An angular momentum analysis of symmetric products of paired nucleon states

J. A. Sheikh, J. Sita, and C. R. Sarma
Department of Physics, Indian Institute of Technology, Bombay 400 076, India

(Received 27 May 1986; accepted for publication 29 October 1986)

Pair correlated angular momentum projected standard Weyl tableau states spanning an m-dimensional paired shell-model space have been obtained for a system of $2N$ identical nucleons. A simple procedure has been developed for carrying out the restriction of the unitary group $U(m)$ to the rotation group $O(3)$ for configurations of both single and multilevel distributions of the single particle states. The mapping of the correlated pair configuration space onto the antisymmetric states allowed by Pauli principle is achieved using the particle antisymmetrizer. The procedure for determining the shell-model Hamiltonian matrix using the above basis is outlined.

I. INTRODUCTION

Extensive studies have been undertaken over the past few years regarding correlations between the phenomenological interacting boson models (IBM-1,2) and the shell-model approaches.[1-4] One of the most useful of such correlations relates to the use of pair-fermion states such as those which occur in broken pair approximation (BPA) methods.[5,6] In these approaches correlated pairs of low angular momentum coupled correlated fermion states ($J = 0,2$) are sought to be related to the s and d bosons of IBM-1 in case of identical even nucleon systems. In most cases Dyson mapping approaches are used to obtain this correlation.[7,8] It was shown in a recent paper[9] that the action of a Dyson mapped Hamiltonian on Dyson mapped boson pair states was equivalent (operationally) to the action of a shell-model Hamiltonian on determinantal products of single particle states. This equivalence was established using the shell-model $|j,m\rangle$ basis. The main feature of IBM and BPA studies is the truncation scheme on the shell-model configuration space based on the use of low angular momentum coupled pair states to generate a physically significant subspace. In this context, the equivalence established using the $|j,m\rangle$ basis[9] is not of much utility since a consistent truncation scheme is not possible in this space.

In the present paper we have attempted to develop a viable procedure for realizing bosonlike products of angular momentum coupled pair states of identical nucleons. A simple method is suggested to obtain good angular momentum states of even nucleon systems. A truncation of the space leads us to consider pair states with $J = 0,2$. A procedure for antisymmetrizing these product states to satisfy the Pauli principle is also proposed. These methods are outlined in Sec. II and illustrated using suitable examples. A brief discussion of the method is presented in Sec. III.

II. ANGULAR MOMENTUM PROJECTED PAIR STATES

Consider an ordered orthonormal set of single particle shell-model basis states spanning a linear vector space V_n,

$$V_n:\left\{|j_a m_a\rangle \,\middle|\, a = 1,...,k;\ -j_a \leqslant m_a \leqslant j_a;\ \sum_{a=1}^{k}(2j_a+1) = n\right\} \tag{1}$$

for a system of $2N$ identical nucleons. In the above set it is assumed that $a < b$ implies $j_a > j_b$ or, if $j_a = j_b$, $m_a > m_b$ and in any product the state $|j_a, m_a\rangle$ precedes $|j_b, m_b\rangle$ as read from left to right. V_n provides the carrier space for the fundamental representation of the unitary group $U(n)$ with its Lie algebra defined by a set of n^2 generators E_{ab} $(a,b = 1,2,...,n)$ satisfying

$$[E_{ab}, E_{cd}] = E_{ad}\delta_{bc} - E_{cb}\delta_{ad} \tag{2}$$

and

$$E^{\dagger}_{ab} = E_{ba}. \tag{3}$$

Using the generators defined by Eqs. (2) and (3), it is possible to express the shell-model Hamiltonian as[10]

$$H = \sum_a \epsilon_a E_{aa} + \frac{1}{2}\sum_{a<b}\sum_{c<d} V_{(ab)(cd)}\,(E_{ac}E_{bd} - \delta_{bc}E_{ad}), \tag{4}$$

where

$$V_{(ab)(cd)} = \tfrac{1}{2}\,(V_{ab,cd} - V_{ab,dc} - V_{ba,cd} + V_{ba,dc}), \tag{5}$$

with

$$V_{ab,cd} = \langle j_a m_a\, j_b m_b | V(1,2) | j_c m_c\, j_d m_d\rangle. \tag{6}$$

For a system of $2N$ identical particles, states allowed by the Pauli principle are the determinantal states spanning the representation $[1^{2N}0]$ of the group $U(n)$,

$$(j_1 m_1 \cdots j_{2N} m_{2N}) \equiv \mathscr{A}\,(|j_1 m_1\rangle \cdots |j_{2N} m_{2N}\rangle), \tag{7}$$

where $\mathscr{A}$ is the particle antisymmetrizer,

$$\mathscr{A} = (2N!)^{-1/2}\sum_{P\in S_{2N}} \delta_P P. \tag{8}$$

There are only three distinct categories of nonzero matrix elements of Eq. (4) over the basis defined by Eq. (7). These are listed for convenience in the Appendix. The usual problem with this approach is that the dimensionality of the space is too large and no consistent scheme of truncation is available on the configuration space.

An alternative to the shell-model approach is to define a space of antisymmetric second rank tensors realized on $V_n \otimes^2$ as

$$V_m:\{\phi_{ab} = -\phi_{ba} = (1/\sqrt{2})(|j_a m_a\rangle|j_b m_b\rangle - |j_b m_b\rangle|j_a m_a\rangle)\,|\,a<b = 1,...,n\}, \tag{9}$$

 0022-2488/87/040751-07$02.50

where V_m is the $[m=n(n-1)/2]$-dimensional space spanned by standard Weyl tableaux (SWT) of the representation $[1^2\dot{0}]$ of $U(n)$. The single particle orbitals defining ϕ_{ab} are assumed ordered in the usual SWT scheme. It then follows that

$$\langle\phi_{ab}|\phi_{cd}\rangle=\delta_{ac}\delta_{bd}, \tag{10}$$

since $a<b$ and $c<d$. For a system of N pairs of identical nucleons, the procedures outlined in an earlier paper[9] lead to Nth rank antisymmetric product states,

$$[(a_1b_1)(a_2b_2)\cdots(a_Nb_N)]=\left(\frac{2^NN!}{(2N)!}\right)^{1/2}\sum_{P\in S_{2N}/S_N\otimes(S_2)^N}\delta_pP(\phi_{a_1b_1}\cdots\phi_{a_Nb_N}), \tag{11}$$

where $a_k=|j_{a_k}m_{a_k}\rangle$, $b_k=|j_{b_k}m_{b_k}\rangle$ and $a_k\neq a_l$, $b_k\neq b_l$ if $k\neq l$. In the above product, for a fixed index a_k the index b_k ranges over $a_k+1\leqslant b_k\leqslant n-2N+2$ so that the number of independent configurations as in Eq. (7) for this value of a_k is $\binom{n-k-1}{2N-2}$ since no index is to occur more than once in a product. The total number of configurations is thus

$$\sum_{k=1}^{n-2n+1}k\binom{n-k-1}{2N-2}=\binom{n}{2N}. \tag{12}$$

Thus the correspondence between configurations of pair states of Eq. (9) and the standard shell-model configuration space $[1^{2N}\dot{0}]$ is one-to-one. That the correspondence is isometric follows using the steps outlined in the earlier paper[9] as

$$[(a_1b_1)(a_2b_2)\cdots(a_Nb_N)]=\mathscr{A}(|a_1\rangle|b_1\rangle|a_2\rangle|b_2\rangle\cdots|a_N\rangle|b_N\rangle). \tag{13}$$

Thus it is quite apparent at this stage that nothing is to be gained using the pair basis states of Eq. (9) in preference to the shell-model basis. No simple and consistent prescription is evident at this stage for truncating the configuration space. This leads us to define an alternative basis set on V_m in terms of angular momentum coupled pair states as

$$\phi_{\mu jm}=\sum_{m_a,m_b}\langle j_am_a\,j_bm_b|\mu jm\rangle\phi_{ab}, \tag{14}$$

where $\mu=(j_a,j_b)$ and $\langle j_am_a,j_bm_b|\mu jm\rangle$ are Clebsch–Gordan coefficients (CGC) of O(3). If $j_a=j_b$, a slight modification is to be made on the right-hand side of Eq. (14) with

$$\sum_{m_a,m_b}\to\sqrt{2}\sum_{m_a,m_b(m_a>m_b)}.$$

This occurs since the ϕ_{ab} are SWT basis states. We note that just as the ϕ_{ab} provide a basis on V_m, the $\phi_{\mu jm}$ lead to an alternative basis on the space. Based on the earlier analysis,[9] we note that V_m is the fundamental representation space of a unitary group $U(m)$ whose Lie algebra is defined by a set of m^2 generators $E_{(ab)(cd)}$ $(a<b,\ c<d=1,\ldots,n)$. The set of states

$$\{\phi_{\mu jm}|\mu\equiv(j_a,j_b);\ |j_a-j_b|\leqslant j\leqslant(j_a+j_b);\ -j\leqslant m\leqslant j\}$$

defines an equivalent orthonormal basis on V_m,

$$\langle\phi_{\mu jm}|\phi_{\mu'j'm'}\rangle=\delta_{\mu\mu'}\delta_{jj'}\delta_{mm'}, \tag{15}$$

and can be used to provide a representation of $U(m)$. The unitary transformations of Eq. (14) enable us to define an equivalent set of shift operators on V_m, $E_{\mu jm;\mu'j'm'}$, which are generators of the Lie algebra of $U(m)$ satisfying

$$[E_{\mu jm;\mu'j'm'},\ E_{\mu''j''m'';\mu'''j'''m'''}]=E_{\mu jm;\mu'''j'''m'''}\delta_{\mu'\mu''}\delta_{j'j''}\delta_{m'm''}-E_{\mu''j''m'';\mu'j'm'}\delta_{\mu\mu'''}\delta_{jj'''}\delta_{mm'''}, \tag{16}$$

and

$$E^{+}_{\mu jm;\ \mu'j'm'}=E_{\mu'j'm';\mu jm}, \tag{17}$$

where

$$E_{\mu jm;\mu'j'm'}=\sum_{\substack{m_a,m_b\\ m'_a,m'_b}}\langle j_am_a\,j_bm_b|\mu jm\rangle\times\langle\mu'j'm'|j'_am'_a\,j'_bm'_b\rangle E_{(\mu m_am_b);(\mu'm'_am'_b)}. \tag{18}$$

We now introduce an ordering in the index set $\phi_{\mu jm}$ by assuming that in any tensor product, the states with j (μ ranging) precede those with j' (μ' ranging) if $j>j'$. If $j=j'$, states of a given m precede those with an m' if $m>m'$ as read from left to right. Among the states in a product with a given j, those having $\mu=(j_a,j_b)$ precede those with $\mu'=(j'_a,j'_b)$ if $j_a>j'_a$ or if $j_a=j'_a$ provided $j_b>j'_b$. Given such an ordering, we now consider a set of primitive Nth rank tensor products of $\phi_{\mu jm}$ spanning $V_m^{\otimes N}$ as

$$V_m^{\otimes N}:\left\{(\phi_{\mu_1j_1m_1})^{N_1}\cdots(\phi_{\mu_mj_mm_m})^{N_m}\,\middle|\,N\geqslant N_1,N_2,\ldots,N_m\geqslant0;\ \sum_{a=1}^{m}N_a=N\right\}, \tag{19}$$

where the first N_1 pair states are associated with particle pair indices (1,2), (3,4),..., $(2N_1-1,2N_1)$, the next N_2 pairs are associated with the indices $(2N_1+1,\ 2N_1+2)$,..., $(2N_1+2N_2-1,2N_1+2N_2)$ and so on. The tensor basis set spanning $V_m^{\otimes N}$ provides an alternate Nth rank basis for $U(n)^{\otimes N}$. A reduction of this space into irreducible subspaces of $U(m)$ stable under the generators defined by Eqs. (16)–(18) follows readily on using symmetrized Wigner operators of the appropriate permutation group as outlined in earlier studies.[11] The permutation group of relevance in the present context is S_N defined on pairs of particles. Since we are interested in bosonlike products of pair states, $\phi_{\mu jm}$, we need only consider the symmetric representation $[N,\dot{0}]$ of $U(m)$, we can readily generate the SWT spanning this representation as

$$[(\phi_{\mu_1j_1m_1})^{N_1}(\phi_{\mu_2j_2m_2})^{N_2}\cdots(\phi_{\mu_mj_mm_m})^{N_m}]=\left[N!\prod_{a=1}^{m}N_a!\right]^{1/2}\times\sum_{P\in S_N}P[(\phi_{\mu_1j_1m_1})^{N_1}(\phi_{\mu_2j_2m_2})^{N_2}\cdots(\phi_{\mu_mj_mm_m})^{N_m}], \tag{20}$$

where each $P\in S_N$ is a permutation defined over particle pair indices $(2a-1,2a)$ with $a=1,\ldots,N$.

Good angular momentum projected states for the N-pair system result, on using the restriction $U(m)\downarrow O(3)$, in the rotation subgroup of the unitary group. Instead of using the recoupling algebra for this purpose[12] we will use the more direct procedure based on subducing the algebra of O(3) from the parent group. For any such restriction, the

procedure is straightforward[10,13] and follows the chain $U(m)\supset O(m)\supset O(3)$ for integer angular momenta and leads to

$$J_+ = \sum_{\mu j}\sum_{m=0}^{j-1}[(j-m)(j+m+1)]^{1/2} \times (E_{\mu jm+1;\mu jm} - E_{\mu j-m;\mu j-m-1}), \quad (21)$$

$$J_- = \sum_{\mu j}\sum_{m=0}^{j-1}[(j-m)(j+m+1)]^{1/2} \times (E_{\mu jm;\mu jm+1}E_{\mu j-m-1;\mu j-m}), \quad (22)$$

$$j_0 = \sum_{\mu j}\sum_{m=0}^{j} m(E_{\mu jm;\mu jm} - E_{\mu j-m;\mu j-m}). \quad (23)$$

Using the commutation relations of Eq. (16), it is easy to verify that $J_\pm$ and J_0 satisfy the usual angular momentum commutation relations. The standard procedure for the subgroup adaptation $U(m)\downarrow O(3)$ follows on starting with a set of SWT spanning $[1^N\dot{0}]$ corresponding to a given total M value and defining a linear combination of these states. If this linear combination corresponds to a total J value, $J=M$, then application of J_+ of Eq. (21) to it annihilates it. If the given J is multiplicity-free in $U(m)\downarrow O(3)$, we obtain a sufficient set of equations to determine the state uniquely to within a normalization factor. If multiplicity is present, the linear combination is resolved in an arbitrary but reasonably consistent manner. Instead of carrying through this scheme for general J and M, we consider briefly the truncation schemes as in IBM-1[1,2] or correlated BPA approach.[4–6] A preliminary step in this type of study is to define correlated pair functions as

$$\phi_{;jm} = \sum_{\mu} A_{\mu jm}\phi_{\mu jm}, \quad (24)$$

where the $A_{\mu jm}$ are variation coefficients which are as yet undetermined. The truncation scheme consists now of restricting ourselves to relatively low angular momentum states among those defined by Eq. (24). In the simplest version of this scheme, it is assumed that only correlated pair states with $j=0,2$ are important in the N-pair configuration space. This and the result of Eq. (24), in turn, imply that only pair states $\phi_{\mu jm}$ with $j=0,2$ need be considered in generating the configuration space of $[1^N\dot{0}]$ spanned by SWT of Eq. (20). This leads to a subspace of V_m of considerably reduced dimensionality than $n(n-1)/2$. Thus for the subshells $j=\frac{7}{2},\frac{5}{2},\frac{3}{2},\frac{1}{2}$ we find that V_m is of 190 dimensions whereas the truncated subspace as above is of 44 dimensions.

Equation (24) implies that the configuration space of correlated pair states is a linear combination of SWT defined by Eq. (20) with individual j_a $(a=1,...,m)$ being restricted to values 0 or 2 insofar as the subspace is concerned. In dealing with this restricted configuration space we first assume that all the pairs of particles are distributed in just a pair of single particle levels characterized by a fixed μ value. If we further assume that $j_a=0$ for all N pairs of a given configuration, the only angular momentum projected state possible is $J=M=0$, so that

$$|[(\phi_{\mu 00})^N];\ 0,\ 0> = [(\phi_{\mu 00})^N], \quad (25)$$

where the right-hand side of Eq. (25) is an SWT corresponding to the Nth rank tensor product of $\phi_{\mu 00}$. As a generalization of this result, consider again a fixed μ value and $j_1=j_2=\cdots=j_p=0, j_{p+1}=j_{p+2}=\cdots=j_{p+q}=2$, where $p+q=N$. The highest weight SWT corresponding to this distribution is one with $m_1=m_2=\cdots=m_p=0$, $m_{p+1}=m_{p+2}=\cdots=m_{p+q}=2$. It is also the highest weight state of $O(3)$ corresponding to $J=M=2q$. For notational convenience we represent this SWT as

$$(p\,q\,0\,0\,0\,0) \equiv [(\phi_{\mu 00})^p(\phi_{\mu 22})^q], \quad (26)$$

where the left-hand side shows that $m=0$ is p-fold occupied, $m=2$ is q-fold occupied, and $m=1,0,-1,-2$ are all unoccupied with these values being read from left to right, respectively. SWT's of lower M value follow successively from Eq. (26) by decreasing by unity the entries starting with q and increasing by the same amount to the immediate right position. Thus, for example,

$N=2q-1$: $(p\,q-1\,1\,0\,0\,0)$,

$M=2q-2$: $(p\,q-1\,0\,1\,0\,0)$, $(p\,q-2\,2\,0\,0\,0)$,

etc. A simple procedure has been worked out for listing all the SWT corresponding to a given M value in dictionary ordering for any given p and q. As an illustration let $p=3$ and $q=4$. The SWT's corresponding to this distribution are

$(3\,3\,0\,0\,0\,1)$, $(3\,2\,1\,0\,1\,0)$, $(3\,2\,0\,2\,0\,0)$, $(3\,1\,2\,1\,0\,0)$, $(3\,0\,4\,0\,0\,0)$.

For a given value of M, let S represent the set of SWT's where

$$S:\ \left\{(pq)\equiv(p\,q_1\,q_2\,q_3\,q_4\,q_5)\ \middle|\ q_i\geqslant 0\ \ \forall i;\ \sum_{i=1}^{5} q_i = q;\ \sum_{i=1}^{5} m_i q_i = M;\ m_1=2,...,m_5=-2\right\}. \quad (27)$$

Using this set, we define a linear combination as

$$|[(\mu 0)^p(\mu 2)^q];\ M> = \sum_{(pq)\in S} A_{(pq)}|(pq)>, \quad (28)$$

where $A_{(pq)}$ are as yet undetermined coefficients. Applying J_+ of Eq. (21) to both sides of Eq. (28) and equating the coefficients of resulting linearly independent SWT, $(pq)=(p\,q_1'q_2'q_3'q_4'q_5')$, to zero we can determine the state corresponding to J, $M=J$. As an illustration of this procedure, consider $p=3$, $q=4$, and $J=M=4$. The SWT corresponding to this case have been listed earlier. Using these and setting up the linear combination as in Eq. (28), applying J_+ to this combination, and equating the result to zero, the following relations among the coefficients are obtained;

$$A_2=(1/\sqrt{3})A_1,$$

$$A_4=(1/2\sqrt{2})A_1-(\sqrt{3}/2)A_3,$$

$$A_5=-(3/8)A_1+(3\sqrt{3}/4\sqrt{2})A_3,$$

where the basis states listed earlier have been numbered 1 to 5 in the dictionary ordering. Choosing $A_5=0$ leads to

$$A_2=(1/\sqrt{3})A_1,\quad A_3=(1/\sqrt{6})A_1,\quad A_4=0,$$

so that

$$|[(\mu 0)^3(\mu 2)^4];\ 44\rangle_{(1)}$$
$$= \tfrac{1}{3}[\sqrt{6}(3\,3\,0\,0\,0\,1) + \sqrt{2}(3\,2\,1\,0\,1\,0) + (3\,2\,0\,2\,0\,0)]. \quad (29)$$

The second of this pair of doubly occurring states results on orthogonalizing the general linear combination to this state, yielding

$$\sqrt{6}A_1 + \sqrt{2}A_2 + A_3 = 0,$$

so that the second normalized state is

$$|[(\mu 0)^3(\mu 2)^4];\ 44\rangle_{(2)}$$
$$= (1/3\sqrt{57})[8\sqrt{3}(3\,3\,0\,0\,0\,1) + 8(3\,2\,1\,0\,1\,0) - 32\sqrt{2}(3\,2\,0\,2\,0\,0) - 18\sqrt{2}(3\,1\,2\,1\,0\,0) - 27\sqrt{3}(3\,0\,4\,0\,0\,0)]. \quad (30)$$

