we note that by means of Eq. (2.21), the Ward identity (2.7) can be rewritten in the form

$$\langle T_{zz}\phi_1\cdots\phi_n\rangle - \langle T_{zz}\rangle\langle\phi_1\cdots\phi_n\rangle$$
$$= -\sum_{k=1}^{n}\left[h_k\nabla_{z_k}G^{z_k}{}_{zz}(z_k,z) + G^{z_k}{}_{zz}(z_k,z)\nabla_{z_k}^{(s_k)}\right]\langle\phi_1\cdots\phi_n\rangle$$
$$+\sum_{j=1}^{M_p} h_{zz,}{}^{j}\frac{\partial}{\partial y^j}\langle\phi_1\cdots\phi_n\rangle, \tag{2.30}$$

where the last term (called the Teichmüller term) makes explicit the contribution of deformations. We shall also need the Ward identity with two insertions of T_{zz}: It reads as

$$\langle T_{zz}T_{ww}\phi_1\cdots\phi_n\rangle - \langle T_{zz}\rangle\langle T_{ww}\phi_1\cdots\phi_n\rangle$$
$$= -\frac{c}{24\pi}(\nabla_w)^3 G^{w}{}_{zz}(w,z)\langle\phi_1\cdots\phi_n\rangle - \left[2\nabla_w G^{w}{}_{zz}(w,z) + G^{w}{}_{zz}(w,z)\nabla_w\right]\langle T_{ww}\phi_1\cdots\phi_n\rangle$$
$$-\sum_{k=1}^{n}\left[h_k\nabla_{z_k}G^{z_k}{}_{zz}(z_k,z) + G^{z_k}{}_{zz}(z_k,z)\nabla_{z_k}^{(s_k)}\right]\langle T_{ww}\phi_1\cdots\phi_n\rangle + \sum_{j=1}^{M_p} h_{zz,}{}^{j}\frac{\partial}{\partial y^j}\langle T_{ww}\phi_1\cdots\phi_n\rangle. \tag{2.31}$$

Assume now that T_{zz} and T'_{zz} obey (2.30) and (2.31) and that $\langle T_{zz}\rangle \neq 0$, but $\langle T'_{zz}\rangle = 0$. Applying (2.30) for T_{zz} and T'_{zz} one obtains

$$\langle C_{zz}\phi_1\cdots\phi_n\rangle - \langle C_{zz}\rangle\langle\phi_1\cdots\phi_n\rangle = 0. \tag{2.32}$$

From (2.28) it follows that C_{zz} is a conformal field of weight (2,0). The same is true for T_{zz} provided that $c = 0$. Taking $n = 1$ in (2.32) and substituting $\phi_1 = C_{ww}$ or $\phi_1 = T_{ww}$, in view of (2.28) and (2.29) one obtains

$$\langle C_{zz}C_{ww}\rangle = \langle C_{zz}\rangle\langle C_{ww}\rangle = \langle T_{zz}\rangle\langle T_{ww}\rangle, \tag{2.33}$$
$$\langle C_{zz}T_{ww}\rangle = \langle C_{zz}\rangle\langle T_{ww}\rangle = \langle T_{zz}\rangle\langle T_{ww}\rangle. \tag{2.34}$$

Consider now the identity (2.31) for T'_{zz} and with $n = 0$. Using (2.29) one finds

$$\langle T'_{zz}T'_{ww}\rangle = 0, \tag{2.35}$$

which combined with (2.28), (2.33), and (2.34) implies

$$\langle T_{zz}T_{ww}\rangle - \langle T_{zz}\rangle\langle T_{ww}\rangle = 0. \tag{2.36}$$

Now applying (2.31) for T_{zz} with $n = 0$, one obtains from (2.36) that

$$\left[2\nabla_w G^{w}{}_{zz}(w,z) + G^{w}{}_{zz}(w,z)\nabla_w\right]\langle T_{ww}\rangle$$
$$= \sum_{j=1}^{M_p} h_{zz,}{}^{j}\frac{\partial}{\partial y^j}\langle T_{ww}\rangle, \tag{2.37}$$

which is contradictory because the rhs is holomorphic in z and w, while the lhs is not. On the torus, for example, (2.37) takes the form (we use the notations and conventions of Ref. 9)

$$[\wp(z-w) + 2\eta_1(\tau)]\langle T_{ww}\rangle = -i\pi\frac{\partial}{\partial\tau}\langle T_{ww}\rangle, \tag{2.38}$$

where $\tau = \tau_1 + i\tau_2$ is the Teichmüller parameter, and

$$\eta_1(\tau) = -2i\pi\frac{\partial}{\partial\tau}\log\eta(\tau)$$
$$= (2\pi)^2\left[\frac{1}{24} - \sum_{n=1}^{\infty} n(e^{-2i\pi n\tau} - 1)^{-1}\right], \tag{2.39}$$

where $\wp$ and η are the Weierstrass and Dedekind functions. Equation (2.38) is manifestly inconsistent; $\langle T_{ww}\rangle \neq 0$ by assumption and the lhs has a second-order pole in $z = w$, while the rhs is constant in z.

The content of this section can be summarized as follows. Deformations of the complex structure behave locally as diffeomorphisms.[14] Therefore, the local properties of the fields (short distance leading singularities) do not distinguish between deformations and diffeomorphisms. The difference emerges at the level of vacuum expectation values, where the global properties of the Riemann surface enter through the vacuum state. We have shown in this section that on a genus $p \geqslant 1$ surface, the energy-momentum tensor develops in general a nontrivial vacuum expectation value (2.18) related to the fact that different complex structures correspond to (unitarily) inequivalent field theories. Equations (2.18) and (2.27) follow directly from the conformal Ward identities and there is no field redefinition compatible with conformal invariance and leading to a vanishing $\langle T_{zz}\rangle$.

Analogous statements hold for $T_{\bar{z}\bar{z}}$, which obeys

$$\langle T_{\bar{z}\bar{z}}\rangle = \sum_{j=1}^{M_p} h_{\bar{z}\bar{z},}{}^{j}(\bar{z})\bar{V}_j(y,\bar{y}), \tag{2.40}$$

where $\{h_{\bar{z}\bar{z},}{}^{j}\}$ is a basis of antiholomorphic differentials and

$$\bar{V}_j(y,\bar{y}) \equiv \frac{\partial}{\partial\bar{y}^j}\log Z(y,\bar{y}), \tag{2.41}$$

provided that $Z(y,\bar{y}) \neq 0$.

III. SOME PROPERTIES OF THE VIRASORO ALGEBRA AT NONTRIVIAL GENUS

As already mentioned in Sec. I the correlation functions (2.6) give rise to a representation $\mathscr{I}$ of a Virasoro algebra. In this section we translate Eq. (2.18) in terms of this algebra and derive some properties of the representation $\mathscr{I}$. Take an aribtrary point $P \in \Sigma_p$. Here P belongs to some patch $U \in \mathscr{A}$ and we denote its coordinates by z. We assume that

$$T_{ww}(w) = \sum_{n=-\infty}^{\infty} L_n(z)(w-z)^{-n-2}, \tag{3.1}$$

where $|w-z| = 1$ and the series has convergent matrix elements in $\mathscr{I}$. Equation (3.1) defines the Virasoro generators in P. By means of the conformal Ward indentities (2.30) and (2.31), Eguchi and Ooguri[9] show that

$$[L_m(z), L_n(z)] = (m-n)L_{m+n}(z)$$
$$+ \tfrac{1}{12}c(m^3 - m)\delta_{m+n,0} \tag{3.2}$$

and that the algebra (3.2), originally defined on U, can be consistently extended to a global algebraic structure on Σ_p. In the atlas $\mathcal{A}$ introduced after Eq. (2.18), the transition functions are isometries $z \to f(z)$ and one has

$$L_n(z) \mapsto f'(z)^{-n} L_n(f(z)) + \tfrac{1}{2}(1-n) f''(z) f'(z)^{-n-2} \times L_{n+1}(f(z)) + \cdots, \tag{3.3}$$

where the ellipsis stands for terms proportonal to $L_m(f(z))$ with $m \geqslant n+2$.

Taking the vacuum expectation value of Eq. (3.1) one obtains

$$\langle T_{ww} \rangle = \sum_{n=-\infty}^{\infty} (w-z)^{-n-2} \langle L_n(z) \rangle. \tag{3.4}$$

On the other hand, inserting the expansion

$$h_{ww,}{}^{j}(w) = \sum_{l=0}^{\infty} (w-z)^l h_{zz,}{}^{j}(z)^{(l)} \tag{3.5}$$

into Eq. (2.18) one has

$$\langle T_{ww} \rangle = \sum_{l=0}^{\infty} (w-z)^l \sum_{j=1}^{M_p} h_{zz,}{}^{j}(z)^{(l)} V_j(y, \bar{y}). \tag{3.6}$$

Comparing (3.4) and (3.6) we finally obtain

$$\langle L_n(z) \rangle = 0, \quad \forall n \geqslant -1 \tag{3.7}$$

and

$$\langle L_n(z) \rangle = \sum_{j=1}^{M_p} h_{zz,}{}^{j}(z)^{(-n-2)} V_j(y, \bar{y}), \quad \forall n \leqslant -2. \tag{3.8}$$

Equations (3.7) are common for every genus and are consistent with the requirement of su(1,1) invariance. Before analyzing the infinite set of equations (3.8), which are the novelty for $p \geqslant 1$, we would like to recall some basic facts from the theory of holomorphic quadratic differentials (see, e.g., Ref. 15).

Consider the $M_p \times M_p$ matrix

$$a_{ji}(z) = h_{zz,}{}^{j}(z)^{(i)} \quad (1 \leqslant j \leqslant M_p, 0 \leqslant i \leqslant M_p - 1). \tag{3.9}$$

The determinant

$$W(P) = \det a_{ji} \tag{3.10}$$

is known as the Wronskian of $h_{zz,}{}^{j}$ in the point P. Following the terminology of Ref. 15, P is called a Weierstrass point if $W(P) = 0$. One can show that there are no Weierstrass points on the torus. A point P is called generic if $W(P) \neq 0$. Note that this characterization of the points on Σ_p is independent of the choice of local coordinates because the vanishing of $W(P)$ has this property.

Returning to Eq. (3.8), we assume first that P is a generic point. Since $W(P) \neq 0$, the vacuum expectation values $\langle L_{-n}(z) \rangle$ vanish for $2 \leqslant n \leqslant M_p + 1$ if and only if $V_j(y, \bar{y}) = 0$ for all $1 \leqslant j \leqslant M_p$. In other words, if some $V_j \neq 0$, then there exist some integers s, $2 \leqslant s \leqslant M_p + 1$ such that

$$\langle L_{-s}(z) \rangle \neq 0 \tag{3.11}$$

in a generic point on Σ_p. We emphasize that the above statement is independent of the choice of local coordinates. On the contrary, apart from $s_{\min}$, the values of s for which (3.12) takes place are in general coordinate dependent [see Eq. (3.3)].

Consider now the case when P is a Weierstrass point. Then $W(P) = 0$, but one can show[15] that taking a sufficiently small $n \leqslant -2$ in (3.8), there exists an $M_p \times M_p$ submatrix in the rhs of (3.8) with a nonvanishing determinant. Therefore, the previous conclusions remain true.

Summarizing the above considerations, we have proven the following proposition. Let $\{V_1, \ldots, V_{M_p}\} \neq 0$. Then for every point $P \in \Sigma_p$ (independent of the choice of coordinates) there exist some integers $s \geqslant 2$ such that (3.11) holds.

Equation (3.11) has some important consequences on the representation of the Virasoro algebra associated with the system under consideration. Denote by $\mathcal{J}$ the representation space of (3.2) equipped with a sesquilinear form (inner product) $\langle \cdot, \cdot \rangle$. Then there are two possibilities. Assume first that there exists a unique vacuum state $\Omega \in \mathcal{J}$ such that

$$\langle \Omega, L_n(z) \Omega \rangle = \langle L_n(z) \rangle. \tag{3.12}$$

Taking the complex conjugate of $\langle \Omega, L_s(z) \Omega \rangle$ and using (3.12), (3.11), and (3.7) we deduce that

$$L_{-s}(z) \neq L_s^{\dagger}(z), \tag{3.13}$$

where the dagger represents the Hermitian conjugation with respect to $\langle \cdot, \cdot \rangle$. This means that one of the basic unitarity requirements is violated at least in the vacuum sector $\mathcal{J}_{\text{vac}} \subset \mathcal{J}$.

Consider now Eq. (3.2) with $m = 0$ and $n = -s$ and take its vacuum expectation value. One obtains

$$\langle \Omega, [L_0(z), L_{-s}(z)] \Omega \rangle = s \langle \Omega, L_{-s}(z) \Omega \rangle \neq 0. \tag{3.14}$$

On the other hand, from rigid scale invariance,

$$L_0(z) \Omega = 0 \tag{3.15}$$

and (3.14) takes the form

$$\langle \Omega, L_0(z) L_{-s}(z) \Omega \rangle \neq 0. \tag{3.16}$$

Equations (3.15) and (3.16) imply that

$$L_0(z) \neq L_0^{\dagger}(z). \tag{3.17}$$

A similar argument shows that (3.13) also extends for $s = 1$. It is also worth stressing that different from the $p = 0$ case, not all of the states $\{L_{-n}(z) \Omega : n \geqslant 2\}$ decouple from the vacuum.

The second possibility, explored usually in the operator approach to string theories,[16] is to preserve the Hermiticity property

$$L_{-n}(z) = L_n^{\dagger}(z) \tag{3.18}$$

for all n, but to admit the existence of (at least) two vacuums Ω and $\tilde{\Omega}$ such that

$$\langle \tilde{\Omega}, L_n(z) \Omega \rangle = \langle L_n(z) \rangle. \tag{3.19}$$

Now Eq. (3.11) implies that, in general, at genus $p \geqslant 1$, there exist $s \geqslant 2$ such that

$$L_s(z) \tilde{\Omega} \neq 0 \tag{3.20}$$

and the states $\{L_s(z) \tilde{\Omega}\}$ couple to the vacuum Ω. Therefore, also, in this case $\mathcal{J}_{\text{vac}}$ is not an irreducible highest weight representation in the sense of Ref. 4.

We have shown above that for $p \geqslant 1$ the vacuum sector $\mathcal{J}_{\text{vac}}$ has in general a richer structure than a standard $h = 0$ irreducible highest weight representation; some of the descendent states of the vacuum couple to the vacuum itself. Concerning the construction of the physical state space in

string theory, we recall that there one has to integrate over the compactified moduli space $\bar{\mathscr{M}}_p$. After this integration the decoupling of the aforementioned states is recovered[17] modulo total derivatives. The latter are not *a priori* always innocuous because for $p \geqslant 2$ the space $\bar{\mathscr{M}}_p$ has a nontrivial boundary.

The above pathologies in the representation $\mathscr{F}_{\text{vac}}$, related to a conformal field theory on Σ_p, occur in the points of the moduli space where $\{V_1,...,V_{M_p}\} \neq 0$; the safe points are those in which

$$Z(y,\bar{y}) \neq 0 \tag{3.21}$$

and

$$\frac{\partial}{\partial y^j} Z(y,\bar{y}) = 0, \quad \forall j = 1,...,M_p. \tag{3.22}$$

The set on which Z vanishes has to be examined separately. Assuming that (3.21) holds, we concentrate on Eq. (3.22). This is a local condition in $\mathscr{M}_p$ and consequently, can be formulated on the Teichmüller space

$$\mathscr{T}_p \equiv \text{Met}(\Sigma_p)/\text{Diff}_0(\Sigma_p) \otimes \mathscr{W}, \tag{3.23}$$

where $\text{Diff}_0(\Sigma_p)$ is the subgroup of $\text{Diff}(\Sigma_p)$ connected to the identity. Denote by $\{\tau^j : j = 1,...,M_p\}$ the complex coordinates on $\mathscr{T}_p$. Then Eq. (3.22) is equivalent to

$$\frac{\partial}{\partial \tau^j} Z(\tau,\bar{\tau}) = 0, \quad \forall j = 1,...,M_p, \tag{3.24}$$

where $Z(\tau,\bar{\tau})$ is modular invariant. Consider now a modular transformation

$$\tau \mapsto F(\tau), \tag{3.25}$$

which admits a fixed point $\tau' \in \mathscr{T}_p$ such that

$$\det\left[\frac{\partial}{\partial \tau^j} F^i(\tau') - \delta^i_j\right] \neq 0. \tag{3.26}$$

Then as a consequence of modular invariance,

$$\frac{\partial}{\partial \tau^j} Z(\tau',\bar{\tau}') = 0, \quad \forall j = 1,...,M_p \tag{3.27}$$

and $\langle T_{zz} \rangle$ vanishes on the orbifold point $y' \in \mathscr{M}_p$ corresponding to $\tau' \in \mathscr{T}_p$. We shall see in Sec. IV that the two orbifold points of $\mathscr{M}_1$ satisfy (3.26). We also note that condition (3.26) is universal (model independent), but we are not aware if it holds for all orbifold points of $\mathscr{M}_p$ with $p \geqslant 2$. The tempting conjecture that this is the case needs a separate investigation.

In conclusion, we observe that all the statements of this section hold *mutatis mutandis* for the generators $\bar{L}_n(\bar{z})$ defined by

$$T_{\bar{w}\bar{w}}(\bar{w}) = \sum_{n=-\infty}^{\infty} \bar{L}_n(\bar{z})(\bar{w}-\bar{z})^{-n-2}. \tag{3.28}$$

Using Eq. (2.40) one finds the counterpart of (3.8), which states that for any $n \leqslant -2$ one has

$$\langle \bar{L}_n(\bar{z}) \rangle = \sum_{j=1}^{M_p} h_{\bar{z}\bar{z}}{}^j(\bar{z})^{(-n-2)} \bar{V}_j(y,\bar{y}). \tag{3.29}$$

IV. EXAMPLES ON THE TORUS

We illustrate below the general results of the previous sections on some explicit examples. For simplicity we consider models defined on the torus equipped with the metric $ds^2 = |dz|^2$, where $z = x_1 + \tau x_2$. The different complex structures are described by the Teichmüller parameter $\tau = \tau_1 + i\tau_2$. The fundamental domain $\mathscr{F}_1$ of τ is fixed to be

$$\mathscr{F}_1 \equiv \{-\tfrac{1}{2} \leqslant \tau_1 \leqslant 0, |\tau| \geqslant 1\} \cup \{0 < \tau_1 < \tfrac{1}{2}, |\tau| > 1\}.$$

Here $\mathscr{F}_1$ is a covering of $\mathscr{M}_1$ and τ is a well-defined local coordinate on $\mathscr{M}_1$. With these conventions the Beltrami and quadratic differentials are given by

$$\eta^z{}_{\bar{z}} = i/2\tau_2, \quad h_{zz} = -i, \tag{4.1}$$

while the explicit form of $G^w{}_{zz}(w,z)$ is

$$G^w{}_{zz}(w,z) = \frac{\vartheta_1'(z-w|\tau)}{2\pi\vartheta_1(w-z|\tau)} + \frac{w-z-\bar{w}+\bar{z}}{2\tau_2}. \tag{4.2}$$

Here ϑ_1 is the Jacobi theta function

$$\vartheta_1(z|\tau) = i \sum_{n=-\infty}^{\infty} (-1)^n \exp\left[i\pi(2n-1)z + i\pi\tau\left(n-\frac{1}{2}\right)^2\right] \tag{4.3}$$

and the prime means the derivative with respect to the first argument.

The simplest, but quite instructive, example is the free scalar field with the modular invariant two-point function

$$\langle \varphi(z,\bar{z})\varphi(w,\bar{w}) \rangle = -\frac{1}{2\pi} \log\left|\frac{\vartheta_1(z-w|\tau)}{\eta(\tau)}\right| + \frac{(z_2-w_2)^2}{2\tau_2}. \tag{4.4}$$

The relative partition function is also modular invariant and reads as

$$Z_B(\tau,\bar{\tau}) = 1/\sqrt{\tau_2}\eta(\tau)\,\overline{\eta(\tau)}, \tag{4.5}$$

where

$$\eta(\tau) = \exp\left(\frac{i\pi\tau}{12}\right) \prod_{n=1}^{\infty} [1 - \exp(2in\pi\tau)]. \tag{4.6}$$

From the explicit form of η it follows that $Z_B(\tau,\bar{\tau})$ has no zeros in $\mathscr{F}_1$. Therefore, Eq. (2.24) applies and

$$h_{zz} V(\tau,\bar{\tau}) = -i\frac{\partial}{\partial\tau} \log Z_B(\tau,\bar{\tau}) = \frac{1}{4\tau_2} - \frac{1}{2\pi}\eta_1(\tau), \tag{4.7}$$

with η_1 defined by (2.39).

Let us first check identity (2.27). Taking into account (4.1) and (4.7), Eq. (2.27) is indeed satisfied because

$$\langle T_{zz} \rangle = \lim_{w\to z}\left[\partial_z\partial_w \langle \varphi(z,\bar{z})\varphi(w,\bar{w}) \rangle + \frac{1}{4\pi}(z-w)^{-2}\right] = \frac{1}{4\tau_2} - \frac{1}{2\pi}\eta_1(\tau). \tag{4.8}$$

As already mentioned, the torus has no Weierstass points and from (3.8) and (4.7) one obtains, for the left Virasoro algebra,

$$\langle L_{-2}(z) \rangle = 1/4\tau_2 - (1/2\pi)\eta_1(\tau), \tag{4.9}$$

where the first term on the rhs keeps trace of the zero modes. Analogously, for the right Virasoro algebra one obtains

$$\langle \bar{L}_{-2}(\bar{z}) \rangle = 1/4\tau_2 - (1/2\pi)\eta_1(-\bar{\tau}). \tag{4.10}$$

Consider, also, the vertex operator

$$\Phi = \int d^2z\sqrt{g}:\exp[4i\sqrt{\pi}\varphi(z,\bar{z})]:,$$

which is the integral of a conformal field of weight (1,1). By means of (2.30) one easily obtains

$$\langle T_{zz}\Phi\cdots\Phi\rangle = -\frac{i}{Z_B}\frac{\partial}{\partial\tau}(Z_B\langle\Phi\cdots\Phi\rangle). \tag{4.11}$$

Equation (4.11) is a particular case of the general relation

$$\langle T_{zz}\Phi\cdots\Phi\rangle = \frac{1}{Z}\sum_{j=1}^{M_p} h_{zz,}{}^{j}\frac{\partial}{\partial y^j}(Z\langle\Phi\cdots\Phi\rangle), \tag{4.12}$$

valid for any $p\geqslant 1$.

We focus now on Eqs. (4.9)–(4.11). Denote by $\mathscr{I}_{L(R)}$ the representation of the left (right) Virasoro algebra as introduced by Eq. (3.19). Then because of (4.9) and (4.10), the vacuum sector $\mathscr{I}_L\otimes\mathscr{I}_R$ has all the peculiar properties described in Sec. III. Moreover, according to Eq. (4.11), the new states arising in the vacuum sector couple not only to the vacuum, but also to the sector with conformal weight (1,1).

Let us look now for points in $\mathscr{M}_1$, where the rhs of Eqs. (4.9)–(4.11) vanish. As is well known, the modular group $SL(2,Z)/Z_2$ acts on $\mathscr{T}_1$ with fixed points. Indeed, $\tau' = i$ and $\tau'' = \frac{1}{2}(-1+i\sqrt{3})$ are fixed points for the modular transformations

$$S:\tau\mapsto -1/\tau, \tag{4.13}$$

$$T^{-1}\circ S:\tau\mapsto -1-1/\tau, \tag{4.14}$$

respectively, and correspond to the two orbifold points of $\mathscr{M}_1$. Moreover,

$$\frac{\partial}{\partial\tau}S(\tau')\neq 1,\quad \frac{\partial}{\partial\tau}[T^{-1}\circ S](\tau'')\neq 1.$$

Combining this information with the fact that the expectation values $\langle L_{-2}(z)\rangle$, $\langle\bar{L}_{-2}(\bar{z})\rangle$, and $\langle T_{zz}\Phi\cdots\Phi\rangle$ are derivatives of modular invariant functions, we conclude that the rhs of (4.9)–(4.11) vanish on the orbifold points of $\mathscr{M}_1$. From the explicit form of $\eta_1(\tau)$ [see Eq. (2.39)], it is not difficult to see that there are no other points on $\mathscr{M}_1$ where $\langle L_{-2}(z)\rangle = \langle\bar{L}_{-2}(\bar{z})\rangle = 0$. The point τ' describes a torus with equal radii and without a Dehn twist. It is also worth mentioning that Schroer's investigation[18] of positivity on the torus shows that in τ' positivity is also satisfied.

A brief comment concerning finite temperature two-dimensional field theories is in order. As is well known,[5,19] the theory of a free boson field on the torus is related to the theory of a free boson field on the cylinder at the *finite* (complex) temperature $\beta = -2i\pi\tau$. Indeed, the partition function (4.5) can also be written as

$$Z_B(\tau,\bar{\tau}) = \mathrm{Tr}\, q^{K_0-1/24}\bar{q}^{\bar{K}_0-1/24}, \tag{4.15}$$

where

$$q = e^{2i\pi\tau}.$$

The trace in the rhs of (4.15) is carried out over the state space of the model on the cylinder and consistently K_0 is the energy operator of the Virasoro algebra $\{K_n\}$ associated to this model. From (4.15) it follows that

$$\langle K_0\rangle_\tau = \tfrac{1}{24} + 1/8\pi\tau_2 - (1/4\pi^2)\eta_1(\tau). \tag{4.16}$$

Note that at zero temperature ($\beta=\infty$) the rhs of (4.16) vanishes and scale invariance is restored, as it should be. Inserting (4.9) into (4.16) one finds

$$\langle K_0\rangle_\tau = \tfrac{1}{24} + (1/2\pi)\langle L_{-2}(z)\rangle. \tag{4.17}$$

The presence of two different moments of the energy-momentum tensor in Eq. (4.17) is not contradictory because (4.17) relates expectation values computed in two different field theories. In fact, $\langle K_0\rangle_\tau\neq 0$ signals the breaking of scale invariance at finite temperature for the boson field on the cylinder. On the contrary, the free boson field on the torus is scale invariant and $\langle L_{-2}(z)\rangle\neq 0$ implies noninvariance under deformations of the complex structure.

Several models on the torus share the properties of the free boson field, as described in this section. Among them we mention the toroidal compactification of a free boson field:

$$\tilde{Z}_B(\tau,\bar{\tau}) = Z_B(\tau,\bar{\tau})\sum_{m,n=-\infty}^{\infty}\exp\left[-\frac{\pi}{\tau_2}(m^2+n^2|\tau|^2-2mn\tau_1)\right], \tag{4.18}$$

the ghost $b-c$ system:

$$Z_{gh}(\tau,\bar{\tau}) = \tau_2|\eta(\tau)|^4, \tag{4.19}$$

and the critical Ising model:

$$Z_I(\tau,\bar{\tau}) = (1/2|\eta(\tau)|)(|\vartheta_2(0|\tau)| + |\vartheta_3(0|\tau)| + |\vartheta_4(0|\tau)|), \tag{4.20}$$

where ϑ_ν are the usual theta functions.[15] From (4.5) and (4.19) one recovers the modular invariant string partition function density

$$Z_{str}(\tau,\bar{\tau}) = Z_B(\tau,\bar{\tau})^{26}Z_{gh}(\tau,\bar{\tau}) = [\sqrt{\tau_2}\eta(\tau)\,\overline{\eta(\tau)}]^{-24}, \tag{4.21}$$

which gives rise to the partition function

$$Z_{str} = \int_{\mathscr{T}_1}\frac{d\tau\,d\bar{\tau}}{\tau_2^2}Z_{str}(\tau,\bar{\tau}). \tag{4.22}$$

In the examples we have considered until now, the partition function was nonvanishing at any point of $\mathscr{T}_1$ and we applied Eqs. (2.24) and (2.41). If the partition function vanishes in some points or on the whole $\mathscr{T}_1$, one may compute $V(\tau,\bar{\tau})$ and $\bar{V}(\tau,\bar{\tau})$ directly from the condensates $\langle T_{zz}\rangle$ and $\langle T_{\bar{z}\bar{z}}\rangle$. Consider, for example, the periodic–periodic free spinor field on Σ_1. As is well known, the relative partition function vanishes identically on $\mathscr{T}_1$. Nevertheless, $\langle T_{zz}\rangle$ and $\langle T_{\bar{z}\bar{z}}\rangle$ are nonvanishing and the direct computation[7] shows that, for example,

$$\langle T_{zz}\rangle = \lim_{w\to z}\left[\frac{i}{2}\langle\psi(z,\bar{z})\partial_w\,\psi(w,\bar{w})\rangle - \frac{1}{4\pi}(z-w)^{-2}\right] = \frac{1}{2\pi}\eta_1(\tau) - \frac{1}{4\tau_2}. \tag{4.23}$$

Consequently, the discussion following Eq. (4.8) applies.

V. CONCLUSIONS

In the present paper we have studied the role of complex structure deformations in conformal field theory on a Riemann surface of nontrivial genus. By means of the conformal Ward identities we have shown the following.

(i) The energy-momentum tensor T_{zz} develops in gen-

eral a nontrivial vacuum expectation value $\langle T_{zz} \rangle$. Maintaining conformal invariance, $\langle T_{zz} \rangle$ cannot be absorbed in a redefinition of T_{zz}.

(ii) Deformations lead in general to unitarily inequivalent field theories.

(iii) The representation of the Virasoro algebra, associated to the conformal field theory, is in general not of the familiar form [see Eq. (1.1)] known for $p = 0$.

(iv) Under certain conditions, including modular invariance, $\langle T_{zz} \rangle$ vanishes on the orbifold points of $\mathcal{M}_p$.

We illustrate the above phenomena on the torus. Throughout our investigation we need only local properties of the moduli space, although we use several global properties of the underlying Riemann surface. It is worth stressing that our results apply for any (not necessarily Gaussian) conformal field theory. We also mention that an alternative analysis of the problems we were concerned with in this paper can be performed[20] by means of the Krichever–Novikov algebra.[8] As one should expect, there is complete agreement.

In conclusion, we point out that our investigation has some implications for string theory, where by construction one integrates over all complex structures. In this sense any string theory is a superposition of conformal field theories with different complex structures. According to (ii) the individual field theories are in general not unitarily equivalent and from point (iii) it follows that not all provide standard unitary highest weight representations of the Virasoro algebra. Nevertheless, one expects to recover unitarity after integration over the moduli space.[17] Thus the global structure of $\bar{\mathcal{M}}_p$, as well as the behavior of the partition and correlation functions on the boundary $\partial\bar{\mathcal{M}}_p$, enter the game.[21] In this context the statistical mechanics of conformal invariant systems on a Riemann surface may also require an integration over the moduli space. These problems deserve further investigations.

ACKNOWLEDGMENTS

We wish to thank A. Bondi, L. Bonora, A. Cappelli, and K. Viswanathan for useful discussions.

[1] I. M. Gel'fand and D. B. Fuks, Funk. Anal. Pril. **2**(4), 92 (1968); M. A. Virasoro, Phys. Rev. D **1**, 2933 (1970); S. Mandelstam, Phys. Rep. **13**, 259 (1974); P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 303 (1986).

[2] S. Ferrara, R. Gatto, and A. F. Grillo, Nuovo Cimento A **12**, 959 (1972); S. Fubini, A. J. Hanson, and R. Jackiw, Phys. Rev. D **7**, 1732 (1973).

[3] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U. P., Cambridge, 1987), Vols. 1 and 2 and references therein.

[4] A. Belavin, A. Polyakov, and A. Zamolodchikov, Nucl. Phys. B **241**, 333 (1984); D. Friedan, Z. Qiu, and S. Shenker, Phys. Rev. Lett. **52**, 1375 (1984); P. Goddard, A. Kent, and D. Olive, Phys. Lett. B **152**, 88 (1985); P. Goddard, A. Kent, and D. Olive, Commun. Math. Phys. **103**, 105 (1986); A. Cappelli, C. Itzykson, and J. B. Zuber, Nucl. Phys. B **280**, 445 (1987).

[5] J. L. Cardy, Nucl. Phys. B **270**, 186 (1986); A. W. W. Ludwig and J. Cardy, *ibid.* **285**, 687 (1987); C. Itzykson and J. B. Zuber, *ibid.* **275**, 580 (1986).

[6] V. S. Dotsenko and V. A. Fateev, Nucl. Phys. B **240**, 312 (1984); **235**, 54 (1984); V. G. Knizhnik and A. B. Zamolodchikov, *ibid.* **247**, 105 (1984).

[7] P. Di Francesco, H. Saleur, and J. B. Zuber, Nucl. Phys. B **290**, 527 (1987).

[8] A. A. Beilinson, Yu. I. Manin, and V. V. Schechtman, *Sheaves of the Virasoro and Neveu–Schwarz Algebras*, in Springer Lecture Notes in Mathematics, Vol. 1289, edited by Yu. Manin (Springer, New York, 1987); L. Alvarez-Gaumé, C. Gomez, G. Moore, and C. Vafa, Nucl. Phys. B **303**, 455 (1988); E. Arbarello, C. De Concini, V. Kac, and C. Procesi, Commun. Math. Phys. **117**, 1 (1988); I. M. Krichever and S. P. Novikov, Funk. Anal. Pril. **21**(2), 46, 21 (1987); Funk. Anal. Pril. **21**(4), 47 (1987); L. Bonora, M. Bregola, P. Cotta-Ramusino, and M. Martellini, Phys. Lett. B **205**, 53 (1988); L. Bonora, A. Lugo, M. Matone, and J. Russo, *A Global Operator Formalism on Higher Genus Riemann Surfaces I: b − c Systems*, SISSA preprint, Trieste 1988.

[9] T. Eguchi and H. Ooguri, Nucl. Phys. B **282**, 308 (1987).

[10] A. A. Belavin and V. G. Knizhnik, Sov. Phys. JETP **64**, 214 (1986); P. Nelson, Phys. Rep. **149**, 337 (1987).

[11] N. S. Hawley and M. Schiffer, Acta Math. **115**, 199 (1966).

[12] L. Bers, Bull. Am. Math. Soc. **5**, 131 (1981).

[13] E. Martinec, Nucl. Phys. B **281**, 157 (1987); D. Friedan and S. Shenker, *ibid.* **281**, 509 (1987).

[14] L. Alfors, *Lectures on Quasiconformal Mappings* (Van Nostrand, New York, 1966).

[15] H. Farkas and I. Kra, *Riemann and Surfaces* (Springer, New York, 1980).

[16] C. Vafa, Phys. Lett. B **190**, 47 (1986); L. Alvarez-Gaumé, C. Gomez, and C. Reina, Phys. Lett. B **190**, 55 (1986); L. Alvarez-Gaumé, C. Gomez, and C. Reina, *New Methods in String Theory*, in Proceedings of the Trieste Spring School on Superstrings, CERN preprint CERN-TH 4775/87; N. Ishibashi, Y. Matsuo, and H. Ooguri, Mod. Phys. Lett. A **2**, 119 (1987); Y. Matsuo, Mod. Phys. Lett. A **3**, 1 (1987); R. Dijkgraff, E. Verlinde, and H. Verlinde, Commun. Math. Phys. **115**, 649 (1988).

[17] E. Guadagnini, M. Martellini, and M. Mintchev, in preparation.

[18] B. Schroer, Phys. Lett. B **199**, 183 (1987).

[19] D. Buchholz, G. Mack, and I. T. Todorov, *Finite Temperature 2-Dimensional QFT Models of Conformal Current Algebra*, Lecture at the International School on Group Theoretical Methods, Varna, June 1987.

[20] P. Cotta-Ramusino, M. Martellini, and M. Mintchev, Phys. Lett. B **215**, 331 (1988).

[21] P. Mansfield, Nucl. Phys. B **238**, 551 (1987); H. Sonoda, *ibid.* **281**, 546 (1987); J. Atick, G. Moore, and A. Sen, *Some Global Issues in String Perturbation Theory*, SLAC preprint SLAC-PUB-4463.

A comparison of solution generating techniques for the self-dual Yang–Mills equations. II

Jacek Tafel[a)]
Centre de Recherches Mathématiques, Université de Montréal, C.P.6128, succ. A, Montréal, Quebec H3C 3J7, Canada

(Received 23 May 1989; accepted for publication 16 August 1989)

It is shown that Bäcklund transformations related to the Yang formulation of the self-dual Yang–Mills equations reduce to Zakharov–Shabat transformations when the gauge group is GL(2,C).

I. INTRODUCTION

In Ref. 1 we discussed the relations between the Zakharov–Shabat transformation (ZST) method[2] and the Atiyah–Ward method[3] of solving the self-dual Yang–Mills (SDYM) quations for $G = \mathrm{GL}(2,C)$. Our main result was that all solutions generated by the Atiyah–Ward Ansätze can be obtained from the simplest Ansatz $\mathscr{A}_0$ by means of ZST's. In this paper we investigate some other approaches to the SDYM equations, namely Bäcklund transformations (BT's)[4–6] connected with the Yang equations.[7] We show that for $G = \mathrm{GL}(2,C)$ they reduce to some ZST's (which are also a kind of BT).

This paper is continuation of Ref. 1. We adopt the same formalism and notation. We assume the gauge group to be GL(2,C) and we consider the SDYM equations in the complexified Minkowski space CM,

$$F_{A\dot{C}B}{}^{\dot{C}} = 0, \tag{1}$$

where F is the curvature of a gauge field A and $A,B,\dot{C}$ are spinor indices.

Equations (1) are integrability conditions of the following system of linear equations,[8,9]

$$\lambda^A(\partial_{A\dot{B}} + A_{A\dot{B}})\psi = 0, \tag{2}$$

where $\lambda^A = (1,\lambda)$, λ being a complex parameter. The Zakharov–Shabat method is based on the observation that the Ansatz

$$\psi' = \chi\psi, \tag{3}$$

where χ is a simple meromorphic function of λ (each of χ and χ^{-1} is assumed to have exactly one first order pole in λ), leads to new self-dual solutions from a given one (see Ref. 1 for details).

When referring to an equation in Ref. 1 we will use the notation (1.a), where (a) is the equation number in Ref. 1.

II. BÄCKLUND TRANSFORMATIONS OF POHLMEYER AND PRASAD *et al.*

Yang[7] noticed that Eqs. (1) are equivalent to a single second-order equation for a matrix valued function J. Indeed, it follows from (1), for $A = B$, that

$$A_{1\dot{B}} = J^{-1}\partial_{1\dot{B}}J, \tag{4}$$

$$A_{2\dot{B}} = 0 \tag{5}$$

in some gauge. In virtue of (4) and (5), Eqs. (1) reduce to the Yang equation

$$\partial_2^{\dot{B}}(J^{-1}\partial_{1\dot{B}}J) = 0. \tag{6}$$

This equation admits BT's (for any gauge group) of the form

$$J'^{-1}\partial_{1\dot{B}}J' = J^{-1}\partial_{1\dot{B}}J - \partial_{2\dot{B}}(e^{\alpha}J^{-1}J') - \partial_{1\dot{B}}\alpha \tag{7}$$

and

$$J'^{-1}\partial_{1\dot{B}}J' = J^{-1}\partial_{1\dot{B}}J - \partial_{2\dot{B}}(e^{\beta}J'^{-1}J) - \partial_{1\dot{B}}\beta, \tag{8}$$

where α and β are harmonic functions,

$$\partial_1^{\dot{B}}\partial_{2\dot{B}}\alpha = \partial_1^{\dot{B}}\partial_{2\dot{B}}\beta = 0.$$

If J satisfies (6), then (7) [or (8)] defines a function J', which also satisfies Eq. (6). Hence J' describes a new self-dual field A'. The BT's (7) and (8) are slight generalizations of the transformations found by Pohlmeyer[5] and Prasad *et al.*[6]

It turns out that relations (7) and (8) are satisfied if J' corresponds to a field generated from potentials (4) and (5) by ZST with

$$\chi = 1 + [(\mu - \nu)/(\lambda - \mu)]P, \quad P^2 = P, \quad \mu,\nu \neq 0 \tag{9}$$

or

$$\chi = 1 + [1/(\lambda - \mu)]S, \quad S^2 = 0, \quad \mu \neq 0. \tag{10}$$

To show this let us consider a relation between the functions J and ψ. It follows from (2), (4), and (5) that we can identify

$$J = \psi(0)^{-1}, \tag{11}$$

provided $\psi(\lambda)$ is holomorphic at $\lambda = 0$. Transformation (9) preserves gauge condition (5) [see (1.15)] and does not generate a pole of ψ' at $\lambda = 0$, hence

$$J' = \psi'(0)^{-1} = J[1 + (\mu\nu^{-1} - 1)P]. \tag{12}$$

It follows from (12) that

$$(\mu - \nu)P = \nu(J^{-1}J' - 1) \tag{13}$$

and

$$(\mu - \nu)P = -\mu(J'^{-1}J - 1). \tag{14}$$

Substituting (4), (5), and (13) [or (14)] into the relation

$$A'_{1\dot{B}} = A_{1\dot{B}} - \partial_{2\dot{B}}[(\mu - \nu)P], \tag{15}$$

which follows from (2), (3), and (9) [compare (1.15)], yields formula (7) with $\alpha = \ln\nu$ [or (18) with $\beta = \ln\mu$]. In a similar way we can show that (7) and (8) are satisfied when J' is generated by transformation (10).

The occurence of α and β in (7) and (8) gives the equa-

[a)] On leave of absence from Institute of Theoretical Physics, University of Warsaw, Hoża 69, 00681, Warsaw, Poland.

 0022-2488/90/051234-03$03.00

tions a richness, but it is only an apparent one, for we can always decompose (7) and (8) into the transformations

$$J'^{-1}\partial_{1\dot B}J' = J^{-1}\partial_{1\dot B}J - \partial'_{2\dot B}(J^{-1}J'), \tag{16}$$

$$J'^{-1}\partial_{1\dot B}J' = J^{-1}\partial_{1\dot B}J + \partial_{2\dot B}(J'^{-1}J), \tag{17}$$

respectively, and the trivial transformation

$$J' \to e^h J', \quad \partial_1^{\dot B}\partial_{2\dot B}h = 0, \tag{18}$$

with $h = \alpha$ or $h = \beta$, respectively. Transformation (18) corresponds to adding a solution to A' with values in the center of $gl(2,C)$. Transformation (17) becomes (16) when J and J' are interchanged, hence (17) can be considered as an inverse transformation to (16).

The gauge field A' generated by (16) can always be obtained by ZST (9) or (10) combined with (18). To show this let us note that $J^{-1}J'$, a 2×2 matrix, can be written in the form

$$J^{-1}J' = \nu + (\mu - \nu)P, \text{ where } P^2 = P, \tag{19}$$

or

$$J^{-1}J' = \mu + S, \text{ where } S^2 = 0. \tag{20}$$

We substitute J' calculated from (19) or (20) into (16). In this way we obtain equations for μ,ν and P (or S), which coincide with conditions (1.16)–(1.20) [or their limits under $\nu\to\mu$, $(\mu-\nu)P\to S$] on free parameters of ZST (9) [or (10)]. To complete the proof we note that (19) [or (20)] is equivalent to (13) (or to its limit) modulo transformation (18).

Equation (19) [or (20)] is also equivalent to Eq. (14) (or to its limit) modulo transformation (18). Hence, if J' satisfies (16) then $e^h J'$ satisfies (17) for some harmonic function h.

We summarize this section with the following proposition.

Proposition 1: For the gauge group $GL(2,C)$ the following transformations generate the same set of self-dual solutions from a given solution: (i) the Bäcklund transformations of type (7) and (8); (ii) the Bäcklund transformation (16) combined with (18); (iii) the Zakharov–Shabat transformation of type (9) and (10) combined with (18).

Remark: If the gauge group is not a subgroup of $GL(2,C)$ then transformations (i) may generate more solutions than transformations (iii).

III. BÄCKLUND TRANSFORMATION OF CORRIGAN *et al.*

Historically, the first BT for Eq. (6) was given by Corrigan *et al.*[4] (transformation β). To describe it we assume $\det J = 1$ [i.e., the gauge group is $SL(2,C)$ rather than $[GL(2,C)]$ and we write J in the form

$$J = \varphi^{-1}\begin{bmatrix} 1 & \gamma \\ -\epsilon & \varphi^2 - \epsilon\gamma \end{bmatrix}. \tag{21}$$

The BT is then given by the equations

$$\begin{aligned} &\varphi' = \varphi^{-1}, \\ &\partial_{1\dot B}\epsilon' = \varphi^{-2}\partial_{2\dot B}\gamma, \\ &\partial_{2\dot B}\gamma' = \varphi^{-2}\partial_{1\dot B}\epsilon. \end{aligned} \tag{22}$$

If J satisfies (6) then solutions $\varphi',\epsilon',\gamma'$ of Eqs. (22)–(24) exist and define a new solution J' of (6). Given a self-dual field the corresponding function J is not uniquely defined. It undergoes the following "gauge" transformations:

$$J \to hJg, \tag{23}$$

where

$$\partial_{1\dot B}h = \partial_{2\dot B}g = 0.$$

These transformations in general do not commute with BT (22).

The action of transformation (22) on self-dual fields can also be described in the following way. Given a self-dual field A we find a gauge (not unique) such that

$$A_{1\dot B} = \begin{Bmatrix} a_{\dot B} & 0 \\ b_{\dot B} & -a_{\dot B} \end{Bmatrix}, \quad A_{2\dot B} = \begin{bmatrix} c_{\dot B} & d_{\dot B} \\ 0 & -c_{\dot B} \end{bmatrix}. \tag{24}$$

(For instance, we first find J and decompose it into $D\tilde D^{-1}$, where D is lower triangular and $\tilde D$ is upper triangular. Then we perform a gauge transformation to obtain $A_{1\dot B} = D^{-1}\partial_{1\dot B}D$ and $A_{2\dot B} = \tilde D^{-1}\partial_{2\dot B}\tilde D$.) Having (24) we define a new field A' by

$$A'_{1\dot B} = \begin{bmatrix} -a_{\dot B} & 0 \\ d_{\dot B} & a_{\dot B} \end{bmatrix}, \quad A'_{2\dot B} = \begin{bmatrix} -c_{\dot B} & b_{\dot B} \\ 0 & c_{\dot B} \end{bmatrix}. \tag{25}$$

It is easy to show that ψ' given by

$$\psi' = \chi\psi, \quad \chi = \begin{bmatrix} 0 & \lambda \\ 1 & 0 \end{bmatrix}, \tag{26}$$

where ψ corresponds to (24), satisfies Eqs. (2) with potentials (25). Hence the field A' is necessarily self-dual.

The dressing function χ, given by (26), can be written in the form

$$\chi = g[1 + (\lambda - 1)P], \quad P^2 = P, \tag{27}$$

where g and P do not depend on λ,

$$g = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Formula (27) is characteristic for a particular class of ZST's [(1.12) with $\nu = 0$]. One can easily check that the field (25) can, indeed, be generated from A using the method of Zakharov and Shabat.

Proposition 2: The Bäcklund transformation of Corrigan *et al.* is a particular case of the Zakharov–Shabat transformation.

A question arises whether ZST's different from (9), (10), or (27) (see the list in Ref. 1) are related to BT's similar to (7), (8), or (22). The main difficulty to find such a correspondence is the fact that these ZST's, in general, do not preserve condition (5). On the other hand, the problem may not be very interesting, since any ZST of type (1.10)–(1.14) is a superposition of (9) or (10) with guage and (active) Lorentz transformations. Indeed, if $L{:}M\to M$ is a Lorentz transformation (we could admit conformal transformations in general) and A represents a self-dual field, then the pullback L^*A is also self-dual and the corresponding wave function is $L^*\psi(\lambda')$, where $\lambda'^A = (1,\lambda')$ is proportional to the transformed spinor λ^A. Using this fact we can shift poles, in λ, of the dressing function χ. This means that the BT's we are looking for can be obtained from (7) or (8) with the help

of Lorentz and gauge transformation (in this way we can also obtain the BT of Corrigan *et al.*).

IV. CONCLUDING REMARKS

The SDYM equations admit several solution generating methods. Among purely local ones the Atiyah–Ward method and the Zakharov–Shabat method seem to be quite successful, e.g., SU(2) magnetic monopoles were obtained by both of them.[10,11] In our previous paper[1] we showed that ZST's preserve the space of solutions generated by the Atiyah–Ward Ansätze. In this paper we compared ZST's with the BT's[4–6] related to the Yang formulation (6) of the SDYM equations. These BT's are described by nonlinear equations (7), (8), and (22), respectively. They seem to be much less practical (with the exception of the BT of Corrigan *et al.*) than ZST's, for to apply ZST we have to solve only linear equations (2) related to an initial gauge field A. We showed that, for $G = \mathrm{GL}(2,C)$, all of these BT's can be realized as ZST's combined with transformations (18) corresponding to adding an Abelian solution to a given one (Propositions 1 and 2). Thus, in some sense, we do not have to worry about these BT's. This situation may change for gauge groups, which are not subgroups of $\mathrm{GL}(2,C)$.

ACKNOWLEDGMENTS

I am greatly indebted to John Harnad for interesting discussions and financial support during my stay in Montreal. I also acknowledge the friendly atmosphere at Centre de Recherches Mathématiques. A preliminary version of this work was presented at the XIX International Colloquium on Group Theoretical Methods in Physics (Canada 1988). I thank organizers of this conference and of the preceding school "New Methods in Field and String Theories" for generous hospitality and financial support.

This work is partially supported by the Polish Research Program CPBP 01.03.