The analysis of fixed μ-angular momentum projected states enables us to generalize to states which have distributions spread over various μ values. Let the single particle shell model basis span the set of levels $j_1, j_2, \ldots j_k$. This leads to a set of $k(k+1)/2$ pair state levels $(j_a j_b)\, j_a \geq j_b$. This generalization to multilevel distribution follows readily on defining a shift operator,

$$E_{\mu' j';\mu'' j''} = \sum_{m'=-j'}^{j'} E_{\mu' j' m';\mu'' j'' m''}, \quad (31)$$

where only those j' are considered that satisfy $(j'_a - j'_b) \leq j' \leq j'_a + j'_b$ for a given μ' and similarly for μ''. It can be readily verified that $E_{\mu' j';\mu'' j''}$ commutes with $J_\pm$ and J_0 of Eqs. (21)–(23),

$$[E_{\mu' j';\mu'' j''}, J_\xi] = 0, \quad (32)$$

where $\xi = \pm, 0$. The utility of the operator defined by Eq. (31) follows if we consider the fixed μ configuration of Eq. (28). Applying J_+ to it and equating the result to zero leads to a determination of the coefficients $A_{(pq)}$ so that we obtain the angular momentum projected state as

$$|[(\mu 0)^p(\mu 2)^q];\ JJ\rangle$$
$$= \sum_{(pq)\in S} A^{(J)}_{(pq)} \times [(\phi_{\mu 00})^p (\phi_{\mu 22})^{q_1} (\phi_{\mu 21})^{q_2} \cdots (\phi_{\mu 2-2})^{q_5}]. \quad (33)$$

The procedure used to determine $A^{(J)}_{(pq)}$ is similar to the one outlined in the example. By construction we therefore have the result

$$J_+ |[(\mu 0)^p(\mu 2)^q];\ JJ\rangle = 0. \quad (34)$$

Using the commutation relation of Eq. (32), we note that

$$J_+ E_{\mu' j';\mu j'} |[(\mu 0)^p(\mu 2)^q];\ JJ\rangle = E_{\mu' j';\mu j'} J_+ |[(\mu 0)^p(\mu 2)^q];\ JJ\rangle = 0. \quad (35)$$

This result implies that the application of $E_{\mu' 0;\mu 0}$ and $E_{\mu' 2;\mu 2}$ to Eq. (33) leads to the states

$$|[(\mu 0)^{p-1}(\mu' 0)(\mu 2)^q];\ JJ\rangle$$

and

$$|[(\mu 0)^p(\mu 2)^{q-1}(\mu' 2)];\ JJ\rangle,$$

respectively. As illustrations, the corresponding states are listed below:

$$|[(\mu 0)^{p-1}(\mu' 0)(\mu 2)^q];\ JJ\rangle$$
$$= \sum_{(pq)\in S} A^{(J)}_{(pq)} [(\phi_{\mu 00})^{p-1} \phi_{\mu' 00} (\phi_{\mu 22})^{q_1} \cdots (\phi_{\mu 2-2})^{q_5}] \quad (36)$$

and

$$|[(\mu 0)^p(\mu 2)^{q-1}(\mu' 2)];\ JJ\rangle$$
$$= \frac{1}{\sqrt{q}} \sum_{(pq)\in S} A^{J}_{(pq)} \sum_{i=1}^{5} \sqrt{q_i} [(\phi_{\mu 00})^p (\phi_{\mu 2 m_i})^{q_i - 1} \phi_{\mu' 2 m_i} \times \prod_{\substack{k=1 \\ (k\neq i)}}^{5} (\phi_{\mu 2 m_k})^{q_k}, \quad (37)$$

where $m_i, m_k = 2,1,0,-1,-2$, for $i,k = 1,2,3,4,5$, respectively. Repeating this process for different $E_{\mu' j';\mu'' j''}$ ($j' = 0,2$), the entire multishell configuration space adapted to a total J, $M = J$ can be generated successively if $A^{J}_{(pq)}$ are known for a single shell distribution. This procedure thus enables us to obtain all the configurations of an N-pair system occurring in a given correlated pair configuration distributed over s- and d-type correlated pair states of Eq. (24).

The procedure developed above is not complete since the effect of the Pauli principle on the projected configurations has not been considered. Since the Hamiltonian operator of Eq. (4) is defined in terms shift operators over V_n, we have found it convenient to transform the angular momentum projected basis states defined as polynomials of $\phi_{\mu j m}$ into the uncoupled $(j_a m_a, j_b m_b)$ form. The effect of the particle antisymmetrizer on these product states is to annihilate any product state of the $|j_a m_a\rangle$ basis in which any index is repeated and introduce a phase factor δ_p in obtaining the SWT spanning $[1^{2N}\dot{0}]$ of U(n). Collecting the coefficients of independent SWT of this type occurring in angular momentum projected configuration space basis states, we incorporate the Pauli principle in the formulation. As an illustration of this procedure we present the case of a three-pair system in the single $j = \frac{7}{2}$ subshell. Here $\mu \equiv 1 = (\frac{7}{2}, \frac{7}{2})$ and let us consider the projected state corresponding to $J = M = 0$. Let all pairs have $j_a = 2$ so that the relevant pair states are

$$\begin{aligned}
\phi_{122} &= (1/\sqrt{42})[\sqrt{7}\phi_{16} - \sqrt{15}\phi_{25} + 2\sqrt{5}\phi_{34}], \\
\phi_{121} &= (1/\sqrt{42})[\sqrt{21}\phi_{17} - 4\phi_{26} + \sqrt{5}\phi_{35}], \\
\phi_{120} &= (1/\sqrt{84})[7\phi_{18} - \phi_{27} - 3\phi_{36} + 5\phi_{45}], \\
\phi_{12-1} &= (-1/\sqrt{42})[\sqrt{21}\phi_{28} - 4\phi_{37} + \sqrt{5}\phi_{46}], \\
\phi_{12-2} &= (1/\sqrt{42})[\sqrt{7}\phi_{38} - \sqrt{15}\phi_{47} + 2\sqrt{5}\phi_{56}],
\end{aligned} \quad (38)$$

where on the right-hand side in single particle labels $i = 1,\ldots,8$ have been used to represent $m_a = \frac{7}{2}, \ldots, -\frac{7}{2}$. Using these in the projected state,

$$|[(\phi_{12})^3];\ 0,0\rangle$$
$$= (1/\sqrt{35})2\sqrt{3}[\phi_{122}\phi_{120}\phi_{12-2}] - 3[\phi_{122}(\phi_{12-1})^2] - 3[(\phi_{121})^2\phi_{12-2}] - \sqrt{3}[\phi_{121}\phi_{120}\phi_{12-1}] - \sqrt{2}[(\phi_{120})^3], \quad (39)$$

and antisymmetrizing and renormalizing the state we obtain

the result

$$\mathscr{A}|[(\phi_{12})^3];\ 0,0\rangle = \tfrac{1}{2}\mathscr{A}\{-(1\,2\,3\,6\,7\,8)+(1\,2\,4\,5\,7\,8) -(1\,3\,4\,5\,6\,8)+(2\,3\,4\,5\,6\,7)\}, \tag{40}$$

where the single particle orbitals have been reordered to yield the respective SWT's. If $j_b = \frac{5}{2}$ is another level available for the system, let $i = 9,10,11,12,13,14$ label its $m_b = \frac{5}{2}, \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}, -\frac{5}{2}$ values, respectively. Representing $(\frac{7}{2},\frac{5}{2})$ by $\mu = 2$, the application of $E_{22;12}$ to both Eq. (39) yields

$$\begin{aligned}
&|[(\phi_{12})^2\phi_{22}];\ 0,0\rangle \\
&\quad = (1/\sqrt{35})\sqrt{2}[\phi_{122}\phi_{120}\phi_{22-2}] + 2[\phi_{122}\phi_{220}\phi_{12-2}] + 2[\phi_{222}\phi_{120}\phi_{12-2}] - \sqrt{6}[\phi_{122}\phi_{12-1}\phi_{22-1}] \\
&\qquad - \sqrt{3}[\phi_{222}(\phi_{12-1})^2] - \sqrt{3}[(\phi_{121})^2\phi_{22-2}] - \sqrt{6}[\phi_{221}\phi_{121}\phi_{12-2}] - [\phi_{121}\phi_{120}\phi_{22-1}] \\
&\qquad - [\phi_{121}\phi_{220}\phi_{12-1}] - [\phi_{221}\phi_{120}\phi_{12-1}] - \sqrt{2}[(\phi_{120})^2\phi_{220}].
\end{aligned} \tag{41}$$

The additional pair states required for the above state are

$$\begin{aligned}
\phi_{222} &= (1/3\sqrt{14})[\sqrt{35}\phi_{1\,13} - 2\sqrt{10}\phi_{2\,12} + \sqrt{30}\phi_{3\,11} - 4\phi_{4\,10} + \sqrt{5}\phi_{59}], \\
\phi_{221} &= (1/6\sqrt{14})[5\sqrt{7}\phi_{1\,14} - \sqrt{5}\phi_{2\,13} - \sqrt{30}\phi_{3\,12} + 7\sqrt{2}\phi_{4\,11} - 11\phi_{5\,10} + 5\sqrt{3}\phi_{69}], \\
\phi_{220} &= (1/2\sqrt{21})[5\phi_{2\,14} - \sqrt{15}\phi_{3\,13} + \sqrt{2}\phi_{4\,12} + \sqrt{2}\phi_{5\,11} - \sqrt{15}\phi_{6\,10} + 5\phi_{79}], \\
\phi_{22-1} &= -(1/6\sqrt{14})[5\sqrt{7}\phi_{89} - \sqrt{5}\phi_{7\,10} - \sqrt{30}\phi_{6\,11} + 7\sqrt{2}\phi_{5\,12} - 11\phi_{4\,13} + 5\sqrt{3}\phi_{3\,14}], \\
\phi_{22-2} &= (1/3\sqrt{14})[\sqrt{35}\phi_{8\,10} - 2\sqrt{10}\phi_{7\,11} + \sqrt{30}\phi_{6\,12} - 4\phi_{5\,13} + \sqrt{5}\phi_{4\,14}],
\end{aligned} \tag{42}$$

where $\mu = 2$ has been used to represent $(\frac{7}{2}, \frac{5}{2})$ in the above. Using these and the pair states of Eq. (38) in the projected state of Eq. (41) and proceeding as in Eq. (40) we obtain the normalized state as

$$\begin{aligned}
|\Psi\rangle = &-0.1376(12368\,14) + 0.2132(12378\,13) + 0.1376(12458\,14) + 0.094(12467\,14) - 0.112(12468\,13) \\
&- 0.1557(12478\,12) - 0.1682(12567\,13) + 0.3146(12568\,12) - 0.1557(12578\,11) + 0.2132(12678\,10) \\
&- 0.2102(13457\,14) + 0.0355(13458\,13) + 0.2185(13467\,13) - 0.1168(13468\,12) + 0.3146(13478\,11) \\
&- 0.1030(13567\,12) - 0.1168(13568\,11) - 0.1112(13578\,10) - 0.1376(136789) + 0.1330(14567\,11) \\
&+ 0.0355(14568\,10) + 0.1376(145789) + 0.2752(23456\,14) - 0.1777(23457\,13) + 0.1330(23458\,12) \\
&+ 0.0389(23467\,12) - 0.1030(23468\,11) - 0.1682(23478\,10) + 0.0389(23567\,11) + 0.2185(23568\,10) \\
&+ 0.094(235789) - 0.1777(24567\,10) - 0.2102(245689) + 0.2752(345679).
\end{aligned} \tag{43}$$

A similar approach can be used to obtain the remaining states

$$\mathscr{A}\,|[(\phi_{12})(\phi_{22})^2];\ 00\rangle \text{ and } \mathscr{A}\,|[(\phi_{22})^3];\ 00\rangle$$

of this two-level system as linear combinations of Slater determinantal states as in Eqs. (40) and (43). The results of the Appendix enable us to generate the Hamiltonian matrix elements for each of these defining determinantal states. Weighting these matrix elements with the coefficients with which the corresponding determinantal states occur in the antisymmetrized angular momentum projected configurations, we can obtain the Hamiltonian matrix in terms of these states. A program is at present being developed for implementing the procedure for a general multilevel system of N pairs of identical nucleons. In this program we carry out a linear variation procedure and extract the relevant correlation parameters of Eq. (24) by a least squares method.

III. DISCUSSION

The most tedious aspect of the procedure outlined in Sec. II occurs in antisymmetrizing the projected n-pair states and expressing these as linear combination of single particle determinantal states. This is evident from an examination of Eqs. (40) and (43). A further complication is the need to use the results of the Appendix repeatedly in obtaining the matrix elements of the shell-model Hamiltonian by matching each determinant of the ket state with a similar determinant in the bra vector. We are unable to suggest any simple alternative to this at present. We have, however, reduced the problem to one of bookkeeping by assigning a simple numbering system to the determinantal states. In spite of this difficulty, the present procedure is simpler to implement than, say, the Dyson boson mapping methods[7–9] with their attendant non-Hermiticity problems. An advantage of the present approach is that it lends itself readily to a generalization to correlated pairs other than those with $j = 0$ and 2. A point to be noted in using the shift operator $E_{\mu' j';\mu'' j''}$ to generate multilevel configurations is that it only generates those configurations which have their origin in the single level state

$$|[(\mu j_1)^{N_1}(\mu j_2)^{N_2}\cdots(\mu j_k)^{N_k}];\ JM\rangle.$$

Thus, for example, states such as

$$\begin{aligned}
&|[(\phi_{12})^2\phi_{22}];\ 55\rangle \\
&\quad = (1/\sqrt{3})\{\sqrt{2}[(\phi_{122})^2\phi_{221}] - [\phi_{122}\phi_{222}\phi_{121}]\},
\end{aligned} \tag{44}$$

do not appear in the configuration space. That this is a natural consequence of using the correlated pair formulation as

in IBM or BPA approaches follows readily. The s- and d-type correlated pair status resulting from Eq. (24) are

$$\phi_{;00} = \sum_{\mu} a_{\mu} \phi_{\mu 00}, \tag{45}$$

$$\phi_{;2m} = \sum_{\mu'} b_{\mu'} \phi_{\mu' 2m} \quad (m = 2,1,0,-1,-2). \tag{46}$$

The angular momentum projected states obtained from these are

$$|[(\phi_{;0})^p(\phi_{;2})^q]; JM\rangle = \sum_{m_1,m_2,\ldots,m_q} A_{2m_1,2m_2,\ldots,2m_q} \times [(\phi_{;00})^p(\phi_{;2m_1}\phi_{;2m_2}\cdots\phi_{;2m_q})], \tag{47}$$

where $A^{(JM)}_{2m_1,2m_2,\ldots,2m_q}$ are coefficients determined as in Sec. II. Using the expansions of Eqs. (45) and (46) in the right-hand side of Eq. (47), we obtain

$$|[(\phi_{;0})^p(\phi_{;2})^q]; JM\rangle = \sum_{m_1,m_2,\ldots,m_q} A^{(JM)}_{2m_1,2m_2,\ldots,2m_q} \sum_{\substack{\mu_1,\mu_2,\ldots,\mu_p \\ \mu'_1,\mu'_2,\ldots,\mu'_q}} a_{\mu_1}a_{\mu_2}\cdots a_{\mu_p} b_{\mu'_1} b_{\mu'_2} [\phi_{\mu_1 00}\phi_{\mu_2 00}\cdots\phi_{\mu_p 00}\cdots\phi_{\mu_p 00}\phi_{\mu'_1 2m_1}\phi_{\mu'_2 2m_2}\cdots\phi_{\mu'_q 2m_q}]. \tag{48}$$

On examining this result we find that the left-hand side is symmetrical under the permutations of the first p pair state indices amongst themselves as also the last q indices. This symmetry has to appear on the right-hand side also. The pair functions on the right-hand side carry the triad of indices μ_{i00} and μ'_{i2m_i}. Thus only configurations reflecting this symmetry appear in using correlated pair formulation and not general configurations as in the example given in Eq. (44). In fact, if we let $\mu = 2 \to \mu = 1$ this state vanishes on using the renormalization necessary for $[\phi_{122}\phi_{222}\phi_{121}] \to [(\phi_{122})^2\phi_{121}]$. This result permits us to use the shift operator $E_{\mu' f,\mu'' f'}$ in going from a single- to a multilevel configuration, since it is a symmetric replacement operator.

We are at present concentrating on two aspects of the present approach. A viable program is being developed using relatively simple two-body interactions which would minimize the computer storage problems. A second aspect of the problem under study is to generalize the approach so that we can handle the configuration space of a proton–neutron system. This would require the restriction of the symmetric representation of $U(m_1 + m_2)$ to the products of the symmetric representations of $U(m_1) \times U(m_2)$, where m_1 and m_2 refer to the pair spaces of the proton and neutron pairs, respectively. Computationally this problem is bound to be more formidable than the one we are trying to implement at present.

ACKNOWLEDGMENTS

The authors express their profound thanks to Professor Y. K. Gambhir for suggesting the problem and helping in the various stages of its solution. The authors also thank Dr. R. S. Nikam for the initial discussions on this problem.

APPENDIX:

Three categories of nonzero matrix elements are possible for the Hamiltonian of Eq. (4) over single determinantal states of single particle orbitals. The forms of these matrix elements can be readily established by specializing the results of an earlier study[14] to single column SWT spanning $[1^{2N}0]$ of $U(n)$. These results are summarized in this Appendix.

For notational convenience we represent the determinantal states of Eq. (7) as

$$(N_1 N_2 \cdots N_n) = \mathscr{A}(|1\rangle^{N_1}|2\rangle^{N_2}\cdots|n\rangle^{N_n}), \tag{A1}$$

where the N_k $(k = 1,\ldots,n)$ have the possible values 0 or 1 subject to $\Sigma_{k=1}^{n} N_k = 2N$. The required matrix elements are of the form

$$H_{(N')(N)} \equiv \langle (N'_1 N'_2 \cdots N'_n)|H|(N_1 N_2 \cdots N_n)\rangle. \tag{A2}$$

The following cases occur.

(i) Zero-excitations: In this case $N'_k = N_k$ for all k from 1 to n, and

$$H_{(N)(N)} = \sum_{i=1}^{n} N_i \epsilon_i + \sum_{\substack{k \\ (k>i)}} N_k V_{(ik)(ik)}. \tag{A3}$$

(ii) One-excitations: In this case $N'_k = N_k$ for all $k \neq a$, b and $N'_a = 1$, $N'_b = 0$, $N_a = 0$, $N_b = 1$. The required result is

$$H_{(N')(N)} = \sum_{i=1}^{n} N_i V_{(ia)(ib)} \delta_p, \tag{A4}$$

where δ_p is the signature of the permutation p matching the location of the index b of the ket vector with that of a of the bra vector.

(iii) Two-excitations: In this case $N'_k = N_k$ for all $k \neq a$, b, c, d, and

$$N'_a = N'_b = 1, \quad N'_c = N'_d = 0;$$

$$N_a = N_b = 0, \quad N_c = N_d = 1.$$

This yields

$$H_{(N')(N)} = V_{(ab)(cd)} \delta_p, \tag{A5}$$

where p is the permutation which matches the pair of indices c, d of the ket vector with a and b of the bra vector.

[1] *Interacting Bosons in Nuclear Physics*, edited by F. Iachello (Plenum, New York, 1979).

[2] *Interacting Bose–Fermi Systems in Nuclei*, edited by F. Iachello (Plenum, New York, 1981).

[3] A. Arima and F. Iachello, in *Advances in Nuclear Physics*, edited by J. W. Negele and E. Vogt (Plenum, New York, 1984), Vol. 13, p. 139.

[4] O. Scholten, Phys. Rev. C **28**, 1783 (1983), and references therein.

[5] Y. K. Gambhir, P. Ring, and P. Schuck, Phys. Rev. C **25**, 2858 (1982);

Y. K. Gambhir and J. A. Sheikh, Phys. Rev. C **34**, 2344 (1986).
[6]S. Pittel, P. D. Dual, and B. R. Barrett, Ann. Phys. (NY) **144**, 168 (1982); A. Van Egmond and K. Allaart, Nucl. Phys. A **394**, 173 (1983).
[7]T. Tamura, Phys. Rev. C **28**, 2840 (1983); H. B. Geyer and S. Y. Lee, *ibid.* **26**, 642 (1982); H. B. Geyer, C. A. Engelbrecht, and F. J. W. Hahne, *ibid.* **33**, 1041 (1986); D. Bonatsos and A. Klein, University of Pennsylvania preprint UPR-0014 NT.
[8]Y. K. Gambhir, J. A. Sheikh, P. Ring, and P. Schuck, Phys. Rev. C **31**, 1519 (1985); C. T. Li, *ibid.* **29**, 2309 (1984).
[9]Y. K. Gambhir, R. S. Nikam, C. R. Sarma, and J. A. Sheikh, J. Math. Phys. **26**, 2067 (1985).
[10]M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, London, 1968).
[11]C. R. Sarma and G. G. Sahasrabudhe, J. Math. Phys. **21**, 638 (1980).
[12]P. Ring and P. Schuck, *The Nuclear Many-Body Problem* (Springer, New York, 1980), Chap. 9.
[13]C. R. Sarma, J. Phys. G **7**, 903 (1981).
[14]C. R. Sarma and S. Rettrup, J. Phys. A **13**, 2267 (1980).

Representations of $N=2$ extended supergravity and unitarity conditions in Osp(N,4)

Joris Van der Jeugt[a)]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 1 August 1985; accepted for publication 3 December 1986)

The structure of unitary irreducible representations of the Lie superalgebra Osp (2,4), the algebra of $N = 2$ extended supergravity, is investigated in detail. In particular, four new classes of shortened multiplets are found, and the complete unitarity conditions are given. The shortened multiplets are shown to correspond to atypical infinite-dimensional representations. Finally, unitarity conditions for Osp(N,4), $3 \leqslant N \leqslant 8$, are constructed.