[1] J. Tafel, J. Math. Phys. **30**, 706 (1989).
[2] V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **13**, 166 (1979).
[3] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[4] E. Corrigan, D. B. Fairlie, P. Goddard, and R. G. Yates, Commun. Math. Phys. **58**, 223 (1978).
[5] K. Pohlmeyer, Commun. Math. Phys. **72**, 37 (1980).
[6] M. K. Prasad, A. Sinha, and L. L. Chau Wang, Phys. Rev. Lett. **43**, 750 (1979).
[7] C. N. Yang, Phys. Lett. **38**, 1377 (1977).
[8] R. S. Ward, Phys. Lett. A **61**, 81 (1977).
[9] A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).
[10] E. Corrigan and P. Goddard, Commun. Math. Phys. **80**, 575 (1981).
[11] P. Forgacś, Z. Horvath, and L. Palla, Phys. Lett. B **109**, 200 (1982).

Blow-up of the SU(2,C) Yang–Mills fields

Yisong Yang[a)]
Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, Minnesota 55455

(Received 2 August 1989; accepted for publication 20 December 1989)

It is remarked that, if the gauge group is not compact, finite-time blowing-up Yang–Mills fields over the Minkowski spaces may exist. The example taken is $G = \mathrm{SU}(2,\mathbb{C})$. It is shown that blow-up can occur in an arbitrarily small time interval when the initial data of the fields are suitably chosen.

I. INTRODUCTION

The studies on the global existence of the Yang–Mills (YM) fields over the Minkowski spaces were initiated by Segal[1] where a local existence result is established for the case that the gauge group $G = \mathrm{SU}(2)$. Since then, various investigations have been carried out. In particular, Glassey and Strauss[2] discussed the global existence and long-time behavior of the SU(2) YM fields under the 't Hooft–Polyakov type ansatz; Ginibre and Velo[3] generalized Segal's result to the YM systems coupled with scalar fields and obtained global existence in $2+1$ dimensions; Eardley and Moncrief[4] proved global existence in $3+1$ dimensions by using the "Cronström gauge"; and Burzlaff and O'Murchadha[5] improved the results of Ref. 4 by getting solutions with a better fall-off rate at spatial infinity. All of these studies were restricted to the case where G is a compact Lie group so that a natural Euclidean scalar product can be built up in the Lie algebra of G and it appears that not much has been said for the global existence problem of the YM fields in the Minkowski spaces if G is not compact.

Our purpose in this paper is to remark that finite-time blowing-up YM fields may be expected to evolve from certain initial data when G fails to be compact. The simple example we take is $G = \mathrm{SU}(2,\mathbb{C})$. A class of YM fields will be constructed so that their imaginary components are smooth and compactly supported but blow up in finite time $t > 0$.

II. THE SU(2,C) YM EQUATIONS

Some general mathematical issues of the SU(2,C) gauge theory were first discussed by Wu and Yang.[6] Here, if we denote by σ_a $(a = 1,2,3)$ the Pauli matrices, then the matrix form of the SU(2,C) gauge potential may be written[7]

$$A_\mu = \tfrac{1}{2}\sigma_a A_\mu^a = u_\mu + v_\mu,$$

where

$$u_\mu = \tfrac{1}{2}\sigma_a u_\mu^a, \quad v_\mu = (i/2)\sigma_a v_\mu^a,$$

and the real functions $u_\mu^a = \mathrm{Re}\, A_\mu^a$, $v_\mu^a = \mathrm{Im}\, A_\mu^a$ are called the real and imaginary components of the gauge potential A_μ^a, respectively.

The real and imaginary components of the field strength tensor, say $U_{\mu\nu}^a$ and $V_{\mu\nu}^a$, are expressed by the formulas

$$U_{\mu\nu}^a = \partial_\mu u_\nu^a - \partial_\nu u_\mu^a + \varepsilon_{abc}(u_\mu^b u_\nu^c - v_\mu^b v_\nu^c),$$
$$V_{\mu\nu}^a = \partial_\mu v_\nu^a - \partial_\nu v_\mu^a + \varepsilon_{abc}(u_\mu^b v_\nu^c + v_\mu^b u_\nu^c),$$

which obey the equations of motion of the SU(2,C) YM theory

$$\begin{aligned} \partial^\nu U_{\mu\nu}^a &= \varepsilon_{abc}(U_{\mu\nu}^b u_c^\nu - V_{\mu\nu}^b v_c^\nu), \\ \partial^\nu V_{\mu\nu}^a &= \varepsilon_{abc}(U_{\mu\nu}^b v_c^\nu + V_{\mu\nu}^b u_c^\nu). \end{aligned} \tag{2.1}$$

Take the spherically symmetric ansatz

$$\begin{aligned} u_0^a &= 0, \quad u_i^a = \varepsilon_{ain} x_n \alpha(r,t), \\ v_0^a &= 0, \quad v_i^a = \varepsilon_{ain} x_n \beta(r,t), \end{aligned} \tag{2.2}$$

where $t = x_0$ is the time and $\mathbf{x} = (x_i)_{i=1}^3$ the space variable, respectively, $r = |\mathbf{x}|$, and α and β are real functions.

From (2.2), we obtain various components of the field strength tensor:

$$\begin{aligned} U_{0i}^a &= \varepsilon_{ain} x_n \alpha_t, \\ U_{ij}^a &= -2\varepsilon_{aij}\alpha + \varepsilon_{ijn} x_n x_a (\alpha^2 - \beta^2) \\ &\quad - (\varepsilon_{ain} x_j - \varepsilon_{ajn} x_i) x_n (1/r)\alpha_r, \\ V_{0i}^a &= \varepsilon_{ain} x_n \beta_t, \\ V_{ij}^a &= -2\varepsilon_{aij}\beta + 2\varepsilon_{ijn} x_n x_a \alpha\beta \\ &\quad - (\varepsilon_{ain} x_j - \varepsilon_{ajn} x_i) x_n (1/r)\beta_r. \end{aligned}$$

Hence, after a lengthy calculation, Eqs. (2.1) are reduced to

$$\begin{aligned} \alpha_{tt} &= \alpha_{rr} + (4/r)\alpha_r + 3(r^2\alpha - 1)\beta^2 + 3\alpha^2 - r^2\alpha^3, \\ \beta_{tt} &= \beta_{rr} + (4/r)\beta_r + 6\alpha\beta - 3r^2\alpha^2\beta + r^2\beta^3. \end{aligned} \tag{2.3}$$

In (2.3), if a further ansatz

$$\alpha(r,t) = 1/r^2, \quad \beta(r,t) = (1/r^2)F(r,t) \tag{2.4}$$

is taken, we obtain a real scalar wave equation

$$F_{tt} = F_{rr} + (1/r^2)F(1 + F^2). \tag{2.5}$$

The trivial solution $F = 0$ gives rise to the well-known Wu–Yang monopole.[8] In the subsequent sections, we shall construct a class of nontrivial solutions of Eq. (2.5) which blow up in finite time.

III. LOCAL EXISTENCE

Consider Eq. (2.5) supplemented with the initial data

$$F(r,0) = F_0(r), \quad F_t(r,0) = F_1(r). \tag{3.1}$$

To stay away from the singular point $r = 0$ of Eq. (2.5), we require that

a) Current address: Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131.

 0022-2488/90/051237-03$03.00

$$\text{supp}(F_0),\quad \text{supp}(F_1)\subset[a,b],$$

where $0<a<b<\infty$ are arbitrary. For our purpose, we also assume that F_0, F_1 are C^∞ and take non-negative values.

Equations (2.5)–(3.1) may formally be converted into the following integral form:

$$F(r,t)=\frac{1}{2}(F_0(r+t)+F_0(r-t))+\frac{1}{2}\int_{r-t}^{r+t}F_1(\xi)d\xi+\frac{1}{2}\int_0^t ds\int_{-1}^{1}\frac{t-s}{(r+(t-s)\eta)^2}F(1+F^2)(r+(t-s)\eta,s)d\eta\equiv J(F)(r,t),$$

namely, the solution of Eqs. (2.5)–(3.1) is realized as a fixed point of the operator J.

Let $C_0(\mathbb{R})$ be the space of continuous functions over $\mathbb{R}$ with compact supports and $\|\ \|$ the usual sup norm of the space. For given $T>0$, define

$$\mathscr{X}=\{F(r,t)\,|\,F \text{ is continuous in } (r,t)\in\mathbb{R}\times[0,T] \text{ and supp } (F(\cdot,t))\subset[a-t,b+t]\}$$

with the norm

$$\|F\|_{\mathscr{X}}=\sup_{t\in[0,T]}\|F(\cdot,t)\|.$$

It is straightforward to verify that J maps $\mathscr{X}$ into itself if $T\leqslant a-\delta$ $(0<\delta<a)$ since for $F\in\mathscr{X}$, $F(r+(t-s)\eta,s)\neq 0$ implies $r+(t-s)\eta\geqslant a-s\geqslant a-T\geqslant\delta$.

Consider a ball in $\mathscr{X}$:

$$B_M=\{F\in\mathscr{X}\,|\,\|F\|_{\mathscr{X}}\leqslant M\}\quad \text{with } M>\|F_0\|.$$

Then we have

$$\|J(F)\|_{\mathscr{X}}\leqslant\|F_0\|+T\|F_1\|+(T^2/\delta^2)M(1+M^2).$$

On the other hand, for $F,G\in B_M$, there holds

$$\|J(F)-J(G)\|_{\mathscr{X}}\leqslant(T^2/\delta^2)(1+3M^2)\|F-G\|_{\mathscr{X}}.$$

Therefore, if

$$T<\min\left\{\frac{M-\|F_0\|}{\|F_1\|+M(1+M^2)/\delta^2},\frac{\delta^2}{1+3M^2},1\right\},$$

then $J:B_M\to B_M$ and is a contraction. Hence J has a unique fixed pont $F\in B_M$ which is a smooth solution of Eqs. (2.5)–(3.1) in the time interval $[0,T]$.

The above construction and the continuation method allow us to conclude that the solution $F(r,t)$ exists in a maximal time interval $[0,T_{\max})$ where either $T_{\max}=a$ or $0<T_{\max}<a$ and

$$\lim_{t\to T_{\max}}\|F(\cdot,t)\|=\infty.$$

Our next goal now is to achieve this latter possibility.

IV. A COMPARISON EQUATION

In order to show that Eqs. (2.5)–(3.1) have finite-time blowing-up solutions, it seems convenient to use the comparison equation

$$\begin{aligned}&G_{tt}=G_{rr}+(1/r^2)G^3,\quad t>0,\\&G(r,0)=F_0(r),\quad G_t(r,0)=F_1(r).\end{aligned}\tag{4.1}$$

Lemma 4.1: If the solution $F(r,t)$ of Eqs. (2.5)–(3.1) exists over a time interval $[0,\tau)$ $(\tau\leqslant a)$, so does the (unique) solution $G(r,t)$ of Eqs. (4.1). Moreover

$$F(r,t)\geqslant G(r,t),\quad (r,t)\in\mathbb{R}\times[0,\tau).\tag{4.2}$$

Proof: It suffices to verify that, for any fixed τ_0: $0<\tau_0<\tau$, Eqs. (4.1) have a solution $G(r,t)$ over $\mathbb{R}\times[0,\tau_0]$ and (4.2) holds for $(r,t)\in\mathbb{R}\times[0,\tau_0]$. To this aim, let us introduce an iterative scheme as in Caffarelli and Friedman[9] as follows:

$$G^0(r,t)\equiv 0,$$

$$G^n(r,t)=\frac{1}{2}(F_0(r+t)+F_0(r-t))+\frac{1}{2}\int_{r-t}^{r+t}F_1(\xi)d\xi+\frac{1}{2}\int_0^t ds\int_{-1}^{1}\frac{t-s}{(r+(t-s)\eta)^2}(G^{n-1})^3(r+(t-s)\eta,s)d\eta,\quad n=1,2,\ldots.$$

Since $G^1(r,t)\geqslant G^0(r,t)$, so $G^2(r,t)\geqslant G^1(r,t)$. In this manner, one establishes the monotone chain

$$\begin{aligned}&G^n(r,t)\leqslant G^{n+1}(r,t),\quad (r,t)\in\mathbb{R}\times[0,\tau_0],\\&n=0,1,2,\ldots.\end{aligned}\tag{4.3}$$

On the other hand, using $F(r,t)\geqslant 0$ and $F(r,t)=J(F)(r,t)$, it is seen that

$$\begin{aligned}&G^n(r,t)\leqslant F(r,t),\quad (r,t)\in\mathbb{R}\times[0,\tau_0],\\&n=0,1,2,\ldots.\end{aligned}\tag{4.4}$$

From (4.3) and (4.4) one concludes that $\lim_{n\to\infty}G^n(r,t)$ exists, say $G(r,t)$, which is the solution of (4.1) on $\mathbb{R}\times[0,\tau_0]$ satisying $G(r,t)\leqslant F(r,t),\forall(r,t)\in\mathbb{R}\times[0,\tau_0]$. The proof of the lemma is complete.

V. PROOF OF BLOW-UP

With the results of Sec. IV, the blowing-up solutions can be constructed by taking some well-developed paths. Here we shall use the energy method presented in Reed and Simon.[10]

Let $G(r,t)$ be the solution of Eqs. (4.1) evolving from its initial data at $t=0$. On the domain of existence of the solution, set

$$I(t)=\int_0^\infty (G(r,t))^2\,dr.$$

The uniqueness of the solution and the fact that $G(r,t)\geqslant 0$ imply $I(t)>0$, $t\geqslant 0$, provided F_0 is not identically zero.

Define $P(t)=I(t)^{-1/2}$. We have

$$P'(t)=-I(t)^{-3/2}\int_0^\infty GG_t\,dt,$$

$$P''(t)=3I(t)^{-5/2}\left\{\left(\int_0^\infty GG_t\,dr\right)^2-I(t)\int_0^\infty G_t^2\,dr\right\}+I(t)^{-3/2}\left\{2\int_0^\infty G_t^2\,dr-\int_0^\infty GG_{tt}\,dr\right\}$$

$$\leqslant I(t)^{-3/2}\left\{2\int_0^\infty G_t^2\,dr-\int_0^\infty GG_{tt}\,dr\right\}\equiv I(t)^{-3/2}J(t),$$

by virtue of the Schwarz inequality.

Using Eqs. (4.1) and $\text{supp}(G(\cdot,t))\subset[a-t,b+t]\subset(0,\infty)$, we can calculate that

$$J(t)=\int_0^\infty dr\left\{2G_t^2+G_r^2-\frac{1}{r^2}G^4\right\}. \tag{5.1}$$

On the other hand, along the solution of Eqs. (4.1), the energy

$$E(t)=\frac{1}{2}\int_0^\infty dr\left\{G_r^2+G_t^2-\frac{1}{2r^2}G^4\right\} \tag{5.2}$$

is a conserved quantity, namely, $E(t)\equiv E(0)$. Hence, substituting (5.2) into (5.1), we get

$$J(t)=4E(0)-\int_0^\infty G_r^2\,dr\leqslant 4E(0).$$

Suppose the initial data F_0, F_1 have been chosen in such a way that

$$E(0)=\frac{1}{2}\int_0^\infty dr\left\{F_{0,r}^2+F_1^2-\frac{1}{2r^2}F_0^4\right\}\leqslant 0. \tag{5.3}$$

Then there holds $P''(t)\leqslant I(t)^{-3/2}J(t)\leqslant 0$. Consequently $P'(t)\leqslant P'(0)$ and

$$P(t)\leqslant I(0)^{-1/2}+tP'(0)$$
$$=I(0)^{-3/2}\left(\int_0^\infty F_0^2\,dr-t\int_0^\infty F_0F_1\,dr\right). \tag{5.4}$$

Assume that F_0,F_1 are not orthogonal: $\int F_0F_1\,dr>0$. For any T:$0<T<a$, in order to get a blowing-up solution of Eqs. (4.1) with $T_{\max}<T$, we need only to force the function $P(t)$ vanish at some point $T_0\in(0,T)$ because

$$P(t)^{-1}\leqslant\|G(\cdot,t)\|(b-a+2T)^{1/2}\leqslant\|G(\cdot,t)\|(a+b)^{1/2}.$$

By virtue of the inequalities (5.4) and $P(0)>0$, it is seen that this latter requirement will be met provided

$$\int_0^\infty F_0^2\,dr<T\int_0^\infty F_0F_1\,dr. \tag{5.5}$$

It is not difficult to verify that, with suitable choice of the initial states F_0, F_1, (5.3) and (5.5) may be satisfied simultaneously.

To see this, let $\zeta(r)\geqslant 0$ be a C^∞ function with support in $[a,b]$ but not identically zero. Put

$$F_0(r)=\lambda^2\zeta(r),\quad F_1(r)=\lambda^3\zeta(r),\quad \lambda>0.$$

Then for sufficiently large values of the parameter λ, F_0, F_1 fulfill the requirements of (5.3) and (5.5).

Therefore, from Lemma 4.1, a solution of Eqs. (2.5)–(3.1) that blows up in finite time $t=T_{\max}<T$ is produced.

VI. CONCLUSIONS

We have illustrated by the simple example $G=\mathrm{SU}(2,\mathbb{C})$ that, when the gauge group is not compact, global existence of the YM fields in the Minkowski spaces may fail. We have shown that the imaginary components, which are smooth and of compact supports, of a class of the $\mathrm{SU}(2,\mathbb{C})$ YM fields, blow up in an arbitrarily small finite time interval if the initial data are suitably chosen.

Remark: The real components of the YM fields constructed in this paper have a point singularity at $r=0$ [see (2.4)]. Thus, if one insists on the global smoothness of the full fields, the example here should be interpreted as a construction of finite-time blowing-up fields in the exterior spatial domain $|\mathbf{x}|\geqslant\varepsilon$ ($\varepsilon>0$). Hence, the conclusion may be restated as follows: For the noncompact gauge group $G=\mathrm{SU}(2,\mathbb{C})$, a class of smooth YM fields (with finite energy) over $|\mathbf{x}|\geqslant\varepsilon$ can always be found that evolve from certain initial states and blow up in an arbitrarily short time span.

[1] I. E. Segal, J. Funct. Anal. **33**, 175 (1979).
[2] R. T. Glassey and W. A. Strauss, Commun. Math. Phys. **81**, 171 (1981).
[3] J. Ginibre and G. Velo, Commun. Math. Phys. **82**, 1 (1981).
[4] D. M. Eardley and V. Moncrief, Commun. Math. Phys. **83**, 171, 193 (1982).
[5] J. Burzlaff and N. O'Murchadha, Commun. Math. Phys. **105**, 85 (1986).
[6] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3824 (1975); **13**, 3233 (1976).
[7] A. Actor, Rev. Mod. Phys. **51**, 461 (1979).
[8] T. T. Wu and C. N. Yang, in: *Properties of Matter under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1968), p. 349.
[9] L. A. Caffarelli and A. Friedman, Trans. Am. Math. Soc. **297**, 223 (1986).
[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.

BRS cohomology in string theory: Geometry of Abelization and the quartet mechanism

Henrik Aratyn
Department of Physics, University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680

(Received 18 August 1989; accepted for publication 10 January 1990)

The Abelian representation of the string Becchi–Rouet–Stora (BRS) cohomology in terms of basic operators naturally forming two dual Kugo–Ojima quartets is found. Furthermore, using a gauge-fixing scheme, a contracting homotopy operator associated with the string BRS operator is constructed. This formalism gives the explicit realizations of the physical and unphysical subspaces. The passage between Abelian and non-Abelian quantities is realized geometrically through the use of the moments of the vertex operator, which act as vielbeins between infinite-dimensional algebras. The connection between the Virasoro and spectrum-generating algebras is clarified and the algebraic duality relation between them is uncovered.

I. INTRODUCTION

In the past few years the language and methods of differential geometry have become fundamental tools in modern quantum field theory. In this paper we show how to evaluate the Becchi–Rouet–Stora (BRS) cohomology class in the case of strings by analyzing the associated geometrical structure.

The key geometrical features are defined in Sec. II in the analysis of the specific gauge-fixing procedure for the infinite-dimensional first-class constraint algebras. Accordingly, we consider a dynamical system governed by a set Φ of functions $\{\phi_a\}$ defined on the phase space W with a Poisson bracket $\{\cdot,\cdot\}$. The functions ϕ_a constrain the dynamics to the submanifold $V \subset W$ defined by the conditions $\{\phi_a = 0\}$.

The physical space is now identified with the gauge-equivalence classes defined by the gauge transformations generated by Φ on V. Since here we deal with the first-class constraints, Φ is closed under the Poisson bracket. Especially, we have $\{\phi_a,\phi_b\}_{|V} = 0$.

The above structure has to be completed by the gauge-fixing conditions (called the subsidiary constraints in Dirac's notation) Q^a in order to enable us to choose one element out of a given equivalence class. This is possible if the gauge-fixing conditions intersect each equivalence class uniquely in one point, as ensured by the basic condition $\det\|\{\phi_a,Q^b\}\|_{|V} \neq 0$.

Consequently, we are led to consider the larger set, which consists of the first-class constraints ϕ_a and the associated gauge-fixing conditions Q^a. We now introduce the gauge-fixing scheme, which algebraically realizes the above conditions in the following way:

$$\begin{aligned} &\{\phi_a,\phi_b\}_{|V} = 0 \rightarrow \{\phi_a,\phi_b\} = U^c_{ab}\phi_c\,, \\ &\det\|\{\phi_a,Q^b\}\|_{|V} \neq 0 \rightarrow \{\phi_a,Q^b\} = \omega^b_a - T^b_{ac}Q^c\,, \end{aligned} \tag{1.1}$$

where ω^b_a is a central element and U^c_{ab}, together with T^c_{ab}, are structure constants to be determined. The first equation in (1.1) describes a conventional Lie algebra of first-class constraints, while the second equation ensures the closure of the resulting algebra containing ϕ_a and Q^a. Furthermore, we make the convenient choice that

$$\{Q^a,Q^b\} = 0\,. \tag{1.2}$$

Note that Eqs. (1.1) impose the condition $\det\|\omega^b_a\| \neq 0$.

Consequently, this way of supplementing the first-class constraints by the gauge-fixing conditions leads to the larger, *closed* extended constraint algebra. In this framework gauge fixing leads to a simple parametrization of gauge orbits, with gauge-fixing conditions playing the role of coordinates. One of the nice features of the extended constrained system is that it naturally admits a symplectic formulation with canonical coordinates. We find explicitly Abelian momentum, conjugated to the gauge-fixing condition and associated with the transport along the gauge orbit.

To establish in a transparent way a connection of original constraints to the set of Abelian operators (here we call this procedure Abelization) we begin our construction by defining the canonical pair of conjugated coordinates and the associated Poisson bracket structure. In this setting we introduce the non-Abelian structure by rotating the original coordinates to the "curvilinear" coordinates. Use of the language of symplectic geometry helps to give us a more transparent view of duality between the operator algebras encountered in this study.

Our geometric framework is centered around the study of rotations of the canonical coordinates by generalized vielbeins e^a_i.[1,2] These vielbeins define the passage between Abelian and and non-Abelian quantities for infinite-dimensional constraint algebras and satisfy specific structure equations.

Once the means of connecting Abelian coordinates to the non-Abelian structure are established we then turn in Sec. III to the specific example of the Virasoro algebra. In Sec. III we rely on our recent study[2] showing how the geometrical interpretation of the vertex operator, in terms of the vielbeins satisfying the Maurer–Cartan (MC) structure equations connected with the Virasoro algebra, makes it possible to rewrite several basis objects in string theory in the form of simple Abelian operators. Understanding of the vertex operator as the generating function for vielbeins[2] has made possible technical progress in constructing the set of Abelian string variables including longitudinal and transverse directions and introduced enormous simplification into the study of their mutual relations.

Remarkably, in the context of string theory, the connec-

 0022-2488/90/051240-13$03.00

tion between the Abelian and non-Abelian quantities is fully preserved when we gradually incorporate first transverse directions and then the Batalin–Fradkin–Vilkovisky (BFV)[3] ghosts and normal ordering in the passage to the complete quantum mechanical framework.

This geometrical setting is then used in Sec. IV to characterize the string BRS cohomology in terms of Abelian string operators, which naturally form two mutually dual Kugo–Ojima (KO) quartets.[4,5]

The Abelian structure uncovered in this study leads to determination of the BRS operator as the exterior derivative operator, with the derivative acting along the unphysical gauge orbits in the extended phase space of the BFV (Ref. 3) formalism. Therefore, we are led to the most transparent evaluation of quantum BRS cohomology in the spirit of differential geometry, which is also a motif of a recent study of classical BRS cohomology.[6] The Abelian derivative along the unphysical directions and BFV ghosts are contained in one BRS supermultiplet and cancel each other via Parisi–Sourlas supersymmetry when embedded in the KO[4,5] quartet.

We also explicitly construct a nilpotent contracting homotopy operator in terms of the gauge-fixing conditions. The cohomology calculation is then further clarified by embedding the homotopy operator and the BRS in a framework analogous to supersymmetric quantum mechanics. Consequently, our construction yields a simple derivation of the string BRS cohomology.

The physical subspace is then determined as the transverse gauge-invariant directions, in agreement with the old no-ghost theorem.[7,8] The zero-mode operators associated with the above quartets reproduce the Freeman–Olive[9] oscillator number operators used in a very compact and elegant way by Freeman and Olive in their derivation of the string BRS cohomology. This geometrical construction clearly suggests use of fiber bundle analogies in identifying the gauge orbits and physical base space.

This work represents progress in the directioin of extending the presently available formalism for cohomology calculation in the quantized systems beyond simple Abelian cases. We hope that pointing out the homotopy proof hidden behind the KO quartet method can open a way to treat more complicated and interesting examples than was possible until now.

Notation: The reader should be warned about a few notational changes with respect to Ref. 2. The fundamental difference is that in this paper the symbol P_i describes what was $\mathscr{D}_i$ in Ref. 2 and vice versa. This change is due to an attempt to achieve a more transparent look for several formulas and exhibit the important dual character of our operational construction.

II. GENERAL CONCEPTS OF DUAL ALGEBRAS

A. Coordinates and canonical mapping

Our geometrical approach to Abelization and the related issue of the gauge fixing of the constraint algebras is most simply discussed in terms of canonical variables. Accordingly, we introduce the Abelian coordinates $X_a = (P_a, Q^b)$ and the associated symplectic metric

$$\omega_{ab} = \begin{pmatrix} 0 & \delta_a^b \\ -\delta_a^b & 0 \end{pmatrix}. \tag{2.1}$$

The metric tensor (2.1) gives rise to the Poisson bracket in the following way:

$$\{G,F\} \equiv \omega_{ab} \frac{\partial G}{\partial X_a} \frac{\partial F}{\partial X_b}, \tag{2.2}$$

where G and F are arbitrary functions on the phase space. Consequently, the canonical coordinates satisfy the Heisenberg algebra

$$\{P_a, Q^b\} = \delta_a^b, \tag{2.3}$$

$$\{P_a, P_b\} = \{Q^a, Q^b\} = 0. \tag{2.4}$$

As in Refs. 1 and 2, we are interested in transforming the coordinate X_a according to

$$\exp(-\Phi)\delta_i^a P_a \exp(\Phi) = e_i^a \mathscr{D}_a \equiv P_i, \tag{2.5}$$

$$\exp(-\Phi)\delta_a^i Q^a \exp(\Phi) = E_a^i Q^a \equiv Q^i, \tag{2.6}$$

where the new "derivative" $\mathscr{D}_a \equiv D_a^b P_b$ is defined in terms of the invertible matrix

$$D_a^b = \delta_a^b - T_{ac}^b Q^c. \tag{2.7}$$

Here T_{ac}^b is the structure constant that dictates the form of the above mapping by entering the expression for the generator $\Phi \equiv -Q^a T_{ac}^b Q^c P_b$. Furthermore, the mapping introduces a bridge variable connecting the a and i indices, namely, a vielbein e_i^a and its inverse E_a^i.

Since the above mapping is a canonical transformation, the quantities P_i and Q^i defined on the rhs of (2.5) and (2.6) satisfy the same Heisenberg algebra that is satisfied by P_a and Q^a. Therefore, $\{P_i, Q^j\} = \delta_i^j$, $\{P_i, P_j\} = 0$, and $\{Q^i, Q^j\} = 0$.

From definition (2.7) it is easy to derive the relations

$$\{\mathscr{D}_a, Q^b\} = D_a^b, \tag{2.8}$$

$$\{\mathscr{D}_a, \mathscr{D}_b\} = D_a^c \frac{\partial \mathscr{D}_b}{\partial Q^c} - D_b^c \frac{\partial \mathscr{D}_a}{\partial Q^c}. \tag{2.9}$$

If we work, as in Refs. 1 and 2, with the closed constraint algebra, then we must require

$$\{\mathscr{D}_a, \mathscr{D}_b\} = U_{ab}^c \mathscr{D}_c, \tag{2.10}$$

where U_{ab}^c is a structure constant antisymmetric in the lower indices and associated with the first-class constraint algebra. The consistency requires that

$$D_a^c \frac{\partial \mathscr{D}_b}{\partial Q^c} - D_b^c \frac{\partial \mathscr{D}_a}{\partial Q^c} = U_{ab}^c \mathscr{D}_c, \tag{2.11}$$

which, expanded order by order in Q^a, reproduces the consistency relations[1,2,10]

$$U_{ab}^d T_{dc}^e + T_{ac}^d T_{bd}^e - T_{bc}^d T_{ad}^e = 0, \tag{2.12}$$

$$U_{ab}^c = T_{ab}^c - T_{ba}^c. \tag{2.13}$$

Note that the consistency relation (2.11) can be given the form

$$\mathscr{D} \wedge \mathscr{D} = F \tag{2.14}$$

if we define

$$\mathscr{D} = D_a^b \frac{\partial}{\partial Q^b} dQ^a, \tag{2.15}$$

$$F=\tfrac{1}{2}U^{c}_{ab}\,\mathscr{D}_c\,dQ^a\wedge dQ^b\,, \tag{2.16}$$

where we have associated to Q^b the one-form dQ^b.

As a consequence of definition (2.14) the two-form F satisfies the Bianchi identity

$$\mathscr{D}\wedge F-F\wedge\mathscr{D}=0\,, \tag{2.17}$$

which in component form reads as

$$U^{d}_{ab}U^{e}_{cd}+U^{d}_{ca}U^{e}_{bd}+U^{d}_{bc}U^{e}_{ad}=0\,, \tag{2.18}$$

reproducing the Jacobi relation which would otherwise follow directly from the algebra (2.10).

This framework allows interpretation of $\mathscr{D}$ as a non-Abelian covariant derivative, with the structure constant T^{c}_{ab} playing the role similar to the connection.[1] This interpretation of T^{c}_{ab} will be further strengthened by its appearance in the structure equations.

B. Duality and the structure equations

Since the Poisson bracket structure, as defined by (2.2), leads to $\{P_a,f\}=\partial f/\partial Q^a$ and $\{P_i,f\}=\partial f/\partial Q^i$ for any function $f(Q)$ of the coordinates, the canonical mapping (2.5) and (2.6) can be cast in the simple form

$$Q^i=E^{i}_{a}(Q)Q^a\,,\quad \frac{\partial}{\partial Q^i}=\frac{\partial Q^a}{\partial Q^i}\frac{\partial}{\partial Q^a}\,, \tag{2.19}$$

where we wrote $E^{i}_{a}(Q)$ to indicate that the vielbein, which "rotates" the coordinate Q^a, depends functionally on Q^a (and the structure constant T^{c}_{ab}). Using the fact that e^{a}_{i} is an inverse vielbein to E^{i}_{a} (i.e., $e^{b}_{i}E^{i}_{a}=\delta^{b}_{a}$ and $e^{b}_{i}E^{j}_{b}=\delta^{j}_{i}$) we easily obtain the dual version of Eqs. (2.19):

$$Q^a=e^{a}_{j}(Q)Q^j,\quad \frac{\partial}{\partial Q^a}=\frac{\partial Q^j}{\partial Q^a}\frac{\partial}{\partial Q^j}\,. \tag{2.20}$$

Comparing with (2.5) we find

$$\frac{\partial Q^a}{\partial Q^i}=\frac{\partial e^{a}_{j}}{\partial Q^i}Q^j+e^{a}_{i}=e^{b}_{i}D^{a}_{b}\,, \tag{2.21}$$

where the first term describes departure from the "global rotation," which would take place with constant e^{a}_{i}. Inserting definition (2.7) of D^{a}_{b} on the rhs of (2.21) we obtain

$$\frac{\partial e^{a}_{j}}{\partial Q^i}=-T^{a}_{bc}e^{b}_{i}e^{c}_{j}\,. \tag{2.22}$$

In this way we have derived our first structure equation. Since from Eqs. (2.19) and (2.21) we find that $E^{i}_{a}(\partial/\partial Q^i)=D^{b}_{a}(\partial/\partial Q^b)$ the structure equation (2.22) can be rewritten as

$$D^{d}_{b}\frac{\partial e^{a}_{j}}{\partial Q^d}=-T^{a}_{bc}e^{c}_{j}\,. \tag{2.23}$$

We call this equation[1,2] the MC structure equation. Quite analogously, we can use Eq. (2.20) to derive the corresponding MC equation for the inverse vielbein E^{i}_{a}. We first notice that the quantity that is inverse to $e^{b}_{i}D^{a}_{b}$ is given by $E^{i}_{b}(D^{-1})^{b}_{a}$ and, therefore,

$$\frac{\partial Q^i}{\partial Q^a}=\frac{\partial E^{i}_{b}}{\partial Q^a}Q^b+E^{i}_{a}=E^{i}_{b}(D^{-1})^{b}_{a}\,. \tag{2.24}$$

Multiplying (2.24) on both sides by D^{a}_{b} and using definition (2.7) we arrive at

$$D^{c}_{a}\frac{\partial E^{i}_{b}}{\partial Q^c}=T^{c}_{ab}E^{i}_{c}\,. \tag{2.25}$$

Making use of (2.20) we can rewrite (2.25) in an alternative way as

$$\frac{\partial E^{i}_{b}}{\partial Q^j}==T^{c}_{ab}E^{i}_{c}e^{a}_{j}\,. \tag{2.26}$$

The MC structure equations ensure consistency with the canonical character of the mapping in (2.5) and (2.6); as an example, consider

$$\frac{\partial Q^j}{\partial Q^i}=\frac{\partial E^{j}_{b}}{\partial Q^i}Q^b+E^{j}_{a}\frac{\partial Q^a}{\partial Q^i}\,. \tag{2.27}$$

One can easily show that the rhs of (2.27) correctly reproduces δ^{j}_{i} and ensures that $\{P_i,Q^j\}=\delta^{j}_{i}$.

The fact that the vielbein's are invertible and satisfy the MC equations allows us to in fact express the formalism in terms of the dual variables, with the Poisson brackets defined as

$$\{G,F\}\equiv\delta^{j}_{i}\left(\frac{\partial G}{\partial P_i}\frac{\partial F}{\partial Q^j}-\frac{\partial G}{\partial Q^j}\frac{\partial F}{\partial P_i}\right). \tag{2.28}$$

This Poisson bracket structure follows from the rules governing the way the symplectic tensor ω_{ab} transforms under the mapping of the canonical variables. If the coordinates $X_a=(P_a,Q^b)$ are mapped into $Y_i=\phi_i(X)$, where in our case $Y_i=(P_i,Q^j)$, then the transformed symplectic metric is given by[11]

$$\omega_{ij}=\omega_{ab}\frac{\partial X_a}{\partial Y_i}\frac{\partial X_b}{\partial Y_j}\,. \tag{2.29}$$

The postulate (2.28) is then simply proved by the calculation involving the quantity

$$\delta^{a}_{b}\frac{\partial P_a}{\partial P_i}\frac{\partial Q^b}{\partial Q^j}=\delta^{a}_{b}(D^{-1})^{c}_{a}E^{i}_{c}\frac{\partial Q^b}{\partial Q^j}=\delta^{i}_{j}\,, \tag{2.30}$$

where we used relation (2.21) and $P_a=(D^{-1})^{c}_{a}E^{i}_{c}P_i$.

To obtain a better view of the relations between formalisms based on (P_n,Q^m) and (P_i,Q^j) we will now introduce the auxiliary quantity

$$T_a=-T^{c}_{ab}Q^bP_c\,. \tag{2.31}$$

In terms of T_a we can write

$$\mathscr{D}_a=P_a+T_a\,. \tag{2.32}$$

Here T_a turns out to be very useful because it does not change its functional dependence on the coordinates under the canonical mapping (2.5) and (2.6). This can be seen by noticing that trivially, the generator of the canonical transformation Φ is invariant under (2.5). Since Φ can be written as

$$\Phi=Q^aT_a\,, \tag{2.33}$$

this fixes the way T_a transforms under the canonical mapping to

$$\exp(-\Phi)\delta^{a}_{i}T_a\exp(\Phi)=e^{a}_{i}T_a\equiv T_i\,, \tag{2.34}$$

where on the rhs we have defined T_i, the transform of T_a. We can obtain the alternative derivation of T_i by inserting definition (2.31) on the lhs of (2.34) and performing separately the transformation of Q^b and P_c (see Ref. 2):

$$T_i = -T^k_{ij} Q^j P_k \,, \tag{2.35}$$

where $T^k_{ij} = \delta^a_i \delta^b_j \delta^k_c T^c_{ab}$ differs from T^c_{ab} only in having the indices $(i,j,k,\ldots)$ rather than $(a,b,c,\ldots)$. Note, also, that as a consequence of relations (2.20) and (2.34), we have

$$\Phi = Q^a T_a = Q^i e^a_i E^j_a T_j = Q^i T_i \,. \tag{2.36}$$

The invariance of Φ under the mapping betwen the two sets of Abelian variables (P_a,Q^b) and (P_i,Q^j) leads to a duality between the quantities with a and i indices.

The fact that the operator T_a does not change its functional form when expressed in terms of two different parametrizations (2.31) and (2.35) allows a totally symmetrical dual construction. Namely, we can start with the coordinates (P_i,Q^j) and transform them according to

$$\begin{aligned}&\exp(\Phi)\delta^i_a P_i \exp(-\Phi) \\ &\quad = \exp(\Phi)\delta^i_a \exp(-\Phi) P_b \delta^b_i \\ &\qquad \times \exp(\Phi)\exp(-\Phi) = P_a = E^i_a D^j_i P_j \,,\end{aligned} \tag{2.37}$$

where we have introduced an invertible matrix

$$D^j_i = \delta^j_i + T^j_{ik} Q^k \tag{2.38}$$

and used the relation

$$P_a = \mathscr{D}_a - T_a = E^i_a (P_i - T_i) = E^i_a D^j_i P_j \,. \tag{2.39}$$

Clearly, we can view Eq. (2.37) as identical to the second equation in (2.20), with

$$\frac{\partial Q^i}{\partial Q^a} = \frac{\partial E^i_b}{\partial Q^a} Q^b + E^i_a = E^j_a D^i_j \tag{2.40}$$

instead of (2.24). We can repeat essentially the same step as in (2.40) deriving the structure equations for e^a_i, so as to arrive at

$$\frac{\partial E^i_b}{\partial Q^a} = T^i_{jk} E^k_b E^j_a \,, \quad D^k_j \frac{\partial E^i_b}{\partial Q^k} = T^i_{jk} E^k_b \,. \tag{2.41}$$

Similarly, we find the inverse relation to (2.40) as

$$\frac{\partial Q^a}{\partial Q^i} = \frac{\partial e^a_j}{\partial Q^i} Q^j + e^a_i = e^a_j (D^{-1})^j_i \,, \tag{2.42}$$

leading to the dual MC structure equations in the same way as above:

$$\frac{\partial e^a_i}{\partial Q^b} = -T^j_{ki} E^k_b e^a_j \,, \quad D^j_k \frac{\partial e^a_i}{\partial Q^j} = -T^j_{ki} e^a_j \,. \tag{2.43}$$

As we see when comparing (2.41) and (2.43) with (2.22) and (2.26), the vielbeins e^a_i and E^i_a interchange their roles under the coordinate change $(P_a,Q^b) \to (P_i,Q^j)$ accompanied by $T^c_{ab} \to -T^k_{ij}$.

It is useful to introduce the quantity[1,2]

$$\mathscr{D}_i \equiv e^a_i P_a = D^j_i P_j \,. \tag{2.44}$$

This construction is clearly analogous to the one we have encountered discussing $\mathscr{D}_a = D^b_a P_b$; in fact, we obtain for (P_i,Q^j) the Poisson bracket relations dual to (2.8) and (2.9):

$$\{\mathscr{D}_i, Q^j\} = D^j_i \,, \tag{2.45}$$

$$\{\mathscr{D}_i, \mathscr{D}_j\} = D^k_i \frac{\partial \mathscr{D}_j}{\partial Q^k} - D^k_j \frac{\partial \mathscr{D}_i}{\partial Q^k} = -U^k_{ij} \mathscr{D}_k \,. \tag{2.46}$$

C. Canonical coordinates and the Darboux theorem

Until now we have based our construction on the existence of the momentum P_a conjugated to Q^b. However, our starting point was the extended constraint algebra (1.1) and (1.2) written only in terms of (ϕ_a,Q^b); accordingly, we would like to find canonical coordinates without explicit mention of P_a. Fortunately, the MC structure equations are differential equations only involving Q^b and we can think about P_a as an auxiliary quantity which serves only to make the construction of vielbeins and notion of duality more transparent.

We will now present the explicit construction of the canonical coordinates only in terms of the original constraints (ϕ_a,Q^b) and vielbeins. Define, namely, $\Phi_i \equiv e^a_i \phi_a$. Then it follows from the MC equations (2.23) and (2.25) that Φ_i, together with $Q^i = E^i_a Q^a$, satisfy the Heisenberg algebra

$$\{\Phi_i, Q^j\} = \delta^j_i \,, \quad \{\Phi_i,\Phi_i\} = 0 = \{Q^i,Q^j\} \,. \tag{2.47}$$

It is easy to see that the Poisson brackets defined as

$$\{G,F\} \equiv \delta^j_i \left(\frac{\partial G}{\partial \Phi_i} \frac{\partial F}{\partial Q^j} - \frac{\partial G}{\partial Q^j} \frac{\partial F}{\partial \Phi_i} \right) \tag{2.48}$$

will reproduce relations (1.1) and (1.2). Furthermore, we can define the non-Abelian quantity $\hat{\phi}_i \equiv D^j_i \Phi_j = D^j_i e^a_j \phi_a$ which, together with Q^j, will satisfy the non-Abelian algebra (2.45) and (2.46) dual to (1.1) and (1.2). The passage back to the quantities labeled by a indices can be performed by acting with the matrices $\partial Q^j/\partial Q^a$ as follows:

$$\frac{\partial Q^j}{\partial Q^a} \Phi_j = E^i_a D^j_i \Phi_j = E^i_a \hat{\phi}_i \equiv \Phi_a \,. \tag{2.49}$$

It follows that Φ_a defined in such a way is equal to $(D^{-1})^b_a \phi_a$ and, together with Q^a, constitutes another set of the canonical coordinates satisfying

$$\{\Phi_a, Q^b\} = \delta^b_a \,, \quad \{\Phi_a,\Phi_b\} = 0 = \{Q^a,Q^b\} \,. \tag{2.50}$$

This construction can be considered as a proof of the Darboux theorem, valid for the case of the constraint algebra (1.1) and (1.2) through an explicit construction of the canonical coordinates. For the given gauge-fixing condition satisfying (1.1) and (1.2) we can therefore construct the above Abelian quantities; this construction is valid everywhere the gauge-fixing condition is defined. Especially, if we work with the model with the globally defined gauge-fixing condition (free of Gribov's problem) our construction will yield globally defined canoncial coordinates.

D. Summary

We can now summarize our results about generators of two dual algebras in the following transparent form:

$$\mathscr{D}_i = D^j_i \frac{\partial}{\partial Q^j} = D^j_i P_j = e^a_i P_a = e^a_i \frac{\partial}{\partial Q^a} \,, \tag{2.51}$$

$$\mathscr{D}_a = D^b_a \frac{\partial}{\partial Q^b} = D^b_a P_b = E^j_a P_j = E^j_a \frac{\partial}{\partial Q^j} \,, \tag{2.52}$$

where we used the usual identification $P_a \leftrightarrow \partial/\partial Q^a$ and $P_i \leftrightarrow \partial/\partial Q^i$.

There is also an extension of the above symplectic structure to two dual canonical coordinates $X_a = (\Phi_a,Q^b)$ and

$X_i = (\Phi_i, Q^j)$, defined in terms of the original constraints (ϕ_a, Q^b) as

$$\Phi_a = E^i_a \hat{\phi}_i = E^i_a D^j_i \Phi_j = (D^{-1})^b_a \phi_b \,, \tag{2.53}$$

$$\Phi_i = e^a_i \phi_a = e^a_i D^b_a \Phi_b = (D^{-1})^j_i \hat{\phi}_j \,. \tag{2.54}$$

For the pair of dual coordinates

$$Q^a = e^a_i Q^i \,, \quad Q^i = E^i_a Q^a \,, \tag{2.55}$$

we found the differential equations

$$\frac{\partial Q^a}{\partial Q^i} = e^b_i D^a_b = (D^{-1})^j_i e^a_j \,, \tag{2.56}$$

$$\frac{\partial Q^i}{\partial Q^a} = (D^{-1})^b_a E^i_b = D^i_j E^j_a \tag{2.57}$$

involving the invertible matrices

$$D^b_a = \delta^b_a - T^b_{ac} Q^c \,, \quad D^j_i = \delta^j_i + T^j_{ik} Q^k \,. \tag{2.58}$$

As a result of (2.56) and (2.57) the matrices (2.58) are related to each other through relations such as $D^j_i = e^a_i (D^{-1})^b_a E^j_b$, $D^b_a = E^i_a (D^{-1})^j_i e^b_j$. Equations (2.56) and (2.57) lead to the MC stucture equations

$$D^c_a \frac{\partial E^i_b}{\partial Q^c} = E^j_a \frac{\partial E^i_b}{\partial Q^j} = T^c_{ab} E^i_c \,, \tag{2.59}$$

$$D^c_a \frac{\partial e^b_i}{\partial Q^c} = E^k_a \frac{\partial e^b_i}{\partial Q^k} = -T^b_{ac} e^c_i \,, \tag{2.60}$$

for the vielbeins, which have the dual counterparts

$$D^k_i \frac{\partial E^j_b}{\partial Q^k} = e^c_i \frac{\partial E^j_b}{\partial Q^c} = T^j_{ik} E^k_b \,, \tag{2.61}$$

$$D^k_i \frac{\partial e^b_j}{\partial Q^k} = e^c_i \frac{\partial e^b_j}{\partial Q^c} = -T^k_{ij} e^b_k \,; \tag{2.62}$$

these relations make it possible to express the matices D^b_a in terms of the vielbeins and structure constants alone. For instance, from

$$\frac{\partial E^j_a}{\partial Q^b} = T^j_{ik} E^i_a E^k_b = (D^{-1})^c_b T^d_{ca} E^j_d \tag{2.63}$$

we find

$$(D^{-1})^c_b T^d_{ca} = T^j_{ik} E^i_a E^k_b e^d_j \,. \tag{2.64}$$

In the case of strings relations (2.63) and (2.64) will simplify significantly, leading to the attractive interpretation of the D matrices as the metric tensors.

III. DUAL VIRASORO ALGEBRAS AND THE VERTEX OPERATOR

A. Coordinates and vielbeins

To illustrate the discussion in Sec. II we now consider the specific case of the Virasoro algebra.

The generators of the Virasoro algebra in the classical case, which we will consider in this section, are defined in terms of the harmonic oscillators $\{\alpha^\mu_n, \alpha^\nu_m\} = in\delta(n+m)\eta^{\mu\nu}$ and $\alpha^\mu_0 = p^\mu/\sqrt{2}$, with p^μ the center-of-mass momentum of the string and $\eta^{\mu\nu} = \mathrm{diag}(-+++\cdots)$.

The convention we use in this section is that the indices $(a,b,c,\ldots)$ from Sec. II go over to $(n,m,k,\ldots)$, while the dual indices $(i,j,k,\ldots)$ become $(r,s,t,\ldots)$. Elements of the sets $(n,m,k,\ldots)$ and $(r,s,t,\ldots)$ are arbitrary integers.

We carry our construction from Sec. II to the case of classical strings by defining the appropriate canonical pair (P_n, Q^m), which constitutes a Heisenberg algebra:

$$\{P_n, Q^m\} = \delta(n+m), \quad \{P_n, P_m\} = 0 = \{Q^n, Q^m\}. \tag{3.1}$$

Note that for the case of strings, the conventional delta function δ^m_n goes over to $\delta(n+m)$.

The form of the coordinates Q^n in this framework has been found in Refs. 10–12, where Q^n were recognized as the subsidiary constraints of the Virasoro algebra and expressed in terms of the modes of the Fubini–Veneziano coordinate:

$$Q_\mu(z) = \sqrt{2} x_\mu - i \frac{1}{\sqrt{2}} p_\mu \ln z + i \sum_{n=1}^{\infty} \left(\frac{\alpha_{\mu n} z^{-n}}{n} - \frac{\alpha^\dagger_{\mu n} z^n}{n} \right). \tag{3.2}$$

Explicitly, we have

$$Q^n = \frac{\sqrt{2}}{k\cdot p} \oint \frac{dz}{2\pi i z} z^n k \cdot Q'(z) = \begin{cases} \dfrac{2}{k\cdot p} k\cdot x \,, & \text{if } n = 0 \,, \\ \dfrac{i\sqrt{2}}{k\cdot p} \dfrac{k\cdot \alpha_n}{n} \,, & \text{otherwise} \,, \end{cases} \tag{3.3}$$

where

$$Q'_\mu(z) = Q_\mu(z) + (i/\sqrt{2}) p_\mu \ln z \,, \tag{3.4}$$

x^μ is the center-of-mass position of the string, and k^μ is some arbitrary null vector $k\cdot k = 0$ satisfying the light-cone condition $k\cdot p \neq 0$.