I. INTRODUCTION

Lie superalgebras play an important role in physics, since they are the mathematical foundation of supergravity theories. A complete classification of all simple Lie superalgebras over the complex numbers has been given by Kac.[1] Moreover, some general theory on finite-dimensional irreducible representations (irreps) of classical Lie superalgebras has been developed.[2]

The relevant Lie superalgebra for N-extended supergravity (with anti-de Sitter space-time) is Osp(N,4). Representations of Osp(N,4) have been studied by various authors.[3–5] Here, it is the *infinite-dimensional* representations that are of importance, since they describe particles in supersymmetric field theories. A complete classification of unitary irreducible representations of Osp(N,4) (with lowest energy, or with lowest weight) has been obtained only for $N = 1$ (see Refs. 6 and 7). For $N \geqslant 2$, partial results have been obtained, especially with respect to unitarity conditions. Moreover, the phenomenon of *multiplet shortening* has already been observed.[3,4]

The work presented in this paper is mainly mathematical. We shall first study the unitary irreducible representations of Osp(2,4). Some of the irreps of Osp(2,4) have already been classified by Ceresole *et al.*[5] Here, we obtain the general structure of an unshortened Osp(2,4) irrep, and moreover, we discuss four types of shortened multiplets. A unitary irrep of Osp(2,4) is labeled by its lowest energy E_0, together with the corresponding spin s and hypercharge y_0. We prove that the unitarity conditions are equivalent with

$$E_0 \geqslant |y_0| + s + 1. \tag{1.1}$$

In general, a unitary irrep of Osp(2,4) decomposes in 16 irreps of the even subalgebra SO(3,2)$\times$U(1) (unshortened case). If $E_0 = \pm y_0 + s + 1$ or $E_0 = \pm y_0 - s$, we prove that the representation is shortened and decomposes in only eight subalgebra irreps. Then, it is quite easy to see that the shortened multiplets of the Lie superalgebra are an infinite-dimensional version of the so-called atypical representations[2] in the study of finite-dimensional irreps. As a consequence, also the unitarity conditions can be deduced from atypicality relations. This is what we have performed for the Lie superalgebras Osp(N,4), with $3 \leqslant N \leqslant 8$. These unitarity conditions are in general simpler than the ones for Osp(2,4), since Osp(2,4) is a Lie superalgebra of type I.

The structure of the paper is as follows: in Sec. II the algebra for Osp(2,4) is given in terms of the usual SO(3,2)$\times$U(1) generators M_{AB} and T^{12}, and the eight Majorana spinor charges. Our main technique is that we work, throughout the whole of the paper, in the multiplicity-free basis for SO(3,2). This means that basis states are in first instance not described in the chain SO(3,2) $\supset$U(1)$\times$SU(2), with U(1) giving the energy label and SU(2) the spin label, but they are described in the chain SO(3,2)$\supset$SU(1,1)$\times$SU(1,1). In the latter chain, basis states of SO(3,2) are uniquely determined by the SU(1,1)$\times$SU(1,1) labels. Hence calculations are much easier in this basis. Moreover, since we are finally only interested in the SO(3,2)$\times$U(1) contents of an Osp(2,4) irrep, the internal labeling of SO(3,2) plays no role. Therefore we go over to a new basis for Osp(2,4), in which the subalgebra SU(1,1)$\times$SU(1,1)$\times$U(1) is apparent. Also, the new (anti)commutation relations in this basis are given in Sec. II. In Sec. III, the structure of unitary irreps of Osp(2,4) is investigated. Part of this section is rather technical; making use of the explicit expression of the second-order Casimir operator C_2 of Osp(2,4), some relevant matrix elements of the odd generators are deduced. Also, four types of shortened irreps arise. Section IV deals with the relation between shortened multiplets and atypical representations of the Lie superalgebra. This relation is extended to Osp(N,4) ($3 \leqslant N \leqslant 8$) in Sec. V, where the unitarity conditions are listed.

II. THE Osp(2,4) SUPERALGEBRA

The even part of the algebra of Osp(2,4) is spanned by ten Hermitian SO(3,2) generators $M_{AB} = -M_{BA}$, where $A,B = 0,1,2,3,4$, and by a single Hermitian SO(2) generator $T^{ij} = -T^{ji}$ $(i,j = 1,2)$. The commutation relations are described by[3,4]

$$[M_{AB},M_{CD}] = i(\eta_{BC}M_{AD} - \eta_{AC}M_{BD} - \eta_{BD}M_{AC} + \eta_{AD}M_{BC}),$$
$$\eta_{AB} = \mathrm{diag}(1,-1,-1,-1,1); \tag{2.1}$$
$$[T^{12},M_{AB}] = 0. \tag{2.2}$$

[a)] Research assistant NFWO (Belgium).

 0022-2488/87/040758-07$02.50

The odd elements of the Lie superalgebra are eight Majorana spinor charges Q^i_α $(i=1,2)$, where $\alpha=1,\ldots,4$ is a Dirac index, which satisfy the anticommutation relations

$$\{Q^i_\alpha,\bar{Q}^j_\beta\}=i\delta^{ij}(l^{AB})_{\alpha\beta}M_{AB}+i\delta_{\alpha\beta}T^{ij}. \tag{2.3}$$

The matrices l^{AB} are defined in Ref. 4. The Majorana spinors are usually parametrized as

$$Q^i_\alpha=\begin{pmatrix} a^i_\alpha \\ \epsilon_{\alpha\beta}\bar{a}^i_\beta \end{pmatrix},\quad \epsilon=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.4}$$

where

$$(a^i_\alpha)^\dagger=\bar{a}^i_\alpha. \tag{2.5}$$

It is well known that M_{04} represents the energy operator, and that $M_{ij}=\epsilon_{ijk}J_k$ form the generators of a compact SU(2) subalgebra describing the spin. Moreover, T^{12} can be interpreted as the hypercharge operator. For the remaining commutation relations between even elements and odd elements, we refer to Ref. 4.

For our purposes, it will be useful to describe the SO(3,2) subalgebra in a multiplicity-free basis. As a consequence, irreps of SO(3,2) are not decomposed according to the compact energy-spin subalgebra $U(1)\times SU(2)$, in which chain there is a degeneracy problem, but they are decomposed in a noncompact $SU(1,1)\times SU(1,1)$ basis. There is no labeling problem in $SO(3,2)\supset SU(1,1)\times SU(1,1)$, and hence it makes the analysis of SO(3,2) irreps, and eventually of Osp(2,4) irreps, much easier. On the other hand, it is fairly easy to transfer the results obtained in this noncompact basis, to the basis with the physical energy-spin labels,[8,9] as we shall perform in Sec. III.

The new SO(3,2) basis consists of the $SU(1,1)\times SU(1,1)$ generators $p_0,p_\pm$ and $q_0,q_\pm$, and of the components of a four-dimensional tensor $R_{ij}(i,j=\pm\frac{1}{2})$ with respect to $SU(1,1)\times SU(1,1)$. The connection between the new basis elements and the M_{AB} generators is given by

$$\begin{aligned}
&p_0=\tfrac{1}{2}(M_{04}+M_{12}),\quad p_+=\tfrac{1}{2}(iM_{01}-M_{14}-M_{02}-iM_{24}),\\
&p_-=\tfrac{1}{2}(iM_{01}+M_{14}+M_{02}-iM_{24});\\
&q_0=\tfrac{1}{2}(M_{04}-M_{12}),\\
&q_+=\tfrac{1}{2}(iM_{01}-M_{14}+M_{02}+iM_{24}),\\
&q_-=\tfrac{1}{2}(iM_{01}+M_{14}-M_{02}+iM_{24});\\
&R_{(1/2)(1/2)}=-(1/\sqrt{2})(M_{03}+iM_{34}),\\
&R_{-(1/2)-(1/2)}=(1/\sqrt{2})(-M_{03}+iM_{34}),\\
&R_{(1/2)-(1/2)}=(1/\sqrt{2})(M_{31}-iM_{23}),\\
&R_{-(1/2)(1/2)}=(1/\sqrt{2})(M_{31}+iM_{23}).
\end{aligned} \tag{2.6}$$

We also introduce the notation

$$Y=T^{12}. \tag{2.7}$$

Then, the hermiticity conditions, in the new basis, read

$$\begin{aligned}
&p_0^\dagger=p_0,\quad p_\pm^\dagger=-p_\mp,\quad q_0^\dagger=q_0,\\
&q_\pm^\dagger=-q_\mp,\quad R_{ij}^\dagger=R_{-i-j},\quad Y^\dagger=Y,
\end{aligned} \tag{2.8}$$

and the nonvanishing commutator relations are given by

$$\begin{aligned}
&[p_0,p_\pm]=\pm p_\pm,\quad [p_+,p_-]=2p_0,\\
&[q_0,q_\pm]=\pm q_\pm,\quad [q_+,q_-]=2q_0,\\
&[p_0,R_{ij}]=iR_{ij},\quad [q_0,R_{ij}]=jR_{ij},\\
&[p_\pm,R_{\mp 1/2j}]=R_{\pm(1/2)j},\\
&[q_\pm,R_{i\mp(1/2)}]=R_{i\pm(1/2)},\\
&[R_{(1/2)\pm(1/2)},R_{-(1/2)\mp(1/2)}]=\mp p_0-q_0,\\
&[R_{\pm(1/2)(1/2)},R_{\pm(1/2)-(1/2)}]=\pm p_\pm,\\
&[R_{(1/2)\pm(1/2)},R_{-(1/2)\pm(1/2)}]=\pm q_\pm.
\end{aligned} \tag{2.9}$$

The odd basis elements of Osp(2,4) are given in terms of the components $\overset{(-)}{a}{}^i_\alpha$ of the Majorana spinors:

$$\begin{aligned}
&Q_{(1/2)0}=(i/2)(\bar{a}^1_1+i\bar{a}^2_1),\quad \bar{Q}_{(1/2)0}=(i/2)(\bar{a}^1_1-i\bar{a}^2_1),\\
&Q_{-(1/2)0}=(i/2)(a^1_1+ia^2_1),\quad \bar{Q}_{-(1/2)0}=(i/2)(a^1_1-ia^2_1),\\
&Q_{0(1/2)}=\tfrac{1}{2}(\bar{a}^1_2+i\bar{a}^2_2),\quad \bar{Q}_{0(1/2)}=\tfrac{1}{2}(\bar{a}^1_2-i\bar{a}^2_2),\\
&Q_{0-(1/2)}=-\tfrac{1}{2}(a^1_2+ia^2_2),\\
&\bar{Q}_{0-(1/2)}=-\tfrac{1}{2}(a^1_2-ia^2_2).
\end{aligned} \tag{2.10}$$

The basis elements Q_{ij} (resp. $\bar{Q}_{ij}$) form a four-dimensional tensor representation of the SO(3,2) algebra. They satisfy

$$Q^\dagger_{ij}=-\bar{Q}_{-i-j},\quad (i,j)=(\pm\tfrac{1}{2},0)\text{ or }(0,\pm\tfrac{1}{2}). \tag{2.11}$$

The commutation relations between the $SU(1,1)\times SU(1,1)$ generators and the odd basis elements $\overset{(-)}{Q}{}_{ij}$ are given by

$$\begin{aligned}
&\left[p_0,\overset{(-)}{Q}{}_{ij}\right]=i\overset{(-)}{Q}{}_{ij},\quad \left[p_\pm,\overset{(-)}{Q}{}_{0\pm 1/2}\right]=0,\\
&\left[p_\pm,\overset{(-)}{Q}{}_{\mp(1/2)0}\right]=\overset{(-)}{Q}{}_{\pm(1/2)0},
\end{aligned} \tag{2.12}$$

and analogous expressions for $\left[q,\overset{(-)}{Q}\right]$. The relations (2.13) express that Q_{ij} (resp. $\bar{Q}_{ij}$) raises (resp. lowers) the hypercharge by one unit:

$$[Y,Q_{ij}]=Q_{ij},\quad [Y,\bar{Q}_{ij}]=-\bar{Q}_{ij}. \tag{2.13}$$

The remaining commutation relations between even and odd basis elements are summarized by

$$\begin{aligned}
&[R_{(1/2)\pm(1/2)},Q_{-(1/2)0}]=(1/\sqrt{2})Q_{0\pm(1/2)},\\
&[R_{-(1/2)\pm(1/2)},Q_{(1/2)0}]=-(1/\sqrt{2})Q_{0\pm(1/2)},\\
&[R_{\pm(1/2)(1/2)},Q_{0-(1/2)}]=(1/\sqrt{2})Q_{\pm(1/2)0},\\
&[R_{\pm(1/2)-(1/2)},Q_{0(1/2)}]=-(1/\sqrt{2})Q_{\pm(1/2)0},
\end{aligned} \tag{2.14}$$

and the same relations with Q replaced by $\bar{Q}$.

Finally, the anticommutation relations among the odd elements need to be determined. One finds

$$\begin{aligned}
&\{Q_{0\pm(1/2)},\bar{Q}_{0\pm(1/2)}\}=\pm q_\pm,\\
&\{Q_{\pm(1/2)0},\bar{Q}_{\pm(1/2)0}\}=\pm p_\pm,\\
&\{Q_{0j},\bar{Q}_{i0}\}=(1/\sqrt{2})R_{ij},\\
&\{Q_{i0},\bar{Q}_{0j}\}=(1/\sqrt{2})R_{ij}\quad (i,j=\pm\tfrac{1}{2}),\\
&\{Q_{0\pm(1/2)},\bar{Q}_{0\mp(1/2)}\}=-q_0\pm\tfrac{1}{2}Y,\\
&\{Q_{\pm(1/2)0},\bar{Q}_{\mp(1/2)0}\}=-p_0\pm\tfrac{1}{2}Y.
\end{aligned} \tag{2.15}$$

The Lie superalgebra of Osp(2,4) is now completely deter-

mined by (2.9) and (2.12)–(2.15). Then, one can easily calculate the second-order Casimir operator C_2 of Osp(2,4):

$$C_2 = I_2 + Y - \tfrac{1}{4}Y^2 + \overline{Q}_{(1/2)0}Q_{-(1/2)0} - \overline{Q}_{-(1/2)0}Q_{(1/2)0} + \overline{Q}_{0(1/2)}Q_{0-(1/2)} - \overline{Q}_{0-(1/2)}Q_{0(1/2)}. \quad (2.16)$$

Herein, I_2 is the quadratic Casimir operator of the even subalgebra SO(3,2):

$$I_2 = R_{(1/2)-(1/2)}R_{-(1/2)(1/2)} - R_{(1/2)(1/2)}R_{-(1/2)-(1/2)} + p_+p_- + p_0^2 - 2p_0 + q_+q_- + q_0^2 - q_0. \quad (2.17)$$

III. STRUCTURE OF UNITARY IRREDUCIBLE REPRESENTATIONS OF Osp(2,4) WITH LOWEST ENERGY

Unitary irreducible representations of Osp(2,4), with the Hermiticity relations (2.8) and (2.11), are infinite dimensional. We shall describe the structure of such irreps by giving their decomposition into irreps of the even subalgebra SO(3,2)×U(1). For representations of Osp(2,4) with lowest energy, the energy operator M_{04} of SO(3,2) is bounded from below. The structure of such SO(3,2) irreps is well known,[3,9] and they can be labeled by (E,j), where E is the minimum energy value, and j the corresponding spin quantum number. The labels (E,j) are the natural quantum numbers in the chain SO(3,2)⊃SU(2)×U(1). Alternatively, the same SO(3,2) irrep can be labeled by a couple of numbers $(\bar{p},\bar{q})$, where $\bar{p}$ is the lowest p_0 eigenvalue, and $\bar{q}$ the corresponding q_0 eigenvalue, in the chain SO(3,2) ⊃SU(1,1)×SU(1,1). The relation between the two sets of labels reads

$$2\bar{p} = E - j, \quad 2\bar{q} = E + j. \quad (3.1)$$

In fact, $(\bar{p},\bar{q})$ is the lowest weight if $\{p_0,q_0\}$ is taken as the Cartan subalgebra of SO(3,2), whereas $(E,-j)$ is the lowest weight if $\{M_{04},M_{12}\}$ is chosen as the Cartan subalgebra. Let y be the eigenvalue of the U(1) operator Y, then a representation of Osp(2,4) is characterized by listing the SO(3,2)×U(1) irreps (E,j,y) or $(\bar{p},\bar{q},y)$ in which it decomposes.

The Lie superalgebra of Osp(2,4) contains four odd positive roots, and hence an irreducible representation of Osp(2,4) consist of at most $2^4 = 16$ irreps of the even part SO(3,2)×U(1). This fact follows immediately from the structure of the enveloping algebra of Osp(2,4). Also, the 16 subalgebra irreps in which a representation of Osp(2,4) decomposes are easy to find. Let E_0 be the lowest energy value, s be the corresponding j value, and y_0 be the hypercharge of the irrep (E_0,s). Then the Osp(2,4) representation can be labeled by $(E_0;s;y_0)$, or alternatively by $(P;Q;y_0)$ with

$$2P = E_0 - s, \quad 2Q = E_0 + s, \quad (3.2)$$

and the decomposition of such an irrep in multiplets (E,j,y) or $(\bar{p},\bar{q},y)$ of SO(3,2)×U(1) is given in Table I. This decomposition was already given by Ceresole, Fré, and Nicolai.[5] If we put $s = 1$, then the first two columns of Table I give the hypercharge, energy and spin values of the $N = 2$ graviton multiplets.[5]

A representation $(E_0;s;y_0)$ consisting of 16 subalgebra irreps, as the one in Table I, is called an unshortened representation. When one or several of the subalgebra irreps (E,j,y) are missing in the decomposition of $(E_0;s;y_0)$ one speaks of *multiplet shortening*. This phenomenon has already been considered by Freedman and Nicolai,[3,4] and later by Ceresole *et al.*[5] Multiplet shortening can occur for two reasons. The first origin of shortening can lie in the structure of the irreps of the subalgebra SO(3,2)×U(1). For instance, one condition is that $j \geqslant 0$. In this case, putting $s = 0$ in Table I, the irreps with spin $s - \frac{1}{2}$ or $s - 1$ cannot occur: this gives rise to multiplet shortening.

This can also be illustrated by the Class 2 representations of Osp(1,4) in the classification of Heidenreich[6]: these representations are just truncations of the unshortened Class 4 irreps.

As a second example, it is known that $E = j + \frac{1}{2}$ (for $j = 0,\frac{1}{2}$) and $E = j + 1$ (for $j = 1,\frac{3}{2},\ldots$) are special representations of SO(3,2) (Refs. 3 and 9). Hence when such conditions occur, multiplet shortening may appear also due to the internal structure of SO(3,2) irreps. This is the case for the Class 1 and Class 3 irreps of Osp(1,4) in the previously mentioned paper.[6]

In this paper, we will be concerned with the second origin of multiplet shortening, which lies in the structure of the irreps of Osp(2,4) themselves. Actually, such shortened representations are then the reflection of atypical representations in the infinite-dimensional case, as we shall see in Sec. IV. In the present section, we shall analyze the Osp(2,4) irreps by means of some tensor operator calculus. It will turn out that, if certain conditions are satisfied, the 16 subalgebra irreps of Table I are not connected to each other in an irreducible representation of Osp(2,4), but that only a part of the 16 irreps form an irreducible representation of the Lie superalgebra: then, one deals with a shortened representation. Whether the irreps (E,j,y) in Table I are all part of one unitary irreducible representation of Osp(2,4), can be analyzed by computing the reduced matrix elements of the SO(3,2) tensors Q_{ij} and $\overline{Q}_{ij}$ between states of different SO(3,2)×U(1) representations. This method is closely related to the one used by Inaba *et al.*[7] to investigate Osp(1,4) representations.

TABLE I. Decomposition of a general Osp(2,4) irrep $(E_0;s;y_0)$ [or $(P;Q;y_0)$] into representations (E,j,y) [or $(\bar{p},\bar{q},y)$] of SO(3,2)×U(1).

y value	(E,j)	$(\bar{p},\bar{q})$
$y_0 \pm 2$	(E_0+1,s)	$(P+\frac{1}{2},Q+\frac{1}{2})$
$y_0 \pm 1$	$(E_0+\frac{1}{2},s+\frac{1}{2}),\ (E_0+\frac{1}{2},s-\frac{1}{2})$ $(E_0+\frac{3}{2},s+\frac{1}{2}),\ (E_0+\frac{3}{2},s-\frac{1}{2})$	$(P,Q+\frac{1}{2}),\ (P+\frac{1}{2},Q)$ $(P+\frac{1}{2},Q+1),\ (P+1,Q+\frac{1}{2})$
y_0	$(E_0,s),\ (E_0+1,s+1),\ (E_0+1,s)$ $(E_0+1,s),\ (E_0+1,s-1),\ (E_0+2,s)$	$(P,Q),\ (P,Q+1),\ (P+\frac{1}{2},Q+\frac{1}{2})$ $(P+\frac{1}{2},Q+\frac{1}{2}),\ (P+1,Q),\ (P+1,Q+1)$

The reduced matrix elements of the Q_{ij} and $\overline{Q}_{ij}$ tensor will be determined in the following chain:

$$\begin{array}{ccccccccc} \mathrm{Osp}(2,4) & \supset & \mathrm{SO}(3,2)\times\mathrm{U}(1) & \supset & \mathrm{SU}(1,1)\times\mathrm{SU}(1,1)\times\mathrm{U}(1) & \supset & \mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1). \\ (P;Q;y_0) & & (\bar{p},\bar{q},y) & & p \quad q \quad y & & m_p \quad m_q \quad y \end{array} \tag{3.3}$$

The Osp(2,4) irrep is labeled by $(P;Q;y_0)$, where (P,Q) are the lowest SO(3,2) labels, and y_0 the corresponding hypercharge. Then, $(P;Q;y_0)$ decomposes in some SO(3,2) $\times$U(1) irreps $(\bar{p},\bar{q},y)$, and finally, the states in an SO(3,2) irrep $(\bar{p},\bar{q})$ can be described by[8,9]

$$|(\bar{p},\bar{q}),p,m_p,q,m_q\rangle, \quad \begin{aligned} m_p &= p,p+1,\ldots, \\ m_q &= q,q+1,\ldots, \end{aligned} \tag{3.4}$$

and where the possible values of (p,q) in terms of $(\bar{p},\bar{q})$ are given in Ref. 9. Hence a basis state of Osp(2,4) can be written in the form

$$|(P;Q;y_0)(\bar{p},\bar{q},y),p,m_p,q,m_q\rangle. \tag{3.5}$$

It is in this basis that we shall perform the calculations. Consider the basis state, with highest y value, and denoted by

$$\begin{aligned} &|P+\tfrac{1}{2},Q+\tfrac{1}{2},y_0+2\rangle\rangle \\ &\quad = |(P;Q;y_0)(P+\tfrac{1}{2},Q+\tfrac{1}{2},y_0+2),P+\tfrac{1}{2}, \\ &\qquad P+\tfrac{1}{2},Q+\tfrac{1}{2},Q+\tfrac{1}{2}\rangle. \end{aligned} \tag{3.6}$$

Hereafter, the notation $|\bar{p},\bar{q},y\rangle\rangle$ shall be used to indicate the basis state (3.5) with lowest SO(3,2) weight, i.e., $p=m_p=\bar{p}$ and $q=m_q=\bar{q}$. The eigenvalue of the SO(3,2) Casimir operator I_2 on states of an irrep $(\bar{p},\bar{q})$ is well known:

$$\langle I_2\rangle = \bar{p}^2 - 2\bar{p} + \bar{q}^2 - \bar{q}. \tag{3.7}$$

Consider now the action of the Osp(2,4) Casimir C_2 upon the basis state (3.6). Since all the Q_{ij} operators produce zero terms, one finds the following expression for the C_2 eigenvalue:

$$\langle C_2\rangle = P^2 - P + Q^2 - \tfrac{1}{4}y_0^2. \tag{3.8}$$

Making use of $\{Q_{(1/2)0},\overline{Q}_{-(1/2)0}\} = -p_0 + \frac{1}{2}Y$, the Casimir C_2 can be written in the following form:

$$\begin{aligned} C_2 = I_2 &+ \overline{Q}_{(1/2)0}Q_{-(1/2)0} + Q_{(1/2)0}\overline{Q}_{-(1/2)0} \\ &+ \overline{Q}_{0(1/2)}Q_{0-(1/2)} - \overline{Q}_{0-(1/2)}Q_{0(1/2)} \\ &+ p_0 + \tfrac{1}{2}Y - \tfrac{1}{4}Y^2. \end{aligned} \tag{3.9}$$

The action of the latter form upon the basis state (3.6), produces

$$\begin{aligned} &Q_{(1/2)0}\overline{Q}_{-(1/2)0}|P+\tfrac{1}{2},Q+\tfrac{1}{2},y_0+2\rangle\rangle \\ &\quad = \tfrac{1}{2}(y_0-2P+1)|P+\tfrac{1}{2},Q+\tfrac{1}{2},y_0+2\rangle\rangle. \end{aligned} \tag{3.10}$$

Now we express that the representation must be unitary, this means that we can make use of (2.11) in order to deduce from (3.10) that

$$\begin{aligned} &|\langle\langle P,Q+\tfrac{1}{2},y_0+1|\overline{Q}_{-(1/2)0}|P+\tfrac{1}{2},Q+\tfrac{1}{2},y_0+2\rangle\rangle|^2 \\ &\quad = -\tfrac{1}{2}(y_0-2P+1). \end{aligned} \tag{3.11}$$

Hence, unitarity implies that $2P-y_0-1\geqslant 0$. Moreover, we see that when $y_0-2P+1=0$, the matrix element (3.11) vanishes, and then also the reduced matrix element of $\overline{Q}_{ij}$ between the SO(3,2)$\times$U(1) irreps $(P+\frac{1}{2},Q+\frac{1}{2},y_0+2)$ and $(P,Q+\frac{1}{2},y_0+1)$ becomes zero. This shows that, in the latter case, the two subalgebra irreps do not belong to the same irreducible representation of Osp(2,4).

Let us now continue the analysis by going from $|P,Q+\frac{1}{2},y_0+1\rangle\rangle$ to $|P,Q,y_0\rangle\rangle$. For this purpose, we have to bring C_2 in the following form:

$$\begin{aligned} C_2 = I_2 &+ \overline{Q}_{(1/2)0}Q_{-(1/2)0} - \overline{Q}_{-(1/2)0}Q_{(1/2)0} \\ &+ \overline{Q}_{0(1/2)}Q_{0-(1/2)} + Q_{0(1/2)}\overline{Q}_{0-(1/2)} \\ &+ q_0 + \tfrac{1}{2}Y - \tfrac{1}{4}Y^2. \end{aligned} \tag{3.12}$$

Consider the action of (3.12) on $|P,Q+\frac{1}{2},y_0+1\rangle\rangle$: obviously, the contributions from $\overline{Q}_{(1/2)0}Q_{-(1/2)0}$ and $\overline{Q}_{0(1/2)}Q_{0-(1/2)}$ vanish, and on the other hand $\overline{Q}_{-(1/2)0}Q_{(1/2)0}|P,Q+\frac{1}{2},y_0+1\rangle\rangle$ is easy deduced from (3.11). So, one obtains that

$$\begin{aligned} &Q_{0(1/2)}\overline{Q}_{0-(1/2)}|P,Q+\tfrac{1}{2},y_0+1\rangle\rangle \\ &\quad = \tfrac{1}{2}(y_0-2Q)|P,Q+\tfrac{1}{2},y_0+1\rangle\rangle, \end{aligned} \tag{3.13}$$

or, making use of the Hermiticity conditions,

$$|\langle\langle P,Q,y_0|\overline{Q}_{-(1/2)}|P,Q+\tfrac{1}{2},y_0+1\rangle\rangle|^2 = -\tfrac{1}{2}(y_0-2Q). \tag{3.14}$$

This gives rise to the inequality $2Q-y_0\geqslant 0$, and again, when $2Q-y_0=0$, the subalgebra irreps $(P,Q+\frac{1}{2},y_0+1)$ and (P,Q,y_0) do not belong to the same irreducible representation of Osp(2,4). One can proceed in the same way, and by rewriting the Lie superalgebra Casimir C_2 in some alternative forms, one finds

$$|\langle\langle P,Q+\tfrac{1}{2},y_0-1|\overline{Q}_{0(1/2)}|P,Q,y_0\rangle\rangle|^2 = \tfrac{1}{2}(y_0+2Q), \tag{3.15}$$

$$\begin{aligned} &|\langle\langle P+\tfrac{1}{2},Q+\tfrac{1}{2},y_0-2|\overline{Q}_{(1/2)0}|P,Q+\tfrac{1}{2},y_0-1\rangle\rangle|^2 \\ &\quad = \tfrac{1}{2}(y_0+2P-1). \end{aligned} \tag{3.16}$$

Until now we have determined only four special matrix elements. Detailed calculations, much facilitated in the chain to SU(1,1)$\times$SU(1,1), showed that the computation of the remaining (reduced) matrix elements always gives rise to one of the conditions obtained from (3.11) or (3.14)–(3.16). As a consequence, the representation $(P;Q;y_0)$ of Osp(2,4) is unitary if the following conditions are satisfied:

$$y_0 - 2P + 1 \leqslant 0, \tag{3.17a}$$

$$y_0 - 2Q \leqslant 0, \tag{3.17b}$$

$$y_0 + 2Q \geqslant 0, \tag{3.17c}$$

$$y_0 + 2P - 1 \geqslant 0. \tag{3.17d}$$

This leads to

$$2P \geqslant |y_0| + 1, \quad 2Q \geqslant |y_0|. \tag{3.18}$$

Making use of (3.2), the relations (3.18) can be expressed in terms of the labels $(E_0;s;y_0)$ of an Osp(2,4) irrep:

$$E_0 - s \geqslant |y_0| + 1, \quad E_0 + s \geqslant |y_0|. \tag{3.19}$$

The last condition is redundant, and hence the only unitarity condition is given by

TABLE II. Decomposition of a shortened Osp(2,4) irrep $(E_0;s;y_0)$ with $E_0 = y_0 + s + 1$ [or $(P;Q;y_0)$ with $y_0 - 2P + 1 = 0$] into representations (E,j,y) [or $(\bar{p},\bar{q},y)$] of SO(3,2)×U(1).

y value	(E,j)	$(\bar{p},\bar{q})$
y_0+1	$(E_0+\frac{1}{2},s+\frac{1}{2})$	$(P,Q+\frac{1}{2})$
y_0	(E_0,s), $(E_0+1,s+1)$, (E_0+1,s)	(P,Q), $(P,Q+1)$, $(P+\frac{1}{2},Q+\frac{1}{2})$
y_0-1	$(E_0+\frac{1}{2},s+\frac{1}{2})$, $(E_0+\frac{1}{2},s-\frac{1}{2})$, $(E_0+\frac{3}{2},s+\frac{1}{2})$	$(P,Q+\frac{1}{2})$, $(P+\frac{1}{2},Q)$, $(P+\frac{1}{2},Q+1)$
y_0-2	(E_0+1,s)	$(P+\frac{1}{2},Q+\frac{1}{2})$

$$E_0 > |y_0| + s + 1. \tag{3.20}$$

This is completely in agreement with Ref. 5.

As we have seen before, the irrep of Osp(2,4) will be shortened if the equality in one of the four conditions (3.17) is satisfied. Hence there are four types of shortened multiplets:

$$E_0 = y_0 + s + 1, \tag{3.21a}$$

$$E_0 = y_0 - s, \tag{3.21b}$$

$$E_0 = -y_0 - s, \tag{3.21c}$$

$$E_0 = -y_0 + s + 1. \tag{3.21d}$$

In the first case, all matrix elements having $(y_0 - 2P + 1)$ as a factor vanish, and various reduced matrix elements of Q_{ij} or $\bar{Q}_{ij}$ between states of SO(3,2)×U(1) irreps $(\bar{p},\bar{q},y)$ become zero, such as in (3.11). Then the irreducible representation of Osp(2,4) consists of only eight subalgebra irreps, instead of the 16 subalgebra irreps of Table I. These eight SO(3,2)×U(1) irreps are given in Table II. Tables III–V give the structure of the Osp(2,4) irreps in the case that (3.21b)–(3.21d) are satisfied: again there are only eight subalgebra irreps.

A special case occurs when two boundary conditions are satisfied simultaneously. It is easy to see that only (3.21a) and (3.21d) can occur at the same time if we also require (3.20), namely for

$$y_0 = 0, \quad E_0 = s + 1. \tag{3.22}$$

Then, all reduced matrix elements with a factor $(y_0 - 2P + 1)$ or a factor $(y_0 + 2P - 1)$ vanish. Consequently, it follows from Tables II and V that the irrep (3.22) is "ultrashort," and contains only four SO(3,2)×U(1) irreps ($s = 0,\frac{1}{2},1,\ldots$):

$$\begin{aligned}(E_0 = s+1; s; y_0 = 0) \to\ & (E = s+\tfrac{3}{2}, s+\tfrac{1}{2}, y = 1)\\ & \oplus (E = s+1, s, y = 0)\\ & \oplus (E = s+2, s+1, y = 0)\\ & \oplus (E = s+\tfrac{3}{2}, s+\tfrac{1}{2}, y = -1).\end{aligned} \tag{3.23}$$

IV. THE RELATION BETWEEN ATYPICAL REPRESENTATIONS AND SHORTENED REPRESENTATIONS

Finite-dimensional irreducible representations of Lie superalgebras have been studied by Kac.[2] In this reference, the important distinction between typical and atypical representations is introduced. A typical representation of a Lie superalgebra $L = L_{\bar{0}} \oplus L_{\bar{1}}$ decomposes in general in 2^M irreps of the even subalgebra $L_{\bar{0}}$, where M is the number of positive odd roots. Atypical representations have not yet been studied in general, but in several case studies[10–15] one notices that they always decompose in less than 2^M irreps of $L_{\bar{0}}$, hence they can be seen as shortened representations.

In supersymmetry or supergravity theories however, one is only interested in infinite-dimensional unitary representations of a Lie superalgebra. Again, there is no general theory of infinite-dimensional irreps of Lie superalgebras, but there are some case studies available.[6,10,14,15] The Lie superalgebras Osp(1,2) (Ref. 10) and Osp(1,4) (Ref. 6) have no atypical representations, since they are of type B(0,n). But the superalgebras Osp(3,2) (Ref. 14) and Osp(4,2) (Ref. 15) do have atypical representations: it was shown that infinite-dimensional atypical representations exist, and moreover, that the unitarity conditions could be deduced from the atypicality conditions.

Let us now consider the case of Osp(2,4). According to

TABLE III. Decomposition of a shortened Osp(2,4) irrep $(E_0;s;y_0)$ with $E_0 = y_0 - s$ [or $(P;Q;y_0)$ with $y_0 - 2Q = 0$] into representations (E,j,y) [or $(\bar{p},\bar{q},y)$] of SO(3,2)×U(1).

y value	(E,j)	$(\bar{p},\bar{q})$
y_0+1	$(E_0+\frac{1}{2},s-\frac{1}{2})$	$(P+\frac{1}{2},Q)$
y_0	(E_0,s), (E_0+1,s), $(E_0+1,s-1)$	(P,Q), $(P+\frac{1}{2},Q+\frac{1}{2})$, $(P+1,Q)$
y_0-1	$(E_0+\frac{1}{2},s+\frac{1}{2})$, $(E_0+\frac{1}{2},s-\frac{1}{2})$, $(E_0+\frac{3}{2},s-\frac{1}{2})$	$(P,Q+\frac{1}{2})$, $(P+\frac{1}{2},Q)$, $(P+1,Q+\frac{1}{2})$
y_0-2	(E_0+1,s)	$(P+\frac{1}{2},Q+\frac{1}{2})$

TABLE IV. Decomposition of a shortened Osp(2,4) irrep $(E_0;s;y_0)$ with $E_0 = -y_0 - s$ [or $(P;Q;y_0)$ with $y_0 + 2Q = 0$] into representations (E,j,y) [or $(\bar{p},\bar{q},y)$] of $SO(3,2)\times U(1)$.

y value	(E,j)	$(\bar{p},\bar{q})$
y_0+2	(E_0+1,s)	$(P+\frac{1}{2},Q+\frac{1}{2})$
y_0+1	$(E_0+\frac{1}{2},s+\frac{1}{2}),\ (E_0+\frac{1}{2},s-\frac{1}{2}),\ (E_0+\frac{3}{2},s-\frac{1}{2})$	$(P,Q+\frac{1}{2}),\ (P+\frac{1}{2},Q),\ (P+1,Q+\frac{1}{2})$
y_0	$(E_0,s),\ (E_0+1,s),\ (E_0+1,s-1)$	$(P,Q),\ (P+\frac{1}{2},Q+\frac{1}{2}),\ (P+1,Q)$
y_0-1	$(E_0+\frac{1}{2},s-\frac{1}{2})$	$(P+\frac{1}{2},Q)$

Kac,[2] an irreducible irrep $(a_1;a_2;a_3)$ of $C(2) = \mathrm{Osp}(2,4)$ is atypical if one of the following conditions is fulfilled:

$$a_1 = 0, \quad a_1 - a_2 - 1 = 0, \qquad (4.1)$$
$$a_1 - a_2 - 2a_3 - 3 = 0, \quad a_1 - 2a_2 - 2a_3 - 4 = 0,$$

otherwise it is typical. The conditions (4.1) can be deduced from

$$(\Lambda + \rho,\alpha) = 0, \quad \alpha \in \bar{\Delta}_1^+, \qquad (4.2)$$

with Λ the highest weight of the irrep, ρ half the sum of positive even roots minus half the sum of positive odd roots of $C(2)$, $\bar{\Delta}_1^+$ the set of odd positive roots α for which 2α is not an even root, and $(\ldots,\ldots)$ the bilinear form in weight space induced from a fixed nondegenerate bilinear form on the Lie superalgebra. The highest weight Λ can be written as

$$\Lambda = (c,\lambda_1,\lambda_2) = (a_1 - a_2 - a_3, (a_2 + a_3)/2, a_3/2). \qquad (4.3)$$

Herein c, λ_1, and λ_2 are the eigenvalues of the three diagonal $U(1)$ operators in the chain $\mathrm{Osp}(2,4) \supset U(1)\times SO(3,2) \supset U(1)\times U(1)\times U(1)$. Instead of using the highest weight, one can also express the atypicality by means of the lowest weight $-\Lambda$:

$$(-\Lambda - \rho,\alpha) = 0, \quad \alpha \in \bar{\Delta}_1^+. \qquad (4.4)$$

This gives rise to the four equations

$$-c + 2\lambda_1 + 4 = 0, \quad -c + 2\lambda_2 + 3 = 0,$$
$$-c - 2\lambda_1 = 0, \quad -c - 2\lambda_2 + 1 = 0. \qquad (4.5)$$

All this holds for finite-dimensional irreducible representations with lowest weight $-\Lambda$. But consider now the case of an infinite-dimensional representation with lowest weight $-\Lambda$. Then, part of the theory of Kac[2] can be repeated, and the representation may again be called atypical if one of the conditions (4.4) is fulfilled. According to Table I, the lowest y value is $y_0 - 2$, and the corresponding $(\bar{p},\bar{q})$ values are $P + \frac{1}{2}$, $Q + \frac{1}{2}$. Hence in that case the lowest weight is

$$-\Lambda = (-c, -\lambda_1, -\lambda_2) = (y_0 - 2, P + \tfrac{1}{2}, Q + \tfrac{1}{2}). \qquad (4.6)$$

With this identification, the relations (4.5) read

$$y_0 - 2P + 1 = 0, \quad y_0 - 2Q = 0,$$
$$y_0 + 2P - 1 = 0, \quad y_0 + 2Q = 0. \qquad (4.7)$$

But these are precisely the conditions we have deduced in Sec. III. Hence we have shown that the atypicality conditions, transferred to infinite-dimensional representations with lowest weight, give rise to a set of conditions which correspond precisely to the ones for a shortened representation for the Lie superalgebra Osp(2,4). In fact, we can say that shortened representations are "atypical infinite-dimensional" irreps, whereas unshortened representations are typical.

V. UNITARY CONDITIONS FOR Osp(N,4) ($3<N<8$)

An atypicality condition is a linear combination of labels put equal to zero. In Sec. III we have seen that they are precisely these linear combinations which appear in the expressions of reduced matrix elements of odd tensors between states of subalgebra irreps. Hence when those linear combinations are identically zero, the representation is shortened. But, moreover, these linear combinations also appear when positivity conditions (deduced from Hermiticity relations for the generators) are required. As a consequence, these atypicality conditions are also the boundary conditions for unitarity.

We have worked this out for the infinite-dimensional representations (with lowest weight) of Osp(N,4) ($3\leqslant N\leqslant 8$). In this case, an irrep of Osp(N,4) is characterized by its lowest energy value E_0, the corresponding spin s, and the Dynkin labels $(a_1,a_2,\ldots)$ of the finite-dimensional irrep of SO(N). This irrep of Osp(N,4) contains $(E_0,s)\times(a_1,a_2,\ldots)$ as an irreducible $SO(3,2)\times SO(N)$ component. In Table VI we have summarized the unitarity conditions. These turn out to be simpler than in the case of Osp(2,4). This is because Osp(2,4) is a Lie superalgebra of type I, whereas Osp(N,4) ($N\geqslant 3$) is a Lie superalgebra of type II [$B(m,2)$ or $D(m,2)$].

TABLE V. Decomposition of a shortened Osp(2,4) irrep $(E_0;s;y_0)$ with $E_0 = -y_0 + s + 1$ [or $(P;Q;y_0)$ with $y_0 + 2P - 1 = 0$] into representations (E,j,y) [or $(\bar{p},\bar{q},y)$] of $SO(3,2)\times U(1)$.

y value	(E,j)	$(\bar{p},\bar{q})$
y_0+2	(E_0+1,s)	$(P+\frac{1}{2},Q+\frac{1}{2})$
y_0+1	$(E_0+\frac{1}{2},s+\frac{1}{2}),\ (E_0+\frac{1}{2},s-\frac{1}{2}),\ (E_0+\frac{3}{2},s+\frac{1}{2})$	$(P,Q+\frac{1}{2}),\ (P+\frac{1}{2},Q),\ (P+\frac{1}{2},Q+1)$
y_0	$(E_0,s),\ (E_0+1,s+1),\ (E_0+1,s)$	$(P,Q),\ (P,Q+1),\ (P+\frac{1}{2},Q+\frac{1}{2})$
y_0-1	$(E_0+\frac{1}{2},s+\frac{1}{2})$	$(P,Q+\frac{1}{2})$

TABLE VI. Unitarity conditions for irreps $(E_0,s;a_1,a_2,\ldots)$ of Lie superalgebras Osp$(N,4)$.