We now complete the picture by defining the conjugated operators P_n. It is easy to construct P_n explicitly, as was done in Refs. 12 and 2, with the result

$$P_n = (1/\sqrt{2}) \alpha^\mu_n \bar{p}^\nu \eta_{\mu\nu} \,, \tag{3.5}$$

where $\bar{p}^\mu$ is another null vector, beside k^μ, explicitly given by

$$\bar{p}^\mu \equiv p^\mu - \tfrac{1}{2} p^2 [k^\mu/(k\cdot p)] \,, \tag{3.6}$$

with the properties

$$\bar{p}^2 = 0, \quad \bar{p}\cdot k = p\cdot k \neq 0, \quad \bar{p}^\mu p_\mu = \tfrac{1}{2} p^2 \,. \tag{3.7}$$

Once the canonical pair (P_n, Q^m) is found for the case of the classical strings we can proceed as in Sec. II and construct the new non-Abelian structure. For this purpose we need to introduce the appropriate structure constants: They will be provided by the Virasoro algebra generated by the Virasoro operators, which here take the form

$$L_n = \frac{1}{2} \sum_m \alpha^\mu_m \alpha^\nu_{n-m} \eta_{\mu\nu} \,. \tag{3.8}$$

The classical Virasoro algebra is then given by

$$\{L_n, L_m\} = U^p_{nm} L_p \,, \tag{3.9}$$

with the structure constant

$$U^p_{nm} = i(n-m)\delta(n+m-p) \tag{3.10}$$

carrying a factor of i because here we are dealing with the classical algebra. One finds that the unique solution to the Jacobi identities (2.12) and (2.13) is provided by

$$T^{p}_{nm} = -im\delta(n+m-p)\,. \tag{3.11}$$

Accordingly, we can introduce the matrix D^{m}_{n} as

$$D^{m}_{n} = D(m-n) = \delta(n-m) - T^{m}_{np}Q^{-p} = \delta(n-m) - i(n-m)Q^{n-m} \tag{3.12}$$

and use it to construct the "non-Abelian" derivative $\mathscr{D}_n \equiv D^{m}_{n}P_m$. The Jacobi relations ensure that $\mathscr{D}_n$ actually satisfies the Virasoro algebra and, of course, also the relation

$$\{\mathscr{D}_n, Q^m\} = \delta(n+m) - T^{-m}_{np}Q^{-p} = D^{-m}_{n}\,. \tag{3.13}$$

The explicit calculation reveals that[12]

$$\mathscr{D}_n = P_n - T^{m}_{np}Q^{-p}P_m = \frac{1}{2}\sum_m \alpha^{\mu}_{m}\alpha^{\nu}_{n-m}\Pi_{\mu\nu}\,, \tag{3.14}$$

where

$$\Pi^{\mu\nu} = [1/(k\cdot p)](k^{\mu}\bar{p}^{\nu} + \bar{p}^{\mu}k^{\nu}) \tag{3.15}$$

is a projection operator onto the longitudinal directions on the $k-p$ plane since $\Pi^{\nu}_{\mu}k_{\nu} = k_{\mu}$ and $\Pi^{\nu}_{\mu}p_{\nu} = p_{\mu}$. Hence we recognize in $\mathscr{D}_n$ the two-dimensional Virasoro operators which generate diffeomorphisms on the $k-p$ plane. This follows from the fact that the gauge generator $\mathscr{D}_n$ is built from the Abelian canonical operators (P_n, Q^m) only and these operators are confined to the longitudinal directions. We can therefore alternatively denote $\mathscr{D}_n$ by

$$L^{\text{long}}_{n} \equiv \mathscr{D}_n = D^{m}_{n}P_m \tag{3.16}$$

and correspondingly, split the original Virasoro operators into longitudinal and transverse parts as

$$L_n = L^{\text{long}}_{n} + L^{\text{tr}}_{n} = \mathscr{D}_n + L^{\text{tr}}_{n}\,, \tag{3.17}$$

where

$$L^{\text{tr}}_{n} = \frac{1}{2}\sum_m \alpha^{\mu}_{m}\alpha^{\nu}_{n-m}\epsilon_{\mu\nu}\,, \tag{3.18}$$

with $\epsilon_{\mu\nu}$ a projection onto the directions orthogonal to the $k-p$ plane:

$$\epsilon_{\mu\nu} \equiv \eta_{\mu\nu} - \Pi_{\mu\nu}\,. \tag{3.19}$$

We note that since clearly, $\{L_n, Q^m\} = \{L^{\text{long}}_{n}, Q^m\}$, the Virasoro algebra can be extended to the form

$$\{L_n, L_m\} = U^{k}_{nm}L_k\,, \tag{3.20}$$

$$\{L_n, Q^m\} = \delta(n+m) - T^{-m}_{nk}Q^{-k} = D^{-m}_{n}\,, \tag{3.21}$$

$$\{Q^n, Q^m\} = 0\,. \tag{3.22}$$

We now have all the ingredients necessary to carry the canonical transformation (2.5) and (2.6) over to the case of strings:

$$\exp(-\Phi)\delta^{n}_{r}P_n\exp(\Phi) = e^{n}_{r}\mathscr{D}_n = e^{n}_{r}L^{\text{long}}_{n}\,, \tag{3.23}$$

$$\exp(-\Phi)\delta^{r}_{n}Q^n\exp(\Phi) = E^{r}_{n}Q^n = Q^r\,. \tag{3.24}$$

In Ref. 2 it was pointed out that the vertex operator defined by

$$V(rk,z) = \exp\left(\frac{i\sqrt{2}}{k\cdot p}rk\cdot Q(z)\right) = z^{r}\exp\left(\sum_{n=-\infty}^{\infty} irQ^n z^{-n}\right) = \sum_{m=-\infty}^{\infty} e^{m}_{r}z^m \tag{3.25}$$

is a generating function for the vielbeins e^{n}_{r}. As a consequence of relation (3.25) we find the Poisson bracket

$$\{L_n, e^{m}_{r}\} = \{L^{\text{long}}_{n}, e^{m}_{r}\} = i(m-n)e^{m-n}_{r} = -T^{m}_{nk}e^{k}_{r}\,, \tag{3.26}$$

which reproduces the MC structure equation (2.23) in this new framework. Several other consequences following from the vertex representation of the vielbein are listed in the Appendix.

It turns out that the vertex is also the generating function for the inverse vielbein E^{r}_{n}, which provides the following relation:

$$e^{n}_{r} = (r/n)E^{r}_{-n}\,. \tag{3.27}$$

This in turn ensures the second MC structure equation (2.25):

$$\{L_n, E^{r}_{m}\} = \{L^{\text{long}}_{n}, E^{r}_{m}\} = -imE^{r}_{m+n} = T^{k}_{nm}E^{r}_{k}\,. \tag{3.28}$$

We can use this information to find the vielbein representation of the D matrices. Our starting point is relation (2.58) for the inverse matrix $(D^{-1})^{n}_{m} = (D^{-1})(n-m)$; inserting the explicit structure constants we obtain

$$(D^{-1})^{n}_{m} = (D^{-1})(n-m) = (r/m)E^{-r}_{m}e^{n}_{r} = e^{n}_{r}e^{-m}_{-r}\,, \tag{3.29}$$

from which we deduce that

$$D^{n}_{m} = D(n-m) = E^{r}_{-n}E^{-r}_{m}\,. \tag{3.30}$$

Hence the D matrices are being reproduced by the standard geometric construction of the metric tensors from vielbeins.[2] The D matrices satisfy the following Poisson bracket relations with the Virasoro operators:

$$\{L_n, D^{k}_{m}\} = -T^{k}_{ml}D^{l}_{n}\,, \tag{3.31}$$

$$\{L_n, (D^{-1})^{k}_{m}\} = T^{l}_{nm}(D^{-1})^{k}_{l} - U^{k}_{nl}(D^{-1})^{l}_{m}\,. \tag{3.32}$$

Note that for $(D^{-1})(m) = (D^{-1})^{m}_{0}$ Eq. (3.32) reproduces the familiar result[13]

$$\{L_n, (D^{-1})(m)\} = -U^{m}_{nl}(D^{-1})(l) = i(m-2n)(D^{-1})(m-n)\,. \tag{3.33}$$

B. The dual operators

As we have learned in Sec. II, given the Abelian momentum operator P_n we can associate to it, by multiplication with e^{n}_{r}, a generator of the non-Abelian algebra in the dual space

$$\mathscr{D}_r \equiv P_n e^{n}_{r} = [(\bar{p}\cdot\alpha_n)/\sqrt{2}]\,e^{n}_{r}\,. \tag{3.34}$$

It follows that $Q^r = Q^{-m}E^{r}_{m}$ is the dual coordinate which, together with $\mathscr{D}_r$, satisfies

$$\{\mathscr{D}_r, Q^s\} = \delta(s+r) + T^{-s}_{rt}Q^{-t} \equiv D^{-s}_{r}\,, \tag{3.35}$$

which defines the matrix $D^{s}_{r} = \delta(r-s) - i(s-r)Q^{r-s}$ as the dual counterpart of D^{n}_{m}. The definition of the longitudinal vector vertex operator $\mathscr{D}_r$ (Ref. 7) is based on our general algebraic construction (2.47), which also provides an alternative expression in terms of the dual Abelian derivative. Namely, we introduce $P_r \equiv e^{n}_{r}\mathscr{D}_n = e^{n}_{r}D^{m}_{n}P_m$, which is conjugated to Q^s according to $\{P_r, Q^s\} = \delta(r+s)$. We then have

$$\mathscr{D}_r = D_r^s P_s \,. \tag{3.36}$$

Accordingly $\mathscr{D}_r$ satisfies an algebra dual to the original one, with identical structure constants (up to a sign), but with indices defined on the dual space.

$$\{\mathscr{D}_r,\mathscr{D}_s\} = i(s-r)P_n e^n_{r+s} = -U^t_{rs}\mathscr{D}_t \,. \tag{3.37}$$

Note that the quantities introduced above are clearly longitudinal. To introduce the transverse directions, we need the polarization vectors ϵ^i_μ, satisfying

$$k\cdot\epsilon^i = p\cdot\epsilon^i = 0, \quad \sum_{i=1}^{d-2} \epsilon^i_\mu \epsilon^i_\nu = \epsilon_{\mu\nu}, \quad \epsilon^i_\mu \epsilon^j_\nu \eta^{\mu\nu} = \delta^{ij} \,. \tag{3.38}$$

We can then follow standard procedure and define the transverse DDF operators as[14]

$$A^i(r) \equiv \sum_{n=-\infty}^{\infty} \epsilon^i\cdot\alpha_n e^n_r \,. \tag{3.39}$$

Since $\epsilon^i\cdot k = 0$ the DDF operators commute with all the longitudinal operators which are functions of the coordinate Q^n, such as vielbeins, etc.: Their basic properties involve the relation

$$\{A^i(r),A^j(s)\} = ir\delta^{ij}\delta(r+s) \tag{3.40}$$

and the fact that they commute with the Virasoro operators:

$$\{L_n,A^i(r)\} = T^k_{nm}\epsilon^i\cdot\alpha_k e^m_r - T^m_{nk}\epsilon^i\cdot\alpha_m e^k_r = 0 \,. \tag{3.41}$$

In the old dual models, property (3.41) led to the widespread use of DDF operators for defining physical states.[14,8,7]

We will now associate to the DDF operators the dual transverse Virasoro generators by the definition

$$\hat{L}^{\rm tr}_r = \frac{1}{2}\sum_{s=-\infty}^{\infty}\sum_{i=1}^{d-2} A^i(r-s)A^i(s) \,. \tag{3.42}$$

It is a matter of simple verification to prove that $\hat{L}^{\rm tr}_r$ in fact satisfies the dual Virasoro algebra

$$\{\hat{L}^{\rm tr}_r,\hat{L}^{\rm tr}_s\} = U^t_{rs}\hat{L}^{\rm tr}_t \tag{3.43}$$

as a consequence of relation (3.40). It is important to note that one can establish the following relation between two mutually dual transverse Virasoro operators:

$$\hat{L}^{\rm tr}_r = D^s_r e^n_s L^{\rm tr}_n = e^m_r (D^{-1})^n_m L^{\rm tr}_n \,, \tag{3.44}$$

where we have used the vielbein relation[2]

$$e^k_{r-s}e^m_s = D^t_r e^{m+k}_t \,, \tag{3.45}$$

which follows from the vertex representation (3.25). We can rewrite expression (3.36), defining $\mathscr{D}_r$ in a new way:

$$\begin{aligned}\mathscr{D}_r &= D^s_r P_s + D^s_r e^n_r L^{\rm tr}_n - \hat{L}^{\rm tr}_r \\ &= \Phi_r + T^s_{rt}Q^{-t}\Phi_s - \frac{1}{2}\sum_{s=-\infty}^{\infty} A^i(r-s)A^i(s) \\ &= D^s_r\Phi_s - \hat{L}^{\rm tr}_r \,,\end{aligned} \tag{3.46}$$

where we have introduced the operator $\Phi_r \equiv L_n e^n_r = P_r + e^n_r L^{\rm tr}_n$, analogous to Φ_i defined in Sec. II C. Here Φ_r is clearly an extension of the longitudinal operator P_r which also includes the transverse directions in a way that preserves the dual Heisenberg algebra [compare with Eq. (2.47)]

$$\{\Phi_r,Q^s\} = \delta(r+s), \quad \{\Phi_r,\Phi_s\} = 0 = \{Q^r,Q^s\} \tag{3.47}$$

and ensures that Φ_r, as well as Q^s, have zero Poisson brackets with $A^i(r)$.

We will now define the dual Virasoro operators by summing the purely longitudinal operator $\mathscr{D}_r$ and the transverse operator $\hat{L}^{\rm tr}_r$. The obvious result from (3.46) is

$$\hat{L}_r = \mathscr{D}_r + \hat{L}^{\rm tr}_r = D^s_r\Phi_s \,, \tag{3.48}$$

which corresponds to $\hat{\phi}_i$ in Sec. II C. Equation (3.48) defines the decomposition of the dual Virasoro operator into longitudinal and transverse modes, analogous to the decomposition of the original modes in (3.17). One fundamental difference is that we group together operators satisfying the Poisson bracket relations (3.37) and (3.43), which differ by the minus sign. However, the relation $\{\mathscr{D}_r,\hat{L}^{\rm tr}_s\} = -U^t_{rs}\hat{L}^{\rm tr}_t$ ensures consistency and leads to the following dual Virasoro algebra:

$$\{\hat{L}_r,\hat{L}_s\} = -U^t_{rs}\hat{L}_t \,. \tag{3.49}$$

We note that although we recognized $\mathscr{D}_r$ as the longitudinal part of the dual Virasoro operator, $\mathscr{D}_r$ does not commute with the transverse part as a result of the presence of vielbeins in $\hat{L}^{\rm tr}_r$.

There exists a relation inverse to (3.44) given by

$$L^{\rm tr}_n = D^m_n E^r_m \hat{L}^{\rm tr}_r = E^s_n (D^{-1})^r_s \hat{L}^{\rm tr}_r \,. \tag{3.50}$$

The proof involves the relation (Ref. 2) $E^r_m E^s_{n-m} = D^k_n E^{r+s}_k$ and

$$\begin{aligned}L^{\rm tr}_n &= \frac{1}{2}\sum_{m=-\infty}^{\infty} (\epsilon^i\cdot\alpha_m)(\epsilon^i\cdot\alpha_{n-m}) \\ &= \frac{1}{2}\sum_{m=-\infty}^{\infty} A^i(s)A^i(r)E^{-s}_m E^{-r}_{n-m} \,.\end{aligned} \tag{3.51}$$

Hence the longitudinal Virasoro operator $L^{\rm long}_n = \mathscr{D}_n$ can be rewritten analogously to (3.46) as

$$\begin{aligned}L^{\rm long}_n &= D^m_n P_m = D^m_n P_m + D^m_n E^r_m \hat{L}^{\rm tr}_r - L^{\rm tr}_n \\ &= D^m_n \Phi_m - L^{\rm tr}_n \,,\end{aligned} \tag{3.52}$$

where we have defined [as in (2.49)]

$$\Phi_n \equiv P_n + E^r_n \hat{L}^{\rm tr}_r = E^r_n(\mathscr{D}_r + \hat{L}^{\rm tr}_r) = E^r_n \hat{L}_r \,. \tag{3.53}$$

Equation (3.53) therefore leads to

$$L_n = L^{\rm long}_n + L^{\rm tr}_r = D^m_n \Phi_m \,, \tag{3.54}$$

which can be inverted to find an alternative expression for Φ_m of the form $\Phi_n = (D^{-1})^m_n L_m$. Expression (3.54) is useful in deriving the Heisenberg algebra dual to (3.47):

$$\begin{aligned}&\{\Phi_n,Q^m\} = (D^{-1})^k_n\{L_k,Q^m\} = \delta(n+m) \,, \\ &\{\Phi_n,\Phi_m\} = 0 = \{Q^n,Q^m\} \,,\end{aligned} \tag{3.55}$$

where we have used the bracket relations (3.21) and (3.32). It is also clear that Φ_n has a zero Poisson bracket with the DDF operators $A^i(r)$.

We conclude that $\Phi_n = E^r_n \hat{L}_r$ is an extension of the "Abelian derivative" $P_n = E^r_n \mathscr{D}_r$, which includes the transverse directions in a way that does not change the bracket relations with the longitudinal quantities depending on Q^m, but ensures commutativity with the transverse DDF operators. Note that exactly the same can be said about

$\Phi_r = e_r^n L_n$ being an extension of $P_r = e_r^n \mathscr{D}_n$, as well as its relation to the DDF operators.

Let us now consider the Poisson brackets

$$\{L_n, \Phi_m\} = \{D_n^k \Phi_k, \Phi_m\} = T_{nm}^k \Phi_k , \tag{3.56}$$

$$\{\hat{L}_r, \Phi_s\} = -T_{rs}^t \Phi_t , \tag{3.57}$$

which lead to the interesting results

$$\{L_n, \Phi_{m=0}\} = 0, \quad \{\hat{L}_r, \Phi_{s=0}\} = 0 \tag{3.58}$$

for any n and r. Substituting $E_0^r = \delta(r)$ and $\mathscr{D}_{r=0} = \frac{1}{4}p^2$ we therefore find that

$$\Phi_{m=0} = D^{-1}(n) L_n = E_0^{-r}(\mathscr{D}_r + \hat{L}_r^{\mathrm{tr}}) = \frac{1}{4}p^2 + \frac{1}{2}\sum_{s=-\infty}^{\infty} A^i(-s) A^i(s) \tag{3.59}$$

commutes with all Virasoro operators and is in fact the classical analog of the Brink–Olive operator E (here shifted by L_0) found in Ref. 13. Similarly, $\Phi_{r=0}$ can be easily recognized as

$$\Phi_{r=0} = e_{r=0}^n L_n = L_0 \tag{3.60}$$

as a result of $e_0^n = \delta(n)$.[2]

We can summarize the relevant formulas describing Φ's operators in the following way:

$$\Phi_n = E_n^r \hat{L}_r = E_n^r D_r^s \Phi_s = (D^{-1})_n^m L_m , \tag{3.61}$$

$$\Phi_r = e_r^n L_n = e_r^n D_n^k \Phi_k = (D^{-1})_r^s \hat{L}_s , \tag{3.62}$$

which clearly correspond to relations (2.53) and (2.54) for the special case of the Virasoro algebra. In a sense the expressions $L_n^{\mathrm{long}} = D_n^m P_m$ and $\mathscr{D}_r = D_r^s P_r$ are generalized to $L_n = D_n^m \Phi_m$ and $\hat{L}_r = D_r^s \Phi_r$ by incorporating the transverse directions. As a consequence of Eqs. (3.61) and (3.62) we can cast the relation between Φ_r and Φ_n in the form analogous to (2.19) and (2.20):

$$\Phi_r = \frac{\partial Q^n}{\partial Q^r} \Phi_n , \quad \Phi_n = \frac{\partial Q^r}{\partial Q^n} \Phi_r . \tag{3.63}$$

The zero modes are simply

$$\Phi_{r=0} = L_0, \quad \Phi_{m=0} = \hat{L}_0 , \tag{3.64}$$

with $\Phi_{m=0}$ commuting with all the Virasoro operators L_n. Since $\hat{L}_0 = \frac{1}{4}p^2 + \frac{1}{2}\Sigma_{-\infty}^{\infty} A^i(-s) A^i(s)$, we note that $\Phi_{m=0}$ is practically the counting operator of the transverse modes only. Meanwhile, $\Phi_{r=0} = \frac{1}{4}p^2 + \frac{1}{2}\Sigma_{m\neq 0} \alpha_{-m}^{\mu} \alpha_m^{\nu} \eta_{\mu\nu}$ counts all modes. This asymmetry between the above zero-modes of mutually dual operators will have an important consequence for the determination of the physical modes, as we will see in Sec. IV.

The formalism clearly exhibits the duality, which is most compactly expressed by

$$L_n \leftrightarrow \hat{L}_r, \quad T_{nm}^k \leftrightarrow -T_{r-s}^{-t}, \quad e_r^m \leftrightarrow E_m^r . \tag{3.65}$$

IV. BRS COHOMOLOGY IN TERMS OF ABELIAN STRING OPERATORS

In Sec. III we have seen how geometric construction based on the rotation of canonical coordinates by vielbeins has led to two mutually dual classical Virasoro operators. In this section we show how to extend this construction to the quantum case and then we turn to the investigation of its physical applications. Two related technical issues have to be addressed in this connection; one is how to deal with the normal ordering and another is how to add the ghosts in order to take care of the anomaly terms.

In the classical formalism all extended phase space variables including ghosts enter the formalism of Sec. II on equal footing with the canonical coordinates (P_n, Q^m). For example, the generator of Φ of the canonical mapping can be extended to $\Phi \equiv Q^{-n} T_n$, with T_n (Refs. 1 and 10) defined in terms of the variables (P_n, Q^n) and the BFV ghosts defined as $T_n \equiv -T_{nm}^k (Q^{-m} P_k + \eta^{-m} \mathscr{P}_k)$. As shown in Ref. 1, the BFV ghosts transform as

$$\begin{aligned} &\exp(-\Phi) \delta_r^n \mathscr{P}_n \exp(\Phi) = e_r^n \mathscr{P}_n \equiv \mathscr{P}_r , \\ &\exp(-\Phi) \delta_n^r \eta^n \exp(\Phi) = E_n^r \eta^n \equiv \eta^r . \end{aligned} \tag{4.1}$$

If we define an Abelian nilpotent operator as

$$\Omega_A \equiv \eta^{-n} P_n , \tag{4.2}$$

it therefore transforms, under the canonical mapping, as

$$\exp(-\Phi) \Omega_A \exp(\Phi) = \eta^{-n}(P_n - Q^{-m} T_{nm}^k P_k - \eta^{-m} T_{nm}^k \mathscr{P}_k) \equiv \Omega^{\mathrm{long}} . \tag{4.3}$$

The operator (4.3) looks similar to the complete quantum BRS operator:

$$\begin{aligned} \Omega &= :\eta^{-n}[(L_n - \delta_{n,0}) - \eta^{-m} T_{nm}^k \mathscr{P}_k]: \\ &= :\eta^{-n}[(L_n^{\mathrm{tr}} - \delta_{n,0}) + P_n - Q^{-m} T_{nm}^k P_k \\ &\quad - \eta^{-m} T_{nm}^k \mathscr{P}_k]: . \end{aligned} \tag{4.4}$$

We note that the bosonic Abelian operators (P_n, Q^n) associated with the longitudinal directions and the ghosts $(\eta^n, \mathscr{P}_n)$ enter expressions (4.3) and (4.4) in a very symmetric way. This suggests that the Parisi–Sourlas supersymmetry can be used to provide an understanding of the structure of the BRS cohomology classes via a cancellation between the above unphysical modes.

Our strategy will be to recover the quantum versions of the operators introduced in Sec. III directly from the BRS operator. This automatically takes care of the normal ordering problem, as well as ensures correct incorporation of the ghost variables. The strong resemblance between operators (4.3) and (4.4) suggests that the basic ingredients of our geometric construction can be used to explain the structure of the BRS operator and analyze its action in the geometrical setting. As we will show below, our approach will also shed light on the structure of the physical subspace and the associated gauge orbits.

Therefore, our starting point is the BRS quantization procedure, with the physical states defined by the BRS condition

$$\Omega |\mathrm{phys}\rangle = 0 . \tag{4.5}$$

From the nilpotency of Ω it follows that the physical states are defined up to BRS-exact states:

$$|\mathrm{phys}\rangle \sim |\mathrm{phys}\rangle + \Omega |\mathrm{something}\rangle . \tag{4.6}$$

Accordingly, physical states are represented by the cohomology classes of Ω.

In Ref. 5, Kato and Ogawa applied BRS symmetry to the study of covariant quantization of string theories. Conse-

quently, the structure of the physical states was determined and identified with the purely transverse (DDF) states.[14] Kato and Ogawa's analysis was based on the formalism introduced by KO (Ref. 4) in which the ghosts and unphysical modes form a so-called quartet and disappear effectively from the physical subspace in a Fock space representation. Note that in Ref. 5 construction of the quartet did not identify uniquely the members of the quartet, as will be done here, but involved a mathematical induction proof based on expansion in a scaling parameter defined specifically for this purpose.

Further clarification of the structure of the physical states, which used results contained in Ref. 5, was later given in Ref. 15.

There have been several other independent studies of the BRS cohomology associated with the string theory. A mathematical study of the semi-infinite cohomology of the Virasoro algebras was presented in Ref. 16. In the less formal development[17] the physical state, as defined by the old no-ghost theorem,[8] was put in correspondence with the state defined by (4.5) and embedded in the enlarged Fock space.

A different and elegant analysis of the cohomology of Ω was presented by Freeman and Olive in Ref. 9 and based on the special Ω-exact Hermitian operators: The crucial observation here was that the physical state must be in the kernel of these operators.

For more recent studies of the BRS cohomology the reader is referred to Refs. 18 and 19. Here we describe the KO quartet procedure in terms of the mutually dual Abelian operators and show that they define a proper conceptual framework for discussing the BRS cohomology of the string.

Let us first introduce quantum versions of the fundamental objects presented in Secs. II and III.

We start with the gauge-fixing conditions Q^n defined in terms of the harmonic oscillators a_n^μ satisfying $[a_n^\mu, a_m^\nu] = n\delta(n+m)\eta^{\mu\nu}$ and given by

$$Q^0 = 2i\frac{k\cdot x}{k\cdot p}, \quad Q^n = -\frac{\sqrt{2}}{n}\frac{k\cdot a_n}{k\cdot p}, \quad n\neq 0; \tag{4.7}$$

they enter the extended Virasoro algebra, which this time contains an anomaly term:

$$\begin{aligned}[L_n, L_m] &= U^p_{nm} L_p + (D/12)\, n^3\delta(n+m) \\ &\quad - [(D-24)/12] n\delta(n+m), \\ [L_n, Q^m] &= \delta(n+m) - T^{-m}_{np} Q^{-p}, \quad [Q^n, Q^m] = 0.\end{aligned} \tag{4.8}$$

The structure constants U^p_{nm} and T^p_{nm} are given by

$$\begin{aligned}U^p_{nm} &= (n-m)\delta(n+m-p), \\ T^p_{nm} &= -m\delta(n+m-p).\end{aligned} \tag{4.9}$$

As before, the momentum operator defined as $P_n = (1/\sqrt{2})\,\bar{p}\cdot a_n$ is conjugate to the Q^m's:

$$[P_n, Q^m] = \delta(n+m), \quad [P_n, P_m] = 0 = [Q^n, Q^m]. \tag{4.10}$$

Moreover, the formalism introduces two bridge variables between the Abelian and non-Abelian quantities: the vielbein e^n_r and its inverse E^r_n. Both e^n_r and E^r_n depend only on the variables Q^n and the constants T^k_{nm} and turn out to be moments of the vertex operator.[2] The vielbeins satisfy the fundamental MC equations

$$\begin{aligned}[L_n, e^m_r] &= [L^{\text{long}}_n, e^m_r] = -T^m_{nk} e^k_r, \\ [L_n, E^r_m] &= T^k_{nm} E^r_k,\end{aligned} \tag{4.11}$$

which are essential, as we have seen, for the above transition from the Abelian to the non-Abelian basis.

A. Quartet mechanism

Our strategy will be to associate the unphysical modes with the KO quartet. This can be illustrated easily while working with the quantum theory described by the Abelian BRS operator $\Omega_A = \eta^{-n} P_n$ from (4.2), which is clearly nilpotent and quadratic in the mode variables. The procedure can be performed along the lines of Refs. 5 and 6. The relevant quantities are described by

$$\begin{aligned}&[\Omega_A, Q^n] = \eta^n, \quad \{\Omega_A, \eta^n\} = 0, \quad \{\Omega_A, \mathscr{P}_n\} = P_n, \\ &[\Omega_A, P_n] = 0, \quad [\Omega_A, a^i_n] = 0, \quad (i = 1,\ldots,D-2).\end{aligned} \tag{4.12}$$

The quartet consists of the modes $(P_n, Q^n, \eta^n, \mathscr{P}_n)$. Grouping the quartet together with the transverse modes a^i_n in one multiplet $\phi_\rho = (a^i_n, Q^n, P_n, \eta^n, \mathscr{P}_n)$ we can define the metric tensor $\eta_{\rho\sigma}$ as

$$\eta_{\rho\sigma} \equiv [\phi_\rho, \phi^\dagger_\sigma], \tag{4.13}$$

where ρ and σ label the $D-2$ transverse directions as well as the unphysical gauge coordinates of the quartet.

Explicitly, the metric tensor takes the form

$$\eta_{\rho\sigma} = \begin{array}{c} \\ a^i_n \\ Q^n \\ P_n \\ \eta_n \\ \mathscr{P}_n \end{array} \begin{array}{c} \begin{array}{ccccc} a^j_{-m} & -Q^{-m} & P_{-m} & \eta^{-m} & \mathscr{P}_{-m} \end{array} \\ \begin{pmatrix} \delta^{ij}\delta_{nm} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\delta_{nm} & 0 & 0 \\ 0 & -\delta_{nm} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{nm} \\ 0 & 0 & 0 & \delta_{nm} & 0 \end{pmatrix} \end{array}. \tag{4.14}$$

Note that Q^n is anti-Hermitian and therefore, $Q^{n\dagger} = -Q^{-n}$. We now proceed as in Refs. 4 and 5 by introducing the family of projection operators. The projection operator onto the subspace spanned by the physical transverse modes is given by

$$P^{(0)} = \sum_k (a^{i_1\dagger}_{n_1} \cdots a^{i_k\dagger}_{n_k} |0\rangle\langle 0| a^{i_k}_{n_k} \cdots a^{i_1}_{n_1}). \tag{4.15}$$

The remaining projection operators with $n \geqslant 1$ project onto the nth unphysical mode sector and are defined recursively as

$$\begin{aligned}P^{(n)} = (1/n)(&-P^\dagger_k P^{(n-1)} Q^k - Q^{k\dagger} P^{(n-1)} P_k \\ &+ \eta^{k\dagger} P^{(n-1)} \mathscr{P}_k + \mathscr{P}^\dagger_k P^{(n-1)} \eta^k).\end{aligned} \tag{4.16}$$

Each of these projection operators (with $n \geqslant 1$) is BRS exact as a result of

$$P^{(n)} = \{\Omega_A, R^{(n)}\}, \tag{4.17}$$

with

$$R^{(n)} = -(1/n)(\mathscr{P}_k^{\dagger} P^{(n-1)} Q^k + Q^{k\dagger} P^{(n-1)} \mathscr{P}_k) . \tag{4.18}$$

Using the assumption of asymptotic completeness we can rewrite the solution of the Abelian BRS condition $\Omega_A|\psi\rangle = 0$ as

$$\begin{aligned} |\psi\rangle &= \sum_{n=0}^{\infty} P^{(n)}|\psi\rangle \\ &= P^{(0)}|\psi\rangle + \sum_{n=1}^{\infty} \{\Omega_A, R^{(n)}\}|\psi\rangle \\ &= P^{(0)}|\psi\rangle + \Omega_A|\chi\rangle . \end{aligned} \tag{4.19}$$

This decomposition is the main result of the KO quartet mechanism and shows how the unphysical quartet modes decouple from the physical component $P^{(0)}|\psi\rangle$ as BRS-exact states. Although Ω_A is given by $\Omega_A = [\Omega, :\Sigma_m Q^{-m} P_m :]$, relating the cohomology of the Abelian BRS operator Ω_A to the cohomology of the original BRS operator Ω is not simple.

Here we will instead recover the Abelian structure inside the BRS operator Ω by using the vielbeins which satisfy the MC equations (4.11). Then we will show how the Abelian quantities from Sec. III form two KO quartets.

We now start the main construction of this section. We first introduce the two Abelian quantities[2]

$$Q^r \equiv Q^{-m} E^r_m, \quad \mathscr{P}_r \equiv \mathscr{P}_n e^n_r, \tag{4.20}$$

which are clearly self-commuting (anticommuting). We again need to introduce the transverse DDF operators[14] $A^i(r) \equiv \Sigma_{n=-\infty}^{\infty} \epsilon^i \cdot \alpha_n e^n_r$, which have the same form as in (3.39). Note that since ϵ^i is orthogonal to k, normal ordering is unnecessary in the definition of $A^i(r)$. As before, the DDF operators commute with the vielbeins and the Virasoro operators. One finds easily that the DDF operators satisfy the quantum version of the algebra (3.40) (Ref. 2):

$$[A^i(r), A^j(s)] = r\delta^{ij}\delta(r+s) . \tag{4.21}$$

We will embed our first quartet construction in the dual space.[2] Explicitly, we form the quartet as follows:

$$\begin{aligned} &[\Omega, Q^r] \equiv \eta^r, \quad \{\Omega, \eta^r\} = 0, \quad \{\Omega, \mathscr{P}_r\} \equiv \tilde{\Phi}_r, \\ &[\Omega, \tilde{\Phi}_r] = 0, \quad [\Omega, A^i(r)] = 0 \quad (i = 1, \ldots, D-2) . \end{aligned} \tag{4.22}$$

One finds that

$$\begin{aligned} \eta^r &= \eta^{-n} E^r_n , \\ \tilde{\Phi}_r &= (\mathscr{L}_n - \delta_{n,0}) e^n_r + \mathscr{P}_n \eta^{-m} T^n_{mk} e^k_r , \end{aligned} \tag{4.23}$$

where $\mathscr{L}_n - \delta_{n,0} \equiv \{\Omega, \mathscr{P}_n\}$ are the modified BRS-exact Virasoro operators that satisfy the centerless algebra for the nilpotent Ω. It is interesting to note that the fermionic members η^r and $\mathscr{P}_r$ also appear in the Abelization procedure of Batalin and Fradkin[20]; see, also the discussion in Ref. 2.

We can introduce a more economical expression for $\tilde{\Phi}_r$ by introducing the normal ordering of the first and second terms in (4.23) while taking into account

$$L_n e^n_r = e^n_r L_n - T^n_{nk} e^k_r ,$$

$$\mathscr{P}_n \eta^{-m} T^n_{mk} e^k_r = -\eta^{-m} \mathscr{P}_n T^n_{mk} e^k_r + T^n_{nk} e^k_r . \tag{4.24}$$

Clearly, the constants on the rhs of Eqs. (4.24) cancel and we arrive at

$$\tilde{\Phi}_r = :((L_n - \delta_{n,0}) - \eta^{-m} \mathscr{P}_k T^k_{nm}) e^n_r : . \tag{4.25}$$

Remarkably, (4.25) differs only by the normal ordering from the classical expression $\Phi_r + \eta^{-s}\{\Phi_r, \mathscr{P}_s\}$ for the "pure-gauge" Abelian operator found in Ref. 2.

By definition, $\tilde{\Phi}_r$ is also the BRS-exact operator; furthermore, we have

$$[\tilde{\Phi}_r, \tilde{\Phi}_s] = \{\Omega, [\tilde{\Phi}_r, \mathscr{P}_s]\} = 0 \tag{4.26}$$

since

$$[\tilde{\Phi}_r, \mathscr{P}_s] = -T^n_{mk} e^k_s \mathscr{P}_n e^m_r + T^n_{mk} e^k_s \mathscr{P}_n e^m_r = 0$$

as a result of the MC equations (4.11). One can also show easily that $[\tilde{\Phi}_r, \eta^s] = 0$ and

$$[\tilde{\Phi}_r, Q^s] = \delta(r+s) . \tag{4.27}$$

Hence the Heisenberg algebra originally introduced in Sec. II has maintained its form despite the incorporation of transverse direction in Sec. III and the incorporation of BFV ghosts and normal ordering in this section.

Similarly, we can find the way to generalize the Heisenberg algebra (3.55) or (3.10). This will amount to constructing the dual KO quartet. Obviously, we have to start with the original gauge-fixing operator Q^n and proceed as follows:

$$\begin{aligned} &[\Omega, Q^n] = D^{-n}_m \eta^{-m} \equiv \hat{\eta}^n, \quad \{\Omega, \hat{\eta}^n\} = 0 , \\ &\{\Omega, \hat{\mathscr{P}}_n\} \equiv \tilde{\Phi}_n , \quad [\Omega, \tilde{\Phi}_n] = 0 , \end{aligned} \tag{4.28}$$

where we have defined the variable conjugated to $\hat{\eta}^n$ as

$$\hat{\mathscr{P}}_n \equiv (D^{-1})^k_n \mathscr{P}_k . \tag{4.29}$$

Clearly, $\{\hat{\mathscr{P}}_n, \hat{\eta}^m\} = \delta(n+m)$. This definition of $\hat{\mathscr{P}}_n$ leads to the expression

$$\begin{aligned} \tilde{\Phi}_n = (D^{-1})^k_n (\mathscr{L}_k - \delta_{k,0}) + T^l_{mn} (D^{-1})^k_l \eta^{-m} \mathscr{P}_k \\ - U^k_{ml} (D^{-1})^l_n \eta^{-m} \mathscr{P}_k . \end{aligned} \tag{4.30}$$

Using similar considerations as in (4.24) we arrive at the more simple normal ordered expression

$$\tilde{\Phi}_n = :(D^{-1})^k_n (L_k - \delta_{k,0}) + T^l_{mn} (D^{-1})^k_l \eta^{-m} \mathscr{P}_k : , \tag{4.31}$$

where in the first term we recognize the normal ordered version of Φ_n from (3.61). Hence, again, our construction yields the ghost and quantum extension of the basic quantities from Sec. III.

It is straightforward to verify that

$$[\tilde{\Phi}_n, \tilde{\Phi}_m] = 0 , \quad [\tilde{\Phi}_n, \hat{\mathscr{P}}_m] = 0 , \quad [\tilde{\Phi}_n, \hat{\eta}^m] = 0 , \tag{4.32}$$

and, of course, $[\tilde{\Phi}_n, Q^m] = \delta(n+m)$. The first identity in (4.32) follows from $[\tilde{\Phi}_n, \tilde{\Phi}_m] = \{\Omega, [\tilde{\Phi}_n, \hat{\mathscr{P}}_m]\}$, which is a consequence of the nilpotency of Ω and the technical identity

$$\begin{aligned} U^k_{pl} (D^{-1})^p_m (D^{-1})^l_n + T^l_{pm} (D^{-1})^k_l (D^{-1})^p_n \\ - T^l_{pn} (D^{-1})^k_l (D^{-1})^p_m = 0 . \end{aligned} \tag{4.33}$$

For the special case $n = 0$ we obtain, from (4.31),

$$\tilde{\Phi}_{m=0} = :(D^{-1})(k)(L_k - \delta_{k,0}):, \tag{4.34}$$

which, of course, must satisfy $[\tilde{\Phi}_n, \tilde{\Phi}_{m=0}] = 0$: Together with the technical identity (4.33), this implies $[L_n, \tilde{\Phi}_{m=0}] = 0$ for the nilpotent BRS operator.

This can also be proved independently. Starting with the basic definition (4.34) and taking care of the normal ordering one can prove the familiar result[13,21]

$$[L_n, \tilde{\Phi}_0] = [(D-26)/12](n^3 - n)D^{-1}(-n). \tag{4.35}$$

Hence for the critical dimension the operator $\tilde{\Phi}_0$ commutes with all the Virasoro operators as the classical operator Φ_0 in Sec. III.

It is now easy to adapt the quartet mechanism using the operators introduced above. Clearly, our discussion suggests the existence of two quartets containing the modes $(\tilde{\Phi}_r, Q^r, \eta^r, \mathscr{P}_r)$ and $(\tilde{\Phi}_n, Q^n, \hat{\eta}^n, \hat{\mathscr{P}}_n)$. The metric tensor can be defined in the same fashion as above:

$$\eta_{\rho\sigma} \equiv [\phi_\rho, \phi_\sigma^\dagger\}, \tag{4.36}$$

with the transverse DDF modes $A^i(r)$ being placed in one multiplet $\phi_\rho = (A^i(r), \tilde{\Phi}_r, Q^r, \eta^r, \mathscr{P}_r)$ or $\phi_\rho = (A^i(r), \tilde{\Phi}_n, Q^n, \hat{\eta}^n, \hat{\mathscr{P}}_n)$ with the members of the dual quartets.

For the quartet in the dual space labeled by r we obtain

$$\eta_{\rho\sigma} = \begin{array}{c} \\ A^i_r \\ Q^r \\ P_r \\ \eta^r \\ \mathscr{P}_r \end{array} \begin{array}{c} \begin{array}{ccccc} A^j_{-s} & -Q^{-s} & P_{-s} & \eta^{-s} & \mathscr{P}_{-s} \end{array} \\ \begin{pmatrix} r\delta^{ij}\delta_{rs} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\delta_{rs} & 0 & 0 \\ 0 & -\delta_{rs} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{rs} \\ 0 & 0 & 0 & \delta_{rs} & 0 \end{pmatrix} \end{array}. \tag{4.37}$$

Now Q^r is anti-Hermitian as a result of $Q^{n\dagger} = -Q^{-n}$ and $(e_r^n)^\dagger = e_{-r}^{-n}$. Correspondingly, we carry the construction of the projection operators (4.15) and (4.16) to the dual space defining

$$P(0) = \sum_k \left(\frac{A^{i_1\dagger}(r_1)}{\sqrt{r_1}} \cdots \frac{A^{i_k\dagger}(r_k)}{\sqrt{r_k}} |0\rangle \times \langle 0| \frac{A^{i_k}(r_k)}{\sqrt{r_k}} \cdots \frac{A^{i_1}(r_1)}{\sqrt{r_1}} \right) \tag{4.38}$$

and for $r \geqslant 1$,

$$P^r = (1/r)(-\tilde{\Phi}_s^\dagger P^{r-1} Q^s - Q^{s\dagger} P^{r-1} \tilde{\Phi}_s + \eta^{s\dagger} P^{r-1} \mathscr{P}_s + \mathscr{P}_s^\dagger P^{r-1} \eta^s). \tag{4.39}$$

Each of the projection operators (4.15) and (4.16) (with $r \geqslant 1$) is BRS exact as a result of

$$P^r = \{\Omega, R^r\}, \tag{4.40}$$

with

$$R^r = -(1/r)(\mathscr{P}_s^\dagger P^{r-1} Q^s + Q^{s\dagger} P^{r-1} \mathscr{P}_s). \tag{4.41}$$

Using the expansions $e_r^n = \delta(n-r) + \cdots$ and $E_n^r = \delta(n+r) + \cdots$ (Ref. 2) in Eqs. (4.38) and (4.39) we see that the completeness assumed above for the original set of projection operators carries over to the dual Fock space projection operators. Hence we can rewrite the solution of the physical state condition $\Omega|\psi\rangle = 0$ as

$$|\psi\rangle = \sum_{r=0}^{\infty} P^r|\psi\rangle = P(0)|\psi\rangle + \sum_{r=1}^{\infty} \{\Omega, R^r\}|\psi\rangle = P(0)|\psi\rangle + \Omega|\chi\rangle. \tag{4.42}$$

This mechanism works the same way for the quartet $(\tilde{\Phi}_n, Q^n, \hat{\eta}^n, \hat{\mathscr{P}}_n)$; the only difference is the form of the projection operators:

$$P^n = (1/r)(-\tilde{\Phi}_m^\dagger P^{n-1} Q^m - Q^{m\dagger} P^{n-1} \tilde{\Phi}_m + \hat{\eta}^{m\dagger} P^{n-1} \hat{\mathscr{P}}_m + \hat{\mathscr{P}}_m^\dagger P^{n-1} \hat{\eta}^m), \tag{4.43}$$

which, again, for $n \geqslant 1$ are BRS exact as a result of

$$P^n = \{\Omega, R^n\}, \tag{4.44}$$

with

$$R^n = -(1/n)(\hat{\mathscr{P}}_m^\dagger P^{n-1} Q^m + Q^{m\dagger} P^{n-1} \hat{\mathscr{P}}_m). \tag{4.45}$$

The final result (4.42) can now be easily obtained along the same line as above.

The above Fock space construction is based on modes with positive and negative frequencies. For $r = 0$ we have $\tilde{\Phi}_{r=0} = \mathscr{L}_0 - 1$, which is a BRS-exact operator and vanishes on the physical states as a result of the mass-shell condition. For the special case $n = 0$ we take $\hat{\mathscr{P}}_0 \equiv ((D^{-1})_0^k - \delta_{k,0})\mathscr{P}_k$, where we have shifted $(D^{-1})(0) \to (D^{-1})(0) - 1 = (D^{-1})(k)D(k)$ and where the last quantity annihilates vacuum. This leads to the Freeman–Olive operator[9]:

$$\hat{E} \equiv \{\Omega, \hat{\mathscr{P}}_{n=0}\} = :(D^{-1})(k)(L_k - \delta_{k,0}): - \mathscr{L}_0. \tag{4.46}$$

Since (4.46) is a BRS-exact operator which counts longitudinal states, we see that any eigenstate will automatically belong to the trivial cohomology classes of Ω.[9]

B. Homotopy construction

It is now easy to put forward the homotopy construction responsible for the successful working of the quartet mechanism. Inspired by the discussion given by Henneaux[18] we now introduce, for the quartet $(\tilde{\Phi}_r, Q^r, \eta^r, \mathscr{P}_r)$ in the dual space, a "conjugated" BRS charge

$$\Omega^\dagger \equiv Q^r \mathscr{P}_{-r}, \tag{4.47}$$

where "conjugation" is used from the point of view of symplectic geometry and not hermiticity.

Clearly, $\Omega^\dagger$ is itself nilpotent and its anticommutator with Ω yields, according to (4.22),

$$\{\Omega, \Omega^\dagger\} = Q^r \tilde{\Phi}_{-r} + \eta^r \mathscr{P}_{-r} = :Q^r \tilde{\Phi}_{-r}: + :\eta^r \mathscr{P}_{-r}: \equiv \mathscr{N}, \tag{4.48}$$

where the "Hermitian" (in the sense of the above conjugation) operator $\mathscr{N}$ counts the modes of ghosts, as well as the unphysical modes associated with excitations generated by Q^r and $\tilde{\Phi}_{-r}$. In fact, $\mathscr{N}$ is an extension of the conventional ghost number operator which includes these extra modes.

Let us now consider the physical state $|\psi\rangle$, which is also an eigenstate of $\mathscr{N}$ with the nonzero eigenvalue n:

$$\Omega|\psi\rangle = 0, \tag{4.49}$$

$$\mathcal{N}|\psi\rangle = n|\psi\rangle\ . \tag{4.50}$$

A simple calculation shows that $\mathcal{N}\Omega^\dagger|\psi\rangle = \Omega^\dagger\mathcal{N}|\psi\rangle = n\Omega^\dagger|\psi\rangle$: This result leads to interpretation of $\Omega^\dagger|\psi\rangle$ as a supersymmetric partner of $|\psi\rangle$ with the same n eigenvalue. Hence we have a doublet

$$\cdot\ |\psi\rangle \underset{\Omega}{\overset{\Omega^\dagger}{\rightleftarrows}} \Omega^\dagger|\psi\rangle \tag{4.51}$$

associated with each nonzero eigenvalue of $\mathcal{N}$.

Analogy with supersymmetric quantum mechanics[22] is at this point completely clear. The origin of this analogy is the supersymmetric algebra satisfied by Ω, $\Omega^\dagger$, and $\mathcal{N}$:

$$\begin{aligned} &\{\Omega,\Omega\} = \{\Omega^\dagger,\Omega^\dagger\} = 0\ , \\ &[\Omega,\mathcal{N}] = [\Omega^\dagger,\mathcal{N}] = 0\ , \quad \{\Omega,\Omega^\dagger\} = \mathcal{N}\ . \end{aligned} \tag{4.52}$$

Accordingly, Ω and $\Omega^\dagger$ play the role of supersymmetry generators and $\mathcal{N}$ plays the role of the Hamiltonian.