N	unitarity condition
3	$E_0 \geqslant s + 2a_1 + 1$
4	$E_0 \geqslant s + 2a_1 + 2a_2 + 1$
5	$E_0 \geqslant s + a_1 + \frac{1}{2}a_2 + 1$
6	$E_0 \geqslant s + a_1 + \frac{1}{2}(a_2 + a_3) + 1$
7	$E_0 \geqslant s + a_1 + a_2 + \frac{1}{2}a_3 + 1$
8	$E_0 \geqslant s + a_1 + a_2 + \frac{1}{2}(a_3 + a_4) + 1$

For $N = 3$, the Dynkin label a_1 of SO(3) is usually replaced by its half-value $j = a_1/2$. Then j can be integral or half-integral, and the unitarity condition becomes

$$E_0 \geqslant s + j + 1, \tag{5.1}$$

in agreement with Ref. 5. Similarly, for $N = 4$, the Dynkin labels (a_1,a_2) of the SO(4) irrep are in general replaced by the labels $(j_1,j_2) = (a_1/2,a_2/2)$ of $\mathrm{SU}(2)\times\mathrm{SU}(2) \approx \mathrm{SO}(4)$. Then, one has

$$E_0 \geqslant s + j_1 + j_2 + 1.$$

Note that in the case of an equality in Table VI, the representation will be shortened.

ACKNOWLEDGMENT

The author would like to thank the referee for pointing out a substantial mistake in an earlier version of the paper.

[1] V. G. Kac, Adv. Math. **26**, 8 (1977).
[2] V. G. Kac, "Representations of classical Lie superalgebras," in *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978).
[3] H. Nicolai, in *Supersymmetry and Supergravity '84*, edited by B. de Witt, P. Fayet, and P. Van Nieuwenhuizen (World Scientific, Singapore, 1984), pp. 368–399.
[4] D. Z. Freedman and H. Nicolai, Nucl. Phys. B **237**, 342 (1984).
[5] A. Ceresole, P. Fré, and H. Nicolai, Class. Quantum. Gravit. **2**, 133 (1985).
[6] W. Heidenreich, Phys. Lett. B **110**, 461 (1982).
[7] I. Inaba, T. Maekawa, and T. Yamamoto, J. Math. Phys. **23**, 954 (1982).
[8] C. Fronsdal, Rev. Mod. Phys. **37**, 221 (1965); Phys. Rev. D **10**, 589 (1974).
[9] N. T. Evans, J. Math. Phys. **8**, 170 (1967).
[10] J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).
[11] J. W. B. Hughes, Ann. Israel Phys. Soc. **3**, 320 (1980).
[12] R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Gen. **16**, 473 (1983).
[13] Q-z. Han and H-z. Sun, Commun. Theor. Phys. **2**, 1137 (1983).
[14] J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).
[15] J. Van der Jeugt, J. Math. Phys. **26**, 913 (1985).

The asymptotic calculation of Wiener integrals occurring in quantum and statistical physics: A divergency problem

Paolo Moretti
Istituto di Ricerca sulle Onde Elettromagnetiche del Consiglio Nazionale delle Ricerche, Via Panciatichi 64, 50127 Firenze, Italy

(Received 18 August 1986; accepted for publication 3 December 1986)

The asymptotic expression of a Wiener functional integral using the eigenvalues of the associated Sturm–Liouville equation is obtained, correcting previous formulas giving a divergent result.

I. INTRODUCTION

In the last few years the functional integration techniques have been widely used in many fields of theoretical physics.[1,2] In spite of their intrinsic difficulty, it is possible to obtain in this way a variety of general results especially when we are concerned with nonlinear problems. For example, we are interested in the study of the statistical mechanics of one-dimensional solitary-wave-bearing scalar fields[3] and quantum tunneling in dissipative systems,[4] in the instanton approach to the particle quantum mechanics and to quantum field theories,[5,6] and in the first quantum corrections to the thermodynamics of nonlinear systems.[7,8]

The main objects we are interested in are the Feynman propagator in its Euclidean version (i.e., for imaginary time) and the partition function; they are essentially Wiener integrals, whose asymptotic form is often needed (the classical $\hbar \to 0$ limit in the propagator, and the low-temperature limit in the partition function). The Laplace method for ordinary integrals can be extended to the Wiener integrals,[9] so that the functional involved in the integration turns out to be a quadratic functional.

II. THE QUADRATIC APPROXIMATION

In order to fix the ideas, let us consider the Euclidean action for a particle[6]:

$$S[x] = \int_0^T \mathscr{L}(x,\dot{x})dt, \quad x(0) = 0, \quad x(T) = X, \tag{1}$$

with

$$\mathscr{L}(x,\dot{x}) = \tfrac{1}{2} m\dot{x}^2 + V(x) \tag{2}$$

and $\dot{x} = dx/dt$ (t = imaginary time), m is the mass of the particle, and $V(x)$ is the potential energy.

The Euclidean propagator takes the form[6]

$$G(X,T) = \int \mathscr{D}x \exp\left\{ -\frac{1}{\hbar} S[x] \right\}, \tag{3}$$

where $\mathscr{D}x$ is the symbolic expression of the differential in function space. In the small-$\hbar$ limit we can write

$$\begin{aligned} &G \simeq P \exp\left\{ -\frac{1}{\hbar} S[\bar{x}] \right\}, \\ &P = \int \mathscr{D}y \exp\left\{ -\frac{1}{2\hbar} \delta^2 S[y] \right\}, \end{aligned} \tag{4}$$

where $\bar{x}$ is the classical trajectory satisfying $\delta S = 0$, $y = x - \bar{x}$, so that $y(0) = y(T) = 0$, and $\delta^2 S$ is the second variation of S in the neighborhood of $\bar{x}$. It is a quadratic functional in $y, \dot{y}$, given by

$$\delta^2 S[y] = \int_0^T \left(\frac{\partial^2 \mathscr{L}}{\partial \dot{x}^2} \dot{y}^2 + 2 \frac{\partial^2 \mathscr{L}}{\partial x \, \partial \dot{x}} y\dot{y} + \frac{\partial^2 \mathscr{L}}{\partial x^2} y^2 \right) dt, \tag{5}$$

where the coefficients are calculated along $\bar{x}(t)$ and therefore are known functions of t. Using the simple form (2) for the Lagrangian, one obtains

$$\begin{aligned} &\delta^2 S[y] = m \int_0^T (\dot{y}^2 + W(t) y^2) dt, \\ &W(t) = \frac{1}{m} \frac{\partial^2 V}{\partial x^2}\bigg|_{x=\bar{x}}. \end{aligned} \tag{6}$$

As it is well known, G can be calculated in terms of the solution of an auxiliary differential equation involving $W(t)$ (Ref. 10). An alternative method is the following: a change of variables is performed by the expansion[11]

$$y(t) = \sum_n^{1 \cdots \infty} a_n y_n(t), \tag{7}$$

where the $y_n(t)$ are the orthonormal set of eigenfunctions of the Sturm–Liouville (Schrödinger-like) equation

$$\left[\frac{d^2}{dt^2} - W(t) + \lambda_n \right] y_n = 0, \quad y_n(0) = y_n(T) = 0. \tag{8}$$

In this way, a function $y(t)$ is represented by the discrete infinity of the coefficients a_n, rather than by the continuous infinity of the numbers $y(t)$, $0 \leqslant t \leqslant T$. It turns out that $\delta^2 S[y]$ is now expressed by[11]

$$\frac{1}{2\hbar} \delta^2 S[y] = \sum_n^{1 \cdots \infty} \lambda'_n a_n^2, \tag{9}$$

with

$$\lambda'_n = \alpha \lambda_n, \quad \alpha = m/2\hbar.$$

If J is the Jacobian of this change of variables, so that

$$\mathscr{D}y = J \prod_{n=1}^{\infty} \frac{da_n}{\sqrt{\pi}}, \tag{10}$$

the resulting asymptotic form of P is

$$P = J \prod_{n=1}^{\infty} (\lambda'_n)^{-1/2}. \tag{11}$$

In the case of negative and zero eigenvalues, this formula can still be applied, when the particular features of $W(t)$ are taken into account.[3,5]

 0022-2488/87/040765-04$02.50

III. THE PROBLEM OF THE DIVERGENCY

The previous arguments are well known, and Eq. (11) is used everywhere without problems. But a remark is necessary; in many applications, only the ratio of two expressions like (11), with different $W(t)$, is needed, so that an essential drawback is not apparent at first sight: this expression for P is in fact a divergent one.

As a concrete example, let us pose

$$V(x) = \tfrac{1}{2} m\omega^2 x^2 \tag{12}$$

so that $W(t) = \omega^2$ [see Eq. (6)]. In this case Eq. (4) is exact, and the standard approach for quadratic functionals gives[10]

$$P = \int \mathscr{D}y \exp\left\{ -\frac{1}{2\hbar}\delta^2 S[y] \right\} = \left(\frac{\alpha\omega}{\pi \sinh(\omega T)} \right)^{1/2}. \tag{13}$$

Let us try to use Eq. (11); from Eq. (8) one has

$$\lambda_n = \omega^2 + n^2\pi^2/T^2. \tag{14}$$

The Jacobian J is easily calculated: we write Eq. (10) in a discrete form, considering $y(t)$ as the limit of the step function $y(t_k)$, $t_k = kT/N$, $k = 1,2,\ldots, N-1$, for $N \to \infty$. The expansion (7) becomes [$y(k) = y(t_k)$, $y_n(k) = y_n(t_k)$]

$$y(k) = \sum_{n=1}^{N-1} a_n y_n(k). \tag{15}$$

Taking the expression of a functional integral as the limit for $N \to \infty$ of an N-dimensional integral[10]

$$\int \mathscr{D}y = \lim_{N\to\infty} A^{-N} \int \cdots \int dy(1)\, dy(2) \cdots dy(N-1),$$
$$A = (\pi T/\alpha N)^{1/2}, \tag{16}$$

and using Eq. (15), from which

$$\int \cdots \int dy(1) \cdots dy(N-1)$$
$$= \det\left(\frac{\partial y(i)}{\partial a_j} \right) \int \cdots \int da_1 \cdots da_{N-1}, \tag{17}$$

where

$$\det\left(\frac{\partial y(i)}{\partial a_j} \right) = \det \begin{pmatrix} y_1(1) & \cdots & y_1(N-1) \\ \vdots & \ddots & \vdots \\ y_{N-1}(1) & \cdots & y_{N-1}(N-1) \end{pmatrix}$$
$$= \Delta(N), \tag{18}$$

we obtain

$$J = \lim_{N\to\infty} J(N), \quad J(N) = A^{-N} \pi^{(N-1)/2} \Delta(N). \tag{19}$$

Here $\Delta(N)$ is calculated by multiplying the matrix in Eq. (18) by its transpose, and it is easily seen that it can be put in the form

$$\Delta(N) = (N/T)^{(N-1)/2} f(N), \tag{20}$$

where $\lim_{N\to\infty} f(N) = 1$, using the orthonormality of the y_n.

This is not a novel result[12] but, putting these expressions in Eq. (11), one has

$$P = \lim_{N\to\infty} J(N) \prod_{n=1}^{N-1} (\lambda'_n)^{-1/2}$$
$$= \lim_{N\to\infty} \left(\frac{\pi T}{\alpha N} \right)^{-N/2} \pi^{(N-1)/2} \left(\frac{N}{T} \right)^{(N-1)/2} \alpha^{-(N-1)/2}$$
$$\times \prod_{n=1}^{N-1} \left(\omega^2 + \frac{n^2\pi^2}{T^2} \right)^{-1/2}$$
$$= \lim_{N\to\infty} \left(\frac{\pi T}{\alpha N} \right)^{-N/2} \pi^{(N-1)/2} \left(\frac{N}{T} \right)^{(N-1)/2} \alpha^{-(N-1)/2}$$
$$\times \left(\frac{T}{\pi} \right)^{N-1} \frac{1}{(N-1)!} \prod_{n=1}^{N-1} \left(1 + \frac{\omega^2 T^2}{n^2\pi^2} \right)^{-1/2},$$

so that, accounting for Stirling's formula, P diverges to zero as

$$\left[\lim_{N\to\infty} \left(\frac{\pi}{2} \right)^{1/2} \left(\frac{e}{\pi} \right)^N \right] \left(\frac{\alpha\omega}{\pi \sinh(\omega T)} \right)^{1/2}. \tag{21}$$

This meaningless result shows clearly that Eq. (11) is not correct in the present form and a careful investigation is needed. Oddly in many important papers this formula is written without any explanation[3,5,6,13]; even if in the final result only the ratio of two functional integrals appears, it is necessary to show that the divergent factor is the same, irrespectively of the form of $W(t)$.

The key of the question is the following: Eqs. (7) and (8) can be used only if the limit $N \to \infty$ is performed, i.e., when $x(t)$, $y(t)$, are continuous functions in the interval $(0,T)$. On the contrary, since Eq. (11) is to be calculated as

$$P = \lim_{N\to\infty} J(N) \prod_{n=1}^{N-1} (\lambda'_n)^{-1/2}, \tag{22}$$

the λ_n to be considered are the eigenvalues of a $(N-1)\times(N-1)$ matrix, rather than of a differential equation, and are therefore functions of N. In other words, denoting by $\lambda_n(\infty)$ the eigenvalues of Eq. (8), one has

$$P = \lim_{N\to\infty} J(N) \prod_{n=1}^{N-1} (\lambda'_n(N))^{-1/2}$$
$$\neq \lim_{N\to\infty} J(N) \prod_{n=1}^{N-1} (\lambda'_n(\infty))^{-1/2}. \tag{23}$$

Our aim is to express P by the $\lambda_n(\infty)$ introducing an "effective" Jacobian $\tilde{J}(N)$ valid for every $W(t)$.

IV. THE CORRECT EXPRESSION FOR P AND THE "EFFECTIVE" JACOBIAN

We begin to write P in the usual "finite-mesh" form[10]

$$P = \lim_{N\to\infty} P_N,$$
$$P_N = A^{-N} \int \cdots \int dy_1 \cdots dy_{N-1} \tag{24}$$
$$\times \exp\left\{ -\alpha\epsilon \left[\sum_{j=0}^{N-1} \left(\frac{y_{j+1} - y_j}{\epsilon} \right)^2 + W_j y_j^2 \right] \right\},$$

where $\epsilon = T/N$, y_k is the same as $y(k)$ in Eq. (16) and $W_j = W(t_j)$, $t_j = \epsilon j$. The argument in the exponential is nothing but $(-(1/2\hbar)\delta^2 S(N))$ and can be written in the form[10]

$$\frac{1}{2\hbar}\delta^2 S(N)=\frac{\alpha}{\epsilon}\sum_{i,j=1}^{N-1}A_{ij}y_i y_j, \tag{25}$$

where

$$(A)=(A_0)+\epsilon^2(W). \tag{26}$$

Here (A_0) and (W) are matrices given by

$$(A_0)=\begin{pmatrix} 2,-1,0,0,\ldots,0 \\ -1,2,-1,0,\ldots,0 \\ \ddots \\ 0,0,\ldots,-1,2 \end{pmatrix},\quad (W)=\begin{pmatrix} W_1,0,0,\ldots,0 \\ 0,W_2,0,\ldots,0 \\ \ddots \\ 0,\ldots,W_{N-1} \end{pmatrix}. \tag{27}$$

Using the unitary matrix whose elements are[12]

$$B_{lm}=(2/N)^{1/2}\sin(\pi lm/N), \tag{28}$$

we go to the new coordinates a_k defined in

$$y_n=\sum_k B_{nk}a_k, \tag{29}$$

so that

$$P_N=A^{-N}\int\cdots\int da_1\cdots da_{N-1}\exp\left\{-\frac{\alpha}{\epsilon}\sum_{i,j=1}^{N-1}H_{ij}a_i a_j\right\}, \tag{30}$$

where

$$H_{ij}=(B^{\dagger}AB)_{ij}=\lambda_i^0\delta_{ij}+\epsilon^2 W_{ij},\quad W_{ij}=\sum_{l=1}^{N-1}W_l B_{il}B_{lj} \tag{31}$$

and

$$\lambda_k^0=4\sin^2(k\pi/2N) \tag{32}$$

are the eigenvalues of (A_0) (Ref. 12).

Posing $a_k\sqrt{\epsilon}=b_k$, it follows that

$$P_N=A^{-N}\epsilon^{-(N-1)/2}\int\cdots\int db_1\cdots db_{N-1}\times\exp\left\{-\frac{\alpha}{\epsilon^2}\sum_{i,j=1}^{N-1}H_{ij}b_i b_j\right\} \tag{33}$$

and we eventually obtain[10]

$$P_N=\{A^{-N}\epsilon^{-(N-1)/2}\pi^{(N-1)/2}\}\{\det[(\alpha/\epsilon^2)H_{ij}]\}^{-1/2}. \tag{34}$$

Denoting by h_i the eigenvalues of (H), we can perform the well-known expansion [see Eq. (31)]

$$h_i=\lambda_i^0+\epsilon^2 W_{ii}+\epsilon^4\sum_{j\neq i}\frac{|W_{ij}|^2}{\lambda_i^0-\lambda_j^0}+\cdots. \tag{35}$$

This expansion, of course, converges, accounting for the expression of the λ_k^0 [Eq. (32)] and of the W_{ij}, given by Eq. (31):

$$W_{ij}=\sum_{l=1}^{N-1}W_l B_{il}B_{lj}\underset{N\to\infty}{\to}\frac{1}{T}\int_0^T W(t)\left[\cos\left(\frac{\pi t(i-j)}{T}\right)-\cos\left(\frac{\pi t(i+j)}{T}\right)\right]dt \tag{36}$$

and representing essentially the Fourier components of $W(t)$. In this way, we can write

$$\det\left(\frac{\alpha}{\epsilon^2}H_{ij}\right)=\prod_{k=1}^{N-1}\left(\frac{\alpha}{\epsilon^2}h_k\right)=\prod_{k=1}^{N-1}\frac{\alpha}{\epsilon^2}\lambda_k^0\prod_{k=1}^{N-1}\left(1+\frac{\epsilon^2}{\lambda_k^0}W_{kk}+\sum_{j\neq k}\frac{\epsilon^2}{\lambda_k^0}\frac{\epsilon^2}{\lambda_k^0-\lambda_j^0}|W_{kj}|^2+\cdots\right). \tag{37}$$

The last step lies in performing the limit $N\to\infty$ in Eq. (34); inserting the explicit expression of λ_k^0 in Eq. (37), it is easily seen that in the second product the (lim) and (π) operations can be interchanged:

$$\frac{\lambda_k^0}{\epsilon^2}=\frac{4N^2}{T^2}\sin^2\left(\frac{k\pi}{2N}\right)\underset{N\to\infty}{\to}\frac{k^2\pi^2}{T^2}=\lambda_k^0(\infty). \tag{38}$$

The $\lambda_k^0(\infty)$ are the eigenvalues of the equation

$$\left[\frac{d^2}{dt^2}+\lambda_k^0(\infty)\right]y(t)=0,\quad y(0)=y(T)=0. \tag{39}$$

In this way we can write

$$\lim_{N\to\infty}\det[(\alpha/\epsilon^2)H_{ij}]=\lim_{N\to\infty}\prod_{k=1}^{N-1}\frac{\alpha}{\epsilon^2}\lambda_k^0\prod_{k=1}^{N-1}\left(1+\frac{W_{kk}}{\lambda_k^0(\infty)}+\sum_{j\neq k}\frac{1}{\lambda_k^0(\infty)}\times\frac{|W_{kj}|^2}{\lambda_k^0(\infty)-\lambda_j^0(\infty)}+\cdots\right)$$

$$=\lim_{N\to\infty}\prod_{k=1}^{N-1}\alpha\left(\frac{\lambda_k^0/\epsilon^2}{\lambda_k^0(\infty)}\right)\prod_{k=1}^{N-1}\left(\lambda_k^0(\infty)+W_{kk}+\sum_{j\neq k}\frac{|W_{kj}|^2}{\lambda_k^0(\infty)-\lambda_j^0(\infty)}+\cdots\right).$$

The last expansion is nothing but $\lambda_k(\infty)$, the k th eigenvalue of Eq. (8), and the limit becomes

$$\lim_{N\to\infty}\prod_{k=1}^{N-1}\frac{\lambda_k^0/\epsilon^2}{\lambda_k^0(\infty)}\prod_{k=1}^{N-1}\lambda_k'(\infty),\quad \lambda_k'(\infty)=\alpha\lambda_k(\infty). \tag{40}$$

We see therefore that P is written in a form similar to that given in Eq. (22); the first factor in brackets in Eq. (34) is nothing but the Jacobian $J(N)$ [see Eq. (19)], but the second one differs from the product of the $(\lambda_k'(\infty))^{-1/2}$ in the factor

$$\prod_{k=1}^{N-1}\epsilon\left(\frac{\lambda_k^0(\infty)}{\lambda_k^0}\right)^{1/2}=\prod_{k=1}^{N-1}\frac{k\pi}{2N\sin(k\pi/2N)}. \tag{41}$$

Using the trigonometric formula[14]

$$\prod_{k=1}^{N-1}\sin\left(\frac{k\pi}{2N}\right)=\frac{\sqrt{N}}{2^{N-1}}$$

and Stirling's approximation, we obtain

$$\prod_{k=1}^{N-1}\epsilon\left(\frac{\lambda_k^0(\infty)}{\lambda_k^0}\right)^{1/2}\underset{N\to\infty}{\to}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\pi}{e}\right)^N. \tag{42}$$

From Eqs. (19), (34), (40), and (42), the final result is

$$P = \lim_{N\to\infty} \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{\pi}{e}\right)^{N} J(N) \prod_{k}^{N-1} (\lambda'_k(\infty))^{-1/2}$$

$$= \lim_{N\to\infty} \tilde{J}(N) \prod_{k=1}^{N-1} (\lambda'_k(\infty))^{-1/2}. \tag{43}$$

In other words, Eq. (22) is applicable, provided that $J(N)$ is replaced by the "effective" Jacobian

$$\tilde{J}(N) = (2/\pi)^{1/2}(\pi/e)^{N} J(N), \tag{44}$$

irrespective of the particular form of $W(t)$. For example, if $W(t) = \omega^2$, Eq. (44) gives the correct result for P [see Eq. (21)].