As elementary supersymmetric quantum mechanics teaches us, we will have pairing of the states corresponding to the nonzero eigenvalues of $\mathcal{N}$, with the exception of the zero eigenvalue. However, since $\mathcal{N}$ is a BRS-exact operator it follows that

$$\Omega\Omega^\dagger|\psi\rangle = \mathcal{N}|\psi\rangle = n|\psi\rangle \tag{4.53}$$

and for $n \neq 0$ the state $|\psi\rangle$ is BRS exact. Consequently, the physical states must have $n = 0$.

Clearly, in this construction the nilpotent $\Omega^\dagger$ plays the role of the contracting homotopy operator. Therefore, we have understood the essence of the KO quartet mechanism as the construction of the contracting homotopy. This not only represents an abstract refinement of the existing method, but more important, provides a possible conceptual extension of the cohomology derivation to the quantum models, which are not Abelian or not readily Abelianized along the path described in Secs. II and III.

C. Comments

One recognizes in relation (4.42) and the discussion around (4.53) the no-ghost theorem[8] in the version adopted to the extended Fock space. We can rewrite it compactly as

$$|\text{phys}\rangle = |\text{DDF}\rangle + \Omega|\chi\rangle\ . \tag{4.54}$$

Equation (4.54) shows that any physical state can be decomposed into transverse DDF states with no ghost excitations and the unphysical BRS null states. Note that the presence of the ghost zero mode introduces a vacuum doubling: For a discussion of this and, also, the case $p = 0$ we refer the reader to, for instance, Ref. 15. In the present work we clearly covered the case $p \neq 0$ only.

We have seen that the algebra satisfied by the original harmonic oscillators α_n^μ and the ghosts decomposed into two separate algebras under transformation to the dual space. One part contains the unphysical gauge directions and transformed ghosts, namely the first-class constraints $\tilde{\Phi}_r$, the longitudinal subsidiary constraints Q^s and $(\eta^r,\mathcal{P}_r)$, or the multiplet $(\tilde{\Phi}_n,Q^n,\hat{\eta}^n,\hat{\mathcal{P}}_n)$. The other part consists of the purely transverse second-class DDF operators $A^i(r)$. This suggests a fiber bundle picture of phase space,[2] with the physical space being the base manifold and the conjugate pair $(\tilde{\Phi}_r,Q^r)$ or $(\tilde{\Phi}_n,Q^n)$ parametrizing the gauge orbit.

It is interesting to observe connection of our approach to some other earlier proofs of the no-ghost theorem. Comparing with Ref. 17 one sees that the change of the basis of the string Fock space corresponding to going from the operators L_n and $K_n = (k\cdot p/\sqrt{2})(\delta_{n,0} - nQ^n)$ to $\tilde{\Phi}_r$ and Q^r essentially diagonalizes the Thorn matrix $\mathcal{M}^P$ and accordingly, simplifies the structure of the Fock space. Since $\tilde{\Phi}_r$ contains the longitudinal vector vertex operator $\mathcal{D}_r = P_n e_r^n$ (Refs. 2 and 7) one notes the connection to the proof based on the spectrum generating algebra.[7] We note, however, that the use of ghosts was essential here for the identification of the null-norm states.

It is easy to generalize the quartet mechanism to the case of the extended phase space encountered in the study of OSP(1,1|2) symmetry. In this case the extra pairs (π,λ) and $(\bar{\mathcal{P}},\bar{\eta})$ constitute a separate quartet of unphysical modes which cancel each other.

ACKNOWLEDGMENTS

I would like to thank V. Guggenheim, M. Henneaux, L. Lambe, M. Mayer, J. Stasheff, and S. Solomon for valuable discussions.

This work was supported in part by the U.S. Department of Energy, Contract No. DE-FG02-84ER40173.

APPENDIX: VIELBEIN-VERTEX RELATIONS

In this Appendix, we derive a number of technical identities involving the vielbein and its inverse. The basic identity is

$$V(rk,z) = \exp\left(\frac{i\sqrt{2}}{k\cdot p}\, rk\cdot Q(z)\right) = \sum_{m=-\infty}^{\infty} e_r^m z^{+m}\ . \tag{A1}$$

A similar equation relates the inverse vielbein to the dual vertex operator:

$$\begin{aligned} \tilde{V}(mk,z) &= \exp\left(-m\sum_{r=-\infty}^{\infty} Q^{-r}z^r + m\ln z\right) \\ &= \sum_{r=-\infty}^{\infty} E_m^r\, z^{-r}\ . \end{aligned} \tag{A2}$$

As before, $Q^r = E_m^r Q^{-m}$. This supports our observation that the inverse vielbein E_m^r, as a function of Q^r, resembles e_r^n in terms of Q^n. This observation is clearly supported by our discussion in Sec. II, which is centered around dual MC structure equations.

As shown in Ref. 2, these moments are related by

$$e_r^m = (r/m)E^r{}_{-m}\ . \tag{A3}$$

It will be useful to note certain special cases in which the vielbeins have one index equal to zero. One finds, using (A1) and (A2), the explicit results

$$e_0^n = \delta(n), \quad e_r^0 = \delta(r) + rQ^r = D(r) \tag{A4}$$

and

$$E_0^r = \delta(r), \quad E_n^0 = \delta(n) - nQ^n = D(-n)\ . \tag{A5}$$

It follows by simple complex analysis from (A1) that

$$e_r^{n-m}e_s^m = e_{r+s}^n \tag{A6}$$

and similarly, from (A2),

$$E_n^{r-s}E_m^s = E_{m+n}^r . \tag{A7}$$

Another set of important identities involves products of vielbeins and inverse vielbeins: In this case we obtain

$$E_{-m}^r e_s^{n-m} = E_{-n}^{r+s}, \quad E_n^r e_{s-r}^m = e_s^{m-n} . \tag{A8}$$

For the special case $n = 0$, (A8) simplifies to

$$E_m^r e_s^m = E_0^{r+s} = \delta(r+s) , \tag{A9}$$

while setting $s = 0$ we obtain

$$e_{-r}^m E_n^r = e_0^{m-n} = \delta(m-n) . \tag{A10}$$

Equations (A9) and (A4) reflect the fact that E_m^r is both a left inverse and right inverse of e_r^m.

Finally, let us also list the important identities used in Sec. III:

$$e_{r-s}^n e_s^m = (\delta_r^t + T_{rs}^t Q^{-s})\, e_t^{n+m} , \tag{A11}$$

$$E_{n-m}^r E_m^s = (\delta_n^k - T_{np}^k Q^{-p}) E_k^{r+s} = D_n^k E_k^{r+s} . \tag{A12}$$

[1]H. Aratyn and R. Ingermanson, Nucl. Phys. B **299**, 507 (1988).
[2]H. Aratyn and R. Ingermanson, Phys. Rev. Lett. **61**, 2050 (1988); Int. J. Mod. Phys. A **4**, 3503 (1989).
[3]E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. B **55**, 224 (1975) and preprint CERN-TH 2332 (1977); I. A. Batalin and G. A. Vilkovisky, *ibid.* B **69**, 309 (1977); M. Henneaux, Phys. Rep. **126**, 1 (1985).
[4]T. Kugo and I. Ojima, Suppl. Progr. Theor. Phys. **66**, 1 (1979).
[5]M. Kato and K. Ogawa, Nucl. Phys. B **212**, 443 (1983).
[6]M. Henneaux and C. Teitelboim, Commun. Math. Phys. **115**, 213 (1988); J. Fisch and M. Henneaux, in *Constraint's Theory and Relativistic Dynamics*, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987), p. 57; M. Henneaux and C. Teitelboim, in *Quantum Field Theory and Quantum Statistics*, edited by I. A. Batalin *et al.* (Hilger, London, 1987), p. 165.
[7]R. C. Brower, Phys. Rev. D **6**, 1655 (1972); R. C. Brower and P. Goddard, Nucl. Phys. B **40**, 437 (1972); see, also, J. H. Schwarz, Phys. Rep. **8**, 269 (1973).
[8]P. Goddard and C. B. Thorn, Phys. Lett. B **40**, 235 (1972).
[9]M. D. Freeman and D. Olive, Phys. Lett. B **175**, 151 (1986).
[10]H. Aratyn, R. Ingermanson, and A. J. Niemi, Phys. Lett. B **194**, 506 (1987); Phys. Lett. B **195**, 149 (1987); Nucl. Phys. B **307**, 157 (1988).
[11]F. A. Berezin, Commun. Math. Phys. **40**, 153 (1975).
[12]H. Aratyn and R. Ingermanson, Class. Quant. Grav. **5**, L213 (1988).
[13]L. Brink and D. Olive, Nucl. Phys. B **56**, 253 (1973).
[14]E. Del Giudice, P. Di Vecchia, and S. Fubini, Ann. Phys. (NY) **70**, 378 (1972).
[15]M. Henneaux, Phys. Lett. B **177**, 35 (1986).
[16]I. Frenkel, H. Garland, and G. Zuckerman, Proc. Natl. Acad. Sci. (USA) **83**, 8442 (1986); see, also, D. Kastler and R. Stora, J. Geom. Phys. **2**, 1 (1985); **3**, 437, 483 (1986); B. Kostant and S. Sternberg, Ann. Phys. **176**, 49 (1987); M. E. Mayer, Acta Phys. Austriaca Suppl. **XXIII**, 481 (1981); M. E. Mayer, in *Fundamental Problems of Quantum Theory*, edited by A. Frigerio and V. Gorini (Plenum, New York, 1986), p. 403.
[17]C. B. Thorn, Nucl. Phys. B **286**, 61 (1987).
[18]M. Henneaux, Ann. Phys. (N.Y.) **194**, 281 (1989); M. Henneaux, in *Quantum Mechanics of Fundamental Systems*, edited by C. Teitelboim (Plenum, New York, 1988).
[19]S. Hwang and R. Marnelius, Gothenburg preprint 88-35; CERN preprint CERN-TH.5235/88.
[20]I. A. Batalin and E. S. Fradkin, Nucl. Phys. B **279**, 514 (1987).
[21]H. Neuberger, Phys. Lett. B **188**, 214 (1987).
[22]E. Witten, Nucl. Phys. B **188**, 513 (1981); **207**, 253 (1982).

Application of optimal control theory for selective vibrational excitation in molecules modeled as harmonic physical systems

J. G. B. Beumee
Department of Mathematics, Twente University of Technology, Enschede, The Netherlands
H. Rabitz
Department of Chemistry, Frick Laboratories, Princeton University, Princeton, New Jersey 08544-1009

(Received 4 January 1989; accepted for publication 10 January 1990)

The design of optimal electromagnetic fields producing selective vibrational excitation in molecules modeled as harmonic physical systems is shown to be equivalent to minimizing a quadratic cost functional balancing the energy distribution in the molecule and the fluence of the input. In the control problem, two approaches are employed to insure that the final excitation is attained. One method uses a control strategy that employs a terminal constraint and in the other approach the cost functional is augmented with a terminal cost. The asymptotic form of the state and costate is investigated for both strategies in the limit that the final time approaches infinity, and some mathematical results on the form of the Lagrange parameter are presented for the first type of controller. These two results allow for a detailed discussion on the appropriate choice of practical design constants. For the example of a linear chain molecule, and approximation for the eigenvalues of the Hamiltonian matrix is derived for the limiting case where the weighting on the fluence of the optical field in the cost functional increases to infinity. Also, for the linear chain it is shown that the eigenvalues are bounded and that this bound does not depend on the length of the chain.

I. INTRODUCTION

Recently, control theory has been applied to model selective vibrational excitation in molecules by means of optical fields.[1,2] If the atomic motion is not far from equilibrium, the effective interaction potential becomes approximately harmonic so that the quantal averages behave as a classical harmonic system. The optical field is taken to couple to the bonds through a linear dipole,[2] and the objective is to design a laser pulse in a given time interval such that the target bond in the molecule attains a certain degree of excitation at the end of the interval while the energy distribution throughout the remainder of the molecule is minimal over the entire interval. Simultaneously, the optical field used to control the molecule should have the smallest possible fluence, the time integral of the intensity of the field. To balance the significant internal energy of the molecule and the fluence of the optical field, this paper proposes a cost functional containing a quadratic cost of the state (average positions and momenta) and the fluence. The optical field will be called optimal if it minimizes this cost functional while producing the desired excitation. To insure this excitation at the final time, it is possible to use a controlling strategy specifying the final state exactly or a final term may be added to the cost functional specifying the penalty for not attaining the required excitation. Both approaches lead to LQ (linear quadratic) optimal control methods that are well known in the literature.[3–5]

The reduction of the design of a laser pulse for generating selective vibrational excitations in a molecule to a linear optimal control problem is the main contribution of this paper. Though the techniques used here are known in the control literature, this application gives rise to mathematically interesting limiting procedures and design problems. A recent study used the harmonic approximation of the molecule but introduced a cost functional that was not quadratic.[2] As a result, they were unable to make use of the large quantity of analytical results known for linear optimal control. The cost functional introduced in this paper is a modification of the one employed in Ref. 2 and recent numerical results suggest that the qualitative difference in the optimal fields derived from both cost functionals is negligible[6] under appropriate circumstances. Other studies in the mathematics literature that are useful in this context concentrate on quantum control in general,[7] and recently quantum control of a molecule by means of optical fields was attempted.[1] Although a special case, the calculations in the present paper can also be viewed as fully quantal through the use of Ehrenfest's theorem.[2]

In Sec. II, the appropriate cost functionals are introduced and the optimal field is expressed in terms of the solution to the Hamiltonian equations under the assumption of controllability of the state (average positions and momenta) of the individual atoms. Since these functionals are quadratic, it is formally possible to establish a relationship between the optimal electromagnetic field and the state of the molecular system using the solution to the matrix Ricatti equation.[5] However, the rapidity of the molecular motion does not allow sufficient time to measure the molecular state and use this information to determine the optimal field. In other words, the extremely short vibrational time scales make a physical feedback mechanism impossible; hence, the construction of the optimal electromagnetic field is an open loop control strategy. Another practical aspect is that the intensity of the required optimal field tends to be very high and, to alleviate this problem, the weighting of the fluence in the cost functional may be increased. Therefore, in Sec. III, the behavior of the optimal field and the associated position and

 0022-2488/90/051253-08$03.00

momenta is investigated in the limit that this weighting approaches infinity. The eigenvalues of the Hamiltonian matrix can then be approximated, which is illustrated for the case of a model linear homogeneous chain molecule. This section also shows a bound for the eigenvalues of the linear chain that is independent of the size of the chain. In Sec. IV, the asymptotic form for the state and optical field of the harmonic system in the limit that the final time approaches infinity is derived, which shows that the choice of an appropriate time interval is tightly connected with the question of the weighting. These asymptotic expansions contain a Lagrange parameter that depends on a negative definite matrix satisfying a Riccati equation, and in this section it will be shown that this matrix is well defined and approaches a limit as the final time increases to infinity. A discussion on the appropriate choice of design parameters concludes this section. Finally, in Sec. V, our conclusions will be presented and some future research indicated.

II. FORMULATION

If the motion of the individual atoms in a molecule is not far from equilibrium, the effective interaction potential is approximately quadratic,[8] in which case the quantal averages of the system can be represented as a classical harmonic mass–spring system.[2] In this section, two different control strategies are introduced for the selective vibrational excitation in a molecule based on quadratic cost functionals balancing the energy of the system and the fluence of the optical field. The homogeneous linear chain introduced in this section serves as the canonical example.

More specifically, let the average bond-length displacement coordinates q and their associated momentum p be denoted by $q^T = (q_1(t),...,q_n(t))$ and $p^T = (p_1(t),...,p_n(t))$, respectively; then the molecular Hamiltonian H for this system is given by[2]

$$H = \tfrac{1}{2}p^T G p + \tfrac{1}{2}q^T F q - u(t) b_1^T q, \tag{2.1}$$

where $F_{ij} = \delta_{ij} k_i$, k_i being the force constant of the ith bond. Due to the use of bond-length coordinates, the mass tensor G becomes a tridiagonal matrix

$$G_{ij} = \delta_{ij}(\mu_{i+1} + \mu_i) - \delta_i(j+1)\mu_i - \delta_{i(j-1)}\mu_{i+1} = G_{ji}, \tag{2.2}$$

with $\mu_i = 1/m_i$, m_i being the mass of the ith atom.[2] The external optical field $u(t)$ couples to the bond q in the form of a linear dipole with b_1 being the dipole derivative vector associated with the molecule. Using (2.1), the equations of motion for the state $x(t)^T = (q(t)^T, p(t)^T)$ become

$$\frac{d}{dt}x(t) = \frac{d}{dt}\begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} 0 & G \\ -F & 0 \end{bmatrix}\begin{bmatrix} q(t) \\ p(t) \end{bmatrix} + \begin{bmatrix} 0 \\ b_1 \end{bmatrix} u(t) = Ax(t) + bu(t), \tag{2.3}$$

which is the familiar form of a linear system with system maxtrix A, input matrix $b^T = (0, b_1^T)$, and input $u(t)$. Reference to b as an input matrix in the present molecular control context is done to keep within standard control nomenclature.

The objective is to design a minimum fluence optical field $u(t)$ that directs the system in such a fashion that, at some final time T, a specific excitation of one or more target bonds is produced while a minimum of energy is deposited in the remaining bonds throughout the time interval. Quantifying the simultaneous costs of this control strategy, we propose a cost functional consisting of a quadratic measure of the excitation of the whole system and a quadratic measure of the fluence of the optical field. Hence, the following cost functional $\Phi(x,u)$ is adopted:

$$\Phi(x,u) = \frac{1}{2}\int_0^T [x(t)^T Q x(t) + r u(t)^2]\,dt, \tag{2.4}$$

where Q is a diagonal matrix of size $2n$ with constant γ_1 on the upper left-hand block and γ_2 on the lower right-hand block. For more physical intuition on this functional, see Ref. 2. The second integral in (2.4) is referred to as the fluence of the field, and the design constant r determines the ratio between the two contributions to the cost of guiding the system to its desired final excitation. One way to guarantee the specified vibrational excitation at time T is to require that $x(T)$ satisfies a terminal condition; another possibility is to add a quadratic terminal cost to $\Phi(x,u)$ that depends on the difference between $x(T)$ and the desired excitation. Finding an optimal electromagnetic field therefore reduces to one of the following optimal control problems.

(i) Minimize $\Phi(x,u)$ with respect to the otpical field u (input) subject to the equations of motion (2.3) and the final conditon $Hx(T) = \chi$. Appropriate choices for H and χ then correspond to the desired excitation. For the sake of simplicity, it will be assumed that H is a vector since the generalization to multiple constraints on $x(T)$ is straightforward.

(ii) Minimize the cost functional $\Phi(x,u) + \frac{1}{2}(x(T) - \xi)^T S_f (x(T) - \xi)$ with respect to u subject to the equations of motion (2.3). The matrix S_f in the final cost should be chosen such that designated entries of $x(T)$ are close to the corresponding entries in the target state ξ. Whereas the advantage of the terminal constraint controller (i) is that at the final time T the exact target state is reached, the cost functional in (ii) has the advantage that it can readily be generalized to the associated stochastic optimal control problem. Since, the real molecules in the presence of collisions, the equations of motion (2.3) are a simplification, this generalization is important.

For the terminal constraint controller it is well known that the optimal field can be obtained by introducing the $2n$-dimensional Lagrangian function $\lambda(s)$ (the costate) and the Lagrange parameter η such that [3–5]

$$\Phi(x,u,\lambda,\eta) = \frac{1}{2}\int_0^T [x(t)^T Q x(t) + r u(t)^2]\,dt - \int_0^T \lambda(t)^T \left[\frac{d}{dt}x(t) - Ax(t) - bu(t)\right]dt + \eta[Hx(T) - \chi]. \tag{2.5}$$

The minimum of (2.4) is then obtained from the saddle point of (2.5), with can be determined from the requirement that the first variations of $x(t)$, $u(t)$, $\lambda(t)$, and η vanish. The optimal input then equals $(u(t) = -r^{-1}b^T\lambda(t)$, where

$$\frac{d}{dt}\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \begin{bmatrix} A & -br^{-1}b^T \\ -Q & -A^T \end{bmatrix}\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}, \tag{2.6}$$

with $\lambda(T) = H^T\eta$ and $x(0) = x_0$. Hence, the calculation of the optimal field reduces to solving this two-point boundary-value problem. In the case of the second controller (ii), it is found that the state $x(t)$ and costate $\lambda(t)$ also satisfy Eq. (2.6), but then the final condition for the costate is given by $\lambda(T) = S_f(x(T) - \xi)$ (Ref. 5). In accordance with control theory practice,the matrix in (2.6) will be referred to as the Hamiltonian matrix, which should not be confused with the molecular Hamiltonian H in (2.1).

It is well known that an eigenvalue λ of the Hamiltonian matrix in (2.6) is accompanied by an eigenvalue $-\lambda$, which follows from the analytical form of its determinant.[5] For the present choice of Q, a theorem by Kučera[9] implies that all eigenvalues of this matrix must have a nonzero real part if the pair (A,b) is controllable, and that purely imaginary eigenvalues correspond to uncontrollable modes of the system (2.3).[9] The system (2.3) is defined as controllable if its state $x(t)$ is able to reach any point in R^n, which is equivalent to the requirement that the matrix $[b\,Ab\,A^2b\cdots A^{2n-1}b]$ has rank $2n$. In the remainder of the paper we assume controllability for the pair (A,b); hence, the Hamiltonian matrix in (2.6) can be decomposed as follows:

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}^{-1}\begin{bmatrix} A & -br^{-1}b^T \\ -Q & -A^T \end{bmatrix}\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = \begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{bmatrix}, \tag{2.7}$$

where the matrix composed of the W_{ij} submatrices contains the $4n$ eigenvectors. The diagonal of the matrix $\mathrm{diag}(\Lambda, -\Lambda)$ on the right-hand side contains the eigenvalues λ_j of the Hamiltonian matrix that have nonzero real parts due to the assumption of controllability of (A,b). Moreover, it will be assumed that the eigenvalues on the diagonal of Λ have positive real parts and, if some of these coincide, then the matrix Λ is in Jordan form. This decomposition is used in the following section while the remainder of this section will concern itself with the eigenvalues of the Hamiltonian matrix.

The characteristic polynomial $\Delta(s)$ of the Hamiltonian matrix is determined as follows[5,9]:

$$\begin{aligned}\Delta(s) &= \det\begin{bmatrix} (sI-A) & br^{-1}b^T \\ Q & (sI+A^T) \end{bmatrix} \\ &= \det(sI-A)\cdot\det[(sI+A^T) \\ &\quad - Q(sI-A)^{-1}br^{-1}b^T] \\ &= \det(sI-A)\cdot\det(sI+A_T) \\ &\quad \times\det[I - Q(sI-A)^{-1}br^{-1}b^T(sI+A^T)^{-1}] \\ &= \det(sI-A)\cdot\det(sI+A^T)[1 - r^{-1}b^T \\ &\quad \times(sI+A^T)^{-1}Q(sI-A)^{-1}b],\end{aligned}$$

and additional calculations show that, in terms of the matrices F and G in (2.3), $\Delta(s)$ can be evaluated as

$$\begin{aligned}\Delta(s) &= b_1^T(s^2I + GF)^{-1} \\ &\quad \times(-\gamma_1 G^2 + \gamma_2 s^2)(s^2I + FG)^{-1}b_1.\end{aligned} \tag{2.9}$$

Since this is a polynomial in terms of s^2, the roots of $\Delta(s)$ must occur in pairs as remarked earlier.

Expressing the input vector b_1 in (2.9) in terms of the eigenvectors of the matrix A yields further information on $\Delta(s)$ since A has certain properties due to its physical origin. Consider the eigenvectors $e_j^T = (v_j^T, w_j^T)$ of A with corresponding eigenvalues ρ_j,

$$\begin{aligned} &Ae_j = \begin{bmatrix} 0 & G \\ -F & 0 \end{bmatrix}\begin{bmatrix} v_j \\ w_j \end{bmatrix} = \rho_j\begin{bmatrix} v_j \\ w_j \end{bmatrix} = \rho_j e_j, \quad j = 1,\ldots,n, \\ &GFv_j = -\rho_j^2 v_j = \sigma_j v_j, \quad FGw_j = -\rho_j^2 w_j = \sigma_j w_j, \\ &\quad j = 1,\ldots,n. \end{aligned} \tag{2.10}$$

The eigenvalues σ_j of the matrices FG and GF are real because the matrices F and G are simultaneously diagonalizable[10] and they must be positive corresponding to the oscillatory modes of the original system. Zero or negative eigenvalues correspond to translational or unstable modes, respectively, which are not consistent with the known oscillatory motion of a molecule. If all eignvalues of GF are different, then the eigenvectors v_j, $j = 1,\ldots,n$, are the perpendicular principal axes of the oscillations and w_j, $j = 1,\ldots,n$, are the left eigenvectors of GF.[8,10] Notice that they are not simultaneously normalizable.[10] Now let σ_j, $j = 1,\ldots,n$, be the positive real eigenvalues of GF; then the eigenvalues of A become $\rho_j = i\sqrt{\sigma_j}$, $\rho_{j+n} = -\rho_j$, $i = 1,\ldots,n$, with the associated eigenvectors $e_j^T = ((Gw_j)^T, \rho_j w_j^T))$, $j = 1,\ldots,2n$. Assume that $\{w_j\}_{j=1,n}$ is normalized and that the vector b_1 in b in (2.3) can be expanded in the form $b_1 = \Sigma\alpha_j w_j$; then substitution into (2.9) yields

$$\begin{aligned}\Delta(s) &= \prod_{j=1}^{n}(s^2+\sigma_j)^2 \\ &\quad \times\left[1 + r^{-1}\sum_{k=1}^{k=n}\alpha_k^2\left\{\frac{\gamma_1\sigma_k^2(v_k^T v_k) - \gamma_2 s^2}{(s^2+\sigma_k)^2}\right\}\right],\end{aligned} \tag{2.11}$$

If some of the eigenvalues of GF coincide, then the sets of eigenvectors $\{v_j\}_{j=1,n}$ and $\{w_j\}_{j=1,n}$ are not orthogonal, in which case (2.11) would become a more complicated summation.

Expression (2.11) shows that, if $\alpha_p = 0$ for some index p, then one of the roots of $\Delta(s)$ is given by $s^2 = -\sigma_p$, implying that a purely imaginary eigenvalue of A coincides with an eigenvalue of the Hamiltonian matrix. From the theorem by Kučera,[9] it follows then that this eigenvalue corresponds to an uncontrollable mode, which in turn implies that (A,b) is not controllable contrary to our assumption. Hence, all coefficients α_j are nonzero. Conversely, if $\alpha_j \neq 0$ for all j, then substituting $s = i\beta$ (β being real) into (2.11) shows that $\Delta(i\beta) > 0$, so indeed none of the eigenvalues of the Hamiltonian matrix resides on the imaginary axis.

An important example treated in detail in the next section is provided by the case where the system reduces to a linear homogeneous chain ($m_j = m = 1/\mu$, $k_j = \delta$, $j = 1,\ldots,n$). Then, $F = \delta I$ and

$$G=\mu\begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & & \ddots & & & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}. \tag{2.12}$$

The eigenvalues σ_j become $\sigma_j=\delta\theta_j$, where $\theta_j=4\mu\sin[j\pi/2(n+1)]^2$, $j=1,\ldots,n$, are the eigenvalues of G. Some calculation shows that (2.11) now reduces to

$$\begin{aligned}\Delta(s)&=\prod_{j=1}^{n}(s^2+\delta\theta_j)^2\\&\quad\times\left[1+r^{-1}\sum_{k=1}^{k=n}\alpha_k^2\left\{\frac{\gamma_2\theta_k^2-\gamma_2 s^2}{(s^2+\delta\theta_k)^2}\right\}\right]\\&=a(s)+r^{-1}d(s),\end{aligned} \tag{2.13}$$

where

$$\begin{aligned}a(s)&=\prod_{j=1}^{n}(s^2+\delta\theta_j)^2,\\d(s)&=\sum_{k=1}^{k=n}\alpha_k^2(\gamma_1\theta_k^2-\gamma_k s^2)\left\{\prod_{\substack{j=1\\j\neq k}}^{n}(s^2+\delta\theta_j)^2\right\}.\end{aligned} \tag{2.14}$$

Notice that $0<\sigma_j=\delta\theta_j<4\delta\mu$, independent of the size of the chain.

The following section will be concerned entirely with the properties of the eigenvalues of the Hamiltonian matrix associated with the optimal control problem of the linear homogeneous chain.

III. THE LINEAR CHAIN

The roots of the characteristic polynomial (2.13) cannot be determined analytically, but certain approximations can be made. In this section the behavior of the roots of (2.13) will be investigated in the limit as r approaches zero, and for the physically interesting limit as r becomes large, a perturbation expression for the real part of the eigenvalues of (2.13) will be obtained. Moreover, an upper bound to the moduli of the roots of $a(s)+r^{-1}d(s)$ will be derived, which is independent of the size of the linear chain.

First consider the limit that r becomes small. Since $\alpha_j\neq 0$, it is true that $\Delta(s)=0$ is equivalent to

$$\sum_{j=1}^{j=n}\alpha_j^2\left\{\frac{\gamma_1\theta_j^2-\gamma_2 s^2}{(s^2+\delta\theta_j)^2}\right\}=-r \tag{3.1}$$

or

$$d(s)=-ra(s). \tag{3.2}$$

As r decreases, the solutions to Eq. (3.2) typically fall into two categories. A set of $4n-2$ solutions approaches the roots of the polynomial $d(s)$ and the remaining two solutions have moduli that tend to infinity in Butterworth patterns.[4,5] In the present case, the latter can be found by assuming that s is large and using the asymptotic expansion of the left-hand side of (3.1). This yields

$$s_{1,2}\approx\pm r^{-1/2}\sqrt{\gamma_2|b_1|^2}, \tag{3.3}$$

where $|b_1|^2$ is the magnitude of the vector b_1 in b given by $|b_1|^2=\Sigma\alpha_j^2$. For small r, optical fields of almost arbitrary intensity are allowed so that $x(t)$ will be very rapidly directed from its initial condition x_0 to the zero state (no cost), and then close to the final time the system will be equally rapidly sent to its desired final condition. Equation (3.2) implies that in this limit only one particular oscillatory mode of the physical system is damped severely while the remaining modes approach the transfer zeros [the roots of $d(s)$] of the system.[4]

Numerical simulations show that, for the two control strategies above, the intensity of the optimal electromagnetic field is generally considerable.[1,2,6] This problem may be alleviated by increasing the parameter r and thereby placing more of a penalty on the fluence of the optical field, which in turn will reduce the required intensity. Now, in the limit of r approaching infinity, the roots λ_j of $a(s)+r^{-1}d(s)$ coincide with the roots of $a(s)$ so that, to first order, $\lambda_j=\nu_j$, $j=1,\ldots,2n$, where $\nu_j=i\sqrt{\sigma_j}=i\sqrt{\delta\theta_j}=-\nu_{n+j}$, $j=1,\ldots,n$, which provides $2n$ eigenvalues. For large but finite r, the ν_j will have a small perturbation $\Delta\nu_j$, which can be calculated as follows. Let

$$a(\nu_j+\Delta\nu_j)+(1/r)d(\nu_j+\Delta\nu_j)=0,\quad j=1,\ldots,2n, \tag{3.4}$$

which, expanded in a Taylor series, is equivalent to

$$\begin{aligned}&a(\nu_j)+a'(\nu_j)\Delta\nu_j+\tfrac{1}{2}a''(\nu_j)(\Delta\nu_j)^2+(1/r)d(\nu_j)\\&\quad+(1/r)\Delta\nu_j d'(\nu_j)+\cdots=0,\end{aligned} \tag{3.5}$$

for all j. [Here, $a'(\ )$ indicates the derivative of the function $a(\ )$.] Since $a(\nu_j)=a'(\nu_j)=0$, this equation becomes to leading order in $r^{-1/2}$:

$$\begin{aligned}&\tfrac{1}{2}a''(\nu_j)(\Delta\nu_j)^2+r^{-1}d(\nu_j)=0,\quad j=1,\ldots,2n,\\&\Delta\nu_j=\pm\sqrt{[-2d(\nu_j)]/[ra''(\nu_j)]},\quad j=1,\ldots,n,\end{aligned} \tag{3.6}$$

and further calculation shows that this equals

$$\Delta\nu_j=\pm\tfrac{1}{2}|\alpha_j|\sqrt{(\gamma_1|\theta_j|+\delta\gamma_2)/r},\quad j=1,\ldots,n, \tag{3.7}$$

where

$$\theta_j=4\mu\sin[j\pi/2(n+1)]^2,\quad \alpha_j=b_1^T w_j,\quad j=1,\ldots,n.$$

Notice that $\Delta\nu_j$ is entirely real and that the set $\pm\nu_j\pm|\Delta\nu_j|$, $j=1,\ldots,n$, provides $4n$ different (approximate) eigenvalues for the Hamiltonian matrix. The first-order perturbation $\Delta\nu_j$ also depends linearly on the coefficients $\alpha_j=b_1^T w_j$ and vanishes as $\alpha_j\to 0$. It is true in general that $\lambda_j\to\pm i\sqrt{\delta\theta_j}$ if α_j vanishes, which implies that further terms in the perturbation expansion also depend on α_j. In the present case, all α_j are nonzero because of the controllability of (A,b), so that $\Delta\nu_j\neq 0$, $j=1,\ldots,n$.

One more fact can be shown about the spectrum of the Hamiltonian matrix of the linear chain, which can be surmised from the form of Eq. (3.1). Since $0<\sigma_j=\delta\theta_j<4\delta\mu$ or $0<\theta_j<4\mu$, for all j, and because the function on the left-hand side in (3.1) is inversely proportional to s^2, it is to be expected that no solutions to (3.1) exist if s is large enough. This approach proves correct and is formalized in the following theorem.

Theorem 3.1: Let $0<\theta_j<4\mu=C$ and $|b_1|^2=\Sigma\alpha_j^2$, then the eigenvalues λ_j of the Hamiltonian matrix, $\Delta(\lambda_j)=0$, are such that

$$|\lambda_j|^2 \leqslant \delta C + (1/2r)|b_1|^2\gamma_2 + (1/2r)|b_1| \times \sqrt{|b_1|^2\gamma_2^2 + 4r(\gamma_1 C^2 + \gamma_2 \delta C)}\,. \tag{3.8}$$

The proof will be presented after this simple lemma:

Lemma 1: For all real and positive θ_j it is true that

$$\left\langle \frac{\gamma_1\theta_j^2 - \gamma_2 s^2}{(s^2+\delta\theta_j)^2} \right\rangle \leqslant \frac{\gamma_1\theta_j^2 + \gamma_2\delta\theta_j}{(|s|^2 - \delta\theta_j)^2} + \frac{\gamma_2}{|(|s|^2 - \delta\theta_j)|}\,. \tag{3.9}$$

Proof:

$$\left\langle \frac{\gamma_1\theta_j^2 - \gamma_2 s^2}{(s^2+\delta\theta_j)^2} \right\rangle = \left| \frac{\gamma_1\theta_j^2 + \gamma_2\delta\theta_j}{(s^2+\delta\theta_j)^2} - \frac{\gamma_2}{(s^2+\delta\theta_j)} \right| \leqslant \left\langle \frac{\gamma_1\theta_j^2 + \gamma_2\delta\theta_j}{(s^2+\delta\theta_j)^2} \right\rangle + \left\langle \frac{\gamma_2}{(s^2+\delta\theta_j)} \right\rangle. \tag{3.10}$$

Since θ_j is real and positive, we have that

$$\begin{aligned} |(s^2+\delta\theta_j)^2| &= |(s^2+\delta\theta_j)|^2 \\ &= |s|^4 + 2\delta\theta_j\,\mathrm{Re}(s^2) + \delta^2\theta_j^2 \\ &\geqslant |s|^4 - 2\delta\theta_j|s|^2 + \delta^2\theta_j^2 \\ &= (|s|^2 - \delta\theta_j)^2, \end{aligned} \tag{3.11}$$

where $Re(s^2) \geqslant -|s|^2$ was utilized. Hence, $|(s^2+\delta\theta_j)|^{-1} \leqslant |(|s|^2 - \delta\theta_j)|^{-1}$, which can be substituted into (3.9). This yields inequality (3.9) and proves the lemma.

Proof of Theorem 3.1: The polynomial $\Delta(s)$ in (2.13) cannot have any roots in the domain of the argument s where both $|s|^2 > \delta C$ so that $|a(s)| \neq 0$, and where simultaneously s satisfies the inequality

$$\left| r^{-1} \sum_{j=1}^{j=n} \alpha_j^2 \left\{ \frac{\gamma_1\theta_j^2 - \gamma_2 s^2}{(s^2+\delta\theta_j)^2} \right\} \right| < 1, \tag{3.12}$$

which is obvious from (3.1). If we let $|s|^2 > \delta C$, then the left-hand side of (3.11) can be majorized as follows:

$$\begin{aligned} &\left| r^{-1} \sum_{j=1}^{j=n} \alpha_j^2 \left\{ \frac{\gamma_1\theta_j^2 - \gamma_2 s^2}{(s^2+\delta\theta_j)^2} \right\} \right| \\ &\leqslant r^{-1} \sum_{j=1}^{j=n} \alpha_j^2 \left| \left\{ \frac{\gamma_1\theta_j^2 - \gamma_2 s^2}{(s^2+\delta\theta_j)^2} \right\} \right| \\ &\leqslant r^{-1} \sum_{j=1}^{j=n} \alpha_j^2 \left\{ \frac{\gamma_1\theta_j^2 + \gamma_2\delta\theta_j}{(|s|^2 - \delta\theta_j)^2} + \frac{\gamma_2}{|(|s|^2 - \delta\theta_j)|} \right\} \\ &\leqslant r^{-1} \sum_{j=1}^{j=n} \alpha_j^2 \left\{ \frac{\gamma_1\theta_j^2 + \gamma_2\delta\theta_j}{(|s|^2 - \delta C)^2} + \frac{\gamma_2}{(|s|^2 - \delta C)} \right\} \\ &\leqslant r^{-1} \sum_{j=1}^{j=n} \alpha_j^2 \left\{ \frac{\gamma_1 C^2 + \gamma_2\delta C}{(|s|^2 - \delta C)^2} + \frac{\gamma_2}{(|s|^2 - \delta C)} \right\} \\ &= r^{-1}|b_1|^2\,\frac{\gamma_1 C^2 + \gamma_2|s|^2}{(|s|^2 - \delta C)^2} = f(|s|), \end{aligned} \tag{3.13}$$

since $|(|s|^2 - \delta\theta_j)|^{-2} \leqslant |(|s|^2 - \delta C)|^{-2}$ for $|s|^2 > \delta C$. For the second inequality, the lemma was employed. It is easy to see that the function $f(|s|)$ in (3.12) is monotonically decreasing if $|s|^2 > \delta C$ so that the domain where $\Delta(s)$ cannot have any roots is given by $F = \{s | f(|s|) < 1,\ |s|^2 > \delta C\}$. Since the solution to $f(s'') = 1$ has one positive solution that equals

$$\begin{aligned} s'' &= \delta C + (1/2r)|b_1|^2\gamma_2 + (1/2r)|b_1| \times \sqrt{|b_1|^2\gamma_2^2 + 4r(\gamma_1 C^2 + \gamma_2\delta C)} \\ &= \Gamma(C,r), \end{aligned} \tag{3.14}$$

and since $\Gamma(C,r) \geqslant \delta C$, it is clear that the domain F reduces to $F = \{s | |s|^2 > \Gamma(C,r)\}$. As a result, $|\lambda_j|^2 \leqslant \Gamma(C,r)$ for all j, which proves the theorem.

Since, for the linear chain $0 < \theta_j < 4\mu = C$ for all n, the right-hand side in (3.8) does not depend on the size of the chain and therefore also $|\lambda_j| \leqslant \Gamma(C,r)$ for all j and n. Physically, this result puts a bound on the rapidity at which a molecule can respond to external fields. Note that this bound depends on the field as well as on the molecule.

The next section returns to the general case and investigates the asymptotic behavior of $x(t)$ and $\lambda(t)$ in the optimally controlled harmonic system as the final time T becomes large. Some applications to the linear chain model will be presented.

IV. INFINITE HORIZON CONTROL

To determine the asymptotic behavior of the optimal field and the associated state as the final time increases, the eigenvector decomposition of the Hamiltonian matrix (2.7) is employed. The state $x(t)$ and costate $\lambda(t)$ can then be expressed in terms of positive and negative exponentials of the matrix Λ, which has nonzero real diagonal entries due to the controllability of (A,b). As a result, in passing to the limit $T \to \infty$, the negative exponentials vanish. This technique has been used to determine the behavior of the infinite horizon optimal controller for other types of cost functionals.[5] The remainder of the section is devoted to showing that the asymptotic form of the Lagrange parameter η in the terminal constraint controller is well defined.

Consider first the terminal constraint controller and introduce $z_1(t)$ and $z_2(t)$ such that

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}^{-1} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}. \tag{4.1}$$

Then, from (2.6) and (2.7), it is clear that

$$\frac{d}{dt}\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix}, \tag{4.2}$$

which has solutions $z_1(t) = e^{\Lambda t} z_1(0)$ and $z_2(t) = e^{-\Lambda t} z_2(0)$. The initial conditions $z_1(0)$ and $z_2(0)$ can be evaluated from the final conditions $Hx(T) = \chi$, $\lambda(T) = H^T\eta$, and $x(0) = x_0$. Specifically, using (4.1),

$$x(0) = x_0 = W_{11} z_1(0) + W_{12} z_2(0), \tag{4.3a}$$

$$Hx(T) = \chi = HW_{11} e^{\Lambda T} z_1(0) + HW_{12} e^{-\Lambda T} z_2(0), \tag{4.3b}$$

$$\begin{aligned} \lambda(T) = H^T\eta &= W_{21} z_1(T) + W_{22} z_2(T) \\ &= W_{21} e^{\Lambda T} z_1(0) + W_{22} e^{-\Lambda T} z_2(0). \end{aligned} \tag{4.3c}$$

Solving for $z_2(0)$ using (4.3a) and (4.3c) yields

$$z_2(0) = [W_{21} e^{\Lambda T} W_{11}^{-1} W_{12} - W_{22} e^{-\Lambda T}]^{-1} \times [-H^T\eta + W_{21} e^{\Lambda T} W_{11}^{-1} x_0], \tag{4.4}$$

so that $z_1(0)$ can be determined from (4.3a). Notice that the matrices W_{ij} are all invertible, which can be seen from the

representation of the (unique) positive definite and negative definite solution of the algebraic Ricatti equation in terms of the W_{ij} matrices. Moreover, in (4.4) the inverse of the matrix combination on the right-hand side is well defined since for large enough T the second term in the combination vanishes, leaving the first term, which is invertible. For large T the determinant of the matrix combination is therefore nonzero and, since it is an analytical function, it must be almost everywhere nonzero.

As T becomes large, the matrix term $e^{-\Lambda T}$ in (4.4) becomes small so that $z_1(0)$ and $z_2(0)$ reduce to

$$z_2(0) = W_{12}^{-1}x_0 - W_{12}^{-1}W_{11}e^{-\Lambda T}W_{21}^{-1}H^T\eta + O(e^{-2\Lambda t}),$$
$$z_1(0) = e^{-\Lambda T}W_{21}^{-1}H^T\eta + O(e^{-2\Lambda T}), \tag{4.5}$$

where $O(\ )$ indicates matrix combinations containing $e^{-2\Lambda T}$. This can be shown from expanding the inverse matrix combination in (4.4) in terms of matrices containing $e^{-\Lambda T}$. Substituting (4.5) into (4.3b), the parameter η can now be determined:

$$\chi = HW_{11}e^{\Lambda T}z_1(0) + HW_{12}e^{-\Lambda T}z_2(0)$$
$$= HW_{11}W_{21}^{-1}H^T\eta + O(e^{-\Lambda T}) \tag{4.6}$$

or

$$\eta = \chi/(HW_{11}W_{21}^{-1}H^T) + O(e^{-\Lambda T}), \tag{4.7}$$

so that η approaches a constant as T increases. Notice that here the assumption of H being a vector was utilized. Finally, substituting $z_1(0)$, $z_2(0)$, and η into (4.1) yields

$$x(t) = W_{11}e^{\Lambda(t-T)}W_{21}^{-1}H^T\cdot(\chi/HW_{11}W_{21}^{-1}H^T)$$
$$+ W_{12}e^{-\Lambda t}W_{12}^{-1}x_0 + O(e^{-\Lambda T}),$$
$$u(t) = -(1/r)b^TW_{21}e^{\Lambda(t-T)}W_{21}^{-1}H^T$$
$$\times(\chi/HW_{11}W_{21}^{-1}H^T)$$
$$-(1/2r)b^TW_{22}e^{-\Lambda t}W_{12}^{-1}x_0 + O(e^{-\Lambda T}). \tag{4.8}$$

This result shows that the behavior of the input $u(t)$ and the state $T_0^{-1} = \min_j \mathrm{Re}(\lambda_j)$, where the λ_j are the eigenvalues of the Hamiltonian matrix in Λ, and let c' be a constant such that $e^{-c'}$ is physically negligible. Then, in the region where $0 \leqslant t \leqslant c'T_0$, the first terms in Eq. (4.8) for $x(t)$ and $\lambda(t)$ vanish while, in the region where $0 \leqslant (T-t) \leqslant c'T_0$, it is found that the first terms dominate and the second terms become negligible. In the intermediate interval, both $x(t)$ and $u(t)$ are essentially zero. The input $u(t)$ in the time interval $[0,c'T_0]$ is designed to take the initial condition x_0 and reduce it to the null state $x(t) \approx 0$, thereby extracting all energy from the physical system. In the region where $0 \leqslant T-t \leqslant c'T_0$, the reverse process occurs and the input transfers the system from $x(T-c'T_0) \approx 0$ to a state $x(T)$ such that $Hx(T) = \chi$. Notice that the final time T must be of order $T \approx 2c'T_0$ for the asymptotic approximation (4.8) to be valid.

A similar asymptotic limit can be obtained for the state and the optimal field in the case of the cost functional (ii) of Sec. II. Then

$$x(t) = -W_{11}e^{\Lambda(t-T)}(W_{21} - M_0W_{11})^{-1}M_0\xi$$
$$+ W_{12}e^{-\Lambda t}W_{12}^{-1}x_0 + O(e^{-\Lambda T}),$$
$$u(t) = -(1/r)b^TW_{21}e^{\Lambda(t-T)}(W_{21} - M_0W_{11})^{-1}M_0\xi$$
$$-(1/2r)b^TW_{22}e^{-\Lambda t}W_{12}^{-1}x_0 + O(e^{-\Lambda T}). \tag{4.9}$$

The asymptotic form displays the same behavior as in the case of the terminal constraint controller and again the same three time domains can be identified. In this regard as in other aspects, the behavior of the two controlling strategies is very similar. Notice that as M_0 becomes large, $x(T)$ approaches the target state ξ.

Returning to the terminal constraint controller, it is seen that the relative length of the three time regions characterized by the behavior of $x(t)$ and $\lambda(t)$ in (4.8) depends on the real part of the eigenvalue matrix Λ. In the limit that r is large, the real parts of the eigenvalues of the Hamiltonian matrix are approximately given by Eq. (3.7), so in that case we have

$$T_0 \approx \min_j \left[\tfrac{1}{2}|\alpha_j|\sqrt{(\gamma_1\theta_j + \delta\gamma_2)/r}\,\right]. \tag{4.10}$$

Since T_0 is inversely proportional to $r^{-1/2}$, it is clear that increasing r means that the length of the intervals $[0,c'T_0]$ and $[T-c'T_0,T]$ increases as $r^{-1/2}$ so that the physical process takes a longer time to progress from state x_0 to $x(t) \approx 0$ and back to the final state $x(T)$. If r becomes so large that $2c'T_0 \gg T$, then the distinction between the three regions disappears and the cost of the functional increases. A more thorough discussion on this will be presented below. Even though (4.10) holds for linear chains, this perturbation analysis is applicable to all harmonic physical systems so that it is true in general that $T_0 \propto r^{-1/2}$. The precise constant of proportionality will of course depend on the physical system under consideration and has to be determined numerically.

More information on the asymptotic solution for the Lagrange parameter η for the terminal constraint controller can be derived from the full solution to Eq. (2.6) with the appropriate final conditions. Using the transition matrix, the state $x(t)$ and costate $\lambda(t)$ can be determined as follows:

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \exp\left[t\begin{bmatrix} A & -br^{-1}b^T \\ -Q & -A^T \end{bmatrix}\right]\begin{bmatrix} x(0) \\ \lambda(0) \end{bmatrix}$$
$$= \begin{bmatrix} \phi_{11}(t,0) & \phi_{12}(t,0) \\ \phi_{21}(t,0) & \phi_{22}(t,0) \end{bmatrix}\begin{bmatrix} x(0) \\ \lambda(0) \end{bmatrix}, \tag{4.11}$$

with final conditions

$$Hx(T) = \chi = H\phi_{11}(T,0)x_0 + H\phi_{12}(T,0)\lambda(0), \tag{4.12a}$$

$$\lambda(T) = H^T\eta = \phi_{21}(T,0)x_0 + \phi_{22}(T,0)\lambda(0), \tag{4.12b}$$

in terms of the partition matrices of the transition matrix. At this point we will assume that $\phi_{22}(t,0)$ is invertible so that it can be shown from (4.12b) that $\lambda(0) = \phi_{22}(T,0)^{-1}[H^T\eta - \phi_{12}(T,0)x_0]$, which substituted into (4.12a) yields

$$H\phi_{12}(T,0)\phi_{22}(T,0)^{-1}H^T\eta$$
$$=\chi+\{\phi_{22}(T,0)^{-1}\phi_{12}(T,0)-H\phi_{11}(T,0)\}x_0,$$
$$\eta=\frac{\chi+[\phi_{22}(T,0)^{-1}\phi_{12}(T,0)-H\phi_{11}(T,0)]x_0}{H\phi_{12}(T,0)\phi_{22}(T,0)^{-1}H^T} \quad (4.13)$$
$$=\frac{\chi+Z(T)x_0}{HK(T)H^T},$$

with obvious notation.

Now the following can be shown.

Theorem 4.1: Assume that $\phi_{22}(T,0)$ is invertible and let $K(T)=\phi_{12}(T,0)\phi_{22}(T,0)^{-1}$ be the matrix in the denominator of the expression for the Lagrange parameter η in (4.13); then $K(T)$ is strictly negative definite, $K(T)<0$, and $HK(T)H^T\downarrow HW_{11}W_{21}^{-1}H^T$ as T approaches infinity.

Proof: The arguments for the proof make use of the Riccati equation and are based on techniques in Ref. 5. Since the transition matrix of the Hamiltonian matrix in (4.11) satisfies the equation

$$\frac{d}{dt}\begin{bmatrix}\phi_{11}(t,0) & \phi_{12}(t,0)\\ \phi_{21}(t,0) & \phi_{22}(t,0)\end{bmatrix}$$
$$=\begin{bmatrix}A & -br^{-1}b^T\\ -Q & -A^T\end{bmatrix}\begin{bmatrix}\phi_{11}(t,0) & \phi_{12}(t,0)\\ \phi_{21}(t,0) & \phi_{22}(t,0)\end{bmatrix}, \quad (4.14)$$

the partition matrices $\phi_{12}(t,0)$, $\phi_{22}(t,o)$ satisfy

$$\frac{d}{dt}\phi_{12}(t,0)=A\phi_{12}(t,0)-br^{-1}b^T\phi_{22}(t,0),$$
$$\frac{d}{dt}\phi_{22}(t,0)=-Q\phi_{12}(t,0)-A^T\phi_{22}(t,0), \quad (4.15)$$

from which it is established that

$$\frac{d}{dt}\phi_{22}(t,0)^{-1}=\phi_{22}(t,0)^{-1}Q\phi_{12}(t,0)\phi_{22}(t,0)$$
$$+\phi_{22}(t,0)^{-1}A^T. \quad (4.16)$$

This is now used to show that the matrix $K(t)$ satisfies the following Ricatti equation:

$$\frac{d}{dt}K(t)=\frac{d}{dt}\phi_{12}(T,0)\phi_{22}(T,0)^{-1}$$
$$+\phi_{12}(T,0)\frac{d}{dt}\phi_{22}(T,0)^{-1}$$
$$=AK(t)+K(t)A^T+K(t)QK(t)-br^{-1}b^T, \quad (4.17)$$

with $K(0)=0$ since $\phi_{12}(0,0)=0$. We now introduce the vector $(v(t)^T,y(t)^T)^T$ satisfying Eq. (2.6) with the two-point boundary conditions $v(0)=0$, $y(T)=H^T$, and let $v(t)=K(t)y(t)$. Then it is found that

$$-HK(T)H^T$$
$$=-H\phi_{12}(T,0)\phi_{22}(T,0)^{-1}H^T$$
$$=-\int_0^T\frac{d}{dt}\{y(t)^TK(t)y(t)\}dt$$
$$=-\int_0^T\{[-Qv(t)-A^Ty(t)]^TK(t)y(t)\}dt$$
$$-\int_0^T\{y(t)^TK(t)[-Qv(t)-A^Ty(t)]$$
$$-\int_0^T y(t)^T[AK(t)+K(t)A^T$$
$$+K(t)QK(t)-br^{-1}b^T]y(t)\}$$
$$=\int_0^T\{v(t)^TQv(t)+y(t)^Tbr^{-1}b^Ty(t)\}dt$$
$$=\int_0^T\{v(t)^TQv(t)+r^{-1}[b^Ty(t)]^2\}dt>0, \quad (4.18)$$

which holds for all final conditions $y(T)=H^T$ and all times T. The last term is strictly positive since the only solution rendering this integral zero is $(v(t)^T,y(t)^T)^T=(0,0)$ for all t, but this solution does not satisfy the boundary conditions. Since $(v(t)^T,y(t)^T)^T$ satisfies Eq. (2.6) and (A,b) is controllable, the last integral in (4.18) is bounded from above and increases as T increases. Obviously then this integral approaches a finite limit, so indeed

$$0>HK(T)H^T\downarrow HK(\infty)H^T. \quad (4.19)$$

The fact that $K(\infty)=W_{11}W_{21}^{-1}$ can be shown from the analytical solution for $K(t)$ in terms of the eigenvectors and eigenvalues (with nonzero real parts) of the Hamiltonian matrix and the fact that in the limit as $T\to\infty$ the negative exponentials vanish.

This result can be used to determine the behavior of $x(t)$ and $\lambda(t)$ in the various limits of the design constants T and r. From (4.11), (4.13), and the expression for $\lambda(0)$, it is clear that $x(t)$ and $\lambda(t)$ are proportional to η, which in turn is inversely proportional to $K(T)$. For fixed r, the theorem above shows that $|K(T)|$ increases if T increases; hence, from (4.8), it follows that the excursions of $x(t)$ and the peak intensity of the optical field decrease. Once $T\approx 2c'T_0$, the behavior of $x(t)$ and $\lambda(t)$ can be divided into the three regions introduced before and, for even larger T, both $x(t)$ and $\lambda(t)$ remain unchanged except that the middle time region where $x(t)\approx 0$ and $\lambda(t)\approx 0$ increases. For such large final time we have that $K(T)\approx K(\infty)$ and the shape of $x(t)$ and $\lambda(t)$ hardly changes, so that the total cost of this optimal field for such large times remains constant. If, on the other hand, $T\downarrow 0$, then $K(T)\approx 0$ so that $\eta\uparrow\infty$ and obviously then, also, $\max_t|u(t)|\uparrow\infty$. In a very short time interval only an extremely intense input could send the system from the initial state to the desired final state. Conversely, if T is fixed and r is small enough that $T\approx 2c'T_0$, then the asymptotic approximation (4.8) holds for $x(t)$ and $\lambda(t)$ and changes in r effect only the coefficients of the matrix Λ. As r increases this does not hold anymore and, in fact, for very large r we have that $K(T)\propto r^{-1}$ so that $\eta\propto r$, while further investigation reveals that in this case generally $\max_t|u(t)|\uparrow\infty$. This property for the matrix $K(T)$ can be shown from a first-order perturbation analysis of the partition matrix $\phi_{12}(T,0)$ or more directly from the behavior of the solution of the Riccati equation (4.17) for large r. The intensity of the required field under these circumstances must increase because the coupling of the state $x(t)$ with $\lambda(t)$ decreases, yet the final condition $x(T)$ must be attained. The minimal cost strategy therefore suggests that for a fixed parameter r the final time $T\approx 2c'T_0$ is chosen, since then the minimal intensi-

ty is required for the optical field. If a low peak intensity is still difficult to realize, in practice the design parameter r must be increased and the T chosen accordingly. In the case that the parameter r is large, the critical time T_0 can be estimated from (4.10); otherwise numerical determination of the eigenvalues of the Hamiltonian matrix will be needed.

V. CONCLUSIONS

This paper presents two different optimal control strategies to generate selective vibrational excitations in a molecule modeled as a classical harmonic system, or equivalently a quantum harmonic system taken with Ehrenfest's theorem. Both employ a cost functional balancing the fluence of the optical field with the internal energy of the harmonic system, but one strategy requires that the target bond attains the desired excitation exactly (a terminal constraint controller) while the other introduces a final cost for not reaching the desired final state. The behavior of the two controllers was shown to be very similar, which may not be surprising since they were based on the solution to the same Hamiltonian equation. In this paper it is assumed that the positions and momenta of the physical system are controllable by the optical field. In practice, however, molecules generally show many symmetries and the coupling to the optical field is usually such that it does not allow for all possible oscillations. Controllability in that case may not hold but the system is still stabilizable so that most arguments in this paper remain valid.

Some mathematically interesting problems were encountered due to the physical nature of the application. Since the required field intensity tends to be high, the limit of large r in the control strategies becomes important in Sec. III an approximation for the eigenvalues of the Hamiltonian matrix was obtained for this case. To first order, the eigenvalues are the imaginary eigenvalues of the classical system corresponding to the vibrational frequencies in the infrared spectrum and a real part proportional to $r^{-1/2}$, which indicates the time scale needed to provide a satisfactory excitation at the final time. From the asymptotic expressions for the state and costate in Sec. IV for large final time, it is seen that the optimal final time increases with $r^{1/2}$ while the associated optimal electromagnetic field decreases as r^{-1}. In the design of actual experiments, one strategy may be to choose the parameter r so large that the optical field is sufficiently weak with the final time being selected accordingly. These large time approximations are especially useful for numerical purposes as well, since the exact solution to the Hamiltonian equations on long time intervals requires a treatment on several different time scales. The numerical solutions are particularly difficult to obtain if the eigenvalues of the Hamiltonian vary considerably in magnitude. Another mathematical result in Sec. III includes a bound on the eigenvalues of the Hamiltonian matrix for the case where the harmonic system is a linear chain. This bound turns out to be independent of the length of the chain. In the terminal constraint controller in Sec. IV, the Lagrange parameter enforcing the final condition on the state depends on a particular matrix combination that was shown to be negative definite by relating it to the Riccati equation of an entirely different control problem. Stabilizability insures the existence of the Lagrange parameter and using well-known arguments it was shown that this matrix combination approached a limit as the final time increased to infinity.

Approximating a molecule as a harmonic system is only valid if the motion of the individual atoms is not too vigorous so that the interaction potential between the atoms is approximately quadratic. The motion of a real molecule, however, is ruled by the laws of quantum mechanics involving nonlinear potentials where the presence of electrons may also need to be taken into account if the fields become sufficiently intense. The uncertainty in the interaction potential as well as external disturbances may be modeled as a noise term in the equations of motion of the atoms, which suggests generalizing the controlling strategy above to a stochastic linear optimal control problem. Another possibility to compensate for the uncertainty in the equations of motion is to augment the cost functional with an expression depending on the sensitivity of the model. The second controller strategy introduced in this paper can be easily generalized to the stochastic case and may yield useful insight into the feasibility of constructing optimal electromagnetic fields in the presence of uncertainties. A further possibility is to introduce a nonlinear interaction potential in the determinisitic model above rendering the optimal control problem also nonlinear, but many analytical results are known for this case. One aspect concerning the time scale of the control process is that, for large r and correspondingly large T, the collisions of molecules cannot always be excluded. This will introduce impulses to the system in time intervals governed by an exponential law, which turns the control problem into a stochastic one containing Poisson noise, but little is known about this type of linear optimal control. Finally, it would also be interesting to consider the distributed system limit of this problem, where the chain of atoms is so large that it can be adequately modeled as a continuous string.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Office of Naval Research and the Air Force Office of Scientific Research for support of this research. One of the authors would like to thank H. Kwakernaak, H. Nijmijer, A. v.d. Schaft, and J. W. Polderman in the Twente University of Technology for helpful comments.

[1] A. P. Pierce, M. A. Dahleh, and H. Rabitz, Phys. Rev. A **37**, 4950 (1988).
[2] S. Shi, A. Woody, and H. Rabitz, J. Chem. Phys. **88**, 6870 (1988).
[3] A. E. Bryson and Y. C. Ho, *Applied Optimal Control* (Hemisphere, New York, 1975).
[4] T. Kailath, *Linear Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
[5] H. Kwakernaak and R. Sivan, *Linear Optimal Control* (Wiley, New York, 1972).
[6] C. Swieters, J. G. B. Beumee, and H. Rabitz, manuscript in preparation.
[7] A. G. Butkovskii and Yu. I. Samoilenko, "Control of quantum systems," translated from Avtom. Telemekh. **4**, 5–25 (1979).
[8] V. I. Arnold, *Mathematical Methods in Classical Mechanics*, Graduate Text in Mathematics No. 60 (Springer, New York, 1978).
[9] V. Kučera, IEEE Trans. Autom. Control **AC-17**, 344 (1972).
[10] C. Lanczos, *The Variational Principles of Mechanics* (Dover, New York, 1970).

Angular distributions, rotational summations, and integrations in the quadrupole and two-photon photodissociation of molecules[a)]

Shan-Tao Lai and Ying-Nan Chiu[b)]
Center for Molecular Dynamics and Energy Transfer, Department of Chemistry, The Catholic University of America, Washington, DC 20064

(Received 29 June 1989; accepted for publication 13 December 1989)

In this work, the *closed forms* of the angular distributions in quadrupole and two-photon photodissociation of molecules are obtained. These are applicable to the angular factors of other second-order, nonlinear, or quadratic processes.

I. INTRODUCTION

The angular distribution of molecular photodissociated fragments has been of interest experimentally and theoretically for a long time.[1–15] For a one-photon electric-dipole transition process the general formula for the cross section of the observed fragments is

$$I(\Theta) = (I_0/4\pi)[1 + \beta P_2(\cos\Theta)], \tag{1}$$

where Θ is the polar angle between the electric polarization direction of the incident light (taken to be the Z axis) and the observation direction κ (Fig. 1). For rotating molecules, I_0 and β contain not only rovibronic transition matrix elements but also the angular dependence of the fragmentation direction versus the dipole excitation direction.

In obtaining the rotational matrix elements, one not only has to integrate the dipole transition moment over rotation matrices that make up the wave functions for symmetric top or linear molecules (or over spherical harmonics for rigid rotors), but also has to sum over the magnetic quantum number M. Even for a dipole transition for a diatomic molecule such integrations and summations can be difficult and tedious. Recently the present authors devised a general method[16] to perform these integrations and summations beyond the dipole operator. In the present work we propose to apply our general method to derive the angular distribution for the electric quadrupole and nonresonant two-photon transition processes in photodissociation, which may contain azimuthal (Φ) dependence besides the polar (Θ) dependence.[15] We shall apply our method not only to diatomic and linear molecules but also to symmetric top molecules of a specific symmetry (D_{3h}) for which the electronic selection rule over the angular momentum projections on the molecular axis, K (or Λ), will also be considered. In the process of deriving the above-mentioned angular factors, we have obtained many additional formulas for the integration of rotation matrices over tensor operators (beyond the first rank) and summations of various rotational products over M. We give these in Appendices A–D. These general integrations and summations will be useful in other angular dependence studies, for example, in the extension of angular distribution of the electric-dipole excited photoelectrons[17,18] to electric-quadrupole and two-photon excited photoionizations and in other second-order angular correlation studies.

II. CLASSICAL TREATMENT OF ELECTRIC-QUADRUPOLE AND TWO-PHOTON PHOTOFRAGMENTATION OF DIATOMIC MOLECULES

We define the coordinate system for photodissociation in Fig. 1. All of the vector directions and their designations are the same as Chiu's.[15] We assume that a molecule's center of mass is at the origin, XYZ stand for the laboratory frame, Z is the electric polarization direction of the incident light in a dipole transition process (but is taken to be the light propagation direction in quadrupole transition processes), μ is the molecular principal z axis (and may also be considered to be the molecular transition dipole moment direction when a parallel transition occurs), $\mathbf{r}_d$ is the relative motion vector for the two dissociated fragments, and κ is the direction of observation of the fragments. We summarize all of the above definitions as follows:

$$\begin{aligned}
&\cos\vartheta_\mu \equiv \hat{\mu}\cdot\hat{z},\\
&\cos\vartheta_i \equiv \hat{\mu}\cdot\hat{r}_d,\\
&\cos\vartheta \equiv \hat{Z}\cdot\hat{r}_d,\\
&\cos\vartheta' \equiv \hat{\kappa}\cdot\hat{r}_d,\\
&\cos\Theta \equiv \hat{\kappa}\cdot\hat{Z},
\end{aligned} \tag{2}$$

where a caret designates a unit vector.

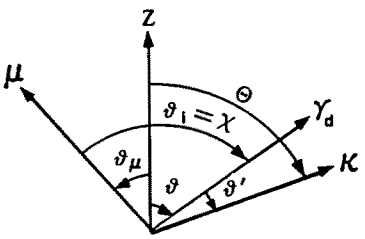


FIG. 1. Coordinate system for photodissociation. Here, **Z** is the electric polarization direction of light, μ is the molecular principal (**z**) axis (μ may be considered the molecular transition moment direction if the latter is along the molecular **z** axis), $\mathbf{r}_d$ is the relative motional vector for the two dissociated fragments, κ is the direction of observation of the fragments. The relationship between different axis systems is comprised of three Euler angles. Thus ϑ_μ is part of $\Omega = (\varphi_\mu \vartheta_\mu \chi_\mu)$. The arrow directions indicate the sign of rotational angles. Self-consistent use of these directional conventions is important.

[a)] Taken in part from a dissertation to be submitted by Shan-Tao Lai in partial fulfillment of the Ph.D. degree at the Catholic University of America.

[b)] To whom requests for reprints should be addressed.

 0022-2488/90/051261-13$03.00

We shall formulate the fragment angular distribution of the photodissociation by a beam of circularly polarized light propagating along Z. This amounts to combination of linearly polarized light with polarization along X and Y but with a 90° phase difference. This discussion will be limited to a one-photon electric-quadrupole transition. The angular factors will be analogous to a two-photon process without a resonant intermediate state. The Euler angles $(\varphi_\mu \vartheta_\mu \chi_\mu)$ describe the molecule-fixed frame relative to the laboratory-fixed frame. The electric vector ε of the light beam has been chosen as along the Z axis of the lab-fixed frame. The angle ϑ_i between the molecular principal axis and $\mathbf{r}_d$ is a very important angle, which contains the dynamical information of the dissociation process. Without loss of generality we can consider that a molecule absorbs a photon in one of the linearly polarized quadrupole mode[19] XZ and dissociates into two fragments by an internal quadrupole d_{z^2} transition operator in molecular coordinates. (This is in contrast to dissociation by the dipole z transition operator.) The probability of absorption is proportional to $|\langle XZ\rangle|^2$. For a two-photon process this may correspond to two beams polarized along X and Z, respectively. For convenience we define

$$H' = r^2 Y_{21}^{\text{space}}, \tag{3a}$$

where H' is an operator for left circularly polarized light defined in terms of the space frame. Also, H' can be expressed in terms of the molecular frame:

$$H' = r^2 \sum_\mu D^2_{\mu 1}(-\Omega_\mu) Y^{\text{mol}}_{2\mu} = r^2 \sum_\mu D^{2*}_{1\mu}(\Omega_\mu) Y^{\text{mol}}_{2\mu}. \tag{3b}$$

We denote $W(\Theta\Phi)$ as the final recoil distribution (or differential cross section) of the fragments in the laboratory fixed frame. It may be expanded in the complete set of spherical harmonics as follows:

$$W(\Theta,\Phi) = \sum_{l,q} B_{lq} Y_{lq}(\Theta,\Phi), \tag{4}$$

where

$$\begin{aligned} B_{lq} &= \langle Y^*_{lq}(\Theta,\Phi)\rangle \\ &= ((2l+1)/4\pi)^{1/2} \\ &\quad \times \sum_{q',s} \langle D^l_{qs}(\varphi_i \vartheta_i \chi_i) D^l_{sq'}(\varphi_\mu \vartheta_\mu \chi_\mu) \\ &\quad \times D^{l*}_{q'0}(\varphi' \vartheta' \chi')\rangle. \end{aligned} \tag{5}$$

Here we have expanded the spherical harmonics in terms of three rotation matrices. Let $I(\Theta\Phi)$ denote the angular distribution of the fragments observed in the laboratory-fixed frame. A particular molecule oriented at the origin with Euler angles $(\varphi_\mu \vartheta_\mu \chi_\mu)$ will make a contribution $|H'|^2 W(\Theta\Phi)$ to $I(\Theta\Phi)$, and the total fragment angular distribution is given by integrating over all molecular orientations:

$$I(\Theta,\Phi) = \int \left| r^2 \sum_\mu D^{2*}_{1\mu}(\Omega_\mu) Y^{\text{mol}}_{2\mu} \right|^2 \cdot W(\Theta,\Phi) d\Omega_\mu. \tag{6}$$

Here the spherical harmonics in molecular coordinates will be understood to represent molecular transition moments. Using the properties of Wigner rotation matrix elements, we obtain

$$\begin{aligned} I(\Theta,\Phi) = 10\pi \sum_{l,q} (2l+1)^{1/2} \cdot \sum_{\mu,\mu'} (-1)^{-\mu'} \langle D^l_{q0}(\varphi_i \vartheta_i \chi_i) D^l_{-\mu'+\mu,0}(\varphi' \vartheta' \chi')\rangle \\ \cdot \begin{pmatrix} 2 & 2 & l \\ \bar{1} & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 2 & l \\ \bar{\mu} & \mu' & -\mu'+\mu \end{pmatrix} Y_{lq}(\Theta,\Phi) \\ \times \sum_L (2L+1)^{1/2} \cdot r^4 \begin{pmatrix} 2 & 2 & L \\ \mu & \bar{\mu}' & \overline{\mu-\mu'} \end{pmatrix} \begin{pmatrix} 2 & 2 & L \\ 0 & 0 & 0 \end{pmatrix} Y^{\text{mol}}_{L,\mu-\mu'}. \end{aligned} \tag{7}$$

We take $\mu = \mu' = 0$ for molecular dissociation with the d_{z^2} internal transition moment. Therefore we obtain the total fragment angular distribution for electronic-quadrupole transition by using left circularly polarized light as follows:

$$\begin{aligned} I(\Theta,\Phi) = I_0 \Bigg\{ 1 + \frac{\sqrt{4\pi}}{7} \sum_q \langle D^2_{q0}(\varphi_i \vartheta_i \chi_i) D^2_{00}(\varphi' \vartheta' \chi')\rangle Y_{2q}(\Theta,\Phi) \\ - \frac{4}{7} \frac{\sqrt{4\pi}}{5} \sum_q \langle D^4_{q0}(\varphi_i \vartheta_i \chi_i) D^4_{00}(\varphi' \vartheta' \chi')\rangle Y_{4q}(\Theta,\Phi) \Bigg\}, \end{aligned} \tag{8}$$

where

$$I_0 = -\sqrt{\pi} r^4 \sum_L (2L+1)^{1/2} \begin{pmatrix} 2 & 2 & L \\ 0 & 0 & 0 \end{pmatrix}^2 Y^{\text{mol}}_{L0}. \tag{9}$$

Here the angle ϑ' contains dynamical information on the dissociation process and the second and fourth Legendre moments may be calculated by the Monte Carlo method or other statistical methods. The angle ϑ_i is determined by the nature of the dissociative electronic transition. In the fast dissociation limit, the molecule dissociates before it has time to rotate significantly, i.e., $\cos\vartheta' = 1$ or, geometrically, the fragment recoil angle is equal to zero with respect to the $\mathbf{r}_d$ axis.

Equation (8) is a general electronic-quadrupole transition angular distribution that shows noncylindrical symmetry. The azimuthal Φ dependence in addition to the polar Θ angle dependence arises from quadrupole (or two differently polarized photons) absorption. Of course, the molecular quadrupole transition operator is not limited to $d_{z^2} = (4\pi/5)^{1/2} Y_{20}$. It will be shown in later quantum mechan-

ical treatments that different molecular quadrupole transitions will have different vibronic selection rules and different angular dependences.

III. QUANTUM MECHANICAL TREATMENT OF ELECTRIC-QUADRUPOLE AND TWO-PHOTON PHOTOFRAGMENTATION OF DIATOMIC MOLECULES

A. Wave functions and quadrupole elements

We use H' to denote the interaction potential between the molecule and the radiation field in the quadrupole mode. Thus for light propagating along Z the interaction can be written as in Eq. (3) for left circularly polarized light.

Let a photon beam $h\nu$ impinge on a target of a randomly oriented molecule (or atom). The molecule is excited from a lower bound state Ψ_i to a higher-lying repulsive sate Ψ_f. The differential cross section $I(\Theta\Phi)$ for the dissociative transition $i\rightarrow f$ is proportional to the square of transition quadrupole matrix element,

$$I(\Theta,\Phi)\sim|\langle\Psi_f|H'|\Psi_i\rangle|^2, \tag{10}$$

where $(\Theta\Phi)$ are polar and azimuthal angles referred to the laboratory frame (Fig. 1). According to the Born–Oppenheimer approximation, for the bound state, the total wave function can be written as a product[7]

$$|\Psi_i\rangle=\psi_i^{(e)}(r_{ji};R)\chi_{vJ}(R)\left(\frac{2J+1}{8\pi^2}\right)^{1/2}D^J_{MK}(\varphi_\mu,\vartheta_\mu,0). \tag{11}$$

Here Ψ_i^e is the electronic wave function that depends on electronic coordinates r_j and on the internuclear separation R, $\chi_{vJ}(R)$ is the wave function for the vibrational state v, and D^J_{MK} denotes a rotational part that expresses the orientation of the diatomic (or linear) molecule. The results of using a single rotation matrix D^J_{MK} will be the same as using a $(1/\sqrt{2})$ normalized linear combination of two rotation matrices for lambda doubling states. Angles $(\varphi_\mu\vartheta_\mu 0)$ specify the orientation of the molecular figure axis along $\vec{\mu}$. Here J is the total angular momentum. Its projection along the Z axis of the space-fixed frame (chosen to lie along $\vec{\epsilon}$) is M and its projection along the figure z axis ($\vec{\mu}$) in the body-fixed frame is K ($K=0$ for Σ states, $+1$ for Π states, etc.).

For the repulsive state wave function chosen to satisfy the proper boundary conditions, we have the expression[7]

$$\langle\Psi_f|=\sum_{J',M'}(2J'+1)i^{J'}\exp(-i\delta_{J'})\Psi_f^{(e)}(r'_j;R)\times\chi_{J'}(kR)D^{J'*}_{M'\Lambda'}(\varphi_\mu,\vartheta_\mu,0)D^{J'}_{M'\Lambda}(\Phi,\Theta,0), \tag{12}$$

where the angles $(\varphi_\mu\vartheta_\mu 0)$ and $(\Phi\Theta 0)$ have the same meaning as before. We now substitute Eqs. (12), (11), and (3) into Eq. (10). The differential cross section may be written in the form

$$I(\Theta,\Phi)=\left|\sum_{J',M'}H'^{fi}_{vJ'}A_{J'}D^{J'}_{M'\Lambda}(\Phi,\Theta,0)\right|^2, \tag{13}$$

where

$$H'^{fi}_{vJ'}=i^{J'}\exp(-i\delta_{J'})\cdot\langle\psi_f^{(e)}(r'_j;R)\chi_{J'}(KR)\times|Y^{\rm mol}_{2\mu}\cdot r^2|\psi_i^{(e)}(r_j;R)\chi_{vJ}(R)\rangle \tag{14}$$

and

$$A_{J'}=(2J'+1)(2J+1)^{1/2}\begin{pmatrix}J' & 2 & J\\ \bar{M}' & \bar{1} & M\end{pmatrix}\times\begin{pmatrix}J' & 2 & J\\ \bar{\Lambda} & \bar{\mu} & K\end{pmatrix}. \tag{15}$$

Equation (13) is the angular distribution of the fragments that arises from one (J,M) initial sublevel. Since the molecular ensemble is randomly oriented, we assume that the magnetic levels are equally populated; then the angular distribution of the fragments is obtained by averaging over the M of the initial state. From Eq. (13),

$$I(\Theta,\Phi)=\sum_M\left|\sum_{J',M'}H'^{fi}_{vJ'}\cdot(2J'+1)\times\begin{pmatrix}J' & 2 & J\\ \bar{M}' & \bar{1} & M\end{pmatrix}\begin{pmatrix}J' & 2 & J\\ \bar{\Lambda} & \bar{\mu} & K\end{pmatrix}\times D^{J'}_{M'\Lambda}(\Phi,\Theta,0)\right|^2. \tag{16}$$

B. $\Sigma\rightarrow\Sigma$ transition due to electric quadrupole

For the $\Sigma\rightarrow\Sigma$ transition, $\mu=0$ and $\Lambda=K=0$; in this case, Eq. (16) becomes

$$I(\Theta,\Phi)=\sum_M\left|\sum_{J',M'}H'^{fi}_{vJ'}\cdot(2J'+1)\times\begin{pmatrix}J' & 2 & J\\ \bar{M}' & \bar{1} & M\end{pmatrix}\begin{pmatrix}J' & 2 & J\\ 0 & 0 & 0\end{pmatrix}\times D^{J'}_{M'0}(\Phi,\Theta,0)\right|^2. \tag{17}$$

For H' given in Eq. (3) the selection rules are $\Delta J=0,\pm 2$, and $M'=M-1$; therefore we obtained the closed form of Eq. (17) as follows (for details, see Appendix B):

$$I(\Theta,\Phi)=I_0(1+b_1\sin^2\Theta+b_2\sin^4\Theta), \tag{18}$$

where

$$I_0=\tfrac{1}{4}[a_1+4(J-1)(J+1)a_2+4J(J+2)a_3+2a_4+2a_5+4a_6], \tag{19}$$

$$b_1=\frac{(4J^2+4J-9)a_1-12(J-2)a_2+12(J+3)a_3-2(2J-7)a_4+2(2J+9)a_5-40a_6}{2a_1+8(J-1)(J+1)a_2+8J(J+2)a_3+4a_4+4a_5+8a_6}, \tag{20}$$

$$b_2=\frac{-3(J-1)(J+2)a_1-3(J-2)(J-3)a_2-3(J+3)(J+4)a_3+5(J-2)a_4-5(J+3)a_5+35a_6}{2a_1+8(J-1)(J+1)a_2+8J(J+2)a_3+4a_4+4a_5+8a_6}, \tag{21}$$

and

$$a_1=\frac{9(J+1)^2}{(2J-1)^2(J-1)^2(2J-3)^2}|H'^{fi}_{v,J}|^2,\quad a_2=\frac{J(J-2)}{(2J+1)^2(2J-1)^2}|H'^{fi}_{v,J-2}|^2,$$
$$a_3=\frac{(J+1)(J+3)}{(2J+3)^2(2J+1)^2}|H'^{fi}_{v,J+2}|^2,$$
$$a_4=\frac{3[J(J+1)(J-1)(J-2)]^{1/2}}{(2J+3)(2J+1)(2J-1)^2}\left[H'^{fi*}_{v,J}H'^{fi}_{v,J-2}+H'^{fi}_{v,J}H'^{fi*}_{v,J-2}\right],\tag{22}$$
$$a_5=\frac{3[J(J+1)(J+2)(J+3)]^{1/2}}{(2J+3)^2(2J-1)(2J+1)}\left[H'^{fi*}_{vJ}H'^{fi}_{v,J+2}+H'^{fi}_{vJ}H'^{fi*}_{v,J+2}\right],$$
$$a_6=\frac{J(J+1)[(J-1)(J-2)(J+2)(J+3)]^{1/2}}{(2J-1)(2J+3)(2J+1)^2}\left[H'^{fi*}_{vJ-2}H'^{fi}_{vJ+2}+H'^{fi}_{v,J-2}H'^{fi*}_{v,J+2}\right].$$

C. Z-polarized light two-photon transition

For Z-polarized light propagating along X or Y, the two-photon interaction operator is

$$ZZ=\frac{4\pi}{3}Y_{10}\cdot Y_{10}$$
$$=\frac{4\pi}{3}\sum_l C^2(11l;000)Y_{l0}\cdot\left(\frac{3\cdot 3}{4\pi(2l+1)}\right)^{1/2}.\tag{23a}$$

Without loss of generality, we shall omit the scalar constant term and consider the angular factors due to $l=2$:

$$H'=r^2Y^{\text{space}}_{20}$$
$$=r^2\sum_\mu D^{2*}_{0\mu}(\Omega_\mu)Y^{\text{mol}}_{2\mu},\tag{23b}$$

Similarly, we obtain the angular distribution of the fragments produced by this two-photon transition by Z-polarized light absorbed by a diatomic molecule:

$$I(\Theta,\Phi)=\sum_M\left|\sum_{J',M'}H'^{fi}_{vJ'}\cdot(2J'+1)\times\begin{pmatrix}J' & 2 & J\\ \bar{M}' & 0 & M\end{pmatrix}\begin{pmatrix}J' & 2 & J\\ \bar{\Lambda} & \bar{\mu} & K\end{pmatrix}\times D^{J'}_{\mu'\Lambda}(\Phi,\Theta,0)\right|^2,\tag{24}$$

where $H'^{fi}_{vJ'}$ has the same form as in Eq. (14). For the $\Sigma\to\Sigma$ two-photon transition, $\mu=0$ and $\Lambda=K=0$. Therefore Eq. (24) is reduced to

$$I(\Theta,\Phi)=\sum_M\left|\sum_{J',M'}H'^{fi}_{vJ'}\cdot(2J'+1)\times\begin{pmatrix}J' & 2 & J\\ \bar{M}' & 0 & M\end{pmatrix}\begin{pmatrix}J' & 2 & J\\ 0 & 0 & 0\end{pmatrix}\times D^{J'}_{M'0}(\Phi,\Theta,0)\right|^2.\tag{25}$$

For this, the selection rules for $\Sigma\to\Sigma$ two-photon transition are $\Delta J=0,\ \pm 2$ and $M'=M$. With the help of Eqs. (24) and (32) of Ref. 16, we obtained the closed form of Eq. (25):

$$I(\Theta,\Phi)=I_0[1+\beta P_2(\cos\Theta)+\gamma P_4(\cos\Theta)],\tag{26}$$

where

$$I_0=(a_0+b_0+c_0)d_0,$$
$$a_0=12\cdot(J-1)(2J-1)(2J-3)(2J+3)^2|H'^{fi}_{v,J-2}|^2,$$
$$b_0=5\cdot(J+1)(2J+1)(2J-3)(J^2+J+6)|H'^{fi}_{v,J}|^2,\tag{27}$$
$$c_0=12\cdot(J-1)(2J+5)(2J-1)(2J+3)^2|H'^{fi}_{v,J+2}|^2,$$
$$d_0=[40(2J+1)(2J-1)^2(2J+3)^2(2J-3)]^{-1};$$
$$\beta=\frac{40}{7}\frac{(a_1+b_1+c_1+d_1+e_1)}{(a_0+b_0+c_0)},$$
$$a_1=3(J+1)(J-1)(2J+3)^2(2J-3)H'^{fi}_{J-2,J},$$
$$b_1=18(J+1)(J+2)(2J-1)^2(2J-3)H'^{fi}_{J+2,J},$$
$$c_1=3(J-1)(J-2)(2J-3)(2J+3)^2|H'^{fi}_{v,J-2}|^2,$$
$$d_1=-(J+1)(2J+1)(2J-3)\times(4J^2+4J-15)|H'^{fi}_{v,J}|^2,$$
$$e_1=3(J-1)(2J+5)(2J+3)(J+3)\times(2J-1)|H'^{fi}_{v,J+2}|^2;\tag{28}$$
$$\gamma=\frac{18}{7(2J+1)}\frac{(a_2+b_2+c_2+d_2+e_2+f_2)}{(a_0+b_0+c_0)},$$
$$a_2=10(J+1)(J-1)(J-2)(2J+1)(2J-3)\times(2J+3)H'^{fi}_{J-2,J},$$
$$b_2=40(J+1)(J+2)(J+3)(2J+1)(2J-1)\times(2J-3)H'^{fi}_{J+2,J},$$
$$c_2=35(J+1)(J-1)(J+2)(2J-1)(2J-3)\times(2J+3)H'^{fi}_{J+2,J-2},\tag{29}$$
$$d_2=3(J-1)(J-2)(J-3)(2J-3)\times(2J+3)^2|H'^{fi}_{v,J-2}|^2,$$
$$e_2=12(J+1)(J-1)(J+2)(2J-3)\times(2J+1)^2|H'^{fi}_{v,J}|^2,$$
$$f_2=3(J-1)(J+3)(J+4)(2J-1)(2J+5)\times(2J+3)|H'^{fi}_{v,J+2}|^2;$$

and

$$H'^{fi}_{J-2,J}\equiv H'^{fi*}_{v,J-2}\cdot H'^{fi}_{v,J}+H'^{fi}_{v,J-2}\cdot H'^{fi*}_{v,J},$$
$$H'^{fi}_{J+2,J}\equiv H'^{fi*}_{v,J+2}\cdot H'^{fi}_{v,J}+H'^{fi}_{v,J+2}\cdot H'^{fi*}_{v,J},\tag{30}$$
$$H'^{fi}_{J+2,J-2}\equiv H'^{fi*}_{v,J+2}\cdot H'^{fi}_{v,J-2}+H'^{fi}_{v,J+2}\cdot H'^{fi*}_{v,J-2}.$$

In this section we have elucidated the general rotational dependence of the one-photon quadrupole and two-photon photofragmentation in terms of molecular vibronic matrix elements of diatomic and linear molecules. In the next section we shall give the angular distribution for photofragmentation of symmetric top molecules by the above mechanism.

IV. ELECTRIC-QUADRUPOLE AND ONE- AND TWO-PHOTON PHOTOFRAGMENTS OF ROTATING LINEAR AND SYMMETRIC TOP MOLECULES OF D_{3h} SYMMETRY

We still use the same coordinate system as in Fig. 1. The linear and nonlinear polyatomic molecules are located at the origin. All of the vectors have been defined in Sec. II. The total cross section $\sigma_L(\Gamma_\gamma)$ or the probability of the observation is proportional to

$$\sigma_L^{(N)}(\Gamma_\gamma) = \frac{1}{8\pi^2}\cdot\frac{1}{2J+1}\sum_M \left|\int \langle F,f|M_L^{(N)}|J,i\rangle d\Omega_\mu\right|^2 = \frac{1}{8\pi^2}\cdot\frac{1}{2J+1}\sum_M \left|\int \langle F|J\rangle\cdot[M_L^{(N)}]d\Omega_\mu\right|^2 \tag{31}$$

where $N=1$ or 2 denotes a one- or two-photon process, respectively, L stands for the 2^L multipole transition, and Γ_γ is the irreducible symmetry of the transition density. $[M^{(N)}]$ is the vibronic transition matrix element.[15] f and i are the final and initial vibronic states, respectively, $|J\rangle$ represents the rotational wave functions of the rotating parent molecule, and the magnetic quantum number M represents the substate of the rotating parent molecule.

A. One-photon electric-dipole fragmentation

For review and comparison with existing work, we first consider a one-photon electric-dipole process. But we proceed beyond existing work to obtain closed forms and results for symmetric top molecules. The transition moments for Z polarization is given by[15]

$$M_1^{(1)} = K_1^{(1)}\langle Z\rangle = K_1^{(1)}\langle R_{10}\rangle = K_1^{(1)}\langle T_{10}^{\rm lab}(r)\rangle. \tag{32}$$

For the bound state the rotational wave function can be written as follows for a linear or diatomic molecule:

$$|JM\rangle = [(2J+1)/4\pi]^{1/2}Y_{JM}(\vartheta,\varphi). \tag{33}$$

[As an approximation for polyatomic molecules we may also use the same rotational wave function. This approximation is forgivable if the symmetric top molecule in the initial state is assumed to have no internal rotation ($K=0$ in D^J_{MK}); more exactly, we need to use Wigner rotational matrix elements as the wave functions of symmetric top molecules. But the calculation process becomes more tedious and will be considered in future work with specific examples.] The plane wave approximation will be used for the relative motion of the two dissociated fragments. If we use $\mathbf{r}_d$ to represent the relative motional vector for the two fragments, the "normalized" plane wave amplitude along the observation direction κ is[15]

$$\langle F| = (1/\sqrt{4\pi})e^{i\kappa\cdot\mathbf{r}_d} = \frac{1}{\sqrt{4\pi}}\sum_{l=0}^{\infty} i^l(2l+1)j_l(kr_d)P_l(\cos\vartheta') = \sqrt{4\pi}\sum_l\sum_{m'} R_{Kl}(r_d)Y^*_{lm'}(\vartheta,\varphi)Y_{lm'}(\Theta,\Phi), \tag{34}$$

where $R_{kl}(r_d) = i^l j_l(kr_d)$ and we have used the rotational transformation to show the relationship between the angles (Fig. 1). We noted that $(\Theta\Phi)$ are the angles of the observation vector κ with respect to the polarization axis Z. The total angular distribution of the fragments is given in the laboratory coordinate system. Here $(\vartheta\varphi)$ are the angles for the orientation of the dissociation vector $\mathbf{r}_d$ with respect to the polarization z axis. They may be related to the molecular orientation angles Ω_μ and molecular internal rotational angles $(\vartheta_i\varphi_i)$ later.

From Eqs. (33) and (34), we obtain

$$\langle F|J\rangle = (2J+1)^{1/2}\sum_{l,m'} R_{Kl}(r_d)(-1)^{m'}\sum_L\left[\frac{(2l+1)(2J+1)}{4\pi(2L+1)}\right]^{1/2}\cdot C(lJL;-m',M,-m'+M) \times C(lJL;000)\cdot\sum_q D^{L*}_{-m'+M,q}(\Omega_\mu)Y_{Lq}(\vartheta_i\varphi_i)Y_{lm'}(\Theta,\Phi). \tag{35}$$

Here we have used the coupling of the spherical harmonics and we have transformed $Y_{L,-m'+m}(\vartheta,\varphi)$ to show its relationship (Fig. 1) to the molecular orientation Ω_μ and $(\vartheta_i\varphi_i)$. We also need to transform the electric dipole transition moment for Z polarization from the lab frame to the molecular frame, i.e.,

$$M_1^{(1)} = K_1^{(1)}\sum_{q'} D^{1*}_{0q'}(\Omega_\mu)\langle T^{\rm mol}_{1q'(r)}\rangle. \tag{36}$$

Substituting Eqs. (36) and (35) into Eq. (31),

$$\sigma_L^{(N)}(\Gamma_\gamma) = \sqrt{\pi}(2J+1)[K_1^{(1)}]^2\sum_M\left|\sum_{J'} R_{KJ'}(r_d)\cdot(2J'+1)^{1/2}\begin{pmatrix}J' & J & 1\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}J' & J & 1\\ \bar{M} & M & 0\end{pmatrix}Y_{J'M}(\Theta,\Phi)\right|^2 \times\sum_{q,q'}(-1)^{q'}\sum_L(2L+1)^{1/2}\begin{pmatrix}1 & 1 & L\\ q & \bar{q}' & -q+q'\end{pmatrix}\begin{pmatrix}1 & 1 & L\\ 0 & 0 & 0\end{pmatrix}Y_{L,q-q'}(\vartheta_i\varphi_i)\langle T^{\rm mol}_{1q}(r)\rangle^*\langle T^{\rm mol}_{1q}(r)\rangle, \tag{37}$$

where $l=J\pm1$, with the help of Eqs. (24) and (24c) of Ref. 16 and the formula of the Clebsch–Gordan coefficients.[20] We finally obtain, for a one-photon electric-dipole transition,

$$\sigma_L^{(N)}(\Gamma_\gamma) = (1/4\pi)A_0\sigma_0^{(1)}[1+\beta P_2(\cos\Theta)], \tag{38}$$

where

$$\sigma_0^{(1)} = [1/(2J+1)][(J+1)|R_{K,J+1}(r_d)|^2 + J|R_{K,J-1}(r_d)|^2], \tag{39}$$

$$A_0 = \frac{1}{9}\sqrt{\pi}\,[K_1^{(1)}]^2(2J+1)\left\{\left[\sqrt{\frac{1}{4\pi}} + \sqrt{\frac{4}{5}}\,Y_{20}(\varphi_i\vartheta_i)\right]|\langle T_{20}^{\rm mol}(r)\rangle|^2 + \sqrt{\frac{1}{5}}\,[|\langle T_{11}^{\rm mol}(r)\rangle|^2 + |\langle T_{1\bar{1}}^{\rm mol}(r)\rangle|^2]\right.$$
$$\cdot\left[Y_{00}(\vartheta_i,\varphi_i) - \frac{1}{\sqrt{5}}\,Y_{20}(\vartheta_i\varphi_i)\right] - \sqrt{\frac{6}{5}}\,[Y_{2\bar{2}}(\vartheta_i\varphi_i)\langle T_{11}^{\rm mol}(r)\rangle^*\langle T_{1\bar{1}}^{\rm mol}(r)\rangle$$
$$\left. + Y_{22}(\vartheta_i\varphi_i)\langle T_{1\bar{1}}^{\rm mol}(r)\rangle^*\langle T_{11}^{\rm mol}(r)\rangle]\right\}, \tag{40}$$

$$\beta = (J+1)(J+2)|R_{K,J+1})(r_d)|^2 + J(J-1)|R_{K,J-1}(r_d)|^2 + 3J(J+1)$$
$$\times[R_{K,J+1}^*(r_d)R_{K,J-1}(r_d) + R_{K,J+1}(r_d)R_{K,J-1}^*(r_d)]/$$
$$(2J+1)[(J+1)|R_{K,J+1}(r_d)|^2 + J|R_{K,J-1}(r_d)|^2]\,. \tag{41}$$

For D_{3h} symmetry molecular photodissociation, $T_{10}(r)\subset A''$, $T_{1\pm1}(r)\subset E'$[15]. Therefore for one-photon electric-dipole excitation the angular dependences of the cross section for the observation of photofragments from a rotating parent molecule of D_{3h} point group symmetry are given by

$\sigma_1^{(1)}(A_1'\to E')$

$$= \frac{1}{36\sqrt{\pi}}(2J+1)[K_1^{(1)}]^2\sigma_0[1+\beta P_2(\cos\Theta)]\left\{\left[Y_{00}(\vartheta_i\varphi_i) - \frac{1}{\sqrt{5}}\,Y_{20}(\vartheta_i\varphi_i)\right][|\langle T_{11}^{\rm mol}(r)\rangle|^2 + |\langle T_{1\bar{1}}^{\rm mol}(r)\rangle|^2]\right.$$
$$\left. - \sqrt{\frac{6}{5}}\,[\langle T_{11}^{\rm mol}(r)\rangle^*\langle T_{1\bar{1}}^{\rm mol}(r)\rangle Y_{2\bar{2}}(\vartheta_i\varphi_i) + \text{c.c.}]\right\} \tag{42}$$

and

$$\sigma_1^{(1)}(A_1'\to A_2'') = \frac{1}{36\sqrt{\pi}}(2J+1)[K_1^{(1)}]^2\sigma_0[1+\beta P_2(\cos\Theta)]\left[Y_{00}(\vartheta_i\varphi_i) + \frac{2}{\sqrt{5}}\,Y_{20}(\vartheta_i\varphi_i)\right]\cdot|\langle T_{10}^{\rm mol}(r)\rangle|^2. \tag{43}$$

Other allowed dipole transitions can be obtained from Eqs. (38)–(41).