V. CONCLUDING REMARKS

We have given here an expression for the asymptotic form of integrals in function space, in terms of the eigenvalues of the Schrödinger-like equation associated with the second variation of the functional. This expression corrects previous formulas and shows that the eigenvalues of this equation can still be used (without solving the more difficult eigenvalue problem for the functional in discrete form), provided that an "effective" Jacobian is introduced. This could appear only a conceptual improvement, since in actual calculations the ratio of such integrals is especially needed; but, in order to obtain the correct result, it is important to show that the change in the Jacobian is common to all the forms of functions $W(t)$.

[1] *Path Integrals and Their Applications in Quantum, Statistical and Solid State Physics*, edited by G. J. Papadopulos and J. T. Devreese (Plenum, New York, 1978).
[2] L. S. Shulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[3] K. M. Leung, Phys. Rev. B **26**, 226 (1982).
[4] D. Waxman and A. J. Leggett, Phys. Rev. B **32**, 4450 (1985).
[5] C. G. Callan and S. Coleman, Phys. Rev. D **16**, 1762 (1977).
[6] S. Coleman, in *The Whys of Subnuclear Physics*, Erice 1977, edited by A. Zichichi (Plenum, New York, 1979).
[7] R. P. Feynman, *Statistical Mechanics* (Benjamin, Reading, MA, 1972).
[8] R. Giachetti and V. Tognetti, Phys. Rev. Lett. **55**, 912 (1985).
[9] M. Schilder, Trans. Am. Math. Soc. **125**, 63 (1966).
[10] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
[11] Ref. 2, Chap. 12.
[12] Ref. 2, Chap. 14.
[13] E. Gildener and A. Patrascioiu, Phys. Rev. D **16**, 423 (1977).
[14] L. B. W. Jolley, *Summation of Series* (Chapman and Hall, London, 1925), pp. 170–171.

Comment on "Compact expression for Löwdin's alpha function" [J. Math. Phys. 26, 940 (1985)]

Noboru Suzuki[a)]
Department of Materials Science, The University of Electro-Communications, Chofu-Shi, Tokyo 182, Japan

(Received 18 October 1985; accepted for publication 3 December 1986)

Recently Antone [J. Math. Phys. **26**, 940 (1985)] has presented a compact expression for Löwdin's α function, and expected that, in calculation on the electronic properties of molecules and solids, it is the most convenient of all the expressions available in literature. But, in the present paper, it is shown that the expression for the α function obtained from the expansion formula of Silverstone and Moats [Phys. Rev. A **16**, 1731 (1977)] is remarkably more compact and, therefore, more convenient than Antone's expression. Though not visibly, it is proved with the aid of some manipulation that the formula of Silverstone and Moats can be reduced to the well-known expansion formula of a solid harmonic. In the special case where the center of the function to be expanded and that of the α function both are located on the z axis, some mistakes are found that lead to Antone's expression for the α function.

I. INTRODUCTION

A function of the form $F(R) = f(R)Y_L^M(\theta_R,\phi_R)$ is expanded in terms of radial functions and spherical harmonics at a new origin, a position from which is denoted by polar coordinates (r, θ, ϕ). Here the new origin is assumed to be located at the position displaced from the old origin by $-\mathbf{a}$, i.e., $\mathbf{R} = \mathbf{r} - \mathbf{a}$. Usually, this expansion radial function is called Löwdin's α function.

Recently, Antone has presented a compact expression for the α function, and expected that, in calculations on the electronic properties of molecules and solids, it is more convenient than any other expression available to date.[1] But, in his paper, it is not realized that the expansion formula[2] of Silverstone and Moats provides a very simple expression for the α function. Certainly, when $f(R) = R^L$, their formula is not reduced visibly to the well-known formula of a solid harmonic.[3] But, with the aid of some manipulation, the reduction can be accomplished. This will be shown in Sec. II. In Antone's paper, in the special case where the two origins both are placed on the same z axis, there are a few mistakes in the course which led to his expression for the α function. These will be pointed out clearly in Sec. III. Also, in Sec. III, the relation between the expression for the α function derived from the formula of Silverstone and Moats and the one given by Antone, will be manifested, and the latter expression in the above case will be given in an explicit form. In the last section, Sec. IV, will be described how much better the former expression is than the latter.

II. REDUCTION OF THE FORMULA OF SILVERSTONE AND MOATS TO THE EXPANSION FORMULA OF A SOLID HARMONIC

Silverstone and Moats obtained an expansion formula of $f(R)Y_L^M(\theta_R,\phi_R)$ by means of the Fourier-transform convolution theorem for overlap integrals.[2] The formula may be written as

$$f(R)Y_L^M(\theta_R,\phi_R) = \sum_{l=0}^{\infty} \sum_{\lambda=|L-l|}^{L+l} \sum_{m=-l}^{l} \left[\frac{(2l+1)(2\lambda+1)}{4\pi(2L+1)}\right]^{1/2} \times C(l\lambda L; m\, M-m) C(l\lambda L; 0\,0) v_{l\lambda L}(r,a) \times Y_l^m(\theta,\phi) Y_\lambda^{M-m}(\theta_a,\phi_a), \tag{1}$$

where

$$v_{l\lambda L}(r,a) = 2\pi(-1)^{(L-l+\lambda)/2} a^{-1} \sum_{K=0}^{(L+l+\lambda)/2} \sum_{k=0}^{[(L+l+\lambda)/2]-K} \frac{(L-K-\tfrac{1}{2})!(l-k-\tfrac{1}{2})!(-(L+l-\lambda)/2+K+k-\tfrac{1}{2})!}{K!k!((L+l+\lambda)/2-K-k)!} \times \left(\frac{r}{a}\right)^{2k-l-1} \int_{|r-a|}^{r+a} dR \left(\frac{R}{a}\right)^{2K-L+1} f(R). \tag{2}$$

Here $C(l\lambda L; m\, M-m)$ and $C(l\lambda L; 0\,0)$ are the Clebsch–Gordan coefficients.[4] The factorial for a half-integer has been defined by Eq. (1) of Ref. 5.

When $f(R) = R^L$, Eq. (1) should be reduced to the famous expansion formula of a solid harmonic.[3] This is proved in the following. Setting $f(R) \equiv R^L$ in Eq. (2) and then putting $R/a \equiv S$ and $r/a \equiv s$, we obtain

a) Present address: Institute for Communication Sciences, The University of Electro-Communications, Chofu-Shi, Tokyo 182, Japan.

 0022-2488/87/040769-05$02.50

$$v_{l\lambda L}(r,a)=2\pi(-1)^{(L-l+\lambda)/2}a^{L}\sum_{K=0}^{(L+l+\lambda)/2}\sum_{k=0}^{[(L+l+\lambda)/2]-K}\frac{(L-K-\tfrac{1}{2})!(l-k-\tfrac{1}{2})!(-(L+l-\lambda)/2+K+k-\tfrac{1}{2})!}{K!k!((L+l+\lambda)/2-K-k)!}$$
$$\times s^{2k-l-1}\int_{|s-1|}^{s+1}dS\,S^{2K+1}. \tag{3}$$

The definite integral in Eq. (3) proves to be expressed in terms of power series of s as

$$\int_{|s-1|}^{s+1}dS\,S^{2K+1}=\sum_{p=0}^{K}\frac{K!(K+\tfrac{1}{2})!}{p!(p+\tfrac{1}{2})!(K-p)!(K-p+\tfrac{1}{2})!}s^{2p+1} \tag{4}$$

by carrying out the integration over S and then manipulating the result. Introducing Eq. (4) into Eq. (3), changing the summation order some times, and then replacing the summation indices by new ones, we arrive at

$$v_{l\lambda L}(r,a)=2\pi(-1)^{(L-l+\lambda)/2}a^{L}\sum_{n=0}^{(L+l+\lambda)/2}I_{\lambda,n}s^{2n-l}, \tag{3'}$$

where

$$I_{\lambda,n}=\sum_{p=0}^{(L+l+\lambda)/2-n}\sum_{q=0}^{n}\frac{(-(L+l-\lambda)/2+n+p-\tfrac{1}{2})!(p+q+\tfrac{1}{2})!(L-p-q-\tfrac{1}{2})!(l-n+q-\tfrac{1}{2})!}{p!(p+\tfrac{1}{2})!((L+l+\lambda)/2-n-p)!q!(q+\tfrac{1}{2})!(n-q)!}. \tag{5}$$

Here the use of the following relation obtained by setting $\mu\equiv\max\{q,L+1\}$, $\rho\equiv\min\{q,L+1\}$, and $\nu\equiv -L+p-\tfrac{1}{2}+\rho$ in Eq. (B1) of Ref. 5:

$$(p+q+\tfrac{1}{2})!(L-p-q-\tfrac{1}{2})!$$
$$=(-1)^{L-p-q}\frac{(p+q+\tfrac{1}{2})!}{(-L+p+q-\tfrac{1}{2})!}$$
$$=(-1)^{L-q-q}(p+\tfrac{1}{2})!q!(L+1)!\sum_{t=0}^{\min\{q,L+1\}}\left[t!(q-t)!(L+1-t)!\left(-L+p-\frac{1}{2}+t\right)!\right]^{-1}, \tag{6}$$

leads to another expression for $I_{\lambda,n}$,

$$I_{\lambda,n}=(-1)^{L}(L+1)!\sum_{t=0}^{\min\{n,L+1\}}\frac{J_t^{(1)}J_t^{(2)}}{t!(L+1-t)!}, \tag{5'}$$

where the abbreviations,

$$J_t^{(1)}=\sum_{p=0}^{(L+l+\lambda)/2-n}(-1)^{p}$$
$$\times\frac{(-(L+l-\lambda)/2+n+p-\tfrac{1}{2})!}{p!((L+l+\lambda)/2-n-p)!(-L+p-\tfrac{1}{2}+t)!} \tag{7}$$

and

$$J_t^{(2)}=\sum_{q=t}^{n}(-1)^{q}\frac{(l-n+q-\tfrac{1}{2})!}{(q+\tfrac{1}{2})!(n-q)!(q-t)!} \tag{8}$$

have been used. Therefore if either of $J_t^{(1)}$ and $J_t^{(2)}$ vanishes, so does their product. Now we carry out the summation of Eqs. (7) and (8). Equation (7) is rewritten as

$$J_t^{(1)}=(-1)^{-(L+l-\lambda)/2+n}$$
$$\times\sum_{p=0}^{(L+l+\lambda)/2-n}\left[p!\left(\frac{L+l-\lambda}{2}-n-\frac{1}{2}-p\right)!\right.$$
$$\left.\times\left(\frac{L+l+\lambda}{2}-n-p\right)!\left(-L+t-\frac{1}{2}+p\right)!\right]^{-1}. \tag{7'}$$

Putting $(L+l-\lambda)/2-n-\tfrac{1}{2}\equiv\mu$, $(L+l+\lambda)/2-n\equiv\rho$, and $-L+t-\tfrac{1}{2}\equiv\nu-\rho$ in Eq. (B1) of Ref. 5, one finds that $\mu+\nu=l-2n+t-1$ and $\mu+\nu-\rho=(-L+l-\lambda)/2-n+t-1$. Since $t\leqslant n$ and $\lambda\geqslant|L-l|$, obviously $\mu+\nu-\rho<0$. Therefore from the note on Eq. (B1) of Ref. 5, if $\mu+\nu=l-2n+t-1\geqslant 0$, $J_t^{(1)}$ vanishes. Otherwise, i.e., for $2n-l-t\geqslant 0$, $J_t^{(1)}$ is summed to

$$J_t^{(1)}=(-1)^{\lambda}((L-l+\lambda)/2+n-t)!$$
$$\times[((L+l-\lambda)/2-n-\tfrac{1}{2})!$$
$$\times((-L+l+\lambda)/2-n+t-\tfrac{1}{2})!$$
$$\times((L+l+\lambda)/2-n)!(2n-l-t)!]^{-1}. \tag{7''}$$

On the other hand, $J_t^{(2)}$ may be rewritten as

$$J_t^{(2)}=(-1)^{n-l}$$
$$\times\sum_{q=0}^{n-t}\left[q!\left(n-l-t-\frac{1}{2}-q\right)!\right.$$
$$\left.\times(n-t-q)!\left(t+\frac{1}{2}+q\right)!\right]^{-1}. \tag{8'}$$

When $n-l-t-\tfrac{1}{2}$, $n-t$, and $t+\tfrac{1}{2}$ are set equal to μ, ρ, and $\nu-\rho$, respectively, in Eq. (B1) of Ref. 5, it turns out

that $\mu+\nu=2n-l-t$ and $\mu+\nu-\rho=n-l$. Because $\mu+\nu=2n-l-t$ must be larger than or equal to zero in order that $J_t^{(1)}$ takes a nonzero value, the condition that $\mu+\nu-p=n-l\geqslant 0$ is necessary for $J_t^{(2)}$ to remain nonzero. For $n-l\geqslant 0$, $J_t^{(2)}$ is summed to

$$J_t^{(2)}=(-1)^{n-l}\times\frac{(2n-l-t)!}{(n-l-t-\frac{1}{2})!(n+\frac{1}{2})!(n-t)!(n-l)!}. \quad (8'')$$

Introduction of Eqs. (7'') and (8'') into Eq. (5'), and then insertion of the relation

$$\frac{((L-l+\lambda)/2+n-t)!}{(n-l-t-\frac{1}{2})!(n-t)!}=\left(\frac{L+l+\lambda}{2}+\frac{1}{2}\right)!\left(\frac{L-l+\lambda}{2}\right)!\times\sum_{p=0}^{n-t}\left[p!\left(\frac{L+l+\lambda}{2}+\frac{1}{2}-p\right)!\times(n-t-p)!\left(-l-\frac{1}{2}+p\right)!\right]^{-1}, \quad (9)$$

which is derived by setting $\mu\equiv(L+l+\lambda)/2+\frac{1}{2}$, $\nu\equiv n-l-t-\frac{1}{2}$, and $\rho\equiv n-t$ in Eq. (B1) of Ref. 5, into the resulting equation lead to

$$I_{\lambda,n}=(-1)^n\frac{(L+1)!((L+l+\lambda)/2+\frac{1}{2})!((L-l+\lambda)/2)!}{((L+l-\lambda)/2-n-\frac{1}{2})!((L+l+\lambda)/2-n)!(n+\frac{1}{2})!(n-l)!}\times\sum_{p=0}^{n}\left[p!\left(\frac{L+l+\lambda}{2}+\frac{1}{2}-p\right)!\left(-l-\frac{1}{2}+p\right)!\right]^{-1}\times\sum_{t=0}^{\min\{n-p,L+1\}}\left[t!(L+1-t)!(n-p-t)!\left(\frac{-L+l+\lambda}{2}-n-\frac{1}{2}+t\right)!\right]^{-1}. \quad (5'')$$

Letting $\max\{n-p,L+1\}\equiv\mu$, $\min\{n-p,L+1\}\equiv\nu$, and $(-L+l+\lambda)/2-n-\frac{1}{2}\equiv\nu-\rho$ in Eq. (B1) of Ref. 5, one finds that the sum over t is equal to

$$\frac{((L+l+\lambda)/2-p+\frac{1}{2})!}{(L+1)!((-L+l+\lambda)/2-p-\frac{1}{2})!(n-p)!((L+l+\lambda)/2-n+\frac{1}{2})!}.$$

Substituting this result into Eq. (5''), we arrive at

$$I_{\lambda,n}=(-1)^n\frac{((L+l+\lambda)/2+\frac{1}{2})!((L-l+\lambda)/2)!}{((L+l-\lambda)/2-n-\frac{1}{2})!((L+l+\lambda)/2-n)!(n+\frac{1}{2})!(n-l)!((L+l+\lambda)/2-n+\frac{1}{2})!}\times\sum_{p=0}^{n}\left[p!\left(\frac{-L+l+\lambda}{2}-\frac{1}{2}-p\right)!(n-p)!\left(-l-\frac{1}{2}+p\right)!\right]^{-1}. \quad (5''')$$

If $(-L+l+\lambda)/2-\frac{1}{2}$, n, and $-l-\frac{1}{2}$ are set μ, ρ, and $\nu-\rho$, respectively, in Eq. (B1) of Ref. 5, then $\mu+\nu=(-L-l+\lambda)/2+n-1$ and $\mu+\nu-\rho=(-L-l+\lambda)/2-1$. Since $\lambda\leqslant L+l$, obviously $\mu+\nu-\rho<0$. Therefore from the note on Eq. (B1) of Ref. 5, if $\mu+\nu=(-L-l+\lambda)/2+n-1\geqslant 0$, the sum in Eq. (5''') vanishes. Otherwise, i.e., for $(L+l-\lambda)/2-n\geqslant 0$, the sum proves to be equal to

$$(-1)^n\frac{((L+l-\lambda/2))!}{((-L+l+\lambda)/2-\frac{1}{2})!(n-l-\frac{1}{2})!n!((L+l-\lambda)/2-n)!}.$$

Connecting $(L+l-\lambda)/2-n\geqslant 0$ with the condition $n-l\geqslant 0$ in order that $J_t^{(1)}J_t^{(2)}$ takes a nonzero value, we find that the permissible values of n are limited to the range from l to $(L+l-\lambda)/2$. On the other hand, λ runs from $|L-l|$ to $L+l$ by step 2, and thus the value of λ which gives $I_{\lambda,n}$ a nonzero value is restricted only to $L-l$. Therefore the allowed value of n is found to be only l.

Consequently, Eq. (5''') is rewritten as

$$I_{\lambda,n}=\delta_{\lambda,L-l}\delta_{n,l}\frac{(L+\frac{1}{2})!}{(l+\frac{1}{2})!(L-l+\frac{1}{2})!}=\delta_{\lambda,L-l}\delta_{n,l}\frac{2(2L+1)!}{(2l+1)!(2L-2l+1)!}\frac{l!(L-l)!}{L!}. \quad (5'''')$$

Then introduction of the last equality of Eq. (5'''') into Eq. (3') yields the single-term expression for $v_{lL}(r,a)$,

$$v_{lL}(r,a)=\delta_{\lambda,L-l}4\pi(-1)^{L-l}a^L\times\frac{(2L+1)!}{(2l+1)!(2L-2l+1)!}\frac{l!(L-l)!}{L!}s^l. \quad (3'')$$

Further, by putting Eq. (3'') and the expression for $C(l\,L-l\,L;0\,0)$,

$$C(l\,L-l\,L;0\,0)=[(2l)!(2L-2l)!/(2L)!]^{1/2}L!/l!(L-l)!, \quad (10)$$

in Eq. (1), we obtain a simple expression for $R^L\times Y_L^M(\theta_R,\phi_R)$:

$$R^L Y_L^M(\theta_R,\phi_R)=\sum_{l=0}^{L}\sum_{m=M-L+l}^{M+L-l}(-1)^{L-l}\times\left[\frac{4\pi(2L+1)!}{(2l+1)!(2L-2l+1)!}\right]^{1/2}\times C(l\,L-l\,L;m\,M-m)\times r^l Y_l^m(\theta,\phi)a^{L-l}Y_{L-l}^{M-m}(\theta_a,\phi_a). \quad (11)$$

This is no other than the expansion formula of a solid harmonic.[3] Here note that, if the new origin is placed at the position displaced from the old origin by **a** (in Antone's notation of Ref. 1, $\mathbf{r}_2$), not $-\mathbf{a}$, as done in Ref. 1, the sign $(-1)^{L-l}$ in Eq. (11) does not appear.

When $f(R)$ is a power of R, $v_{l\lambda L}(r,a)$ can be also expressed by Eq. (27) of Ref. 6, which is derived by integrating the right-hand side of Eq. (5) of Ref. 2 in the different order from the order in which the integration led to Eq. (2). The use of Eq. (27) of Ref. 6 leads directly to Eq. (11), which should be noted.

III. RELATION BETWEEN THE EXPRESSION FOR THE α FUNCTION DERIVED FROM THE FORMULA OF SILVERSTONE AND MOATS AND THE ONE GIVEN BY ANTONE

Comparing the expansion formula [Eq. (1)] of Silverstone and Moats with Eq. (1) of Ref. 1, which defines the α function, denoted by $\alpha(l,\lambda,L,r,a,f)$ [in Antone's notation, $\alpha(l_1,l_2,L,r_1,r_2,f)$], one finds that generally the α function can be also expressed in the form

$$\alpha(l,\lambda,L,r,a,f)=[(2l+1)(2\lambda+1)/4\pi(2L+1)]^{1/2}C(l\lambda L;0\,0)\times v_{l\lambda L}(r,a). \quad (12)$$

This is the very result obtained by applying the Fourier-transform convolution theorem to the integral over Ω_1 in Eq. (12) of Ref. 1, which gives an integral form of the α function, and by then using the orthonormality of the spherical harmonics defined in the angular coordinates Ω_2.

In Ref. 1 the explicit form of the α function is given by Eq. (23) or (24). Its use is valid for a general case, but inappropriate for the special case where $\mathbf{r}_2=(r_2,0,\phi_2)$ [in the present notation $\mathbf{a}=(a,0,\phi_a)$]. In the special case, the choice of $m=L$ and $m_1=l_1$ (in the present notation, $M=L$ and $m=l$) in Eq. (1) of Ref. 1, which led to Eq. (12) of Ref. 1, cannot be allowed unless $L=l_1$ (in the present notation, $L=l$), because of relation (6) of Ref. 1. Furthermore, in the above case, such an integral over θ_2 (in the present notation θ_a) as seen in Eq. (34) of Ref. 1, never yields because $\theta_2=0$, and, therefore, the relation expressed by Eq. (30) of Ref. 1, cannot be utilized. This means that, in the special case, $\alpha(l_1,l_2,L,r_1,r_2,f)$ remains nonzero even if $l_2\neq L-l_1$ (in the present notation, $\lambda\neq L-l$) as in many other cases.

In the above case, the α function derived in Antone's manner, may be expressed as

$$\alpha^{(A)}(l,\lambda,L,r,a,f)=(4\pi/(2\lambda+1))^{1/2}[C(l\lambda L;L_<\,0)]^{-1}\times\int_0^{\pi}\sin\theta\,d\theta\int_0^{2\pi}d\phi\times Y_l^{L_<*}(\theta,\phi)f(R)Y_L^{L_<}(\theta_R,\phi), \quad (13)$$

where $L_<$ denotes the lesser of L and l. Here the superscript (A) has been attached to the symbol α to distinguish the α function given by Eq. (13) from the one expressed by Eq. (12). In any case the expression for the α function, Eq. (12), is unchanged unless the form of $f(R)$ varies. On the contrary, the expression derived in Antone's manner, Eq. (23) or (24) of Ref. 1, varies case by case. For instance, $\alpha^{(A)}(l,\lambda,L,r,a,f)$ expressed by Eq. (13), is not equal to $\alpha(l,\lambda,L,r,a,f)$. The relation between them can be obtained by applying the Fourier-transform convolution theorem to the integral in Eq. (13). It may be written as

$$\alpha^{(A)}(l,\lambda,L,r,a,f)=\sum_{\mu=|L-l|}^{L+l}{}^{2}\left(\frac{2\mu+1}{2\lambda+1}\right)^{1/2}\frac{C(l\mu L;L_<\,0)}{C(l\lambda L;L_<\,0)}\times\alpha(l,\mu,L,r,a,f). \quad (14)$$

Needless to say, the explicit form of $\alpha^{(A)}(l,\lambda,L,r,a,f)$ can be obtained by carrying out the same procedure as used in Ref. 1. The course is, however, very complicated. In contrast, introduction of the explicit form of $C(l\mu L;0\,0)$ and the expression for $v_{l\lambda L}(r,a)$ given by Eq. (2), into Eq. (12) leads to the explicit form of $\alpha^{(A)}(l,\lambda,L,r,a,f)$ more easily. It is written as

$$\alpha^{(A)}(l,\lambda,L,r,a,f)=(-1)^{|L-l|}\frac{(L_<+\frac{1}{2})!}{(2L_<+1)!}\times\left[4\pi\frac{(2l+1)(L+l-\lambda)!(-|L-l|+\lambda)!(L+l+\lambda+1)!}{(2\lambda+1)(|L-l|+\lambda)!}\right]^{1/2}\times\sum_{K=0}^{L+l}\sum_{k=0}^{L+l-K}\frac{(L-K-\frac{1}{2})!(l-k-\frac{1}{2})!(-L_<+K+k-\frac{1}{2})!}{K!k!(L+l-K-k)!}\times\left(\frac{r}{a}\right)^{2k-l-1}\int_{|r-a|}^{r+a}dR\left(\frac{R}{a}\right)^{2K-L+1}f(R). \quad (15)$$

It goes without saying that the respective sums of $C(l\lambda L;M\,0)((2\lambda+1)/4\pi)^{1/2}\alpha^{(A)}(l,\lambda,L,r,a,f)$ and $C(l\lambda L;M\,0)((2\lambda+1)/4\pi)^{1/2}\alpha(l,\lambda,L,r,a,f)$ over λ prove to be equal to another α function $(1/r)\alpha_l(f\,L\,M\,|\,l)$ which has been defined by Eq. (2.1) of Ref. 7.

IV. CONCLUDING REMARKS

Certainly, in the case of $f(R)=R^L$, Antone's expression for the α function, Eq. (23) or (24) of Ref. 1, has the advantage of being reduced visibly to the expansion formula

of a solid harmonic over the expression derived from the formula [Eq. (12)] of Silverstone and Moats. But the former expression is rather complicated, and so, in calculation on the electronic properties of molecules and solids, it does not seem to be more convenient than any other appearing in the literature, because the radial part of the atomic functions used frequently in the calculation, never take the form of R^L. On the other hand, the latter expression is very simple, thanks to the application of the Fourier-transform convolution theorem for overlap integrals. No more simplified expression than it can be obtained unless the explicit form of $f(R)$ is given. Comparison of the forms of the two expressions leads readily to the conclusion that the latter expression is much more compact than the former. Since the simplicity of the latter expression is independent of the form of $f(R)$, it can be expected to be utilized conveniently whenever one needs the calculation of the α function.

[1]A. A. Antone, J. Math. Phys. **26**, 940 (1985).
[2]H. J. Silverstone and R. K. Moats, Phys. Rev. A **16**, 1731 (1977).
[3]For example, see p. 151 of D. M. Brink and G. R. Satchler, *Angular Momentum* (Oxford U. P., London, 1968).
[4]For example, M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[5]N. Suzuki, J. Math. Phys. **26**, 3193 (1985).
[6]K. G. Kay, H. D. Todd, and H. J. Silverstone, J. Chem. Phys. **51**, 2359 (1969).
[7]N. Suzuki, J. Math. Phys. **25**, 1133 (1984).

A class of self-dual solutions for SU(2) gauge fields on Euclidean space

Bernard Leaute
Institut Henri Poincaré, Laboratoire de Physique Théorique, 11, rue P. et M. Curie, 75231 Paris Cedex 05, France

Gilbert Marcilhacy
Université Paris VI, Laboratoire de Physique Théorique, 4, Place Jussieu, 75252 Paris Cedex 05, France

(Received 6 February 1986; accepted for publication 26 November 1986)

By applying the method of separation of variables a particular class of solutions, depending on the four variables (x_μ) of Yang's equation in the R gauge for self-dual SU(2) gauge fields on Euclidean four-dimensional flat space, is obtained. These solutions are parametrized by a particular form of the fifth Painlevé transcendent. For a specific choice of certain constants, there is degeneration, and the solutions functionally depend on elementary functions.

I. INTRODUCTION

Yang[1] has reduced the system of equations for self-dual SU(2) gauge fields on Euclidean four-dimensional flat space to the following equations:

$$\Phi(\Phi_{\xi\bar{\xi}}+\Phi_{\eta\bar{\eta}})-\Phi_\xi\Phi_{\bar{\xi}}-\Phi_\eta\Phi_{\bar{\eta}}+\Psi_\xi\bar{\Psi}_{\bar{\xi}}+\Psi_\eta\bar{\Psi}_{\bar{\eta}}=0\,, \tag{1a}$$

$$\Phi(\Psi_{\xi\bar{\xi}}+\Psi_{\eta\bar{\eta}})-2\Psi_\xi\Phi_{\bar{\xi}}-2\Psi_\eta\Phi_{\bar{\eta}}=0 \quad \text{(and the complex conjugate)}\,, \tag{1b}$$

where Φ denotes a real function, Ψ a complex function of the variables ξ, $\bar{\xi}$, η, $\bar{\eta}$, the overbar denotes complex conjugation, and

$$\sqrt{2}\xi=x_1+ix_2\,,\quad \sqrt{2}\eta=x_3-ix_4\,. \tag{2}$$

Various solutions to these equations have already been proposed in the first place by Yang himself, thanks to an appropriate ansatz. Ray[2] carrying on this way has broadened the scope on the results established by Yang.

By a method founded on the separation of variables we have also obtained one class of solutions connected with particular form of the fifth Painlevé transcendent.[3] The aim of the present paper is to show that the same process can be extended in a such manner that the solutions calculated depend effectively on the four variables $x_\mu(x_1,x_2,x_3,x_4)$. We can, following Takeno,[4] give another formulation of the system (1), useful for the development of the process. Let

$$\Phi\Delta\Phi-(\nabla\Phi)^2+\nabla\Psi\cdot\nabla\bar{\Psi}=-i\mathscr{P}(\Psi,\bar{\Psi})\,, \tag{3a}$$

$$\Phi\Delta\Psi-2\nabla\Phi\cdot\nabla\Psi=2i\mathscr{P}(\Psi,\Phi) \text{ and c.c.}, \tag{3b}$$

where ∇ and Δ are, respectively, the gradient and Laplacian operators in the Euclidean four-dimensional flat space metric and the symbol $\mathscr{P}$ is defined by

$$\mathscr{P}(\alpha,\beta)=(\alpha_{x_1}\beta_{x_2}-\alpha_{x_2}\beta_{x_1})-(\alpha_{x_3}\beta_{x_4}-\alpha_{x_4}\beta_{x_3}) \tag{4}$$

for each pair of functions α and β of the variables (x_1,x_2,x_3,x_4). The system (3a) and (3b) can be written in terms of real functions by means of the change

$$\Psi=\chi\exp(i\lambda)\,, \tag{5}$$

which leads to

$$\Phi\Delta\Phi-(\nabla\Phi)^2+(\nabla\chi)^2+\chi^2(\nabla\lambda)^2=2\chi\mathscr{P}(\lambda,\chi)\,, \tag{6a}$$

$$\Phi[\Delta\chi-\chi(\nabla\lambda)^2]-2\nabla\chi\cdot\nabla\Phi=-2\chi\mathscr{P}(\lambda,\Phi)\,, \tag{6b}$$

$$\Phi(\chi\Delta\lambda+2\nabla\chi\cdot\nabla\lambda)-2\chi\nabla\lambda\cdot\nabla\Phi=2\mathscr{P}(\chi,\Phi)\,. \tag{6c}$$

We observe that if the phase function λ is taken as constant and if $\mathscr{P}(\chi,\Phi)=0$ the system is reduced to

$$(\mathrm{Re}\,\mathscr{E})\Delta\mathscr{E}=(\nabla\mathscr{E})^2 \text{ and c.c.}, \tag{7}$$

where it is set

$$\mathscr{E}\equiv\Phi+i\chi\,. \tag{8}$$

In the case where the functions Φ and χ depend only on the variables $\rho\equiv(x_1^2+x_2^2)^{1/2}$ and $z=x_3$, (7) is a form of the Ernst system.[5] The link between the Ernst system and particular solutions of the SU(2) self-dual gauge fields equations has been well noted by Witten.[6]

The class of solutions of (6a)–(6c), which we propose to determine here, is of the more general type; they involve a nonconstant phase function and depend on the variables (ρ,x_3,x_4).

II. DETERMINATION OF THE SOLUTIONS RESEARCHED

A. General analysis

Following the method of separation of variables employed here, we investigate solutions of the system (6a)–(6c) in the form

$$\begin{aligned}\Phi(x_\mu)&=F(\rho)K(x_3,x_4)\,,\\ \Psi(x_\mu)&=G(\rho)L(x_3,x_4)\,,\end{aligned} \tag{9}$$

or, in terms of purely real functions as in Ref. 3

$$\begin{aligned}\Phi(x_\mu)&=f(\rho)U(x_3,x_4)\,,\\ \chi(x_\mu)&=[g^2(\rho)+h^2(\rho)]^{1/2}U(x_3,x_4)\,,\\ \lambda(x_\mu)&=\arctan(g/h)+V(x_3,x_4)\,,\end{aligned} \tag{10}$$

where f, g, and h are functions with real values of the variable $\rho\equiv(x_1^2+x_2^2)^{1/2}$ while U and V are also real functions of which will be fixed in the next development. The insertion of these expressions in (6a)–(6c) gives

 0022-2488/87/040774-03$02.50

$$f(f''+(1/\rho)f')-f'^2+g'^2+h'^2+f^2[U^{-1}(U_{x_3x_3}+U_{x_4x_4})$$
$$-U^{-2}(U_{x_3}^2+U_{x_4}^2)]+(g^2+h^2)[(U^{-1}U_{x_3}-V_{x_4})^2+(U^{-1}U_{x_4}+V_{x_3})^2]=0, \tag{11a}$$
$$f[gg''+hh''+(1/\rho)(gg'+hh')]-2f'(gg'+hh')f(g^2+h^2)[U^{-1}(U_{x_3x_3}+U_{x_4x_4})-U^{-2}(U_{x_3}^2+U_{x_4}^2)]$$
$$-f(g^2+h^2)[(U^{-1}U_{x_3}-V_{x_4})^2+(U^{-1}U_{x_4}+V_{x_3}^2)^2]=0, \tag{11b}$$
$$hg''-zh''+(1/\rho)(hg'-gh')-2(f'/f)(hg'-gh')-(g^2+h^2)(V_{x_3x_3}+V_{x_4x_4})=0, \tag{11c}$$

with the notations $f'=df/d\rho$, $g'=dg/d\rho$,... .

This last equation can be separated only if the function $V(x_3,x_4)$ is bilinear in x_3 and x_4. However, the examination of the separability of the previous Eqs. (11a) and (11b) shows that in fact the functions U and V must necessarily be written as

$$U=\exp(ax_3+bx_4), \tag{12}$$
$$V=cx_3+dx_4, \tag{13}$$

a, b, c, and d being real constants (nonzero). It follows from this result that by a simple linear combination of (11a) and (11b) we may deduce the following equation:

$$f[(ff''+gg''+hh'')+(1/\rho)(ff'+gg'+hh')]$$
$$-f(f'^2-g'^2-h'^2)-2f'(gg'+hh')=0,$$

which leads to the first integral

$$(f^2+g^2+h^2)'=(A/\rho)f^2 \quad (A\neq 0), \tag{14}$$

where A is a constant of integration. Equation (11c) also is easily integrated under the form

$$hg'-gh'=(B/\rho)f^2 \quad (B\neq 0), \tag{15}$$

where B is a new constant of integration.

Now it is suitable to take these results into account in Eq. (11a), which, consequently to (12) and (13), can be written as

$$f\left(f''+\frac{1}{\rho}f'\right)-f'^2+g'^2+h'^2$$
$$+(g^2+h^2)[(a-d)^2+(b+c)^2]=0. \tag{16}$$

The effective insertion of (14) and (15) in this equation leads, after a rather tedious computation, to a nonlinear differential equation of the third order governing the function $Q\equiv f^2+g^2+h^2$. This equation is reducible to one of the second order for the function $X=Q'/Q$, and it is written

$$X''+\frac{1}{\rho}X'-\left[\frac{1}{X}-\frac{1}{2(A-X)}\right]X'^2$$
$$+\frac{X^2}{\rho^2}\left[\frac{2B^2}{A-X}+\frac{A-X}{2}\right]$$
$$+2[(a-d)^2+(b+c)^2](A-X)=0. \tag{17}$$

Making the changes

$$w=1-X/A, \quad \text{with } A\neq 0 \text{ and } x=\rho^2. \tag{18}$$

We finally find the equation

$$\frac{d^2w}{dx^2}+\frac{1}{x}\frac{dw}{dx}-\left(\frac{1}{2w}+\frac{1}{w-1}\right)\left(\frac{dw}{dx}\right)^2$$
$$-\frac{(1-w)^2}{2x^2}\left(\frac{A^2}{4}w+\frac{B^2}{w}\right)$$
$$-\frac{(a-d)^2+(b+c)^2}{2x}w=0, \tag{19}$$

in which we recognize a particular form of the nonlinear differential equation defining the *fifth Painlevé transcendent*.[7] With the values of the parameters

$$\alpha=-\frac{A^2}{8}, \quad \beta=-\frac{B^2}{2},$$
$$\gamma=-\tfrac{1}{2}[(a-d)^2+(b+c)^2], \quad \delta=0, \tag{20}$$
$$\text{with } a-d \text{ and } b+c\neq 0.$$

Comparing this with the previous work[3] we see that the dependence on x_4, introduced here, does not change the particular form of the Painlevé equation that we obtain in both situations. The difference lies in the expression of the parameter γ which involves the constants a, b, c, and d, which appear in (12) and (13). From their respective diverse values various original particular cases will result which will be examined subsequently.

Returning to Eqs. (14) and (15), it is possible to obtain functions f, g, and h parametrized by $w(\rho)$, the Painlevé transcendent defined in (19). We then have

$$f=[A(1-w)]^{1/2}\exp\int_{\rho_0}^{\rho}\frac{A(1-w)}{2\rho}d\rho,$$

and

$$g^2+h^2=Aw\exp\int_{\rho_0}^{\rho}\frac{A(1-w)}{\rho}d\rho,$$
$$h+ig=(g^2+h^2)^{1/2}\exp\left\{i\int_{\rho_0}^{\rho}\frac{B(1-w)}{\rho w}d\rho\right\}.$$

Now we are in possession of all elements to give entire expressions for the functions Φ and Ψ, which were introduced in the beginning, as functionals of the fifth Painlevé transcendent defined by (19); we find the following results:

$$\Phi(x_\mu)=[A(1-w)]^{1/2}$$
$$\times\exp\left\{ax_3+bx_4+\int_{\rho_0}^{\rho}\frac{A(1-w)}{2\rho}d\rho\right\}, \tag{21}$$

$$\Psi(x_\mu)=(Aw)^{1/2}\exp\left\{(a+ic)x_3+(b+id)x_4\right.$$
$$\left.+\int_{\rho_0}^{\rho}\left(\frac{A}{2}+i\frac{B}{w}\right)\frac{1-w}{\rho}d\rho\right\}, \tag{22}$$

for the general case

$$A \text{ and } B\neq 0, \quad a,b,c,d\neq 0,$$
$$a-d\neq 0, \quad b+c\neq 0.$$

Because of the presence of the square roots the result is valid only for the intervals of the variable ρ where $w(\rho)$ is such as

$$A>Aw>0.$$

B. Particular cases

They correspond to the particular values of the constants A, B, a, b, c, and d that we excluded in the above analysis.

The three cases defined by (i) $A=0, B\neq 0$; (ii) $A\neq 0$, $B=0$; and (iii) $A=B=0$ could be discussed in a similar manner to that used in the analogous cases encountered in Ref. 3. We observe that the corresponding solutions $\Phi(x_\mu)$, $\Psi(x_\mu)$ still depend on the particular forms $w(\rho)$ of the fifth Painlevé transcendent—*without any possible degeneration*—the parameter $\gamma = -[(a-d)^2+(b+c)^2]/2$ being by assumption nonzero.

The expression of this parameter in function of the constants $a, b, c,$ and d, introduced in (12) and (13), is a specific aspect of this study in comparison with Ref. 3. Indeed we see that $\gamma=0$, when $a=d$ and $b=-c$ and, in this case, by means of the change of variable

$$u=\log x \Leftrightarrow u=\log\rho^2, \tag{23}$$

Eq. (19) (with $\gamma=0$) can be brought back to the form

$$\frac{d^2w}{du^2}-\left(\frac{1}{2w}+\frac{1}{w-1}\right)\left(\frac{dw}{du}\right)^2 - (1-w)^2\left(\frac{A^2}{8}w+\frac{B^2}{2w}\right)=0. \tag{24}$$

We recognize here a particular form of the equation of the type 38 of the Painlevé and Gambier classification.[7] Thus the following first integral can immediately be written

$$\left(\frac{dw}{du}\right)^2=(w-1)^2\left(\frac{A^2}{4}w^2+2Kw-B^2\right), \tag{25}$$

where K is a new constant of integration. To pursue the calculation it is necessary to consider the trinom

$$(A^2/4)+2K-B^2.$$

Several situations occur following its value. We obtain the following results: If

$$(A^2/4)+2K-B^2>0,$$

we have

$$\lambda(u-u_0)=\log[\mu w+\nu - [(A^2/4)w^2+2Kw-B^2)^{1/2}]/(w-1) \tag{26}$$

with the notations

$$\lambda\equiv\left(\frac{A^2}{4}+2K-B^2\right)^{1/2},\quad \mu\equiv\frac{(A^2/4)+K}{\lambda},\quad \nu\equiv\frac{K-B^2}{\lambda};$$

if

$$(A^2/4)+2K-B^2<0,$$

$$\lambda_1(u-u_0)=\arcsin\left\{\frac{(A^2/4+K)(w-1)-\lambda_1^2}{(K^2+A^2B^2/4)^{1/2}|w-1|}\right\}, \tag{27}$$

with $\lambda_1\equiv[-(A^2/4+2K-B^2)]^{1/2}$;

or if

$$A^2/4+2K-B^2=0,\quad \text{with } A^2/4+K\neq 0,$$

$$u-u_0=-[((A^2/4)w^2+2Kw-A^2)^{1/2}\times((A^2/4+K)(w-1))^{-1}] \tag{28}$$

or

$$u-u_0=\frac{2\epsilon}{B^2+A^2/4}\left(\frac{(A^2/4)w+B^2}{(w-1)}\right)^{1/2},\quad \epsilon\equiv\pm 1, \tag{28'}$$

where $u_0\equiv\log\rho_0^2$ is an integration constant.