B. For one-photon quadrupole transition (left-circular polarized light)

The transition moments for X polarization (or Z polarization) with Z propagation (or X propagation) is given by[15]

$$M_2^{(1)} = K_2^{(1)}\langle(xz+zx)\rangle = K_2^{(1)}\{\langle T_{21}^{\rm lab}(r)\rangle - \langle T_{2\bar{1}}^{\rm lab}(r)\rangle\} = K_2^{(1)}\sum_{q'}\,[D_{1q'}^{2*}(\Omega_\mu) - D_{1\bar{q}'}^{2*}(\Omega_\mu)]\langle T_{2q'}^{\rm mol}(r)\rangle. \tag{44}$$

To simplify the derivation, we consider only left circularly polarized light, which is a normalized combination of linearly polarized light:

$$M_2^{(1)} = K_2^{(1)}\sum_{q'} D_{1q'}^{2*}(\Omega_\mu)\langle T_{2q'}^{\rm mol}(r)\rangle. \tag{45}$$

Substituting Eqs. (45) and (35) into Eq. (31), we obtain

$$\sigma_2^{(1)}(\Gamma_\gamma) = \sqrt{\pi}\,(2J+1)[K_2^{(1)}]^2\sum_M\left|\sum_{J'}R_{KJ'}(r_d)(2J'+1)^{1/2}\begin{pmatrix}J' & J & 2\\ \bar{M}+1 & M & \bar{1}\end{pmatrix}\begin{pmatrix}J' & J & 2\\ 0 & 0 & 0\end{pmatrix}\cdot Y_{J'M-1}(\Theta,\Phi)\right|^2$$
$$\times\sum_{q,q'}\sum_L(-1)^{q'}(2L+1)^{1/2}\begin{pmatrix}2 & 2 & L\\ \bar{q} & q' & q-q'\end{pmatrix}\begin{pmatrix}2 & 2 & L\\ 0 & 0 & 0\end{pmatrix}\langle T_{2q}^{\rm mol}(r)\rangle^*\langle T_{2q'}^{\rm mol}(r)\rangle Y_{L,\bar{q}+q'}(\vartheta_i\varphi_i)\,, \tag{46}$$

where the closed forms of the angular part are exactly the same as in Eq. (17) except that we have changed H'^{fi} into $R_{KJ'}(r_d)$. Equation (46) can be calculated further. The total cross section of one-photon quadrupole transition with left circularly polarized light has the form

$$\sigma_2^{(1)}(\Gamma_\gamma)=\sqrt{\pi}\,(2J+1)[K_2^{(1)}]^2\sum_M\left|\sum_{J'}R_{KJ'}(r_d)(2J'+1)^{1/2}\begin{pmatrix}J' & J & 2\\ \bar{M}+1 & M & \bar{1}\end{pmatrix}\begin{pmatrix}J' & J & 2\\ 0 & 0 & 0\end{pmatrix}Y_{J'M-1}(\Theta,\Phi)\right|^2$$
$$\times\left\{\frac{1}{5}\sum_{q=-2}^{2}|\langle T_{2q}^{\rm mol}(r)\rangle|^2Y_{00}(\vartheta_i\varphi_i)+\frac{1}{35}[2|\langle T_{20}^{\rm mol}(r)\rangle|^2+|\langle T_{2\bar{1}}^{\rm mol}(r)\rangle|^2\right.$$
$$+|\langle T_{21}^{\rm mol}(r)\rangle|^2-2|\langle T_{22}^{\rm mol}(r)\rangle|^2-2|\langle T_{2\bar{2}}^{\rm mol}(r)\rangle|^2]Y_{20}(\vartheta_i\varphi_i)$$
$$-\frac{1}{35}[2(\langle T_{2\bar{2}}^{\rm mol}(r)\rangle\langle T_{20}^{\rm mol}(r)\rangle^*+\langle T_{20}^{\rm mol}(r)\rangle\langle T_{20}^{(\rm mol)}(r)\rangle+\sqrt{6}\langle T_{21}^{\rm mol}(r)\rangle^*\langle T_{2\bar{1}}^{\rm mol}(r)\rangle]Y_{2\bar{2}}(\vartheta_i\varphi_i)$$
$$-\frac{1}{35}[2(\langle T_{2\bar{2}}^{\rm mol}(r)\rangle^*\langle T_{20}^{\rm mol}(r)\rangle+\langle T_{20}^{\rm mol}(r)\rangle^*\langle T_{22}^{\rm mol}(r)\rangle)+\sqrt{6}\langle T_{2\bar{1}}^{\rm mol}(r)\rangle^*\langle T_{21}^{\rm mol}(r)\rangle]Y_{22}(\vartheta_i\varphi_i)$$
$$+\frac{1}{7}\sqrt{\frac{2}{5}}\left[\sqrt{\frac{1}{6}}\langle T_{20}^{\rm mol}(r)\rangle^*\langle T_{2\bar{2}}^{\rm mol}(r)\rangle-\frac{2}{3}\langle T_{21}^{\rm mol}(r)\rangle^*\langle T_{2\bar{1}}^{\rm mol}(r)\rangle+\frac{1}{\sqrt{6}}\langle T_{22}^{\rm mol}(r)\rangle^*\langle T_{20}^{\rm mol}(r)\rangle\right]Y_{4\bar{2}}(\vartheta_i\varphi_i)$$
$$+\frac{1}{7}\sqrt{\frac{2}{5}}\left[\sqrt{\frac{1}{6}}\langle T_{2\bar{2}}^{\rm mol}(r)\rangle^*\langle T_{20}^{\rm mol}(r)\rangle-\frac{2}{3}\langle T_{2\bar{1}}^{\rm mol}(r)\rangle\langle T_{21}^{\rm mol}(r)\rangle+\frac{1}{\sqrt{6}}\langle T_{20}^{\rm mol}(r)\rangle^*\langle T_{22}^{\rm mol}(r)\rangle\right]Y_{42}(\vartheta_i\varphi_i)$$
$$\left.+\frac{1}{3}\sqrt{\frac{2}{35}}[\langle T_{22}^{\rm mol}(r)\rangle^*\langle T_{2\bar{2}}^{\rm mol}(r)\rangle Y_{4\bar{4}}(\vartheta_i\varphi_i)+\text{c.c.}]\right\}. \tag{47}$$

For special cases, we consider the example of molecules of the D_{3h} point group symmetry under one-photon quadrupole transition and fragmentation. The transition operators in this group have symmetry[14] as follows:

$$T_{20}(r)\subset A_1',\quad T_{2\pm1}(r)\subset E'',\quad T_{2\pm2}(r)\subset E'. \tag{48}$$

Let us assume an initial molecular state $|i\rangle\subset A_1'$, E', or E''. The final state could belong to A_1', E', or E''. Therefore the total angular dependence of the cross section for the observation of photofragments from a rotating parent molecule of the D_{3h} point group symmetry for one-photon electric-quadrupole excitation is given by

$$\sigma_2^{(1)}(E''\to E'')=\sigma_2^{(1)}(E'\to E')$$
$$=\sqrt{\pi}(2J+1)[K_2^{(1)}]^2\sum_M\left|\sum_{J'}R_{KJ'}(r_d)(2J'+1)^{1/2}\begin{pmatrix}J' & J & 2\\ \bar{M}+1 & M & \bar{1}\end{pmatrix}\begin{pmatrix}J' & J & 2\\ 0 & 0 & 0\end{pmatrix}Y_{J'M-1}(\Theta,\Phi)\right|^2$$
$$\times\left\{\tfrac{1}{5}[Y_{00}(\vartheta_i\varphi_i)+\tfrac{2}{7}Y_{20}(\vartheta_i\varphi_i)+\tfrac{2}{7}Y_{40}(\vartheta_i\varphi_i)]|\langle T_{20}^{\rm mol}(r)\rangle|^2\right.$$
$$+\tfrac{1}{5}[Y_{00}(\vartheta_i\varphi_i)-\tfrac{2}{7}Y_{20}(\vartheta_i\varphi_i)+\tfrac{1}{21}Y_{40}(\vartheta_i\varphi_i)][|\langle T_{22}^{\rm mol}(r)\rangle|^2+|\langle T_{2\bar{2}}^{\rm mol}(r)\rangle|^2]$$
$$-\tfrac{2}{35}[(|\langle T_{2\bar{2}}^{\rm mol}(r)\rangle\langle T_{20}^{\rm mol}(r)\rangle^*+\langle T_{2\bar{2}}^{\rm mol}(r)\rangle^*\langle T_{20}^{\rm mol}(r)\rangle)Y_{2\bar{2}}(\vartheta_i\varphi_i)+\text{c.c.}]$$
$$+\frac{1}{7\sqrt{15}}[(\langle T_{20}^{\rm mol}(r)\rangle^*\langle T_{2\bar{2}}^{\rm mol}(r)\rangle+\langle T_{22}^{\rm mol}(r)\rangle^*\langle T_{20}^{\rm mol}(r)\rangle)Y_{4\bar{2}}(\vartheta_i\varphi_i)+\text{c.c.}]$$
$$\left.+\frac{1}{3}\sqrt{\frac{2}{35}}[\langle T_{22}^{\rm mol}(r)\rangle^*\langle T_{2\bar{2}}^{\rm mol}(r)\rangle Y_{4\bar{4}}(\vartheta_i\varphi_i)+\text{c.c.}]\right\}, \tag{49}$$

$$\sigma_2^{(1)}(A_1'\to E'')=\sqrt{\pi}(2J+1)[K_2^{(1)}]^2\sum_M\left|\sum_{J'}R_{KJ'}(r_d)(2J'+1)^{1/2}\begin{pmatrix}J' & J & 2\\ \bar{M}+1 & M & \bar{1}\end{pmatrix}\begin{pmatrix}J' & J & 2\\ 0 & 0 & 0\end{pmatrix}Y_{J'M-1}(\Theta,\Phi)\right|^2$$
$$\times\left\{\tfrac{1}{5}[Y_{0,0}(\vartheta_i\varphi_i)+\tfrac{1}{7}Y_{20}(\vartheta_i\varphi_i)-\tfrac{4}{21}Y_{40}(\vartheta_i\varphi_i)][|\langle T_{21}^{\rm mol}(r)\rangle|^2+|\langle T_{2\bar{1}}^{\rm mol}(r)\rangle|^2]\right.$$
$$-\tfrac{\sqrt{6}}{35}[\langle T_{21}^{\rm mol}(r)\rangle^*\langle T_{2\bar{1}}^{\rm mol}(r)\rangle Y_{2\bar{2}}(\vartheta_i\varphi_i)+\langle T_{2\bar{1}}^{\rm mol}(r)\rangle^*\langle T_{21}^{\rm mol}(r)\rangle Y_{22}(\vartheta_i,\varphi_i)]$$
$$\left.-\frac{2}{21}\sqrt{\frac{2}{5}}[\langle T_{21}^{\rm mol}(r)\rangle^*\langle T_{2\bar{1}}^{\rm mol}(r)\rangle Y_{4\bar{2}}(\vartheta_i\varphi_i)+\langle T_{2\bar{1}}^{\rm mol}(r)\rangle^*\langle T_{21}^{\rm mol}(r)\rangle Y_{42}(\vartheta_i\varphi_i)]\right\}, \tag{50}$$

$$\sigma_2^{(1)}(A_1'\to E')=\sqrt{\pi}(2J+1)[K_2^{(1)}]^2\sum_M\left|\sum_{J'}R_{KJ'}(r_d)(2J'+1)^{1/2}\begin{pmatrix}J' & J & 2\\ \bar{M}+1 & M & \bar{1}\end{pmatrix}\begin{pmatrix}J' & J & 2\\ 0 & 0 & 0\end{pmatrix}Y_{J'M-1}(\Theta,\Phi)\right|^2$$
$$\times\left\{\tfrac{1}{5}[Y_{00}(\vartheta_i\varphi_i)-\tfrac{2}{7}Y_{20}(\vartheta_i\varphi_i)+\tfrac{1}{21}Y_{40}(\vartheta_i\varphi_i)][|\langle T_{22}^{\rm mol}(r)\rangle|^2+|\langle T_{2\bar{2}}^{\rm mol}(r)\rangle|^2]\right.$$
$$\left.+\frac{1}{3}\sqrt{\frac{2}{35}}[\langle T_{22}^{\rm mol}(r)\rangle^*\langle T_{2\bar{2}}^{\rm mol}(r)\rangle Y_{4\bar{4}}(\vartheta_i\varphi_i)+\langle T_{2\bar{2}}^{\rm mol}(r)\rangle^*\langle T_{22}^{\rm mol}(r)\rangle Y_{44}(\vartheta_i\varphi_i)]\right\}, \tag{51}$$

$$\sigma_2^{(1)}(A_1'\to A_1') = \sqrt{\pi}(2J+1)[K_2^{(1)}]^2 \sum_M \left| \sum_{J'} R_{KJ'}(r_d)(2J'+1)^{1/2} \begin{pmatrix} J' & J & 2 \\ \bar{M}+1 & M & \bar{1} \end{pmatrix} \begin{pmatrix} J' & J & 2 \\ 0 & 0 & 0 \end{pmatrix} Y_{J'M-1}(\Theta\Phi) \right|^2$$

$$\times \left\{ \tfrac{1}{5}[Y_{00}(\vartheta_i\varphi_i) + \tfrac{2}{7}Y_{20}(\vartheta_i\varphi_i) + \tfrac{2}{7}Y_{40}(\vartheta_i\varphi_i)] |\langle T_{20}^{\mathrm{mol}}(r)\rangle|^2, \right. \tag{52}$$

$$\sigma_2^{(1)}(E'\to E'') = \sqrt{\pi}(2J+1)[K_2^{(1)}]^2 \sum_M \left| \sum_{J'} R_{KJ}(r_d)(2J'+1)^{1/2} \begin{pmatrix} J' & J & 2 \\ \bar{M}+1 & M & \bar{1} \end{pmatrix} \begin{pmatrix} J' & J & 2 \\ 0 & 0 & 0 \end{pmatrix} Y_{J'M-1}(\Theta\Phi) \right|^2$$

$$\times \left\{ \tfrac{1}{5}[Y_{00}(\vartheta_i\varphi_i) + \tfrac{1}{7}Y_{20}(\vartheta_i\varphi_i) - \tfrac{4}{21}Y_{40}(\vartheta_i\varphi_i)] [|\langle T_{21}^{\mathrm{mol}}(r)\rangle|^2 + |\langle T_{2\bar{1}}^{\mathrm{mol}}(r)\rangle|^2] \right.$$

$$- \frac{\sqrt{6}}{35} [\langle T_{21}^{\mathrm{mol}}(r)\rangle^* \langle T_{2\bar{1}}^{\mathrm{mol}}(r)\rangle Y_{2\bar{2}}(\vartheta_i\varphi_i) + \langle T_{2\bar{1}}^{\mathrm{mol}}(r)\rangle^* \langle T_{21}^{\mathrm{mol}}(r)\rangle Y_{22}(\vartheta_i\varphi_i)]$$

$$\left. - \frac{2}{21}\sqrt{\frac{2}{5}} [\langle T_{21}^{\mathrm{mol}}(r)\rangle^* \langle T_{2\bar{1}}^{\mathrm{mol}}(r)\rangle Y_{4\bar{2}}(\vartheta_i\varphi_i) + \langle T_{2\bar{1}}^{\mathrm{mol}}(r)\rangle^* \langle T_{21}^{\mathrm{mol}}(r)\rangle Y_{42}(\vartheta_i\varphi_i)] \right\}. \tag{53}$$

The other allowed one-photon electric-quadrupole transitions for other molecules with different symmetries can be obtained from Eq. (47) similarly.

C. Two-photon electric-dipole transition

For two photons with parallel polarization, the transition moment in the lab-fixed frame is[15]

$$M_1^{(2)} = K_1^{(2)}\langle R_{10}R_{10}\rangle = K_1^{(2)}\left\{ -(1/\sqrt{3})\langle T_{00}^{\mathrm{lab}}(rr)\rangle + \sqrt{\frac{2}{3}}\langle T_{20}^{\mathrm{lab}}(rr)\rangle \right\}. \tag{54}$$

In the molecular-fixed frame the transition moment can be written in the form

$$M_1^{(2)} = K_1^{(2)}\left\{ -\frac{1}{\sqrt{3}}\sum_{q'} D_{0q'}^{0*}(\Omega_\mu)\langle T_{0q'}^{\mathrm{mol}}(rr)\rangle + \sqrt{\frac{2}{3}}\sum_{q'} D_{0q'}^{2*}(\Omega_\mu)\langle T_{2q'}^{\mathrm{mol}}(rr)\rangle \right\}. \tag{55}$$

Substituting Eqs. (55) and (35) into Eq. (31), we obtain the angular distribution of a two-photon nonresonant process in the form

$$\sigma_1^{(2)}(\Gamma_\gamma) = 2\pi[K_1^{(2)}]^2 \sum_M \left| -\frac{1}{\sqrt{3}}\sum_{J'} R_{KJ'}(r_d) \begin{pmatrix} J' & J & 0 \\ \bar{M} & M & 0 \end{pmatrix} \begin{pmatrix} J' & J & 0 \\ 0 & 0 & 0 \end{pmatrix} Y_{J'M}(\Theta,\Phi) \right.$$

$$\times \langle T_{00}^{\mathrm{mol}}(rr)\rangle Y_{00}(\vartheta_i\varphi_i) + \sqrt{\frac{2}{15}}\sum_{J'} R_{KJ'}(r_d) Y_{J'M}(\Theta,\Phi)$$

$$\left. \times \begin{pmatrix} J' & J & 2 \\ \bar{M} & M & 0 \end{pmatrix} \begin{pmatrix} J' & J & 2 \\ 0 & 0 & 0 \end{pmatrix} \sum_q \langle T_{2q}^{\mathrm{mol}}(rr)\rangle Y_{2q}(\vartheta_i\varphi_i) \right|^2. \tag{56}$$

This equation can be reduced to the general form

$$\sigma_1^{(2)}(\Gamma_\gamma) = \sigma_0\{1 + \beta\cos^2\Theta + \gamma\cos^4\Theta\}, \tag{57}$$

where

$$\sigma_0 = \sigma_0' a_0, \tag{58}$$

$$\sigma_0' = \frac{[K_1^{(2)}]^2}{6(2J+1)} |R_{vJ}(r_d)|^2 |\langle T_{00}^{\mathrm{mol}}(rr)\rangle|^2 \cdot |Y_{00}(\vartheta_i\varphi_i)|^2, \tag{59}$$

$$a_0 = 1 + [(11J^2+11J-8)a_1 + (3J^2+5J+4)a_2 + (3J^2+J+2)a_3]A$$
$$+ a_4B + a_5C + a_6D - (J-5)a_7E - (J+6)a_8F + 3a_9G, \tag{60}$$

$$a_1 = \frac{J(J+1)|R_{K,J}(r_d)|^2}{8\sqrt{\pi}(2J+3)^2(2J-1)^2\cdot b_0} = \frac{J(J+1)}{8\sqrt{\pi}b_0} \begin{Bmatrix} 0 & J+1 & J-1 \\ 0 & J-1 & J+1 \end{Bmatrix}^4 |R_{K,J}(r_d)|^2, \tag{61}$$

$$b_0 = |R_{KJ}(r_d)|^2 |\langle T_{00}(rr)\rangle|^2 |Y_{00}(\vartheta_i\varphi_i)|^2, \tag{62}$$

$$a_2 = \frac{9(J+1)(J+2)|R_{K,J+2}(r_d)|^2}{32\sqrt{\pi}(2J+5)(2J+3)^2(2J+1)b_0} = \frac{9(J+1)(J+2)}{32\sqrt{\pi}b_0(2J+3)^2} \cdot \begin{Bmatrix} 0 & J+2 & J \\ 0 & J & J+2 \end{Bmatrix}^2 |R_{K,J+2}(r_d)|^2, \tag{63}$$

$$a_3 = \frac{9J(J-1)|R_{K,J-2}(r_d)|^2}{32\sqrt{\pi}(2J+1)(2J-1)^2(2J-3)\cdot b_0} = \frac{9J(J-1)}{32\sqrt{\pi}b_0(2J-1)^2} \begin{Bmatrix} 0 & J & J-2 \\ 0 & J-2 & J \end{Bmatrix}^2 |R_{K,J-2}(r_d)|^2, \tag{64}$$

$$a_4=\frac{J(J+1)|R_{K,J}(r_d)|^2}{2\sqrt{2\pi}(2J+3)(2J-1)\cdot b_0}=\frac{J(J+1)}{2\sqrt{2\pi}b_0}\begin{Bmatrix}0 & J+1 & J-1\\ 0 & J-1 & J+1\end{Bmatrix}^2|R_{K,J}(r_d)|^2, \tag{65}$$

$$a_5=\frac{3(J+1)(J+2)}{4\sqrt{2\pi}(2J+3)(2J+5)^{1/2}(2J+1)^{1/2}\cdot b_0}=\frac{3(J+1)(J+2)}{4\sqrt{2\pi}(2J+3)\cdot b_0}\begin{Bmatrix}0 & J+2 & J\\ 0 & J & J+2\end{Bmatrix}, \tag{66}$$

$$a_6=\frac{3J(J-1)}{4\sqrt{2\pi}(2J-1)(2J+1)^{1/2}(2J-3)^{1/2}\cdot b_0}=\frac{3J(J-1)}{4\sqrt{2\pi}(2J-1)\cdot b_0}\begin{Bmatrix}0 & J & J-2\\ 0 & J-2 & J\end{Bmatrix}, \tag{67}$$

$$a_7=\frac{3J(J+1)(J+2)}{16\sqrt{\pi}(2J+3)^2(2J+1)^{1/2}(2J-1)(2J+5)^{1/2}\cdot b_0}=\frac{3J(J+1)(J+2)}{16\sqrt{\pi}(2J+3)^2(2J-1)\cdot b_0}\cdot\begin{Bmatrix}0 & J+2 & J\\ 0 & J & J+2\end{Bmatrix}, \tag{68}$$

$$a_8=\frac{3J(J-1)(J+1)}{16\sqrt{\pi}(2J+3)(2J+1)^{1/2}(2J-3)^{1/2}(2J-1)^2\cdot b_0}=\frac{3J(J-1)(J+1)}{16\sqrt{\pi}(2J+3)(2J-1)^2\cdot b_0}\cdot\begin{Bmatrix}0 & J & J-2\\ 0 & J-2 & J\end{Bmatrix}, \tag{69}$$

$$a_9=\frac{9\cdot J(J+1)(J-1)(J+2)}{32\sqrt{\pi}(2J+5)^{1/2}(2J+3)(2J+1)(2J-1)(2J-3)^{1/2}\cdot b_0}$$

$$=\frac{9J(J+1)(J-1)(J+2)}{32\sqrt{\pi}\cdot b_0\cdot(2J+3)(2J+1)(2J-1)}\begin{Bmatrix}0 & J+2 & J-2\\ 0 & J-2 & J+2\end{Bmatrix}, \tag{70}$$

$$A=\sum_{q,q'}\left[\frac{1}{\sqrt{5}}\begin{pmatrix}2 & 2 & 0\\ \bar{q} & q' & 0\end{pmatrix}Y_{00}(\vartheta_i\varphi_i)-\sqrt{\frac{2}{7}}\begin{pmatrix}2 & 2 & 2\\ \bar{q} & q' & q-q'\end{pmatrix}Y_{2q'-q}(\vartheta_i\varphi_i)\right.$$

$$\left.-3\sqrt{\frac{2}{35}}\begin{pmatrix}2 & 2 & 4\\ \bar{q} & q' & q-q'\end{pmatrix}Y_{4q'-q}(\vartheta_i\varphi_i)\right]\langle T_{2q}(rr)\rangle^*\langle T_{2q'}(rr)\rangle, \tag{71}$$

$$B=\frac{1}{\sqrt{5}}\sum_q(-1)^q[\langle T_{00}(rr)\rangle^*\langle T_{2q}(rr)\rangle+\langle T_{00}(rr)\rangle\langle T_{2\bar{q}}(rr)\rangle^*]Y_{2q}(\vartheta_i\varphi_i), \tag{72}$$

$$C=\frac{1}{\sqrt{5}}\sum_q(-1)^q[R^*_{K,J}(r_d)R_{K,J+2}(r_d)\langle T_{00}(rr)\rangle^*\langle T_{2q}(rr)\rangle+R_{KJ}(r_d)R^*_{K,J+2}(r_d)\langle T_{00}(rr)\rangle\langle T_{2\bar{q}}(rr)\rangle^*]Y_{2q}(\vartheta_i\varphi_i), \tag{73}$$

$$D=\frac{1}{\sqrt{5}}\sum_q(-1)^q[R^*_{K,J}(r_d)R_{K,J-2}(r_d)\langle T_{00}(rr)\rangle^*\langle T_{2q}(rr)\rangle+R_{K,J}(r_d)R^*_{K,J-2}(r_d)\langle T_{00}(rr)\rangle\langle T_{2\bar{q}}(rr)\rangle^*]Y_{2q}(\vartheta_i\varphi_i), \tag{74}$$

$$E=\sum_{q,q'}[\langle T_{2\bar{q}}(rr)\rangle^*\langle T_{2\bar{q}'}(rr)\rangle R^*_{K,J}(r_d)R_{K,J+2}(r_d)+\langle T_{2q}(rr)\rangle\langle T_{2q'}(rr)\rangle^*R_{K,J}(r_d)R^*_{K,J+2}(r_d)]$$

$$\times\left[\frac{1}{\sqrt{5}}\begin{pmatrix}2 & 2 & 0\\ q & \bar{q}' & 0\end{pmatrix}Y_{00}(\vartheta_i\varphi_i)-\sqrt{\frac{2}{7}}\begin{pmatrix}2 & 2 & 2\\ q & \bar{q}' & q'-q\end{pmatrix}Y_{2q-q'}(\vartheta_i\varphi_i)+3\sqrt{\frac{2}{35}}\begin{pmatrix}2 & 2 & 4\\ q & \bar{q}' & q'-q\end{pmatrix}Y_{4q-q'}(\vartheta_i\varphi_i)\right], \tag{75}$$

$$F=\sum_{q,q'}[\langle T_{2\bar{q}}(rr)\rangle^*\langle T_{2\bar{q}'}(rr)\rangle R^*_{K,J}(r_d)R_{K,J-2}(r_d)+\langle T_{2q}(rr)\rangle\langle T_{2q'}(rr)\rangle^*R_{K,J}(r_d)R^*_{K,J-2}(r_d)]$$

$$\times\left[\frac{1}{\sqrt{5}}\begin{pmatrix}2 & 2 & 0\\ q & \bar{q}' & 0\end{pmatrix}Y_{00}(\vartheta_i\varphi_i)-\sqrt{\frac{2}{7}}\begin{pmatrix}2 & 2 & 2\\ q & \bar{q}' & q'-q\end{pmatrix}Y_{2q-q'}(\vartheta_i\varphi_i)\right.$$

$$\left.+3\sqrt{\frac{2}{35}}\begin{pmatrix}2 & 2 & 4\\ q & \bar{q}' & q'-q\end{pmatrix}Y_{4q-q'}(\vartheta_i\varphi_i)\right], \tag{76}$$

$$G=\sum_{q,q'}[\langle T_{2\bar{q}}(rr)\rangle^*\langle T_{2\bar{q}'}(rr)\rangle R^*_{K,J+2}(r_d)R_{K,J-2}(r_d)+\langle T_{2q}(rr)\rangle\langle T_{2q'}(rr)\rangle^*R_{K,J+2}(r_d)R^*_{K,J-2}(r_d)]$$

$$\times\left[\frac{1}{\sqrt{5}}\begin{pmatrix}2 & 2 & 0\\ q & \bar{q}' & 0\end{pmatrix}Y_{00}(\vartheta_i\varphi_i)-\sqrt{\frac{2}{7}}\begin{pmatrix}2 & 2 & 2\\ q & \bar{q}' & q'-q\end{pmatrix}Y_{2q-q'}(\vartheta_i\varphi_i)\right.$$

$$\left.+3\sqrt{\frac{2}{35}}\begin{pmatrix}2 & 2 & 4\\ q & \bar{q}' & q'-q\end{pmatrix}Y_{4q-q'}(\vartheta_i\varphi_i)\right], \tag{77}$$

$$\beta=(1/a_0)[\{-6(5J^2+5J-12)a_1+2(J+3)(J-4)a_2+2(J-2)(J+5)a_3\}A-3a_4B$$

$$-3a_5C-3a_6D-6(J+7)a_7E-6(J-6)a_8F-30a_9G], \tag{78}$$

$$\gamma = (1/a_0)[\{27(J-1)(J+2)a_1 + 3(J+3)(J+4)a_2 + 3(J-2)(J-3)a_3\}A + 15(J+3)a_7E + 15(J-2)a_8F + 35a_9G]. \tag{79}$$

Here, we have omitted the superscript "mol" for each $T_{ij}(rr)$ and used some of the formulas of Appendix B and the formulas of $3j$ symbols.[20] The applications to some specific point group symmetry of this formula will be found in another paper.[21]

APPENDIX A: THE SUMMATIONS OF FINITE ROTATION MATRIX ELEMENTS

With the help of Eqs. (21), (23), (31), and (32) of Ref. 16, we obtain the following summations:

$$\sum_m m|d^j_{m',m+1}(\beta)|^2 = m'\cos\beta - 1, \tag{A1}$$

$$\sum_m m|d^j_{m',m-1}(\beta)|^2 = m'\cos\beta + 1, \tag{A2}$$

$$\sum_m m^2|d^j_{m',m+1}(\beta)|^2 = [(j/2)(j+1)\sin^2\beta + 1] + m'[(m'/2)(3\cos^2\beta - 1) - 2\cos\beta], \tag{A3}$$

$$\sum_m m^2|d^j_{m',m-1}(\beta)|^2 = [(j/2)(j+1)\sin^2\beta + 1] + m'[(m'/2)(3\cos^2\beta - 1) + 2\cos\beta], \tag{A4}$$

$$\sum_m m d^j_{m'-2,m}(\beta) d^j_{m'+2,m}(\beta) = 0, \tag{A5}$$

$$\sum_m m^3|d^j_{m',m+1}(\beta)|^2 = [-\tfrac{3}{2}j(j+1)\sin^2\beta - 1] + m'\{(3+m'^2)\cos\beta - \tfrac{3}{2}m'(3\cos^2\beta - 1) + \tfrac{1}{2}[3j(j+1) - 5m'^2 - 1]\sin^2\beta\cos\beta\}, \tag{A6}$$

$$\begin{aligned}\sum_m m^4|d^j_{m',m+1}(\beta)|^2 = {} & [\tfrac{3}{8}j^2(j+1)^2\sin^4\beta + \tfrac{1}{4}j(j+1)\sin^2\beta(3\cos^2\beta - 1) + 3j(j+1)\sin^2\beta + 1] \\ & + (m'^2/8)\sin^2\beta(5\cos^2\beta - 1)(6j^2 + 6j - 5) + (m'^4/8)(35\cos^4\beta - 30\cos^2\beta + 3) - m' \\ & \times[4(m'^2+1)\cos\beta - 3m'(3\cos^2\beta - 1) + 2(3j^2 + 3j - 5m'^2 - 1)\sin^2\beta\cos\beta].\end{aligned} \tag{A7}$$

We know

$$\begin{aligned}\cos^2\beta\, d^j_{\mu m}(\beta) = {} & \frac{d^{j-2}_{\mu m}(\beta)}{j(j-1)(2j-1)(2j+1)}\{(j^2-m^2)(j^2-\mu^2)[(j-1)^2-m^2][(j-1)^2-\mu^2]\}^{1/2} \\ & + \frac{d^{j-1}_{\mu m}(\beta)}{j(2j+1)(j^2-1)}2\mu m[(j^2-m^2)(j^2-\mu^2)]^{1/2} \\ & + \frac{d^j_{\mu m}(\beta)}{(2j-1)(2j+3)}\left[\frac{6\mu^2m^2}{j(j+1)} + (2j^2+2j-1) - 2(m^2+\mu^2)\right] \\ & + \frac{d^{j+1}_{\mu m}(\beta)}{j(j+1)(j+2)(2j+1)}2\mu m\{[(j+1)^2-m^2][(j+1)^2-\mu^2]\}^{1/2} \\ & + \frac{d^{j+2}_{\mu m}(\beta)}{(j+1)(j+2)(2j+1)(2j+3)}\{[(j+1)^2-\mu^2][(j+1)^2-m^2] \\ & \times[(j+2)^2-\mu^2][(j+2)^2-m^2]\}^{1/2},\end{aligned} \tag{A8a}$$

$$\begin{aligned}\sin\beta\cos\beta\, d^j_{\mu m}(\beta) = {} & \pm\frac{d^{j-2}_{\mu m}(\beta)}{j(j-1)(2j+1)(2j-1)}[(j^2-m^2)(j^2-\mu^2)(j+\mu-1)(j-\mu-1)(j\mp m-1)(j\mp m-2)]^{1/2} \\ & + \frac{d^{j-1}_{\mu\pm1}(\beta)\cdot\mu}{j(j^2-1)(2j+1)}[(j\mp m)(j\mp m-1)(j^2-\mu^2)]^{1/2}(j\pm 2m+1) \\ & + \frac{d^j_{\mu\pm1}(\beta)}{(2j-1)(2j+3)}\frac{3\mu^2 - j(j+1)}{j(j+1)}(2m\pm1)[(j\mp m)(j\pm m+1)]^{1/2} \\ & + \frac{d^{j+1}_{\mu\pm1}(\beta)\cdot\mu}{j(j+1)(2j+1)(j+2)}[(j-\mu+1)(j+\mu+1)(j\pm m+1)(j\pm m+2)]^{1/2}(j\mp 2m)\end{aligned}$$

$$\mp\frac{d^{j+2}_{\mu m\pm 1}(\beta)}{(j+1)(j+2)(2j+1)(2j+3)}\{[(j+1)^2-\mu^2][(j+2)^2-\mu^2]$$
$$\times[(j+1)^2-m^2](j\pm m+2)(j\pm m+3)\}^{1/2}. \tag{A8b}$$

Using Eqs. (A8a) and (A8b) and letting $\mu=0$, with the help of the above equations we obtain

$$\sum_m(2m-1)(j+m)[(j+m-1)(j+m-2)(j-m)(j-m+1)]^{1/2}d^{j-2}_{0m-1}(\beta)d^{j}_{0m-1}(\beta)$$
$$=\frac{1}{4}j(j+1)(j-1)[-5(j-2)\sin^4\beta+2(2j-7)\sin^2\beta+4], \tag{A9}$$

$$\sum_m(2m-1)(j-m+1)[(j+m+1)(j+m)(j-m+3)(j-m+2)]^{1/2}d^{j+2}_{0m-1}(\beta)d^{j}_{0m-1}(\beta)$$
$$=\frac{1}{4}j(j+1)(j+2)[5(j+3)\sin^4\beta-2(2j+9)\sin^2\beta+4], \tag{A10}$$

$$\sum_m m[(j+m)(j-m+1)]^{1/2}d^{j}_{0m}(\beta)d^{j}_{0m-1}(\beta)=\frac{1}{2}j(j+1)\sin\beta\cos\beta, \tag{A11}$$

$$\sum_m[(j+m)(j-m+1)]^{1/2}d^{j}_{0m}(\beta)d^{j}_{0m-1}(\beta)=0, \tag{A12}$$

$$\sum_m(1-2m)^2(j+m)(j-m+1)|d^{j}_{0m-1}(\beta)|^2=\frac{1}{2}j(j+1)[-3(j-1)(j+2)\sin^4\beta+(4j^2+4j-9)\sin^2\beta+2], \tag{A13}$$

$$\sum_m(j^2-m^2)(j+m-1)(j+m-2)|d^{j-2}_{0m-1}(\beta)|^2$$
$$=\frac{1}{8}j(j-1)[-3(j-2)(j-3)\sin^4\beta-12(j-2)\sin^2\beta+8(j-1)(j+1)], \tag{A14}$$

$$\sum_m(j-m+2)(j-m+5)(j-m+1)(j+m+1)|d^{j+2}_{0m-1}(\beta)|^2$$
$$=\frac{1}{8}(j+1)(j+2)[-3(j+3)(j+4)\sin^4\beta+12(j+3)\sin^2\beta+8j(j+2)], \tag{A15}$$

$$\sum_m[(j^2-m^2)(j+m-1)(j+m-2)(j-m+2)(j-m+3)(j+m+1)(j-m+1)]^{1/2}d^{j-2}_{0m-1}(\beta)d^{j+2}_{0m-1}(\beta)$$
$$=\frac{j}{8}(j+1)(j-1)(j+2)[35\sin^4\beta-40\sin^2\beta+8]$$
$$=j(j+1)(j-1)(j+2)P_4(\cos\beta). \tag{A16}$$

From Eq. (24a) of Ref. 16, by changing j to $j-1$, we obtain

$$\sum_m[(j+m)(j+m-1)(j-m)(j-m-1)]^{1/2}d^{j}_{m'm}(\beta)d^{j-2}_{m'm}(\beta)$$
$$=\frac{[j(j-1)-m'(m'-1)][j(j-1)-m'(m'+1)]}{2[(j+m')(j+m'-1)(j-m')(j-m'-1)]^{1/2}}(3\cos^2\beta-1). \tag{A17}$$

Changing j to $j+2$ for the above equation, we obtain

$$\sum_m[(j+m+2)(j+m+1)(j-m+2)(j-m+1)]^{1/2}d^{j+2}_{m'm}(\beta)d^{j}_{m'm}(\beta)$$
$$=\frac{[(j+1)(j+2)-m'(m'-1)][(j+1)(j+2)-m'(m'+1)]}{2[(j+m'+2)(j+m'+1)(j-m'+2)(j-m'+1)]^{1/2}}(3\cos^2\beta-1). \tag{A18}$$

Since[20]

$$D^{j}_{m0}(\alpha\beta 0)=(-1)^m(4\pi/2j+1)^{1/2}Y_{jm}(\beta\alpha)=e^{im\alpha}d^{j}_{m0}(\beta), \tag{A19}$$

all of the above equations (after setting $m'=0$) may be converted to summations over m with respect to spherical harmonics. All of the above equations are useful for derivations of the closed forms of quadrupole transitions.

APPENDIX B: THE PROOF OF EQS. (18)–(22)

For the $\Sigma \rightarrow \Sigma$ electric-quadrupole transition, $\mu = 0$ and $\Lambda = K = 0$. Then Eq. (16) reduces to

$$I(\Theta,\Phi) = \sum_{M} \left| \sum_{J',M'} H'^{fi}_{vJ'} (2J'+1) \times \begin{pmatrix} J' & 2 & J \\ \bar{M}' & \bar{1} & M \end{pmatrix} \begin{pmatrix} J' & 2 & J \\ 0 & 0 & 0 \end{pmatrix} D^{J'}_{M'0}(\Phi,\Theta,0) \right|^2. \tag{17}$$

Since

$$D^{J}_{M0}(\Phi,\Theta,0) = (-1)^{M} (4\pi/(2J+1))^{1/2} Y_{J,M}(\Theta,\Phi),$$

Eq. (17) becomes

$$\begin{aligned} I(\Theta,\Phi) &= \sum_{M} \left| \sum_{J'} H'^{fi}_{vJ'} (4\pi)^{1/2} (2J'+1)^{1/2} \begin{pmatrix} J' & 2 & J \\ \bar{M}+1 & \bar{1} & M \end{pmatrix} \begin{pmatrix} J' & 2 & J \\ 0 & 0 & 0 \end{pmatrix} Y_{J',M-1}(\Theta,\Phi) \right|^2 \\ &= 4\pi \sum_{M} \Big| H'^{fi}_{vJ} (2J+1)^{1/2} \begin{pmatrix} J & 2 & J \\ \bar{M}+1 & \bar{1} & M \end{pmatrix} \begin{pmatrix} J & 2 & J \\ 0 & 0 & 0 \end{pmatrix} Y_{JM-1}(\Theta,\Phi) \\ &\quad + H'^{fi}_{v,J-2} (2J-3)^{1/2} \begin{pmatrix} J-2 & 2 & J \\ \bar{M}+1 & \bar{1} & M \end{pmatrix} \begin{pmatrix} J-2 & 2 & J \\ 0 & 0 & 0 \end{pmatrix} Y_{J-2,M-1}(\Theta,\Phi) \\ &\quad + H'^{fi}_{v,J+2} (2J+5)^{1/2} \begin{pmatrix} J+2 & 2 & J \\ \bar{M}+1 & \bar{1} & M \end{pmatrix} \begin{pmatrix} J+2 & 2 & J \\ 0 & 0 & 0 \end{pmatrix} Y_{J+2,M-1}(\Theta,\Phi) \Big|^2. \end{aligned}$$

Using some of the formulas of Appendix A and the formulas of $3j$ symbols of Ref. 20, we obtain Eqs. (18)–(22).

APPENDIX C: ANGULAR DISTRIBUTION OF THE PHOTOFRAGMENTS OF RIGID ROTORS: QUADRUPOLE TRANSITION

We follow Zare's[22] arguments for the quadrupole transition. The total distribution of rigid rotor axes is

$$P_J(\vartheta) = \sum_{M''} |\langle J''M'',20|JM\rangle|^2 |Y_{JM}(\vartheta\varphi)|^2, \tag{C1}$$

where $|J''M''\rangle$ is the initial state of the rotor, and the selection rules are $\Delta J = 0, \pm 2$, that is, $J = J'' - 2$ for the O branch, $J = J''$ for the Q branch, and $J = J'' + 2$ for the S branch.

For Q branch absorption,

$$\begin{aligned} P_{J''}(\vartheta) &= \sum_{M''} \frac{9M''^4 + J''^2(J''+1)^2 - 6M''^2 J''(J''+1)}{J''(J''+1)(2J''-1)(2J''+3)} |Y_{J''M''}(\vartheta,\varphi)|^2 \\ &= \frac{(2J''+1)(J''^2+J+6)}{32\pi(2J''-1)(2J''+3)} \left[1 - \frac{8(4J''^2+4J''-15)}{7(J''^2+J''+6)} P_2(\cos\vartheta) + \frac{216(J''-1)(J''+2)}{35(J''^2+J''+6)} P_4(\cos\vartheta) \right]. \end{aligned} \tag{C2}$$

For the O branch,

$$P_{J''-2}(\vartheta) = \frac{2J''-3}{20\pi} \left[1 + \frac{10(J''-2)}{7(2J''-1)} P_2(\cos\vartheta) + \frac{9(J''-2)(J''-3)}{14(2J''+1)(2J''-1)} P_4(\cos\vartheta) \right]. \tag{C3}$$

For the S branch,

$$P_{J''+2}(\vartheta) = \frac{2J''+5}{20\pi} \left[1 + \frac{10(J''+3)}{7(2J''+3)} P_2(\cos\vartheta) + \frac{9(J''+3)(J''+4)}{14(2J''+1)(2J''+3)} P_4(\cos\theta) \right], \tag{C4}$$

where we have used Eqs. (24) and (32) of Ref. 16 (let $m' = 0$, and then use the relationship between spherical harmonics and the rotation matrix).

APPENDIX D: INTEGRATIONS OF FINITE ROTATION MATRIX ELEMENTS

Since[20]

$$d^{j_1}_{\mu_1 m_1}(\beta) d^{j_2}_{\mu_2 m_2}(\beta) = (-1)^{\mu_2 - m_2} \sum_{j_3} (2j_3+1) \begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \bar{\mu}_2 & \mu_3 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & \bar{m}_2 & m_3 \end{pmatrix} d^{j_3}_{\mu_3 m_3}(\beta), \tag{D1}$$

if we let $\mu_1 = \mu_2 = \mu$, we then obtain

$$d^{j_1}_{\mu m_1}(\beta) d^{j_2}_{\mu m_2}(\beta) = \sum_{j_3} (-1)^{\mu - m_2} (2j_3+1) \begin{pmatrix} j_1 & j_2 & j_3 \\ \mu & \bar{\mu} & 0 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & \bar{m}_2 & m_3 \end{pmatrix} (-1)^{m_3} d^{j_3}_{m_3 0}(\beta). \tag{D2}$$

Also,

$$d^{j}_{m0}(\beta) = (-1)^{m} [(j+m)!/(j-m)!]^{1/2} P_j^{-m}(\cos\beta). \tag{D3}$$

Using Eq. (D3), we convert Eq. (D2) into the following form:

$$d^{j_1}_{\mu m_1}(\beta)d^{j_2}_{\mu m_2}(\beta)=\sum_{j_3}(-1)^{\mu-m_2}(2j_3+1)\begin{pmatrix}j_1 & j_2 & j_3\\ \mu & \bar{\mu} & 0\end{pmatrix}\begin{pmatrix}j_1 & j_2 & j_3\\ m_1 & \bar{m}_2 & m_3\end{pmatrix}\left[\frac{(j_3+m_3)!}{(j_3-m_3)!}\right]^{1/2}P_{j_3}^{-m_3}(\cos\beta). \tag{D4}$$

Multiplying the above equation by $\sin^{n-1}\beta\,d\beta$ and integrating from 0 to π with respect to β on both sides of Eq. (D4), we obtain

$$\int_0^\pi d^{j_1}_{\mu m_1}(\beta)\sin^{n-1}\beta\,d^{j_2}_{\mu m_2}(\beta)d\beta=(-1)^{\mu-m_2}\sum_{j_3}(2j_3+1)\left[\frac{(j_3+m_3)!}{(j_3-m_3)!}\right]^{1/2}\begin{pmatrix}j_1 & j_2 & j_3\\ \mu & \bar{\mu} & 0\end{pmatrix}\begin{pmatrix}j_1 & j_2 & j_3\\ m_1 & \bar{m}_2 & m_3\end{pmatrix}$$
$$\times\frac{2^{-m_3}\cdot\pi\cdot\Gamma(\frac{1}{2}n+\frac{1}{2}m_3)\cdot\Gamma(\frac{1}{2}n-\frac{1}{2}m_3)}{\Gamma(\frac{1}{2}+\frac{1}{2}n+\frac{1}{2}j_3)\cdot\Gamma(\frac{1}{2}n-\frac{1}{2}j_3)\cdot\Gamma(\frac{1}{2}m_3+\frac{1}{2}j_3+!)\cdot\Gamma(\frac{1}{2}m_3-\frac{1}{2}j_3+\frac{1}{2})}$$
$$[R(n\pm m_3)>0]. \tag{D5}$$

Here, we have used the identity[23]

$$\int_0^\pi(\sin\beta)^{n-1}P_{j_3}^{-m_3}(\cos\beta)d\beta=\frac{2^{-m_3}\cdot\pi\cdot\Gamma(\frac{1}{2}n+\frac{1}{2}m_3)\cdot\Gamma(\frac{1}{2}n-\frac{1}{2}m_3)}{\Gamma(\frac{1}{2}+\frac{1}{2}n+\frac{1}{2}j_3)\cdot\Gamma(\frac{1}{2}n-\frac{1}{2}j_3)\cdot\Gamma(\frac{1}{2}m_3+\frac{1}{2}j_3+1)\cdot\Gamma(\frac{1}{2}m_3-\frac{1}{2}j_3+\frac{1}{2})}$$
$$[R(n\pm m_3)>0], \tag{D6}$$

where the $\Gamma(k)$'s are gamma functions.

In the following we consider two special cases.

Case 1: Let $j_1=j_2=j$, $m_2=\Lambda$, $m_1=-\Lambda$, and $n=2$. Then Eq. (D5) reduces to

$$\int_0^\pi d^{j}_{\mu\bar{\Lambda}}(\beta)d^{j}_{\mu\Lambda}(\beta)\sin\beta\,d\beta=(-1)^{\mu-\Lambda}\sum_{j_3=\text{even}}^{2j}(2j_3+1)\left[\frac{(j_3+2\Lambda)!}{(j_3-2\Lambda)!}\right]^{1/2}\begin{pmatrix}j & j & j_3\\ \mu & \bar{\mu} & 0\end{pmatrix}\begin{pmatrix}j & j & j_3\\ \Lambda & \bar{\Lambda} & 2\Lambda\end{pmatrix}$$
$$\times\frac{2^{-2\Lambda}\cdot\pi\cdot\Gamma(1+\Lambda)\cdot\Gamma(1-\Lambda)}{\Gamma(\frac{3}{2}+\frac{1}{2}j_3)\cdot\Gamma(1-\frac{1}{2}j_3)\cdot\Gamma(\Lambda+\frac{1}{2}j_3+1)\cdot\Gamma(\Lambda-\frac{1}{2}j_3+\frac{1}{2})}\quad[R(2\pm2\Lambda)>0]. \tag{D7}$$

It is very hard to get a closed form for this integration.

Case 2: Let $m_1=m_2=0$ and $n=3$. Equation (D5) becomes (we changed μ to m, j_1 to j, j_2 to j', j_3 to J, for convention

$$\int_{-1}^{1}(1-x^2)^{1/2}P_j^m(x)P_{j'}^m(x)dx=(-1)^m\left[\frac{(j+m)!(j'+m)!}{(j-m)!(j'-m)!}\right]^{1/2}\sum_{J=\text{even}}(2J+1)\begin{pmatrix}j & j' & J\\ m & \bar{m} & 0\end{pmatrix}\begin{pmatrix}j & j' & J\\ 0 & 0 & 0\end{pmatrix}$$
$$\times\pi^2\left[4\Gamma(2+\tfrac{1}{2}J)\cdot\Gamma(\tfrac{3}{2}-\tfrac{1}{2}J)\cdot\Gamma(1+\tfrac{1}{2}J)\cdot\Gamma(\tfrac{1}{2}-\tfrac{1}{2}J)\right]^{-1}, \tag{D8}$$

where $x=\cos\beta$, and $P_j^m(x)$ are the associated Legendre polynomials. This way, we obtained the values of the matrix elements $\langle Y_{j'm}|\sin\beta\,|Y_{jm}\rangle$.

The method is different from and appears to be more general than that of Lin and Koizumi.[6]

[1]R. N. Zare, Ph.D. thesis, Harvard University, 1964.
[2]R. N. Zare, J. Chem. Phys. **47**, 204 (1967).
[3]J. Solomon, J. Chem. Phys. **47**, 889 (1967).
[4]J. Solomon, C. Jonah, P. Chandra, and R. Bersohn, J. Chem. Phys. **55**, 1908 (1971).
[5]C. Jonah, J. Chem. Phys. **54**, 1915 (1971).
[6]S. H. Lin and C. J. Koizumi, Can. J. Phys. **49**, 2425 (1971).
[7]R. N. Zare, Mol. Photochem. **4**, 1 (1972).
[8]G. E. Busch and K. R. Wilson, J. Chem. Phys. **56**, 3638 (1972).
[9]S. C. Yang and R. Bersohn, J. Chem. Phys. **61**, 4402 (1974).
[10]C. Pernot, J. Durup, J. B. Ozenne, J. A. Beswick, P. C. Cosby, and J. T. Moseley, J. Chem. Phys. **71**, 2387 (1979).
[11]G. G. Balint-Kurti and M. Shapiro, Chem. Phys. **61**, 137 (1981).
[12]C. H. Greene and R. N. Zare, Ann. Rev. Phys. Chem. **33**, 119 (1982); J. Chem. Phys. **78**, 6741 (1983).
[13]K. M. Chen and C. C. Pei, Chem. Phys. **76**, 219 (1983).
[14]Y. B. Band, K. F. Freed, and S. J. Singer, J. Chem. Phys. **84**, 3762 (1986).
[15]Y. N. Chiu, J. Chinese Chem. Soc. **36**, 79 (1989).
[16]S. T. Lai and Y. N. Chiu, J. Math. Phys. **30**, 844 (1989).
[17]J. C. Tully, R. S. Berry, and A. J. Dalton, Phys. Rev. **176**, 95 (1968).
[18]A. D. Buckingham, J. Orr, and J. M. Sichel, Philos. Trans. R. Soc. London Ser. A **268**, 147 (1970).
[19]Y. N. Chiu, J. Chem. Phys. **42**, 2671 (1965); **48**, 5702 (1968).
[20](a) A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960). (b) M. R. Rotenberg, M. Bivins, J. R. Metropolis, and J. K. Wooten, Jr., *The 3-j and 6-j Symbols* (MIT, Cambridge, MA, 1959).
[21]S. T. Lai and Y. N. Chiu (in preparation).
[22]R. N. Zare, Ber. Bunsenges. **86**, 422 (1982).
[23]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U.S. Government Printing Office, Washington, DC, 1970).

Evaluations of distribution functions for flexible macromolecules by the saddle-point method

Gaoyuan Wei and B. E. Eichinger
Department of Chemistry, BG-10, University of Washington, Seattle, Washington 98195

(Received 28 June 1989; accepted for publication 3 January 1990)

The method of steepest descents has been applied to the evaluation of distribution functions for flexible macromolecules of arbitrary complexity with the effective potential of the mean force being a quadratic form. Approximate evaluations of the distribution functions of the radius of gyration and the two-dimensional shape distribution functions over the entire variable domains are shown to be both feasible and effective. The asymptotics of the distribution functions are also studied, and a simple asymptotic formula is obtained that is valid for flexible macromolecules confined to a plane and of any structure such that the smallest eigenvalue of the Kirchhoff matrix has odd degeneracy.

I. INTRODUCTION

During the past two decades, much progress has been made on the study of equilibrium statistical mechanics of macromolecules with the advent of powerful mathematical tools such as integrals over Stiefel manifolds.[1-3] As a first-order approximation to a generally complicated many-body interaction potential for a macromolecular system, a quadratic potential has proved adequate for many applications in the field of configuration statistics of macromolecules. This has come to be known as the Gaussian model. In this paper, we evaluate distribution functions for the Gaussian model of flexible macromolecules with the use of a steepest-descents technique.

The distribution functions of interest invariably involve integrals of the form[1-5]

$$\int w(x)|\mathbf{1}+i\mathbf{x}\otimes\tilde{\Lambda}^{-1}|^{-a}\,\mathrm{etr}(i\tilde{\mathbf{S}}_d\mathbf{h}\mathbf{x}\mathbf{h}')d\,\mathbf{h}\,d\mathbf{x},$$

where a is a positive number, k is the dimensionality of the space in which the molecules are imbedded, n denotes the number of vertices of the molecular graph with the mean-square distance between two neighboring vertices being $\langle\ell^2\rangle_0$, $\tilde{\Lambda}=\mathrm{diag}(\tilde{\lambda}_1,\tilde{\lambda}_2,\ldots,\tilde{\lambda}_{n-1})$ and $\tilde{\mathbf{S}}_d=\mathrm{diag}(\tilde{S}_1,\tilde{S}_2,\ldots,\tilde{S}_k)$ with $\tilde{S}_j\tilde{\lambda}_j=n\gamma S_j\lambda_j=nkS_j\lambda_j/2\langle\ell^2\rangle_0$, i.e., $\tilde{S}_j=\gamma S_j/n^{\alpha}$ and $\tilde{\lambda}_j=n^{\alpha+1}\lambda_j$, α being a constant properly chosen for a given type of molecules and taking the value 1 for many relatively simple structures such as linear chains, circular chains, double rings, and stars. The λ_j are the nonzero eigenvalues of the Kirchhoff matrix of the graph, the S_j are the principal components of the gyration tensor $\mathbf{S}$ for the molecule, $\mathbf{h}\in\mathbf{SO}(k)$ spans the special orthogonal group, $\mathbf{h}'$ is the transpose of $\mathbf{h}$, $d\,\mathbf{h}$ is the unnormalized invariant measure on $\mathbf{SO}(k)$, $\mathbf{x}=\mathrm{diag}(x_1,x_2,\ldots,x_k)$ with $-\infty\leqslant x_k\leqslant x_{k-1}\leqslant\cdots\leqslant x_1\leqslant\infty$, $d\mathbf{x}=\prod_{i=1}^{k}dx_i$, and the function $w(\mathbf{x})$ is defined as

$$w(x)=\begin{cases}1, & \text{for } k=1,\\ \prod_{\alpha<\beta}^{k}|x_\alpha-x_\beta|, & \text{for } k>1.\end{cases}$$

For convenience, let κ be the diagonal matrix of all distinct $\tilde{\lambda}_j$ each with degeneracy ω_j, i.e., $\kappa=\mathrm{diag}(\kappa_1,\kappa_2,\ldots,\kappa_p)$ with $\kappa_j<\kappa_{j+1}$. In what follows, we first concentrate on the evaluations of distribution functions of the radius of gyration, and then treat shape distribution functions in two dimensions. In both cases, only the saddle-point method is employed and the asymptotic behavior of these functions is also investigated. We will deal with the 3-D shape distribution functions in a separate paper.

II. DISTRIBUTION FUNCTIONS OF RADIUS OF GYRATION

In the framework of the Gaussian model, the probability distribution of the radius of gyration s^2, $s^2=\mathrm{tr}(\mathbf{S})$, for flexible macromolecules was first studied by Fixman in 1961. He obtained simple expressions for the distribution function for infinitely long linear chains at very large and very small values of the arguments by use of the method of steepest descents.[6] Since then, the distribution functions have been evaluated both approximately and exactly by various techniques for a number of different forms of their representations (see Ref. 2 for a review on this subject). Here we present a method, based on the steepest-descents technique, for approximately evaluating these functions in complete generality. In the following, we start with the case where the eigenvalues or the density functions of the eigenvalues of the Kirchhoff matrix are known, and then work on another case where only the coefficients of the eigenpolynomial of the matrix are needed. Finally, we give a special treatment for linear and circular chains, obtaining independently analytic expressions for the second and fourth moments of the distributions as well as carrying out the steepest-descents calculations of the distribution functions.

The most useful form of the distribution functions of the radius of gyration is[1]

$$P(s^2)ds^2=\left(\frac{ds^2}{2\pi}\right)\int_{-\infty}^{\infty}\exp(i\beta s^2)\prod_{j=1}^{n-1}\left(1+i\frac{\beta}{\gamma n\lambda_j}\right)^{-a}d\beta, \quad (1)$$

where $a=k/2$ with k denoting the dimensionality of the space. In terms of the dimensionless quantity $\tilde{s}^2$, the reduced radius of gyration, Eq. (1) becomes

$$P(s^2)ds^2=\left(\frac{d\tilde{s}^2}{2\pi}\right)\int_{-\infty}^{\infty}\exp(i\,\tilde{s}^2x)\prod_{j=1}^{p}\left(1+i\frac{x}{\kappa_j}\right)^{-a\omega_j}dx. \quad (2)$$

 0022-2488/90/051274-06$03.00

We now evaluate the integral in Eq. (2) by the method of steepest descents along the lines of the calculation of Berlin and Kac in their treatment of the spherical model of a ferromagnet.[7] Let

$$g(x) = i\tilde{s}^2 x - a\sum_{j=1}^{p} \omega_j \ln\left(1 + \frac{ix}{\kappa_j}\right), \tag{3}$$

and x_s be the saddle point corresponding to a given $\tilde{s}^2$. The existence and uniqueness of such a normal saddle point for any $\tilde{s}^2$ in the general case considered here will be discussed shortly. A steepest-descent calculation yields

$$P(s^2)ds^2 \simeq (d\tilde{s}^2/\sqrt{2\pi})e^{g(x_s)}[-g''(x_s)]^{-1/2}, \tag{4}$$

where the saddle point x_s for a given $\tilde{s}^2$ is governed by the equation

$$\tilde{s}^2 = a\sum_{j=1}^{p} \frac{\omega_j}{(\kappa_j + ix_s)}, \tag{5}$$

and by the inequality

$$g''(x_s) = -a\sum_{j=1}^{p} \frac{\omega_j}{(\kappa_j + ix_s)^2} < 0, \tag{6}$$

where $g''(x)$ is the second derivative of $g(x)$ with respect to x. From the above two conditions, one immediately sees that x_s must lie on the imaginary axis. Let $y_s = ix_s$. We then have $y_s = -\kappa_1$ and ∞ corresponding to $\tilde{s}^2 = \infty$ and 0, respectively (see also Ref. 8). For $y_s = 0$, one gets $\tilde{s}^2 = a\Sigma_{j=1}^{p}\omega_j/\kappa_j = \langle\tilde{s}^2\rangle_0$, the second moment of the distribution.[2,9] Furthermore, the condition that $\ln(1 + y_s/\kappa_j)$ be real for all j requires that y_s be greater than $-\kappa_1$. It then follows that there exists a unique saddle point x_s for a given $\tilde{s}^2 \in [0, \infty]$. From Eq. (4), we have

$$P(s^2)ds^2 \simeq (d\tilde{s}^2/\sqrt{2\pi a})e^{y_s\tilde{s}^2}D(y_s)[F_1(y_s)]^{-1/2}, \tag{7}$$

where $D(x)$ and $F_1(x)$ are defined as

$$D(x) = \prod_{j=1}^{p}\left(1 + \frac{x}{\kappa_j}\right)^{-a\omega_j} \tag{8}$$

and

$$F_1(x) = \sum_{j=1}^{p} \frac{\omega_j}{(\kappa_j + x)^2}, \tag{9}$$

respectively. For $\tilde{s}^2 = 0$, $\langle\tilde{s}^2\rangle_0$, and ∞, it is not difficult to show that $P(0) = P(\infty) = 0$ and

$$P(\langle s^2\rangle_0)ds^2 \simeq (d\tilde{s}^2/\sqrt{2\pi})(\langle\tilde{s}^4\rangle_0 - \langle\tilde{s}^2\rangle_0^2)^{-1/2}, \tag{10}$$

where the equality $\langle\tilde{s}^4\rangle_0 - \langle\tilde{s}^2\rangle_0^2 = a\,\Sigma_{j=1}^{p}\omega_j/\kappa_j^2$ has been used.[2,9] To see how good the approximation is, let us consider a simple case: A linear chain of two beads connected by a Hookean spring and imbedded in k-dimensional space. The distribution function for these dumbbell molecules is readily found to be given exactly by

$$P(s^2)ds^2 = d\tilde{s}^2[\kappa_1^a/\Gamma(a)]\tilde{s}^{2(a-1)}e^{-\kappa_1\tilde{s}^2}, \tag{11}$$

while the steepest-descent calculation gives

$$P(s^2)ds^2 \simeq d\tilde{s}^2[\kappa_1^a/\Gamma(a)]C_1(a\omega_1)\tilde{s}^{2(a-1)}e^{-\kappa_1\tilde{s}^2}, \tag{12}$$

where $C_1(x)$ is defined as

$$C_1(x) = (2\pi)^{-1/2}x^{-(x+1/2)}x!e^x, \tag{13}$$

which goes to 1 as $x \to \infty$. One thus sees that the two results become identical for infinitely large a. In a sample calculation made for linear chains with eleven beads in one dimension, good agreement between the exact and the approximate results based on Eqs. (7)–(9) was found.

To obtain asymptotic expressions for the distribution functions, we make use of the fact that the saddle point approaches $y_s = -\kappa_1$ for large values of the radius of gyration.[8,10] Let $y_s = -\kappa_1 + \epsilon$, where ϵ is a small and positive quantity. Equation (5) then yields

$$\epsilon \simeq a\omega_1/\tilde{s}^2, \tag{14}$$

for large $\tilde{s}^2$. We thus have

$$F_1(y_s) \simeq \omega_1/\epsilon^2 \tag{15}$$

and

$$D(y_s) \simeq \left(\frac{\kappa_1}{\epsilon}\right)^{a\omega_1}\prod_{j=2}^{p}\left(1 - \frac{\kappa_1}{\kappa_j}\right)^{-a\omega_j}. \tag{16}$$

Substitution of Eqs. (14)–(16) into Eq. (7) gives

$$P(s^2)ds^2 \sim d\tilde{s}^2\, B_1^a C_2(a\omega_1)\tilde{s}^{2(a\omega_1 - 1)}e^{-\kappa_1\tilde{s}^2}, \tag{17}$$

where $C_2(x)$ is given by

$$C_2(x) = (2\pi)^{-1/2}x^{-x+1/2}e^x, \tag{18}$$

and B_j is defined as

$$B_j = \kappa_j^{\omega_j}\prod_{l\neq j}^{p}\left(1 - \frac{\kappa_j}{\kappa_l}\right)^{\omega_l}. \tag{19}$$

Comparing Eq. (17) with the leading terms in the exact asymptotic expansions of the distribution functions as given by[9]

$$P(s^2)ds^2 \sim d\tilde{s}^2[B_1^a/\Gamma(a\omega_1)]\tilde{s}^{2(a\omega_1 - 1)}e^{-\kappa_1\tilde{s}^2}, \tag{20}$$

we see that they differ in that $1/\Gamma(a\omega_1)$ stands in place of $C_2(a\omega_1)$ as a constant factor in Eq. (20). It is of interest to note that $\Gamma(x)$ approaches $1/C_2(x)$ as $x \to \infty$, implying that the larger k is, the better approximation is the steepest-descents calculation.

When the product in Eq. (1) is expressed in terms of the eigenpolynomial $P_G(2 - \lambda)$ of the molecular graph G with $\lambda \equiv -i\beta/\gamma n$ (Ref. 11), i.e.,

$$P_G(2 - \lambda) = -\lambda\prod_{j=1}^{n-1}(\lambda_j - \lambda) = \sum_{j=0}^{n}\alpha_j\lambda^{n-j}, \tag{21}$$

where the α_j are the coefficients of the eigenpolynomial, we have

$$P(s^2)ds^2 = \frac{d\alpha}{2\pi i}\int_{1-i\infty}^{1+i\infty} e^{-\alpha(1-z)}[P_n(z)]^{-a}dz, \tag{22}$$

where $\alpha \equiv 2\gamma n s^2 \equiv 2n^2\tilde{s}^2$, and the function $P_n(z)$ is given by

$$P_n(z) = \frac{1}{\alpha_{n-1}}\sum_{j=0}^{n-1}\alpha_j[2(1-z)]^{n-j-1}, \tag{23}$$

which may also be expressed in terms of Chebyshev polynomials in z (Refs. 2, 11). As in the previous case, let z_s be the saddle point at which the phase of the contour of integration takes the value $\pi/2$. The steepest-descents calculation on the integral in Eq. (22) then yields

$$P(s^2)ds^2 \simeq (d\tilde{s}^2\sqrt{2\pi a})[P_n(z_s)]^{-a}e^{-a(1-z_s)} \times\left[-\lim_{z\to z_s}\frac{d^2}{dz^2}\ln P_n(z)\right]^{-1/2}, \tag{24}$$

where z_s satisfies the equation

$$\alpha = a \lim_{z \to z_s} \frac{d}{dz} \ln P_n(z), \tag{25}$$

under the constraints

$$\lim_{z \to z_s} \frac{d^2}{dz^2} \ln P_n(z) < 0 \tag{26}$$

and

$$P_n(z_s) > 0. \tag{27}$$

It is to be noticed that in this case the saddle points all lie to the left of $z_s = 1$ on the real axis with $z_s = 1$ corresponding to $\tilde{s}^2 = (-a/n^2)(\alpha_{n-2}/\alpha_{n-1})$, which is simply $\langle \tilde{s}^2 \rangle_0$ (Refs. 9, 11). At $s^2 = \langle s^2 \rangle_0$, one therefore has

$$P(\langle s^2 \rangle_0) ds^2 \simeq d\tilde{s}^2 \frac{n^2}{\sqrt{2\pi a}} \left[\left(\frac{\alpha_{n-2}}{\alpha_{n-1}} \right)^2 - 2 \frac{\alpha_{n-3}}{\alpha_{n-1}} \right]^{-1/2} \tag{28}$$

Comparison of Eq. (28) with Eq. (10) shows that

$$\langle \tilde{s}^4 \rangle_0 - \langle \tilde{s}^2 \rangle_0^2 = \frac{a}{n^4} \left[\left(\frac{\alpha_{n-2}}{\alpha_{n-1}} \right)^2 - 2 \frac{\alpha_{n-3}}{\alpha_{n-1}} \right], \tag{29}$$

which reduces to the result obtained by Yang and Yu for $k = 3$ (Ref. 11).

In the case of linear chains,[2,8,11] one finds the alternative representation

$$P_n(x) = (1/n) U_{n-1}(x), \tag{30}$$

where $U_n(x)$ is the Chebyshev polynomial of the second kind of order n. Equation (24) then becomes

$$P(s^2) ds^2 \simeq \frac{d\alpha}{\sqrt{2\pi a}} \left[\frac{U_{n-1}(z_s)}{n} \right]^{-a} e^{-\alpha(1-z_s)} [A(z_s)]^{-1/2}, \tag{31}$$

where z_s is the solution of the equation

$$\beta \equiv \frac{\alpha}{a} = \frac{n}{1 - z_s^2} \left[\frac{n+1}{n} z_s - \frac{U_n(z_s)}{U_{n-1}(z_s)} \right], \tag{32}$$

with the constraints

$$U_{n-1}(z_s) > 0 \tag{33}$$

and

$$A(z_s) \equiv \beta^2 + \frac{n^2 - 1 - 3\beta z_s}{1 - z_s^2} > 0. \tag{34}$$

Note that $U_n(x) = C_n^1(x)$, where $C_n^\lambda(x)$ is the ultraspherical polynomial with the following properties[12]:

$$\frac{d^k}{dx^k} C_n^\lambda(x) = 2^k \frac{\Gamma(\lambda + k)}{\Gamma(\lambda)} C_{n-k}^{\lambda + k}(x), \tag{35}$$

and

$$C_n^\lambda(1) = \binom{2\lambda + n - 1}{n}. \tag{36}$$

Then, with the limit $z_s \to 1$ taken on the right-hand side of Eq. (32), one has

$$\langle \tilde{s}^2 \rangle_0 = (a/6)(1 - 1/n^2), \tag{37a}$$

or

$$\langle s^2 \rangle_0 = (n \langle \ell^2 \rangle_0 / 6)(1 - 1/n^2). \tag{37b}$$

Similarly, by first evaluating Eq. (31) for $z_s = 1$ and then comparing the result obtained with Eq. (10), we obtain

$$\langle \tilde{s}^4 \rangle_0 - \langle \tilde{s}^2 \rangle_0^2 = (a/180)(2 + 5n^{-2} - 7n^{-4}). \tag{38}$$

Equations (37) and (38) recover the earlier results for $k = 3$ (Refs. 2, 9, 11). For circular chains,[2,11] one has

$$P_n(x) = [U_{n-1}(y)/n]^2, \tag{39}$$

where $y^2 = (1 - x)/2$. It then follows that

$$P(s^2) ds^2 \simeq \frac{d\alpha}{\sqrt{2\pi a}} 2^{3/2} \left[\frac{U_{n-1}(z_s)}{n} \right]^{-2a} \times e^{-2\alpha(1 - z_s^2)} [B(z_s)]^{-1/2}, \tag{40}$$

where z_s is governed by the equality

$$\beta \equiv \frac{2\alpha}{a} = \frac{n}{z_s(1 - z_s^2)} \left[\frac{n+1}{n} z_s - \frac{U_n(z_s)}{U_{n-1}(z_s)} \right], \tag{41}$$

and by the following two inequalities:

$$U_{n-1}(z_s) > 0 \tag{42}$$

and

$$B(z_s) \equiv \beta^2 + \frac{n^2 - 1 + \beta(1 - 4z_s^2)}{z_s^2(1 - z_s^2)} > 0. \tag{43}$$

As in the case of linear chains, we find that for circular chains,

$$\langle \tilde{s}^2 \rangle_0 = (a/12)(1 - 1/n^2) \tag{44}$$

and

$$\langle \tilde{s}^4 \rangle_0 - \langle \tilde{s}^2 \rangle_0^2 = (a/720)(1 + 10n^{-2} - 11n^{-4}), \tag{45}$$

which also reproduce the results for $k = 3$ obtained by other analytic techniques.[2,9,11]

III. SHAPE DISTRIBUTION FUNCTIONS IN TWO DIMENSIONS

Distribution functions for the principal components of the gyration tensor for flexible macromolecules confined to a plane were first investigated by Šolc and Gobush,[13] with the results expressed in the form of a Bessel–Fourier series. A different approach to the solution of the above problem was later taken by one of us, and the distribution function for circular chains was obtained as a double sum of simple integrals.[3] Numerical evaluations of the distribution functions for both linear and circular chains in two dimensions have been reported.[4] In what follows, we first carry out the steepest-descents calculations of the general distribution functions over the entire variable domains, and then apply a two-dimensional saddle-point method to obtain asymptotic expressions for the distribution functions.

The general shape distribution functions in two dimensions take the form[3]

$$P(S_1, S_2) dS_1\, dS_1\, dS_2 = dS_1\, dS_2 \frac{\Delta}{2\pi} \int_{-\infty < \beta_2 < \beta_1 < \infty} \beta_- e^{i\beta_+ s^2} J_0(\Delta\beta_-) \times \prod_{j=1}^{n-1} \left[\left(1 + i\frac{\beta_1}{\gamma n \lambda_j} \right) \left(1 + i\frac{\beta_2}{\gamma n \lambda_j} \right) \right]^{-a} d\beta_1\, d\beta_2, \tag{46}$$

where $a = 1/2$, $\Delta = S_1 - S_2$, $\beta_\pm = (\beta_1 \pm \beta_2)/2$, and

$J_0(x)$ is the Bessel function of the first kind of zero order. Under a linear transformation of the integration variables,[3] Eq. (46) can be rewritten as

$$P(S_1,S_2)dS_1\,dS_2 = d\tilde{S}_1\,d\tilde{S}_2(\tilde{\Delta}/\pi)\,|\kappa|^{2a}\int_0^\infty x_- J_0(\tilde{\Delta}x_-)F(x_-)dx_-$$

where $|\kappa| = \Pi_{j=1}^{p}\kappa_j^{\omega_j}$ and

$$F(x_-) = \int_{-\infty}^{\infty} e^{i\tilde{s}^2 x_+}\prod_{j=1}^{p}[x_-^2 + (\kappa_j + ix_+)^2]^{-a\omega_j}dx_+. \tag{48}$$

We now evaluate $F(x_-)$ by the method of steepest descents. On inspection of the integrand in Eq. (48), one immediately sees that any saddle point, denoted by x'_+, must lie on the imaginary axis and that x'_+ is an even function of x_- for a given $\tilde{s}^2$. Let $y_s = ix'_+$. We then have

$$F(x_-) \simeq (\pi/a)^{1/2}D(x_-,y_s)e^{y_s\tilde{s}^2}[F_2(x_-,y_s)]^{-1/2}, \tag{49}$$

where $D(x,y)$ is given by

$$D(x,y) = \prod_{j=1}^{p}[x^2 + (\kappa_j + y)^2]^{-a\omega_j}, \tag{50}$$

and y_s satisfies the equation

$$\tilde{s}^2 = 2a\sum_{j=1}^{p}\omega_j(\kappa_j + y_s)/[x_-^2 + (\kappa_j + y_s)^2], \tag{51}$$

with $y_s > -\kappa_1$ and subject to the additional condition that

$$F_2(x_-,y_s) \equiv \sum_{j=1}^{p}\omega_j[(\kappa_j + y_s)^2 - x_-^2]/[x_-^2 + (\kappa_j + y_s)^2]^2 > 0. \tag{52}$$

It must be noted that under these conditions $|x_-|$ can never go to ∞ for $\tilde{s}^2 > 0$. We therefore conclude that there exists a unique upper limit for $|x_-|$ beyond which no saddle points can be found. As a result, the semi-infinite integral in Eq. (47) needs only to be evaluated for a finite upper integration limit, say, x_c. The existence and uniqueness of such a limit for any given positive $\tilde{s}^2$ were seen in a sample calculation for circular chains with five beads and will also be noticed in an example to follow.

We further note the integral identity

$$\int_0^\infty dx_-\,x_- J_0(\tilde{\Delta}x_-)F(x_-) = \int_0^{-\infty} dx_-\,x_- J_0(\tilde{\Delta}x_-)F(x_-), \tag{53}$$

which indicates that the contribution to the semi-infinite integral includes two parts: One resulting from the steepest-descent calculation of $F(x_-)$ for $x_- \in [0,\infty]$ and the other for $x_- \in [-\infty,0]$. Since the saddle points are even functions of x_-, one thus need only multiply the semi-infinite integral over $x_- \in [0,\infty]$ by a factor of 2. The resulting expressions for the distribution functions then take the form

$$P(S_1,S_2)dS_1\,dS_2 \simeq d\tilde{S}_1\,d\tilde{S}_2(2/\sqrt{\pi a})\,|\kappa|^{2a}\tilde{\Delta}\int_0^{x_c} dx_-\,x_- J_0(\tilde{\Delta}x_-) \times D(x_-,y_s)e^{\bar{y}_s\tilde{s}^2}[F_2(x_-,y_s)]^{-1/2}. \tag{54}$$

For the special case of circular chains with $n = 3$, one has $\omega_1 = 2$ and $\kappa_1 = 27$. Let $\alpha = \tilde{s}^2(\kappa_1 + y_s)$ and $\beta = \tilde{s}^2 x_-$. Equation (51) then becomes

$$\alpha^2 - 2\alpha + \beta^2 = 0, \tag{55}$$

where $a = 1/2$ has been used. Similarly, one has from Eq. (52),

$$F_2(x_-,y_s) = 2(\tilde{s}^2)^2(\alpha^2 - \beta^2)/(\alpha^2 + \beta^2)^2. \tag{56}$$

The solution of Eq. (55) with the contraints $\alpha > 0$ and $\alpha^2 > \beta^2$ is then shown to be

$$\alpha = 1 + \sqrt{1 - \beta^2}, \tag{57}$$

with $|\beta| \leqslant 1$, i.e., $x_c = 1/\tilde{s}^2$. Substitution of the above results into Eq. (52) gives

$$P(S_1,S_2)dS_1\,dS_2 \simeq d\tilde{S}_1\,d\tilde{S}_2\,\kappa_1^2\tilde{\Delta}(\tilde{S}_1\tilde{S}_2)^{-1/2}e^{-\kappa_1\tilde{s}^2}C_3(r), \tag{58}$$

where $r = \tilde{\Delta}/\tilde{s}^2$ and $C_3(x)$ is given by

$$C_3(x) = \pi^{-1/2}(1 - x^2)^{1/2} \times \int_0^1 J_0(x\beta)\,e^{\alpha}[(\alpha/\beta)^2 - 1]^{-1/2}\,d\beta. \tag{59}$$

A comparison of Eq. (58) with the exact result obtained elsewhere[3,13] indicates that the relative error resulting from the steepest-descent calculation is $|C_3(r) - 1| \equiv \epsilon(r)$. In Fig. 1, we plot both $C_3(r)$ and $\epsilon(r)$ as functions of $r \in [0,1]$. It can be seen from these figures that very good approximations are achieved for a wide range of values of the arguments, with a minimum relative error of 1.2×10^{-5} at $r = 0.31$ and that the approximation is especially good in the asymptotic region of the distribution function, as expected. As $r \to 1$, i.e., $S_2 \to 0$, the error is large because the boundary

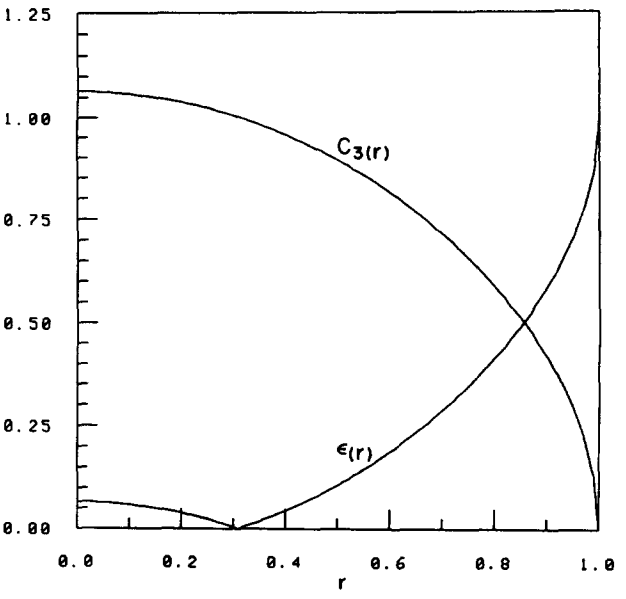


FIG. 1. Plots of $C_3(r)$ and $\epsilon(r)$ as functions of r with $r = \tilde{\Delta}/\tilde{s}^2 = (S_1 - S_2)/(S_1 + S_2)$, where the S_j are the principal components of the gyration tensor for flexible macromolecules. The function $\epsilon(r)$ is defined as $|C_3(r) - 1|$ and $C_3(r)$ is calculated according to Eq. (59). As $r \to 1$, the molecules become linear.

of the domain of the definition of S is reached. In the limiting case where $S_2 = 0$, the integral in Eq. (46) becomes irrelevant as a zero probability density results from the vanishing measure for such a set of configurations.[3] The 100% relative error for the $r = 1$ case is thus traceable to the discontinuity of the measure on the boundary, and is not a matter of concern.

We now turn our attention to the asymptotics of the distribution functions by applying the saddle-point method to the double integral in Eq. (46) or (47). Our approach follows closely that commonly adopted for obtaining the asymptotic expressions of certain integrals over the orthogonal group with real integrands.[14] Define

$$f(x_+,x_-) = i\tilde{s}^2 x_+ + \ln[x_- J_0(\tilde{\Delta}x_-)] - a\sum_{j=1}^{p}\omega_j \ln[x_-^2 + (\kappa_j + ix_+)^2], \quad (60)$$

and let the saddle point, if any, be denoted by $\mathbf{x}' \equiv (x'_+, x'_-)$. Application of the saddle-point method to the double integral in Eq. (47) then gives

$$P(S_1,S_2)dS_1\,dS_2 \simeq d\tilde{S}_1\,d\tilde{S}_2\,2|\kappa|^{2a}\tilde{\Delta}e^{f(x'_+,x'_-)}[-H(x'_+,x'_-)]^{-1/2}, \quad (61)$$

where $H(x'_+,x'_-)$ is the Hessian of $f(x_+,x_-)$ evaluated at $\mathbf{x}'$, that is,

$$H(x'_+,x'_-) = f_{++}(x'_+,x'_-)f_{--}(x'_+,x'_-) - f_{+-}(x'_+,x'_-)f_{-+}(x'_+,x'_-), \quad (62)$$

with $f_{+-} = \partial^2 f/\partial x_+\,\partial x_-$, etc. and $\mathbf{x}'$ satisfies the system of equations

$$\begin{aligned} f_+(x'_+,x'_-) &= 0, \\ f_-(x'_+,x'_-) &= 0, \end{aligned} \quad (63)$$

under the conditions that

$$H(x'_+,x'_-) < 0, \quad (64)$$

and that $\exp[f(x'_+,x'_-)]$ be real. By inspection, we see that x'_+ and x'_- are likely to be imaginary numbers. We therefore let $x_s = ix'_+$ and $y_s = ix'_-$. If x_s and y_s are real for some values of $\tilde{S}_1$ and $\tilde{S}_2$, it follows from Eq. (61) that

$$P(S_1,S_2)dS_1\,dS_2 \simeq d\tilde{S}_1\,d\tilde{S}_2|\kappa|^{2a}\tilde{\Delta}I_0(t)e^{\tilde{s}^2x_s} \times [G(x_s,y_s)]^a[F(x_s,y_s)]^{-1/2}, \quad (65)$$

where $t = \tilde{\Delta}y_s$, $I_n(x)$ is the Bessel function of the second kind of order n, and x_s and y_s are the solutions of the system of equations

$$\begin{aligned} \tilde{s}^2 &= 2a\sum_{j=1}^{p}\omega_j(\kappa_j + x_s)[(\kappa_j + x_s)^2 - y_s^2]^{-1}, \\ 1 + tI_1(t)/I_0(t) &= 2ay_s^2\sum_{j=1}^{p}\omega_j[y_s^2 - (\kappa_j + x_s)^2]^{-1}, \end{aligned} \quad (66)$$

with the constraints

$$G(x_s,y_s) \equiv (-y_s^2)^{1/2a}\prod_{j=1}^{p}[(\kappa_j + x_s)^2 - y_s^2]^{-\omega_j} > 0, \quad (67)$$

and

$$F(x_s,y_s) \equiv a\sum_{j=1}^{p}\omega_j\left\{\frac{t^2 + \beta(1-\beta) - 1}{4y_s^2} \times\left[\frac{1}{(\kappa_j + z_+)^2} + \frac{1}{(\kappa_j + z_-)^2}\right] + \frac{a}{(\kappa_j + z_+)^2}\sum_{k=1}^{p}\frac{\omega_k}{(\kappa_k + z_-)^2}\right\} > 0, \quad (68)$$

where $z_\pm = x_s \pm y_s$, and $\beta = 1 + tI_1(t)/I_0(t) \equiv \beta(t)$. We note that $\beta(x)$ is an even function of x, greater than or equal to 1 and going to x as x becomes infinitely large. As a result, one can never have both $z_+ > -\kappa_1$ and $z_- > -\kappa_1$ for $\tilde{s}^2 > 0$. This is equivalent to saying that the existence and uniqueness of x_s and y_s are not guaranteed. In fact, it is not difficult to show that there is no saddle point lying on the imaginary axis for circular chains with $n = 3$. However, the fact that z_+ and z_- cannot both be greater than $-\kappa_1$ turns out to be exactly what we need to obtain the asymptotic expressions for the distribution functions, as will be seen shortly.

With $a = 1/2$ imposed explicitly, let us first look at the ranges of values of $\tilde{s}^2$ and $\tilde{\Delta}$ that correspond to the saddle points: $z_+ = -\kappa_1 + \epsilon_+$ and $z_- = -\kappa_2 + \epsilon_-$, where ϵ_+ and ϵ_+ are small and positive quantities. Substitution of these equalities into Eq. (66) gives

$$\begin{aligned} \tilde{s}^2 &\simeq \tfrac{1}{2}(\omega_1/\epsilon_+ + \omega_2/\epsilon_-), \\ \frac{\beta}{y_s} &\simeq \tfrac{1}{2}(\omega_1/\epsilon_+ - \omega_2/\epsilon_-). \end{aligned} \quad (69)$$

We thus see that $\tilde{s}^2$ must be large. Noting that $\beta/y_s \simeq \tilde{\Delta}$ for large $\tilde{\Delta}$, we may solve the above system of equations for ϵ_+ and ϵ_-, with the results

$$\begin{aligned} \epsilon_+ &\simeq \omega_1/2\tilde{S}_1, \\ \epsilon_- &\simeq \omega_2/2\tilde{S}_2, \end{aligned} \quad (70)$$

which are consistent with the assumption that ϵ_+ and ϵ_- be small. We have therefore located the ranges of z_+ and z_-, or, equivalently x_s and y_s, which correspond to the asymptotic regions of the distribution functions. Making use of these results, we find

$$G(x_s,y_s) \simeq -B_1 y_s^2 B_2/\epsilon_+^{\omega_1}\epsilon_-^{\omega_2}, \quad (71)$$

which is positive for odd ω_1 as seen from the definition for B_j in Eq. (19), and

$$F(x_s,y_s) \simeq \omega_1\omega_2/4(\epsilon_+\epsilon_-)^2. \quad (72)$$

Noting also that $I_0(x) \sim (2\pi x)^{-1/2}\exp(x)$ for large x, we finally have from Eq. (65),

$$P(S_1,S_2)dS_1\,dS_2 \sim d\tilde{S}_1\,d\tilde{S}_2 C\tilde{S}_1^{(\omega_1-1)/2}\tilde{S}_2^{(\omega_2-1)/2} \times(\tilde{S}_2^{-1} - \tilde{S}_1^{-1})^{1/2}e^{-\kappa_1\tilde{S}_1 - \kappa_2\tilde{S}_2}, \quad (73)$$

for large $\tilde{s}^2$ and $\tilde{\Delta}$. Here, C is a constant given by

$$C = 2^{(\omega_1+\omega_2)/2-2}\pi^{-1/2}\omega_1^{(1-\omega_1)/2}\omega_2^{(1-\omega_2)/2} \times e^{(\omega_1+\omega_2)/2}[B_1(\kappa_1 - \kappa_2)B_2]^{1/2}. \quad (74)$$

Whether or not the exponential factor in C needs to be included depends upon the order of accuracy of Eq. (70). It is gratifying to see that Eq. (73) recovers the asymptotic formula obtained earlier for linear chains by a different ap-

proach.[3,14] In the case of circular chains with odd n greater than 4, for which $\omega_1 = 2$, an asymptotic expression for the distribution function has been obtained elsewhere.[4] Here, the significance of the above result lies in the extension of the types of structures of molecules for which Eq. (73) holds beyond linear and circular chains.

ACKNOWLEDGMENT

We acknowledge the support of this work by Grant No. DE-FG06-84ER45123 from the Department of Energy.

[1]B. E. Eichinger, Macromolecules **10**, 671 (1977).
[2]B. E. Eichinger, Macromolecules **13**, 1 (1980) and references therein.
[3]B. E. Eichinger, Macromolecules **18**, 211 (1985).
[4]L. Y. Shy and B. E. Eichinger, Macromolecules **19**, 838 (1986).
[5]G. Wei and B. E. Eichinger, Macromolecules **22**, 3429 (1989).
[6]M. Fixman, J. Chem. Phys. **36**, 306 (1962).
[7]T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[8]S. R. Coriell and J. L. Jackson, J. Math. Phys. **8**, 1276 (1967).
[9]G. Wei, J. Chem. Phys. **90**, 5873 (1989).
[10]B. E. Eichinger, Macromolecules **14**, 1071 (1981).
[11]Y. Yang and T. Yu, Makromol. Chem. **186**, 513 (1985).
[12]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965); Entries 8.935.1 and 8.937.4, p. 1031.
[13]K. Solc and W. Gobush, Macromolecules **7**, 814 (1974).
[14]R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).

The problem of gravity-gyroscopic waves, which are excited by the oscillations of a curve

V. M. Kharik and U. D. Pletner
Department of Mathematics, Physics Faculty of Moscow State University, 117234 Moscow, USSR

(Received 12 December 1988; accepted for publication 30 August 1989)

The problem of the oscillations of an ideal stratified rotating fluid, which are excited by a curve in the case where the distribution of pressure on both sides of the curve is prescribed, is considered. The solution to the problem is obtained, as well as results about symmetry properties of the operators used for the solution of such problems. The question of the uniqueness of the solution is also considered.

I. INTRODUCTION

This paper continues research started in Refs. 1–5 and is connected with problems relating to the excitation of oscillations in stratified rotating fluids by oscillating curves. These works presented solutions to these problems. Such problems are connected with some questions in cryogenic fluid technology and oceanography.

In this paper, we present the case where the distribution of pressure on both sides of a curve is prescribed. In this case, uncommon boundary conditions have appeared, which include time derivatives. The solution is obtained by using two potentials, which were developed in Ref. 3.

II. THE DEFINITION OF "PROBLEM W"

We shall consider flat movements of an ideal rotating stratified fluid as in Refs. 1–5. The consideration of such problems leads to the equation of gravity-gyroscopic waves in two-dimensional space (see, for example, Ref. 2):

$$\frac{\partial^2}{\partial t^2}\nabla^2 u + \omega_0^2 u_{x_1x_1} + \alpha^2 u_{x_2x_2} = 0, \tag{2.1}$$

where ∇^2 is the Laplace operator in two-dimensional space with variables x_1 and x_2, ω_0^2 is the square of the Wasjal–Brant frequency,[5] and α is the Coriolis parameter. We note that in this paper we shall not discuss physical aspects of Eq. (2.1), suggesting, instead, that the reader see Ref. 6. We only remark that the values ω_0 and α are given constants and $\omega_0 \neq \alpha$. The function $u(x,t)$ [$x = (x_1,x_2)$] is a flow function, and the components of the velocity vector $\vec{v}$ of the fluid particles can be represented by this function in the following expressions:

$$v_1 = -u_{x_2}, \quad v_2 = u_{x_1}.$$

Let us consider the curve

$$\Gamma \equiv \{(x_1,x_2): x_1 = x_1(s), x_2 = x_2(s), s \in [0,l]\},$$

in a fluid whose dynamics are described by Eq. (2.1), and orient the curve Γ by setting its sides Γ^+ and Γ^- in the following way. We denote the tangent vector at a point $x(s) = (x_1(s),x_2(s)) \in \Gamma$ of the curve Γ by $\vec{\tau}_s$, and the normal vector at $x(s) \in \Gamma$ of the curve Γ by $\vec{n}_s$. If we rotate $\vec{\tau}_s$ for $\pi/2$ counterclockwise we obtain $\vec{n}_s$. We shall call the side of the curve Γ that we see by looking toward the vector $\vec{n}_s$ Γ^+ and the opposite side of the curve Γ^-.

We assume that before time $t = 0$ there were no movements of fluid and curve Γ. After time $t = 0$, the pressure distributions on the two sides of the curve Γ are, in general, different. Mathematically it is equivalent to the prescription, on both sides $\Gamma^\pm$, of boundary conditions of the following kind for the function $u(x,t)$:

$$\begin{aligned}(\mathcal{N}_{tx} u(x,t))\big|_{x = x(s) \in \Gamma^\pm} &\equiv \left(\frac{\partial^2}{\partial t^2}\frac{\partial u}{\partial n_s} + \omega_0 \cos(\vec{n}_s x_1)\frac{\partial u}{\partial x_1}\right. \\ &\left. + \alpha^2 \cos(\vec{n}_s x_2)\frac{\partial u}{\partial x_2}\right)\Big|_{x = x(s) \in \Gamma^\pm} = \varphi_\pm(x,t). \end{aligned} \tag{2.2}$$

Our assumptions require that the function $u(x,t)$ satisfy the following initial conditions:

$$u(x,0) = u_t(x,0) = 0. \tag{2.3}$$

To select the unique solution to the problem we have to set, as in Ref. 5, the following conditions of regularity at infinity for the function $u(x,t)$:

$$|D_t^k u| \leqslant A_k(t)/|x|, \quad |D_t^k D_{x_j} u| \leqslant \tilde{A}_k(t)/|x|^2, \tag{2.4}$$

for $|x| = (x_1^2 + x_2^2)^{1/2} \to +\infty$, where

$$D_t^k u \equiv \frac{\partial^k}{\partial t^k} u, \quad k = 1,2;$$

$$D_{x_j} u \equiv \frac{\partial}{\partial x_j} u, \quad j = 1,2;$$

and $A_k(t)$ and $\tilde{A}_k(t)$ are continuous non-negative functions of t.

Since the geometry of the field has singular points at the ends of curve Γ, we naturally assume that the function $u(x,t)$ or its gradient may have singularities in the neighborhoods of the end points of curve Γ. We obtain the following conditions in the neighborhoods of the ends of curve Γ (as, for instance, in Ref. 1) by considering more closely the possible character of these singularities. The function $u(x,t)$ and its derivative $u_t(x,t)$ are bounded in the neighborhoods of the end points of curve Γ. Other kinds of derivatives of this function $D_{x_j} u(x,t)$, $D_t^2 D_{x_j} u(x,t)$ behave like

 0022-2488/90/051280-04$03.00

$$|D_{x_j}u(x,t)|,|D_t^2D_{x_j}u(x,t)|\sim O(r_{1,2}^{-1/2}), \tag{2.5}$$

where $j=1,2$ and $r_{1,2}$ is the distance to the ends of curve Γ.

Summarizing, we may exactly define "problem W."

Problem W: Find the continuous function $u(x,t)$ in the space for $t>0$, which has continuous derivative $u_t(x,t)$ and satisfies Eq. (2.1) in the classical sense, in the space $R^2\setminus\Gamma$ with initial conditions (2.3), boundary conditions (2.2) on the sides of curve Γ, and conditions of regularity at infinity (2.4). Moreover, the function $u(x,t)$ must satisfy conditions (2.5) in the neighborhoods of the end points of curve Γ.

III. THE CLASSICAL SOLUTION TO PROBLEM W

To find the classical solution to problem W we need several important helpful results.

Let us give the following definitions. We say that a function $\nu(s)$ belongs to the $C_{1/2}^{(0,h)}(\Gamma)$, which is given on the curve Γ, if the function $d(s)\cdot\nu(s)\in C^{(0,h)}(\Gamma)$, where $d(s)\equiv|x(s)-x(0)|^{1/2}\times|x(s)-x(l)|^{1/2}$, $x(0)=(x_1(0),x_2(0))$, and $x(l)=(x_1(l),x_2(l))$. The points $x(0)$ and $x(l)$ are end points of the curve Γ, and $|x(s)-x(0)|$ and $|x(s)-x(l)|$ are distances from the point $x(s)=(x_1(s),x_2(s))$ to the end points of the curve Γ. We denote the sets of functions

$$C^{(0)}[0,T;C^{(0,h)}(\Gamma)]\equiv\{\mu(s,t)\in C^{(0)}[0,T;C^{(0,h)}(\Gamma)]:\mu(s,0)=\mu_t(s,0)=0\},$$

$$\overline{C}_0^{(2)}[0,T;C_{1/2}^{(0,h)}(\Gamma)]\equiv\{\mu(s,t)\in C^{(2)}[0,T;C_{1/2}^{(0,h)}(\Gamma)]:\int_{-1}^{1}\mu(\sigma,t)d\sigma=0,\ \forall t\in[0,T]\}.$$

Let us consider the dynamic logarithmic potential and the angle potential for Eq. (2.1),[3]

$$V[\mu](x,t)=\int_\Gamma\mu(s,t)\ln|x-y(s)|ds+\int_0^t\int_\Gamma\mu(s,t-\tau)\frac{1}{\tau}\left[1-\cos\left[\frac{|x-y(s)|_*}{|x-y(s)|}\tau\right]\right]ds\,d\tau, \tag{3.1}$$

$$T[\nu](x,t)=\int_\Gamma\nu(s,t)\Psi(x,s)ds-\int_0^t\int_\Gamma\nu(s,t-\tau)\Phi[\Psi(x,s);\tau]ds\,d\tau, \tag{3.2}$$

where $|x|=(x_1^2+x_2^2)$, $|x|_*=(\alpha^2x_1^2+\omega_0^2x_2^2)^{1/2}$, $y=(y_1(s),y_2(s))\in\Gamma$,

$$\Phi(\xi,t)=\int_0^\xi(\omega_0^2\sin^2\theta+\alpha^2\cos^2\theta)^{1/2}\sin\{(\omega_0^2\sin^2\theta+\alpha^2\cos^2\theta)^{1/2}t\}\,d\theta,$$

and $\Psi(x,s)$ is the kernel of the angle potential, which is defined in the following way[7]:

$$\cos\Psi(x,s)=\frac{x_1-y_1(s)}{|x-y(s)|},\quad \sin\Psi(x,s)=[x_2-y_2(s)]/|x-y(s)|.$$

To assure single-valuedness of the function (3.2) we shall require, as in Ref. 3,

$$\int_{-1}^{1}\nu(s,t)ds=0.$$

We assume that the curve $\Gamma\in A^{(1,\lambda)}$, $0<\lambda<1$ (Ref. 8). We may easily prove the following lemma by using results of the theory of dynamic potentials for Eq. (2.1) developed in Ref. 30.

Lemma 1: If $\Gamma\in A^{(1,\lambda)}$, $\nu(s,t),\mu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$, then we have the following.

(1) The potentials $T[\bar\nu](x,t)$ and $V[\bar\mu](x,t)$ satisfy Eq. (2.1) in a field $R^2\setminus\Gamma$, the initial conditions (2.3), the conditions of regularity at infinity (2.4), and the conditions (2.5) in the neighborhood of the end points of the curve Γ, and are continuous in $R^2\setminus\Gamma$ (the potential $V[\bar\mu](x,t)$ is continuous in R^2).

(2) If a point $x(s)$ is not the end point of the curve Γ, then

$$\lim_{x\to x(s)\in\Gamma^\pm}\mathcal{N}_{tx}T[\bar\mu](x,t)=\frac{\partial}{\partial s}\overline{V}[\mu](x,t)$$

$$=\int_\Gamma\mu(\sigma,t)\frac{\sin\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma+(\alpha^2-\omega_0^2)\int_0^t\int_\Gamma\mu(\sigma,t-\tau)\cos\Psi(s,\sigma)$$

$$\times\sin\Psi(s,\sigma)\sin\{\tau[\omega_0^2\sin^2\Psi(s,\sigma)+\alpha^2\cos^2\Psi(s,\sigma)]^{1/2}\}\frac{\cos\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma\,d\tau,$$

where $\overline{V}[\mu](x,t)$ is the value of the potential $V[\mu]$ on the curve Γ, $\theta(s,\sigma)$ is the angle measured counterclockwise between the vectors $\vec n_s$ and $x(s)y(\sigma)$, to the point

$$|x(s)-y(\sigma)|\sin\theta(s,\sigma)=-\tau_s\cdot x(s)y(\sigma),$$

and

$$\bar\mu(s,t)=\int_0^t(t-\tau)\mu(s,\tau)d\tau;\quad\bar\nu(s,t)=\int_0^t(t-\tau)\cdot\nu(s,\tau)d\tau.$$

Later we shall need one more result, which we formulate in a kind of lemma.

Lemma 2: If $\Gamma\in A^{(1,\lambda)}$, $\nu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$, and a point $x(s)$ is not the end point of the curve Γ, then

$$\lim_{x\to x(s)\in\Gamma^{\pm}} \mathcal{N}_{tx} V[\bar{\nu}](x,t) = -\lim_{x\to x(s)\in\Gamma^{\pm}} \frac{\partial}{\partial\tau_s} T[\tilde{\nu}](x,t) = \pm\pi(E-\omega_0 S_{\omega_0 t*})(E-\alpha S_{\alpha t*})\nu(s,t) + D[\tilde{\nu}](s,t),$$

where

$$D[\tilde{\nu}](s,t) = \int_\Gamma \tilde{\nu}(\sigma,t)\frac{\cos\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma - \int_0^t\int_\Gamma \tilde{\nu}(\sigma,t-\tau)\frac{|x(s)-y(\sigma)|_*}{|x(s)-y(\sigma)|}\sin\left(\tau\frac{|x(s)-y(\sigma)|_*}{|x(s)-y(\sigma)|}\right)\frac{\cos\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma\,d\tau,$$

$$\tilde{\nu}(s,t) = \left(\frac{\partial^2}{\partial t^2}+\omega_0^2\right)\left(\frac{\partial^2}{\partial t^2}+\alpha^2\right)\int_0^t (t-\tau)\,\bar{\nu}(s,\tau)d\tau,$$

$(E-\beta S_{\beta t*})$ is the operator defined by expressions

$$S(\beta t) = \int_0^{\beta t}\frac{J_1(\xi)}{\xi}d\xi, \quad (E-\beta S_{\beta t*})\nu(t) = \nu(t) - \beta\int_0^t S(\beta,(t-\tau))*\nu(\tau)d\tau, \tag{3.3}$$

where $J_1(\xi)$ is the Bessel function of first order.

Proof: Let us consider the system of equations

$$\left(\frac{\partial^2}{\partial t^2}+\omega_0^2\right)\mu_{x_1} = -v_{x_2}, \quad \left(\frac{\partial^2}{\partial t^2}+\alpha^2\right)\mu_{x_2} = v_{x_1}. \tag{3.4}$$

This system was used in Ref. 3 for the construction of the dynamic angle potential and plays the same role for Eq. (2.1) as the Cauchy–Riemann system does for the Laplace equation. One can show by direct calculation that the functions

$$v = V[\bar{\nu}](x,t), \quad u = T[\nu_0](x,t),$$

$$\left\{\nu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)], \quad \nu_0(s,t) = \int_0^t (t-\tau)\bar{\nu}(s,\tau)d\tau\right\},$$

satisfy the system of equations (3.4) in $R^2\setminus\Gamma$. It can be shown by using this fact that, for arbitrary $x\in R^2\setminus\Gamma$,

$$\mathcal{N}_{tx} V[\bar{\nu}](x,t) = -\frac{\partial}{\partial\tau_s} T[\tilde{\nu}](x,t).$$

It is important for later consideration that

$$\tilde{\nu}(s,t)\in C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)].$$

We use this and the results of Ref. 3 to obtain the formula

$$\lim_{x\to x(s)\in\Gamma^{\pm}} \mathcal{N}_{tx} V[\bar{\nu}](x,t) = -\lim_{x\to x(s)\in\Gamma^{\pm}} \frac{\partial}{\partial\tau_s} T[\tilde{\nu}](x,t) = \pm\pi(E-\omega_0 J_{\omega_0 t*})(E-\alpha J_{\alpha t*})\tilde{\nu}(s,t) + D[\tilde{\nu}](s,t).$$

One can show by using a Laplace transformation with respect to t that

$$\pm\pi(E-\omega_0 J_{\omega_0 t*})(E-\alpha J_{\alpha t*})\tilde{\nu}(s,t) = \pm\pi(E-\omega_0 S_{\omega_0 t*})(E-\alpha S_{\alpha t*})\nu(s,t),$$

where the operator $(E-\beta S_{\beta t*})$ is defined by formula (3.3). The lemma has thus been proved.

Lemmas 1 and 2 show some symmetry properties of the operators $V(x,t)$ and $T(x,t)$.

Let us make several remarks. From now on, we shall assume that $\Gamma\in A^{(2,\lambda)}$ and the functions $\varphi_{\pm}(s,t)$ in the boundary conditions (2.2) belong to $C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)]$ and satisfy the following condition of the correspondence:

$$\int_\Gamma [\varphi_+(s,t)-\varphi_-(s,t)]ds = 0, \quad \forall t\geqslant 0. \tag{3.5}$$

We shall look for a solution to problem W of the kind

$$u(x,t) = V[\bar{\nu}](x,t) + T[\bar{\mu}](x,t), \tag{3.6}$$

where $\nu(s,t)$, $\mu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$. According to Lemma 1, the function $u(x,t)$ satisfies all the conditions of problem W except the boundary conditions (2.2). We obtain the following system of integral equations for functions μ and ν by using Lemmas 1 and 2:

$$\frac{\partial}{\partial s}\overline{V}[\mu](s,t) + \pi(E-\omega_0 S_{\omega_0 t*})(E-\alpha S_{\alpha t*})\nu(s,t) + D[\tilde{\nu}](s,t) = \varphi_+(s,t), \tag{3.7}$$

$$\frac{\partial}{\partial s}\overline{V}[\mu](s,t) - \pi(E-\omega_0 S_{\omega_0 t*})(E-\alpha S_{\alpha t*})\nu(s,t) + D[\tilde{\nu}](s,t) = \varphi_-(s,t).$$

By adding and subtracting Eqs. (3.7), we may obtain

$$(E-\omega_0 S_{\omega_0 t*})(E-\alpha S_{\alpha t*})\nu(s,t) = (1/2\pi)(\varphi_+(s,t)-\varphi_-(s,t)), \tag{3.8}$$

$$\frac{\partial}{\partial s}\overline{V}[\mu](s,t) = \tfrac{1}{2}[\varphi_+(s,t)+\varphi_-(s,t)] - D[\tilde{\nu}](s,t). \tag{3.9}$$

We can find the explicit solution to Eq. (3.8). It is

$$\nu(s,t) = (1/2\pi)(E-\omega_0 J_{\omega_* t*})(E-\alpha J_{\alpha t*})\times(\varphi_+(s,t)-\varphi_-(s,t)).$$

This solution belongs to $\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,\alpha)}(\Gamma)]$, because the functions $\varphi_{\pm}(s,t)\in C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)]$; therefore the function

$$\tilde{\nu}(s,t)\in C^{(0)}[0,\infty;C^{(0,\alpha)}(\Gamma)].$$

One can show by using the earlier representation of the oper-

ator $D[\tilde{v}]$ and the fact that $\Gamma \in A^{(2,\lambda)}$ that, for arbitrary functions,

$$\eta(s,t) \in C^{(0)}[0,\infty; C^{(0,h)}(\Gamma)],$$

$$D[\eta](s,t) \in C^{(0)}[0,\infty; C^{(0,\lambda)}(\Gamma)].$$

Thus we reduce the problem of classical solvability to the problem of the solvability of Eq. (3.9), which has the right side from $C^{(0)}[0,\infty; C^{(0,\gamma)}(\Gamma)]$, $\gamma = \min[\alpha,\lambda]$, in the set of functions $\overline{C}_0^{(2)}[0,\infty; C_{1/2}^{(0,h)}(\Gamma)]$. This equation was carefully considered in Ref. 4. Therefore we shall not repeat this work, but we shall formulate the final result.

Lemma 3: Equation (3.9) has the unique solution from the set of functions $\overline{C}_0^{(2)}[0,\infty; C_{1/2}^{(0,h)}(\Gamma)]$ for an arbitrary right side chosen from the set of functions $C^{(0)}[0,\infty; C^{(0,\gamma)}(\Gamma)]$.

In summary, as a result of all the lemmas we obtain the following.

Theorem 1: Problem W has the classic solution (3.6), where

$$\nu(s,t) = (1/2\pi)((E - \omega_0 J_{\omega_0 t^*})(E - \alpha J_{\alpha t^*}) \times [\varphi_+(s,t) - \varphi_-(s,t)],$$

and $\mu(s,t)$ is the solution of Eq. (3.9) from the set of functions $\overline{C}_0^{(2)}[0,\infty; C_{1/2}^{(0,h)}(\Gamma)]$ for arbitrary $\varphi_\pm(s,t) \in C^{(0)}[0,\infty; C^{(0,\alpha)}(\Gamma)]$, which satisfy the conditions of correspondence (3.5).

Let us consider the question of uniqueness of the solution (3.6). We may obtain the energetic relation for Eq. (2.1) by the product of Eq. (2.1) and u_t and by carrying out the integration on some compact field D in R^2, which has the smooth boundary $2D$:

$$\frac{\partial}{\partial t}\left\{\frac{1}{2}\|\nabla u_t\|^2_{L_2(D)} + \frac{\omega_0^2}{2}\|u_{x_1}\|^2_{L_2(D)} + \frac{\alpha^2}{2}\|u_{x_2}\|^2_{L_2(D)}\right\} = \int_{\partial D} (\mathcal{N}_{tx} \cdot u) u_t \, d(\partial D),$$

where $\vec{n}$ in the expression for $\mathcal{N}_{tx}$ is the external normal vector to the boundary of the field D.

Following Ref. 4, we obtain the next theorem.

Theorem 2: Solution (3.6) to problem W is the unique solution.

IV. ANALYSIS OF THE RESULTS

We should note that we have considered the general form of the curve Γ. If some scientists or engineers use the results of this paper and consider the specific forms of the curve, then problem W can be solved more easily.

For example, if we replace the curve Γ by the line

$$\Gamma_0 \equiv \{(x_1,x_2): x_1 = s\cos\varphi,\ x_2 = s\sin\varphi,\ -1 \leqslant s \leqslant 1,\},$$
$$\varphi \in [0,\pi/2],$$

where φ is angle between the axis $0x_1$ and the line Γ_0, and then the second equation of the system (3.8) and (3.9) may be written in the following form:

$$\int_{-1}^{1} \frac{\mu(\sigma,t)}{\sigma - s}\, d\sigma = [\varphi_+(s,t) + \varphi_-(s,t)] \cdot \left(-\frac{1}{2}\right). \tag{4.1}$$

This equation was studied in Refs. 1 and 2 and has the explicit unique solution in the set of the functions $\overline{C}_0^{(2)}[0,\infty; C_{1/2}^{(0,h)}(\Gamma)]$,

$$\mu(s,t) = \frac{1}{2\pi}(1-s^2)^{-1/2}\int_{-1}^{1} \frac{(1-\xi^2)^{1/2}}{\xi - s} \times [\varphi_+(s,t) + \varphi_-(s,t)]\, d\xi. \tag{4.2}$$

One can prove the following theorem by using our remark, Theorem 1, and Theorem 2.

Theorem 3: In the case when the curve Γ is replaced by the line Γ_0, problem W has the explicit unique solution defined by the expression

$$u(x,t) = V[\tilde{\nu}](x,t) + T[\tilde{\mu}](x,t), \tag{4.3}$$

where

$$\nu(s,t) = (1/2\pi)(E - \omega_0 J_{\omega_0 t^*})(E - \alpha J_{\alpha t^*}) \times [\varphi_-(s,t) - \varphi_+(s,t)] \cdot (-1)$$

and $\mu(s,t)$ is defined by expression (4.2), for arbitrary $\varphi_\pm(s,t) \in C^{(0)}[0,\infty; C^{(0,k)}(\Gamma)]$.

ACKNOWLEDGMENTS

The authors are very grateful to Professor S. A. Gabov, of Moscow State University, and to Professor H. T. Williams, of Washington and Lee University, for their attention to this work.

[1] S. A. Gabov, Zh. Vychisl. Mat. Mat. Fiz. **25**, 718 (1985).
[2] S. A. Gabov and U. D. Pletner, Zh. Vychisl. Mat. Mat. Fiz. **25**, 1689 (1985).
[3] S. A. Gabov and U. D. Pletner, Zh. Vychisl. Mat. Mat. Fiz. **27**, 102 (1987).
[4] S. A. Gabov and U. D. Pletner, Zh. Vychisl. Mat. Mat. Fiz. **27**, 711 (1987).
[5] U. D. Pletner, Zh. Vychisl. Mat. Mat. Fiz. **28**, 524 (1988).
[6] S. A. Gabov and A. G. Sveshnikov, *Zadachi Dinamiki Stratifitsirovannykh Zhidkostei* (*Problems of the Dynamics of Stratified Fluids*) (Nauka, Moscow, 1986).
[7] S. A. Gabov, Dokl. Akad. Nauk SSSR **278**, 527 (1984).
[8] Carlo Miranda, *Partial Differential Equations of Elliptic Type*, translated by Z. C. Motteler (Springer, New York, 1970).

The Ginzburg–Landau equations for superconducting films and the Meissner effect

Yisong Yang[a)]
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003

(Received 1 September 1988; accepted for publication 10 January 1990)

The existence of the solutions at the lowest energy level to the Ginzburg–Landau equations describing superconducting films under the influence of an external magnetic field is proved. The condition obtained for maintaining superconducting states does not involve the Ginzburg–Landau parameter, and, more surprisingly, in this condition both the external magnetic field and the thickness of the film play equal parts. The asymptotic behavior of these solutions for small or large values of the external field is also studied, which qualitatively verifies the Meissner effect and quantitatively reproduces the solution found from the classical London theory.

I. INTRODUCTION

In the Ginzburg–Landau theory of superconductivity, the free energy density of a superconductor in the absence of a magnetic field is given in normalized units by the function

$$\mathscr{E}_0 = \tfrac{1}{2}|\nabla\phi|^2 + (\lambda/8)(|\phi|^2 - 1)^2, \tag{1}$$

where $\lambda > 0$ is a dimensionless coupling constant with $\lambda < 1$ and $\lambda > 1$ describing the type I and type II superconductors, respectively. The complex scalar field ϕ is an order parameter so that $|\phi|^2$ gives the relative density of the superconducting electron pairs, called the Cooper pairs, which behave like charged bosonic particles, and the normal and pure superconducting states are characterized accordingly by $|\phi|^2 = 0$ and $|\phi|^2 = 1$. To enlarge the global $U(1)$ symmetry of (1) to a local one, a gauge photon field A has to be introduced and the full free energy density is defined by

$$\mathscr{E} = \tfrac{1}{2}|\operatorname{curl} A|^2 + \tfrac{1}{2}|D_A\phi|^2 + (\lambda/8)(|\phi|^2 - 1)^2, \tag{2}$$

where $D_A\phi = \nabla\phi - iA\phi$. In this model the electric field is absent and the magnetic field is determined through $F \equiv \operatorname{curl} A$.

The equations of motion of (2) are

$$D_A^2\phi + (\lambda/2)(1 - |\phi|^2)\phi = 0 \tag{3}$$

and

$$\operatorname{curl}^2 A + (i/2)(\phi^* D_A\phi - \phi(D_A\phi)^*) = 0; \tag{4}$$

these are the famous Ginzburg–Landau equations.

Equations (3) and (4) were first introduced by Ginzburg and Landau[1] in 1950 in their phenomenological approach to superconductivity close to the transition temperature and later deduced by Gorkov[2] theoretically from his formulation of the Bardeen–Cooper–Schrieffer theory.

The Meissner effect in a superconductor cooled below the transition temperature is characteized by a complete or partial expulsion of the magnetic flux from this superconductor when the external magnetic field is weak; however, the normal state will resume when the external magnetic field is sufficiently strong. The following is a physicist's proof of the celebrated Meissner effect.

The energy density of a superconductor in the presence of a constant external magnetic field H_{ext} is given by

$$\begin{aligned}\tilde{\mathscr{E}} &= \mathscr{E} - (\operatorname{curl} A)\cdot H_{\text{ext}} \\ &= (\tfrac{1}{2}|\operatorname{curl} A|^2 - (\operatorname{curl} A)\cdot H_{\text{ext}}) \\ &\quad + \tfrac{1}{2}|D_A\phi|^2 + (\lambda/8)(|\phi|^2 - 1)^2.\end{aligned} \tag{5}$$

From (5), the equations of motion have two apparent solutions:

(a) $\phi = 0, \quad \operatorname{curl} A = H_{\text{ext}}$,

and

(b) $\phi = e^{i\theta_0}, \quad \theta_0 \in \mathbb{R}^1, \quad A = 0.$

For the solution (a), we find the energy density

$$\tilde{\mathscr{E}}_a = -\tfrac{1}{2}(|H_{\text{ext}}|^2 - (\lambda/4)),$$

and for (b), we find $\tilde{\mathscr{E}}_b = 0$. Therefore if $|H_{\text{ext}}|^2 < (\lambda/4)$ then $\tilde{\mathscr{E}}_a > \tilde{\mathscr{E}}_b$, so the solution (b) is energetically favorable and we are in a superconducting state; while if $|H_{\text{ext}}|^2 > (\lambda/4)$ then $\tilde{\mathscr{E}}_a < \tilde{\mathscr{E}}_b$, so the solution (a) is energetically favorable and we are in the normal state. Hence the Meissner effect follows.

There is a serious gap in the above stated proof: if the solutions (a) and (b) are not energetically stable, they are physically unacceptable. On the other hand, however, a mathematically rigorous proof of the Meissner effect has never appeared so far in the literature. The difficulty lies in the fact that the existence result for energy minimizing solutions of the Ginzburg–Landau equations under the influence of an external magnetic field h has not been established if h is not sufficiently small. For example, Carroll and Glick[3] proved an existence and uniqueness theorem under the condition that both λ and h are small, and more recently, Klimov[4] established some existence results for arbitrary λ but with the hypothesis that the external magnetic field is absent.

We believe if the geometry of a superconductor is simple, the above described difficulty may be overcome. Our present study gives the Meissner effect in superconducting films a mathematical proof using the solutions of the Ginzburg–Landau equations at the lowest energy level. First we

a) Current address: Dept. of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131.

 0022-2488/90/051284-06$03.00

show the existence of the energy minimizing solutions and provide a sufficient condition which keeps the solutions found from being "trivial" (or normal). This is an early glance at why the smallness of the external magnetic field h is important to enable the film to maintain a superconducting state. It is surprising to see that the condition obtained for maintaining superconducting states does not involve the parameter λ, and, in this condition both the thickness of the film and the external magnetic field play *equal* parts. Then we prove that when h is small the solutions are close to the pure superconducting state, and when h is large the solutions will approach the normal state: this furnishes a mathematical proof of the Meissner effect. For large and small values of h, asymptotic expressions for the responding magnetic field in the superconducting film are obtained. Furthermore, a time-evolution model for superconducting films will also be considered.

II. THE GINZBURG–LANDAU EQUATIONS FOR SUPERCONDUCTING FILMS

Consider a superconducting film of thickness $2a$ in a parallel constant magnetic field h. In normalized units the Ginzburg–Landau equations are (Burger[5]; Odeh[6]):

$$\begin{aligned} &\phi''(x) = \lambda\phi(x)(\phi^2(x) - 1) + \lambda(hA(x))^2\phi(x), \\ &A''(x) = \phi^2(x)A(x), \quad -a < x < a; \\ &\phi'(\pm a) = 0, \quad A'(\pm a) = 1, \end{aligned} \tag{6}$$

where, now, ϕ and hA are real scalar fields representing the order parameter and the gauge potential, respectively, $\lambda > 0$ is the usual coupling constant characterizing the types of superconductors, and the induced magnetic field in the film is given by $F(x) = hA'(x)$.

Odeh[6] observed that (6) has a trivial solution $\phi = 0$, $A = x + C$ (where C is an arbitrary constant) and that nontrivial solutions can be produced in a neighborhood of the trivial solution through an implicit function theory approach provided h is slightly below some "critical value" h_0. We note that the order parameter ϕ represents the relative density of the Cooper pairs, and so, the solutions near the normal state $\phi = 0, A = x + C$ may not give physically realistic solutions for the superconducting films. The purpose of this section is to find the energy minimizing solutions of (6) without any restriction to the range of h. We shall see, for small h, these solutions are far from the normal state: this is indeed what really happens in a superconductor.

The energy of the field configuration (ϕ, A) in a superconducting film of thickness $2a$ is given by (cf. Burger[5])

$$\begin{aligned} E(\phi,A) = \frac{1}{2}\int_{-a}^{a} \{(\phi')^2 + \lambda h^2([A'-1]^2 \\ + A^2\phi^2) + (\lambda/2)(\phi^2-1)^2\}dx, \end{aligned} \tag{7}$$

where $(\phi, A) \in W^{1,2}(-a,a)$. It is easily checked that the minimizers of the problem

$$m = \min_{(\phi,A)\in W^{1,2}(-a,a)} E(\phi,A) \tag{8}$$

are the solutions to (6). One can observe that the energy (7) is not good enough to work with because the term $A^2\phi^2$ does not enable us to control the $W^{1,2}(-a,a)$ norm of a minimizing sequence of the problem (8). In order to overcome this difficulty, we consider the modified problems

$$m_\epsilon = \min_{(\phi,A)\in W^{1,2}(-a,a)} E_\epsilon(\phi,A), \tag{9}$$

where $\epsilon > 0$ and

$$\begin{aligned} E_\epsilon \equiv \frac{1}{2}\int_{-a}^{a} \{(\phi')^2 + \lambda h^2([A'-1]^2 + A^2[\phi^2+\epsilon]) \\ + (\lambda/2)(\phi^2-1)^2\}dx. \end{aligned} \tag{10}$$

It is immediate to see that the minimum m_ϵ of E_ϵ is attained in the space $W^{1,2}(-a,a)$.

Indeed, let $(\phi_j, A_j) \in W^{1,2}(-a,a)$ be a minimizing sequence of (9). We can assume

$$E_\epsilon(\phi_j, A_j) < m_\epsilon + 1, \quad j = 1,2,\ldots . \tag{11}$$

From (11) and the Schwarz inequality we easily conclude that $\{(\phi_j, A_j)\}$ is a bounded sequence in $W^{1,2}(-a,a)$. For simplicity we may assume there is a point $(\phi_\epsilon, A_\epsilon) \in W^{1,2}(-a,a)$ such that $(\phi_j, A_j) \overset{w}{\to} (\phi_\epsilon, A_\epsilon)$ in $W^{1,2}(-a,a)$. The compact embedding

$$W^{1,2}(-a,a) \to C^0[-a,a] \tag{12}$$

enables us to conclude that

$$E_\epsilon(\phi_\epsilon, A_\epsilon) \leqslant \liminf_{j\to\infty} E_\epsilon(\phi_j, A_j) = m_\epsilon,$$

and, hence, m_ϵ is attained at $(\phi_\epsilon, A_\epsilon)$. We shall show in the sequel that the minimizers of (8) can be found by taking $\epsilon \to 0^+$.

From the standard regularity theory, $(\phi_\epsilon, A_\epsilon) \in C^\infty[-a,a]$ and solves the problem

$$\begin{aligned} &\phi_\epsilon''(x) = \lambda\phi_\epsilon(x)(\phi_\epsilon^2(x) - 1) + \lambda(hA_\epsilon(x))^2\phi_\epsilon(x), \\ &A_\epsilon''(x) = (\phi_\epsilon^2(x) + \epsilon)A_\epsilon(x), \quad -a < x < a; \\ &\phi_\epsilon'(\pm a) = 0, \quad A_\epsilon'(\pm a) = 1. \end{aligned} \tag{13}$$

Because $A_\epsilon'(\pm a) = 1$, there is a point $x_\epsilon \in (-a,a)$ such that $A_\epsilon''(x_\epsilon) = 0$. Hence, from $(13)_2$, we get $A_\epsilon(x_\epsilon) = 0$. Consequently the Schwarz inequality yields the bound

$$\begin{aligned} |A_\epsilon(x)| &\leqslant \left| \int_{x_\epsilon}^{x} A_\epsilon'(\xi)d\xi \right| \\ &\leqslant \sqrt{2a}\left(\int_{-a}^{a} (A_\epsilon'(\xi))^2 d\xi \right)^{1/2} \\ &\leqslant \sqrt{2a}\left[\left(\int_{-a}^{a} (A_\epsilon'(\xi) - 1)^2 d\xi \right)^{1/2} + \sqrt{2a} \right] \\ &\leqslant \sqrt{2a}\left(\sqrt{\frac{2m_\epsilon}{\lambda h^2}} + \sqrt{2a} \right). \end{aligned} \tag{14}$$

Let $\epsilon_1 < \epsilon_2$ be two positive numbers. It is obvious that

$$m_{\epsilon_1} \leqslant m_{\epsilon_2}. \tag{15}$$

Assume now $\{\epsilon_j\}_{j=1}^{\infty}$ is a sequence of positive numbers satisfying

$$\epsilon_1 > \epsilon_2 > \cdots > \epsilon_j > \cdots, \quad \lim_{j\to\infty} \epsilon_j = 0.$$

From (10), (14), and (15) we readily see that

$\{(\phi_{\epsilon_j}, A_{\epsilon_j})\}$ is a bounded sequence in the space $W^{1,2}(-a,a)$. For simplicity we may assume there is a point $(\tilde{\phi},\tilde{A})\in W^{1,2}(-a,a)$ such that

$$(\phi_{\epsilon_j}, A_{\epsilon_j}) \overset{w}{\to} (\tilde{\phi},\tilde{A}) \text{ in } W^{1,2}(-a,a) \text{ as } j\to\infty. \tag{16}$$

In virtue of the embedding (12) we have

$$(\phi_{\epsilon_j}, A_{\epsilon_j}) \overset{s}{\to} (\tilde{\phi},\tilde{A}) \text{ in } C^0[-a,a] \text{ as } j\to\infty. \tag{17}$$

First we claim that $(\tilde{\phi},\tilde{A})$ is a minimizer of the problem (8).

In fact, for any fixed $(\phi,A)\in W^{1,2}(-a,a)$, we have

$$\begin{aligned}
&\frac{1}{2}\int_{-a}^{a}\left\{(\phi'_{\epsilon_j})^2+\frac{\lambda}{2}(\phi_{\epsilon_j}^2-1)^2+\lambda h^2([A'_{\epsilon_j}-1]^2\right.\\
&\qquad\left.+A_{\epsilon_j}^2[\phi_{\epsilon_j}^2+\epsilon_j])\right\}dx\\
&=E_{\epsilon_j}(\phi_{\epsilon_j},A_{\epsilon_j})\leqslant E_{\epsilon_j}(\phi,A)\\
&=\frac{1}{2}\int_{-a}^{a}\left\{(\phi')^2+\frac{\lambda}{2}(\phi^2-1)^2+\lambda h^2([A'-1]^2\right.\\
&\qquad\left.+A^2[\phi^2+\epsilon_j])\right\}dx.
\end{aligned}\tag{18}$$

Letting $j\to\infty$ and using (16) and (17) in the inequality (18), we obtain $E(\tilde{\phi},\tilde{A})\leqslant E(\phi,A)$. This is the desired comparison. Hence $(\pm\tilde{\phi},\tilde{A})$ are smooth solutions of (6) that minimize the energy (7).

Finally, we want to find some condition under which the minimizers of the problem (8) are not the trivial solutions

$$\phi_0=0, \quad A_0=x+C, \quad C=\text{constants}.$$

To this aim, we have only to achieve the inequality

$$E(\phi_0,A_0)>m. \tag{19}$$

From (7), $E(\phi_0,A_0)=\frac{1}{2}\lambda a$. On the other hand, taking $\phi^0=1, A^0=x$, we have $E(\phi^0,A^0)=\frac{1}{3}\lambda a^3h^2$. Therefore if

$$a^2h^2\leqslant\tfrac{3}{2}, \tag{20}$$

then $E(\phi^0,A^0)\leqslant E(\phi_0,A_0)$. But (ϕ^0,A^0) is not a solution to (6), hence it cannot be an extremal point of E. Thus

$$m<E(\phi^0,A^0)=\tfrac{1}{3}\lambda a^3h^2\leqslant E(\phi_0,A_0)=\tfrac{1}{2}\lambda a, \tag{21}$$

and consequently (19) is achieved.

We summarize the above discussion in the following.

Theorem 2.1: The Ginzburg–Landau equations (6) always have energy minimizing solutions for arbitrary $\lambda>0$ and external magnetic field h. Furthermore, if $a^2h^2\leqslant\frac{3}{2}$, then the energy minimizing solutions are nontrivial (or superconducting).

Remarks: (1) It is surprising to notice that the condition (20) does not involve λ and that both a and h play equal roles in maintaining the film in a superconducting state.

(2) For sufficiently small external magnetic field h, we see from (21) that the energy minimizing solutions of (6) can be far away from the normal state (i.e., from the trivial solutions).

III. THE MEISSNER EFFECT

When a superconductor is cooled below the transition temperature T_c and the applied external magnetic field h is weak, the superconductor will be in a superconducting state; but, however, the superconducting state can be quenched and normal state restored by the application of an external magnetic field exceeding in magnitude a certain value h_c, called the critical or threshold magnetic field. This value h_c obeys approximately the formula

$$h_c\approx h_0(1-(T/T_c)^2), \quad T<T_c,$$

and for different superconducting materials h_0 varies in a very wide range.

In this section we will give the above picture a mathematical description using the energy minimizing solutions of the Ginzburg–Landau equations (6) for superconducting films. Roughly speaking, we shall show that as $h\to 0$ the film is in a pure superconducting state while as $h\to\infty$ the film is in the normal state. Note that here it is impossible to find in a superconducting state the complete expulsion of the magnetic flux from the film.

Recall that the normal and pure superconducting states are given respectively by

$$\phi_0=0, \quad A_0=x+C, \tag{22}$$

and

$$\phi_S^{\pm}=\pm 1, \quad A_S^{\pm}=(\sinh x/\cosh a), \tag{23}$$

where $\cosh x=\frac{1}{2}(e^x+e^{-x})$ and $\sinh x=\frac{1}{2}(e^x-e^{-x})$. Note that the solutions (22) and (23) can all be obtained from solving (6) under the assumption $h=0$.

For given h, let (ϕ_h, A_h) be a minimizer of the problem (8).

Lemma 3.1: $\phi_h\to 0$ uniformly on $[-a,a]$ as $|h|\to\infty$.

Proof: Otherwise, we may assume for definiteness that there is a sequence $\{\phi_{h_j}\}$ and an $\epsilon_0>0$ such that

$$h_j\to\infty \text{ as } j\to\infty \text{ but } \|\phi_{h_j}\|_{C^0[-a,a]}\geqslant\epsilon_0. \tag{24}$$

Let $\phi_1=0, A_1=x$. We have

$$E(\phi_{h_j},A_{h_j})\equiv m_j\leqslant E(\phi_1,A_1)=\tfrac{1}{2}\lambda a, \quad j=1,2,\ldots. \tag{25}$$

Inequality (25) yields the boundedness of $\{\phi_{h_j}\}$ in the space $W^{1,2}(-a,a)$. For simplicity we can assume $\phi_{h_j}\overset{w}{\to}\psi$ for some $\psi\in W^{1,2}(-a,a)$. Thus $\phi_{h_j}\overset{s}{\to}\psi$ in $C^0[-a,a]$ in virtue of the compact embedding (12). From (24) we have

$$\|\psi\|_{C^0[-a,a]}\geqslant\epsilon_0. \tag{26}$$

On the other hand, (25) gives us

$$\int_{-a}^{a}\{(A'_{h_j}-1)^2+A_{h_j}^2\phi_{h_j}^2\}dx\leqslant\frac{a}{h_j^2}\to 0, \quad j\to\infty.$$

Hence

$$\int_{-a}^{a}[(A_{h_j}(x)-A_{h_j}(0)-x)']^2\,dx\to 0, \quad j\to\infty,$$

or equivalently,

$$u_j\equiv A_{h_j}-A_{h_j}(0)\overset{s}{\to}x \text{ in } W^{1,2}(-a,a) \text{ as } j\to\infty.$$

Therefore we have

$$\int_{-a}^{a}A_{h_j}^2\phi_{h_j}^2\,dx=\int_{-a}^{a}[u_j+A_{h_j}(0)]^2\phi_{h_j}^2\,dx$$

$$\geqslant \frac{1}{2}\int_{-a}^{a}(x+A_{h_j}(0))^2\phi_{h_j}^2\,dx - \int_{-a}^{a}(u_j-x)^2\phi_{h_j}^2\,dx, \tag{27}$$

where we have used the simple inequality $2(p+q)^2 \geqslant p^2 - 2q^2$.

Letting $j\to\infty$ in (27) we have

$$\lim_{j\to\infty}\int_{-a}^{a}(x+A_{h_j}(0))^2\phi_{h_j}^2\,dx = 0. \tag{28}$$

By passing to a subsequence if necessary, we may assume $A_{h_j}(0) \to$ some point $\eta\in\mathbb{R}^1$ or $A_{h_j}(0)\to\infty$ as $j\to\infty$. Obviously the latter possibility is prohibited by (26) and (28) and so we may assume $A_{h_j}\to\eta$. From (28) we get

$$\int_{-a}^{a}(x+\eta)^2\psi^2\,dx = 0.$$

Hence $\psi\equiv 0$. This contradicts (26).

Lemma 3.1 is proved.

Now that $\phi_h\to 0$ uniformly on $[-a,a]$ as $|h|\to\infty$, we expect to prove $A_h\to x+C$, $C=$ const as $|h|\to\infty$. In order to avoid the ambiguity arising from the arbitrary constant C, let us consider the normalized field

$$\widetilde{A}_h = A_h - \frac{1}{2a}\int_{-a}^{a}A_h(x)dx.$$

Here, $\widetilde{A}_h$ is a translation of A_h with zero mean value. Since

$$\int_{-a}^{a}((\widetilde{A}_h(x)-x)')^2\,dx = \int_{-a}^{a}(A_h'-1)^2\,dx \leqslant \frac{a}{h^2}\to 0 \text{ as } h\to\infty,$$

therefore we observe that $\widetilde{A}_h(x)\to x$ uniformly on $[-a,a]$ as $|h|\to\infty$.

Consequently we have shown that as $|h|\to\infty$ the rescaled field configurations $(\phi_h,\widetilde{A}_h)$ constructed from the energy minimizing solutions (ϕ_h,A_h) approach the normal state $(0,x)$ uniformly as $|h|\to\infty$.

When $h\to 0$ we expect to have $\phi_h^2\to 1$ uniformly on $[-a,a]$. For otherwise, there is an $\epsilon_0>0$ and a sequence $h_j\to 0$ as $j\to\infty$ such that

$$\|\phi_{h_j}^2 - 1\|_{C^0[-a,a]}\geqslant\epsilon_0, \quad j=1,2,\ldots . \tag{29}$$

By the same reasoning as before we can assume $\phi_{h_j}\xrightarrow{s}\psi$ in $C^0[-a,a]$ for some $\psi\in W^{1,2}(-a,a)$.

Let $\phi_2=1$, $A_2=0$. We have

$$\frac{\lambda}{4}\int_{-a}^{a}(\phi_{h_j}^2-1)^2\,dx \leqslant E(\phi_{h_j},A_{h_j}) \leqslant E(\phi_2,A_2) = \lambda a h_j^2\to 0 \text{ as } j\to\infty.$$

So $\psi\equiv 1$. This contradicts (29).

Lemma 3.2: There exist constants h_0, $K(h_0)>0$ such that

$$\|A_h\|_{C^0[-a,a]}\leqslant K(h_0) \text{ for } |h|<h_0.$$

Proof: Since $\phi_h^2\to 1$ uniformly on $[-a,a]$ as $h\to 0$, there is an $h_0>0$ such that

$$\phi_h^2\geqslant(1/2) \text{ whenever } |h|<h_0. \tag{30}$$

Let $u_h = A_h - x$. Then u_h satisfies

$$u_h'' = \phi_h^2 u_h + x\phi_h^2, \quad -a<x<a; \quad u_h'(\pm a)=0.$$

An integration by parts and an application of the Schwarz inequality yield

$$-\int_{-a}^{a}(u_h')^2\,dx = \int_{-a}^{a}\phi_h^2u_h^2dx + \int_{-a}^{a}x\phi_h^2u_h\,dx \geqslant \frac{1}{2}\int_{-a}^{a}\phi_h^2u_h^2\,dx - \frac{1}{2}\int_{-a}^{a}x^2\phi_h^2\,dx,$$

thus in virtue of (30),

$$\int_{-a}^{a}[(u_h')^2+u_h^2]dx\leqslant 4\int_{-a}^{a}x^2\phi_h^2\,dx, \text{ if } |h|<h_0.$$

From the above inequality and the embedding (12) we see that $\{u_h\mid |h|<h_0\}$, and hence $\{A_h\mid |h|<h_0\}$, is a bounded set in $C^0[-a,a]$. This proves Lemma 3.2.

Using Lemma 3.2 we can deduce the behavior of A_h as $h\to 0$. A_h satisfies the boundary value problem

$$A_h''(x) = A_h(x) + (\phi_h^2-1)A_h, \quad -a<x<a,$$
$$A_h'(\pm a) = 1.$$

Let $G(x|\xi)$ be the Green function of the problem

$$U''(x) - U(x) = f(x), \quad -a<x<a,$$
$$U'(\pm a) = 0.$$

Then we easily find

$$A_h(x) = \frac{\sinh x}{\cosh a} + \int_{-a}^{a}G(x|\xi)(\phi_h^2(\xi)-1)A_h(\xi)d\xi. \tag{31}$$

But $\phi_h^2\to 1$ uniformly on $[-a,a]$ as $h\to 0$ and $\{\|A_h\|_{C^0[-a,a]}\mid |h|<h_0\}$ is a bounded set (cf. Lemma 3.2), consequently we reach from (31) the conclusion

$$A_h(x)\to\frac{\sinh x}{\cosh a} \text{ uniformly on } [-a,a] \text{ as } h\to 0.$$

The above study leads to the following mathematical description of the Meissner effect.

Theorem 3.3: For a given external magnetic field h, let (ϕ_h,A_h) be an energy minimizing solution of the Ginzburg–Landau equations (6) and $(\phi_h,\widetilde{A}_h)$ the associated normalized field configuration.

(a) As $|h|\to\infty$, $(\phi_h,\widetilde{A}_h)$ approach the normal state $(\phi_0,A_0)=(0,x)$ uniformly on $[-a,a]$ and, moreover, if $F_h(x)=hA_h'(x)$ is the induced magnetic field in the superconductor, then

$$\frac{F_h}{h} = 1 + o\left(\frac{1}{|h|}\right). \tag{32}$$

(b) As $h\to 0$, (ϕ_h,A_h) approach the pure superconducting states $(\phi_S^{\pm},A_S^{\pm}) = (\pm 1,\sinh x/\cosh a)$ uniformly on $[-a,a]$ and, moreover,

$$\frac{F_h}{h} = \frac{\cosh x}{\cosh a} + \epsilon(h), \text{ where } \epsilon(t)\to 0 \text{ as } t\to 0. \tag{33}$$

Proof: It suffices to verify (32) and (33).

From $(6)_2$ we have

$$|A_h'(x)-1| = \left|\int_{-a}^{x} \phi_h^2(\xi)A_h(\xi)d\xi\right|$$
$$\leqslant \left(\int_{-a}^{a}\phi_h^2\,d\xi\right)^{1/2}\left(\int_{-a}^{a}\phi_h^2 A_h^2\,d\xi\right)^{1/2}$$
$$\leqslant \frac{2a}{|h|}\|\phi_h\|_{C^0[-a,a]}, \quad x\in[-a,a], \tag{34}$$

since $E(\phi_h,A_h)\leqslant E(\phi_1,A_1)$ implies

$$\int_{-a}^{a}\phi_h^2 A_h^2\,d\xi \leqslant \frac{a}{h^2}.$$

Applying Lemma 3.1 to (34) we reach (32).

Finally (33) follows from Lemma 3.2, $\phi_h^2\to 1$ uniformly on $[-a,a]$, and differentiating (31).

This proves Theorem 3.3.

Remarks: (1) Using the implicit function theorem it is not hard to prove that for small h the energy minimizing solutions of (6) in the neighborhoods of $(\phi_S^{\pm}, A_S^{\pm})$ are unique.

(2) (32) says that the induced magnetic field in the superconductor is approximately proportional to the external field h if h is strong. Thus the superconductor now behaves like a normal conductor.

(3) The first-order approximation of F_h in (33) can be obtained by the classical London theory (cf. Tinkham[7]).

(4) Besides the energy minimizing solutions, (6) may have many other solutions. As a simple example, let us consider the limit case $h=0$. The multiple solutions of (6) can all be produced from the solutions of

$$\phi''(x)=\lambda\phi(x)(\phi^2(x)-1), \quad -a\leqslant x\leqslant a,$$
$$\phi'(\pm a)=0.$$

This equation has at least k distinct pairs of solutions expressed explicitly in terms of the Jacobian elliptic functions if $\lambda a^2 > (k^2\pi^2/4)$.

IV. TIME EVOLUTION

The time dependence of the field configurations in a superconducting film can be switched on according to the convention (cf. e.g., Ambegaokar[8])

$$(\phi_t, A_t) = -\delta E(\phi, A),$$

where δ is the Fréchet derivative. Therefore, supplemented with initial data, we have a time-evolution model governed by the equations

$$\begin{aligned}&\phi_t=\phi_{xx}-\lambda\phi(\phi^2-1)-\lambda h^2A^2\phi,\\ &A_t=A_{xx}-\phi^2A, \quad -a<x<a, \quad t>0;\\ &\phi_x(\pm a)=0, \quad A_x(\pm a)=1, \quad t>0,\\ &\phi=\psi_0(x), \quad A=a_0(x), \quad -a<x<a, \quad t=0.\end{aligned} \tag{35}$$

For greater generality, we allow h to be dependent on $t\geqslant 0$ but uniform in x. In order to have a suitable function space setting, we introduce a translation

$$\alpha = A-x, \quad \alpha_0 = a_0 - x.$$

Now (35) is equivalent to

$$\begin{aligned}&\phi_t=\phi_{xx}-\lambda\phi(\phi^2-1)-\lambda h^2(t)(\alpha+x)^2\phi,\\ &\alpha_t=\alpha_{xx}-\phi^2(\alpha+x), \quad -a<x<a, \quad t>0;\\ &\phi_x(\pm a)=\alpha_x(\pm a)=0, \quad t>0,\\ &\phi=\psi_0, \quad \alpha=\alpha_0, \quad -a<x<a, \quad t=0.\end{aligned} \tag{36}$$

Let $H^k=W^{k,2}(-a,a)$, $H^0=L^2(-a,a)$. Then $T: H^0\to H^0$, where

$$\mathrm{Dom}(T)=\{(\phi,\alpha)\in H^2 \mid \phi_x(\pm a)=\alpha_x(\pm a)=0\}$$

and

$$T(\phi,\alpha)=(-\phi_{xx}, -\alpha_{xx})+(\phi,\alpha)$$

is positive and self-adjoint, hence, a sectorial operator with

$$\mathrm{Dom}(T^{1/2})=H^1$$

(cf. Henry[9]). Consequently if $(\psi_0,\alpha_0)\in H^1$, (36) is uniquely solvable in the weak sense over a small time interval $[0,t_0)$ for some $t_0>0$. Standard parabolic regularity argument shows that the obtained solution $(\phi(t,x),\alpha(t,x))$ is indeed a classical solution on $[0,t_0)\times[-a,a]$. To see that the solution exists for all $t>0$, we have only to show that (ϕ,α) is bounded in H^1 for $t>0$.

Lemma 4.1: There is a constant $K>0$ independent of $t>0$ such that

$$\|\phi\|_{C^0[-a,a]}, \quad \|\alpha\|_{C^0[-a,a]}\leqslant K, \quad t>0.$$

Proof: We shall use the maximum principle for parabolic inequalities to infer the desired pointwise estimates.

First, from $(36)_1$, we have

$$(\phi^2)_t=(\phi^2)_{xx}-2\phi_x^2-2\lambda\phi^2(\phi^2-1)-2\lambda h^2A^2\phi^2$$
$$\leqslant(\phi^2)_{xx}-2\lambda\phi^2(\phi^2-M), \tag{37}$$

where $M=\max\{1,\|\psi_0\|^2_{C^0[-a,a]}\}$. The inequality (37) can be rewritten

$$(\phi^2-M)_t\leqslant(\phi^2-M)_{xx}-2\lambda\phi^2(\phi^2-M).$$

Since $\phi^2-M\leqslant 0$ at $t=0$, $-a<x<a$ and $(\phi^2-M)_x=0$ at $x=\pm a$, $t>0$, we can conclude that $\phi^2-M\leqslant 0$ for all t, x.

Secondly, from $(36)_2$ and $-2x\alpha\phi^2\leqslant a^2\phi^2+\alpha^2\phi^2$, we have

$$(\alpha^2)_t=(\alpha^2)_{xx}-2\alpha_x^2-2\phi^2\alpha^2-2x\alpha\phi^2$$
$$\leqslant(\alpha^2)_{xx}-\phi^2(\alpha^2-a^2)$$
$$\leqslant(\alpha^2)_{xx}-\phi^2(\alpha^2-N^2),$$

where $N=\max\{a,\|\alpha_0\|_{C^0[-a,a]}\}$. By the same reasoning as before we reach $\alpha^2-N^2\leqslant 0$. This completes the proof of Lemma 4.1.

Lemma 4.2: If $h(t)$ is bounded, then there is a constant $K>0$ independent of $t>0$ such that

$$\|\phi_x\|_{L^2(-a,a)}, \|\alpha_x\|_{L^2(-a,a)}\leqslant K.$$

Proof: Define

$$J_0=\int_{-a}^{a}\phi^2(t,x)dx, \quad J_1(t)=\int_{-a}^{a}\phi_x^2(t,x)dx.$$

In virtue of Lemma 4.1, (36), and the Schwarz inequality, it is not hard to find

$$J_0'(t)\leqslant -2J_1(t)+C_0, \quad J_1'(t)\leqslant C_1, \quad t>0,$$

where $C_0, C_1>0$ are two suitable constants independent of $t>0$. Therefore we have

$$J_0'(t) + J_1'(t) \leqslant -2J_1(t) + C_0 + C_1$$
$$\leqslant -(J_0(t) + J_1(t)) + C_2, \quad t > 0, \tag{38}$$

for some constant $C_2 > 0$ (cf. Lemma 4.1). The boundedness of $J_0(t) + J_1(t)$ now follows immediately from solving the differential inequality (38). In particular, the desired integral bound for ϕ_x is obtained. The bound for α_x can similarly be inferred. This proves Lemma 4.2.

The above two lemmas establish the global existence and boundedness of the (unique) solution $(\phi(t,x),\alpha(t,x))$ of the initial value problem (36) for bounded $h(t)$ and $(\psi_0, \alpha_0) \in H^1$. This proves the part (a) of the following.

Theorem 4.3: Suppose $h = h(t)$ is a bounded continuous function on $[0,\infty)$.

(a) If $(\psi_0, a_0) \in W^{1,2}(-a,a)$, then (35) has a unique global classical solution. This solution is pointwise bounded.

(b) If $h(t) \in C^1[0,\infty)$ and $|h(t)|$ is decreasing, then along the solution (ϕ,A) of (35), the energy $E(t) \equiv E(\phi,A)$, $t > 0$ is also decreasing. Moreover, if $h(t) = h_0$ for $t \geqslant \tau_0$, then the solution will approach the set of steady state solutions of (35), i.e., the solutions of (6).

Proof: It remains to verify (b).

Using (35) we have

$$\frac{dE(t)}{dt} = -\int_{-a}^{a} (\phi_t^2 + \lambda h^2(t) A_t^2)dx + \lambda h(t)h'(t)$$
$$\times \int_{-a}^{a} [(A_x - 1)^2 + A^2\phi^2]dx. \tag{39}$$

If $|h(t)|$ is decreasing, then $h(t)h'(t) \leqslant 0$. So $E'(t) \leqslant 0$ due to (39). In the case $h(t) \equiv \text{const}$ for $t > \tau_0$, the ω-limit set of the trajectory $\{(\phi,\alpha)(t)\} \equiv \{(\phi,\alpha)\}_{t \geqslant \tau_0}$ ((ϕ,α) is the solution of (36)) in H^1 is invariant since

$$\{S(t): H^1 \to H^1 | t \geqslant \tau_0\},$$
$$S(t)(\psi_0,\alpha_0) \equiv (\phi,\alpha)(t), \quad (\psi_0,\alpha_0) \in H^1$$

is a dynamical system. From this fact and (39) one finds using a standard argument that the ω-limit set of $\{(\phi,\alpha)(t)\}$ is contained in the set of the steady state solutions of (36).

Theorem 4.3 is proved.

ACKNOWLEDGMENT

The author wishes to thank Professor Joel Spruck for helpful conversations.

[1] V. L. Ginzburg, "On the theory of superconductivity," Nuovo Cimento **2**, 1234 (1955).
[2] L. P. Gorkov, "Microscopic derivation of the Ginzburg–Landau equations in the theory of superconductivity," Sov. Phys. JETP **9**, 1364 (1959).
[3] R. W. Carroll and A. J. Glick, "On the Ginzburg–Landau equations," Arch. Rational Mech. Anal. **16**, 373 (1964).
[4] V. S. Klimov, "Nontrivial solutions of the Ginzburg–Landau equations," Theoret. Math. Phys. **50**, 383 (1982).
[5] J. P. Burger, "Superheating and supercooling in first kind superconductors," in *Superconductivity*, Vol. 1, edited by P. A. Wallace (Gordon and Breach, New York, 1969), pp. 463–488.
[6] F. Odeh, "Existence and bifurcation theorems for the Ginzburg–Landau equations," J. Math. Phys. **8**, 2351 (1967).
[7] M. Tinkham, *Superconductivity* (Gordon and Breach, New York, 1965).
[8] V. Ambegaokar, "Critical currents and the onset of resistance in superconductors," in *Superconductivity*, Vol. 1, edited by P. A. Wallace (Gordon and Breach, New York, 1969), pp. 117–158.
[9] D. Henry "Geometric theory of semilinear parabolic equations," *Lecture Notes in Mathematics*, Vol. 840 (Springer, Berlin, 1981).