In these formulas the function w, researched, appears under the implicit form. In theory, by inversion, we could obtain the explicit formulations. However, the latter would be rather complicated and consequently not very evocative. The last particular case, where $A^2/4+2K-B^2=0$ gives us an explicit result which is easily obtained:

$$w=((A^2/4+B^2)^2(u-u_0)^2+4B^2)\times((A^2/4+B^2)^2(u-u_0)^2-A^2)^{-1},\quad \text{with } u=\log\rho^2. \tag{29}$$

The expressions of the functions $\Phi(x_\mu)$ and $\Psi(x_\mu)$ are formally given by the formulas already noted in (21) and (22) by taking into account the particular situation investigated:

$$\gamma=0,\quad \text{with } a=d \text{ and } b=c.$$

In so doing, we have the result

$$\Phi(x_\mu)=[A(1-w)]^{1/2}\times\exp\left\{ax_3+bx_4+\int\frac{A(1-w)}{2\rho}d\rho\right\}, \tag{30}$$

$$\Psi(x_\mu)=(Aw)^{1/2}\exp\left\{(a-ib)(x_3+ix_4)+\int_{\rho_0}^{\rho}\left(\frac{A}{2}+\frac{iB}{w}\right)\frac{1-w}{\rho}d\rho\right\}. \tag{31}$$

It is essential to point out that the formulas (21) and (22) of Φ and Ψ are in fact parametrized by w, which is the fifth (*nonelementary*) transcendent of Painlevé. Whereas here ($\gamma=0$) the formulas (30) and (31) formally, similar to (21) and (22), depend on the function w which defined either by (26) or by (27), or by (29). This function is an *elementary transcendent* and, in principle, the operations of integration could be performed.

[1] C. N. Yang, Phys. Rev. Lett. **38**, 1317 (1977).
[2] D. Ray, Phys. Lett. B **97** (1) 113 (1980); P. K. Chanda and D. Ray, Phys. Rev. D **31**, 3183 (1985).
[3] B. Léauté and G. Marcilhacy, Phys. Lett. A **93**, 394 (1963).
[4] S. Takeno, Prog. Theor. Phys. **66**, 1250 (1981).
[5] F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[6] L. Witten, Phys. Rev. D **19**, 718 (1979).
[7] P. Painlevé, Acta Math. **25**, 1 (1902); B. Gambier, *ibid.* **33**, 1 (1909); B. F. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).

An example of $\bar{\partial}$ problem arising in a finite difference context: Direct and inverse problem for the discrete analog of the equation $\psi_{xx}+u\psi=\sigma\psi_y$

O. Ragnisco and P. M. Santini
Dipartimento di Fisica, Universita' "La Sapienza," Roma, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Italy

S. Chitlaru-Briggs and M. J. Ablowitz
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

(Received 7 January 1986; accepted for publication 18 November 1986)

The direct and inverse spectral problem for the discrete analog of the equation $\psi_{xx}+u\psi=\sigma\psi_y$ is solved in the framework of "$\bar{\partial}$" theory. The time evolution of the spectral data for the simplest nonlinear differential-difference equations associated to this linear problem is derived.

I. INTRODUCTION

In recent years, there has been considerable interest in the study of exactly solvable nonlinear evolution equations by the method of the inverse scattering transform (IST). The results for one-dimensional partial differential equations and their discrete analogs is by now classical and covered in tests on the subject.[1] On the other hand, the work done on IST for 2 + 1 dimensions has only been satisfactorily understood within the past few years.[2] The prototype problem studied is the Kadomtsev–Petviashvili (KP) equation:

$$(u_t+\sigma uu_x+u_{xxx})_x=-3\sigma^2u_{yy}, \tag{1.1}$$

together with its associated linear problem

$$\sigma\psi_y=\psi_{xx}+u\psi. \tag{1.2}$$

There are two critical choices of the parameter σ: $\sigma=i$ (KPI); $\sigma=1$ (KPII).

Manakov[3] showed that KPI fits within the context of Riemann–Hilbert (RH) theory (i.e., it leads to a nonlocal RH problem). The second case, KPII, was found to lie outside RH theory. It required essential use of the notion of "$\bar{\partial}$" (DBAR) problem. We recall that Beals and Coifman,[4] in their elegant work on systems of ordinary differential equations, noted that the RH problem was, in fact, a special case of the more general notion of a $\bar{\partial}$ problem. The $\bar{\partial}$ problem gives a simple and powerful method by which the underlying inverse spectral problem for the KPII equation (and other analogous equations, like Davey–Stewardson II, modified KPII, ...) can be solved.

In this paper a discrete analog of (1.2) for the case $\sigma=1$ is investigated. To our knowledge this is the first consideration of a discrete multidimensional scattering problem via $\bar{\partial}$ theory. One very important observation is that fully discrete spectral problems virtually always require the use of a $\bar{\partial}$ approach. The reason for this has to do with the fact that discretizations are generally unstable ("ill-posed") as partial difference equations in $\mathbb{Z}^2$ [in analogy with problem (1.2) with $\sigma=1$ for both x and y finite].

Of course, the corresponding one-dimensional discrete problem (i.e., the finite-difference analog of the Schrödinger equation) has been investigated via RH methods by a number of authors.[5] In particular, we refer to the thesis of Sanda L. Chitlaru-Briggs who not only considered this one-dimensional problem, but went beyond to study multidimensional problems such as the one under scrutiny in this paper. Unfortunately, her life was prematurely cut short and her study had to be ended. This article is dedicated to Sanda L. Chitlaru-Briggs.

II. THE DIRECT PROBLEM

We investigate the linear problem

$$\psi(n-1,m)+B(n,m)\psi(n,m)+A(n,m)\psi(n+1,m)=2\psi(n,m+1), \tag{2.1}$$

where $(n,m)\in\mathbb{Z}^2$, and the "potentials" $B, A-1$ vanish sufficiently fast as n and (or) m go to infinity.

This problem, as well as the simplest evolution equations associated with it, has been already introduced in Ref. 6.

It is easy to see that, when $B=0$, the continuum limit of (2.1) is just Eq. (1.2) for $\sigma\in\mathbb{R}$; to perform this limit, set $A(n,m)=\exp[\Delta(V(n+1,m+1)-V(n,m))]$, $x=n\Delta$, $y=(\sigma/2)m\Delta^2$, and let $\Delta\to 0$: one recovers (1.2) with $u=V_x$.

To handle Eq. (2.1), we introduce a function μ, defined as

$$\mu(n,m;z)=\psi(n,m)[\psi^{(0)}(n,m;z)]^{-1}, \tag{2.2}$$

where $\psi^{(0)}$ is a special solution of the "bare" problem associated to (2.1) (i.e., the one corresponding to $B=0, A=1$), given by

$$\psi^{(0)}(n,m;z)=z^{-n}((z+z^{-1})/2)^m. \tag{2.3}$$

The function μ will then satisfy the following equation:

$$z\mu(n-1,m;z)+B(n,m)\mu(n,m;z)+z^{-1}A(n,m)\mu(n+1,m;z)=(z+z^{-1})\mu(n,m+1;z). \tag{2.4}$$

Requiring that, as a function of z, μ satisfies the boundary condition

$$\lim_{|z|\to\infty}\mu(n,m;z)=1, \tag{2.5}$$

 0022-2488/87/040777-04$02.50

Eq. (2.4) is equivalent to the summation equation

$$\mu(n,m;z) = 1 - \sum_{n',m'=-\infty}^{+\infty} G(n-n',m-m';z) \times [B(n',m')\mu(n',m';z) + z^{-1}(A(n',m') - 1)\mu(n'+1,m';z)], \tag{2.6}$$

where the Green's function G is defined as

$$G(n,m;z) = \frac{1}{(2\pi i)^2} \oint_{|z_1|=1} \frac{dz_1}{z_1} \oint_{|z_2|=1} \frac{dz_2}{z_2} \times z_1^n z_2^m \hat{G}(z_1,z_2;z) \tag{2.7}$$

with

$$\hat{G}(z_1,z_2;z) = \left[\frac{z}{z_1} + \frac{z_1}{z} - z_2\left(z + \frac{1}{z}\right)\right]^{-1}. \tag{2.8}$$

From its very definition, it turns out that G enjoys the following symmetry properties:

$$G(n,m;z) = -G(n,m;-z), \tag{2.9a}$$

$$G(n,m;z) = -(-1)^{n+m}G(n,m;z), \tag{2.9b}$$

$$G(n,m;\bar{z}) = \omega_1^{-n}\omega_2^{-m}G(n,m;z), \tag{2.9c}$$

where $\pm\omega_1$, $\pm\omega_2$, defined as

$$\omega_1 = z/\bar{z}, \tag{2.10a}$$

$$\omega_2 = (\bar{z} + 1/2)/(z + 1/z), \tag{2.10b}$$

are the simple pole singularities of G, as a function of z_1 and z_2 on the integration contours. As in the corresponding continuum linear problem (1.2), these singularities are integrable, and by performing the integration with respect to z_2, we get for G the following expression, which clearly shows that G is not an analytic function of $z = r\exp(i\varphi)$ [in Eq. (2.11), $z_1 = \exp(i\vartheta_1)$],

$$\begin{aligned} G(n,m;z) = {} & \frac{1}{2\pi i}(z+z^{-1})^{-m} \oint_{|z_1|=1} \frac{dz_1}{z_1} z_1^n \left(\frac{z_1}{z} + \frac{z}{z_1}\right)^{m-1} \\ & \times \Big\{ \Theta(1-m) - \Theta\left(\frac{\pi}{2}+\varphi\right)\Theta(-\varphi)[\Theta(\vartheta_1+\pi)\Theta(2\varphi-\vartheta_1) + \Theta(\vartheta_1)\Theta(\pi+2\varphi-\vartheta_1)] \\ & - \Theta(\pi+\varphi)\Theta\left(-\varphi-\frac{\pi}{2}\right)[\Theta(-\vartheta_1)\Theta(\vartheta_1-2\varphi-\pi) + \Theta(\pi-\vartheta_1)\Theta(\vartheta_1-2\varphi-2\pi)] \\ & - \Theta(\varphi)\Theta\left(\frac{\pi}{2}-\varphi\right)[\Theta(\pi-\vartheta_1)\Theta(\vartheta_1-2\varphi) + \Theta(-\vartheta_1)\Theta(\vartheta_1-2\varphi+\pi)] \\ & - \Theta\left(-\frac{\pi}{2}+\varphi\right)\Theta(\pi-\varphi)[\Theta(\vartheta_1)\Theta(2\varphi-\pi-\vartheta_1) + \Theta(\vartheta_1+\pi)\Theta(2\varphi-2\pi-\vartheta_1)]\Big\}. \end{aligned} \tag{2.11}$$

The "departure from analyticity" of the function G is measured by its "$\bar{\partial}$" derivative, whose expression is the following one:

$$\frac{\partial G}{\partial \bar{z}} = c(z,\bar{z})(1-(-1)^{n+m})\omega_1^n\omega_2^m, \tag{2.12}$$

where

$$c(z,\bar{z}) = (i/\pi)\,\mathrm{sgn}(\sin 2\varphi)(\bar{z}^2+1)^{-1}. \tag{2.13}$$

Equation (2.12) can be either derived from (2.11) by means of the standard formula

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2\bar{z}}\left[r\frac{\partial}{\partial r} + i\frac{\partial}{\partial \varphi}\right], \tag{2.14a}$$

or, directly from (2.7), taking into account the distribution formula

$$\frac{\partial}{\partial \bar{z}}\left(\frac{1}{z-z_0}\right) = \pi\delta(z-z_0). \tag{2.14b}$$

As in the continuum case, the existence of a connection formula between μ and its "$\bar{\partial}$" derivative plays an essential role in the method. In our case, it has the following expression:

$$\frac{\partial}{\partial \bar{z}}\mu(n,m;z) = \alpha(z)\mu(n,m;\bar{z}) + (-1)^{n+m}\beta(z)\mu(n,m;-\bar{z}). \tag{2.15}$$

The "spectral data" $\alpha(z)$, $\beta(z)$ are related to the potentials through the formulas

$$\alpha(z) = c(z,\bar{z}) \sum_{n,m=-\infty}^{+\infty} \omega_1^{-n}\omega_2^{-m} \times [B(n,m)\mu(n,m;z) + z^{-1}(A(n,m)-1)\mu(n+1,m;z)], \tag{2.16a}$$

$$\beta(z) = -c(z,\bar{z}) \sum_{n,m=-\infty}^{+\infty} (-1)^{n+m}\omega_1^{-n}\omega_2^{-m} \times [B(n,m)\mu(n,m;z) + z^{-1}(A(n,m)-1)\mu(n+1,m;z)]. \tag{2.16b}$$

To prove formulas (2.15), (2.16) it is sufficient to perform the "$\bar{\partial}$" derivative of the summation equation (2.6), taking into account Eq. (2.12) and the symmetry properties (2.9), and then to notice that the lhs and the rhs of (2.15) satisfy the same nonhomogeneous summation equation.

III. THE INVERSE PROBLEM

The main tool for solving the inverse problem, namely for reconstructing the potentials $A(n,m)$ and $B(n,m)$ from the spectral data $\alpha(z)$ and $\beta(z)$, is provided by the generalized Cauchy formula

$$f(z)=\frac{1}{2\pi i}\int\int_D \frac{\partial f/\partial\bar{\zeta}}{\zeta-z}\,d\zeta\wedge d\bar{\zeta}+\frac{1}{2\pi i}\oint_{\partial D}\frac{f(\zeta)}{\zeta-z}\,d\zeta, \tag{3.1}$$

where D is a suitable domain in the z plane and $d\zeta\wedge d\bar{\zeta}=2id\zeta_R d\zeta_I(\zeta=\zeta_R+i\zeta_I=\rho\exp(i\chi))$.

Identifying f with μ, choosing D as the whole complex z plane, and taking into account Eqs. (2.5) and (2.15), formula (3.1) yields the following *linear* integral equation for μ:

$$\mu(n,m;z)=1+\frac{1}{2\pi i}\int\int_D d\zeta\wedge d\bar{\zeta}\,\frac{[\alpha(\zeta)\mu(n,m;\bar{\zeta})+(-1)^{n+m}\beta(\zeta)\mu(n,m;-\bar{\zeta})]}{\zeta-z}. \tag{3.2}$$

Once, through the solution of (3.2), μ is known in the whole complex z plane, one can easily reconstruct the potentials through the formulas

$$B(n,m)=\mu^{(1)}(n,m+1)-\mu^{(1)}(n-1,m), \tag{3.3a}$$

$$A(n,m)=1+\mu^{(2)}(n,m+1)-\mu^{(2)}(n-1,m)+\mu^{(1)}_{(n,m)}(\mu^{(1)}_{(n-1,m)}-\mu^{(1)}_{(n,m+1)}), \tag{3.3b}$$

where $\mu^{(1)}$ and $\mu^{(2)}$ are the leading terms in the asymptotic expansion of μ around infinity, namely

$$\mu^{(1)}(n,m)=\lim_{|z|\to\infty} z(\mu(n,m)-1), \tag{3.4a}$$

$$\mu^{(2)}(n,m)=\lim_{|z|\to\infty} z^2(\mu(n,m)-1-[\mu^{(1)}(n,m)/z]). \tag{3.4b}$$

In terms of the spectral data, they read

$$\mu^{(1)}(n,m)=-\frac{1}{2\pi i}\int\int_D d\zeta\wedge d\bar{\zeta}\,[\alpha(\zeta)\mu(n,m;\bar{\zeta})+(-1)^{n+m}\beta(\zeta)\mu(n,m;-\bar{\zeta})], \tag{3.5a}$$

$$\mu^{(2)}(n,m)=-\frac{1}{2\pi i}\int\int_D d\zeta\wedge d\bar{\zeta}\,\{\zeta\,[\alpha(\zeta)\mu(n,m;\bar{\zeta})+(-1)^{n+m}\beta(\zeta)\mu(n,m;-\bar{\zeta})]\}. \tag{3.5b}$$

IV. SOME ASSOCIATED EVOLUTION EQUATIONS AND THE CORRESPONDING TIME EVOLUTION OF THE SPECTRAL DATA

The simplest nonlinear differential-difference equations solvable through our IST scheme arise as compatibility conditions between (2.1) and

$$\psi_t(n,m)=-z^{-1}\psi(n,m)+G^{(0)}(n,m)\psi(n+1,m), \tag{4.1a}$$

$$G^{(0)}(n,m):=\prod_{j=0}^{\infty}\frac{A(n-j,m+j)}{A(n-1-j,m+j)},$$

$$\psi_t(n,m)=-z^{-2}\psi(n,m)-\tfrac{1}{2}G^{(1)}(n,m)\times[\psi(n,m)-2\psi(n+1,m+1)], \tag{4.1b}$$

$$G^{(1)}(n,m):=\prod_{j=0}^{\infty}\frac{A(n-j,m+1+j)}{A(n-1-j,m+j)},$$

$$\psi_t(n,m)=-(8z^2)^{-1}(z+z^{-1})^2\psi(n,m)-[\tfrac{1}{2}G^{(2)}(n,m)+\tfrac{1}{8}A(n+1,m+2)G^{(3)}(n,m)]\times\psi(n,m)+G^{(2)}(n,m)\psi(n+1,m+1)+G^{(3)}(n,m)\psi(n+1,m+3),$$

$$G^{(2)}(n,m):=\prod_{j=0}^{\infty}\frac{A(n-j,m+1+j)}{A(n-1-j,m+j)}-\frac{1}{4}\sum_{j=0}^{\infty}[A(n-1-j,m+2+j)G^{(3)}(n-2-j,m+j)-A(n-j,m+3+j)G^{(3)}(n-1-j,m+1+j)]\Big[\prod_{s=0}^{\infty}\frac{A(n-j-s,m+2+j+s)}{A(n-1-j-s,m+1+j+s)}\Big]^{-1}, \tag{4.1c}$$

$$G^{(3)}(n,m):=\prod_{j=0}^{\infty}\frac{A(n+1-j,m+3+j)}{A(n-j,m+j)}.$$

The corresponding evolution equations read

$$A_t(n,m)=B(n+1,m)G^{(0)}(n,m+1)-B(n,m)G^{(0)}(n,m),$$

$$B_t(n,m)=G^{(0)}(n,m+1)-G^{(0)}(n-1,m), \tag{4.2a}$$

$$A_t(n,m)=-\tfrac{1}{2}A(n,m)[G^{(1)}(n-1,m)-G^{(1)}(n+1,m)],$$

$$B_t(n,m)=0, \tag{4.2b}$$

$$A_t(n,m)=-\tfrac{1}{2}A(n,m)[G^{(2)}(n-1,m)-G^{(2)}(n+1,m)+\tfrac{1}{2}A(n,m+2)G^{(3)}(n-1,m)-\tfrac{1}{2}A(n+2,m+2)G^{(3)}(n+1,m)], \tag{4.2c}$$

$$B_t(n,m)=0.$$

Equation (4.2a) is clearly a two-dimensional version of the Toda lattice,[7] which is immediately recovered, by assuming that A and B do not depend on m.

Equation (4.2b) is in turn a two-dimensional version of the infinite Volterra system,[8] and finally Eq. (4.2c) is a differential-difference analog of the KPII equation.

The evolution of the spectral data is derived from formulas (4.1) by letting $n,m\to\infty$ and comparing the "$\bar{\partial}$" derivative of Eqs. (4.1) with the time derivative of (2.15). To perform this comparison one has to take into account that, as it can be seen from (2.6), for large n and m μ goes to a constant value as z approaches 0.

The corresponding results are the following.

(i) For Eq. (2a)

$$\alpha_t(z) = (z^{-1} - \bar{z}^{-1})\alpha(z);$$

$$\beta_t(z) = -(z^{-1} + \bar{z}^{-1})\beta(z).$$

(ii) For Eq. (2b)

$$\alpha_t(z)/\alpha(z) = \beta_t(z)/\beta(z) = z^{-2} - \bar{z}^{-2}.$$

(iii) For Eq. (2c)

$$\alpha_t(z)/\alpha(z) = \beta_t(z)/\beta(z)$$
$$= \tfrac{1}{8}[(\bar{z} + \bar{z}^{-1})^2(1 - \bar{z}^{-2})$$
$$- (z + z^{-1})^2(1 - z^{-2})].$$

A more systematic investigation of the class of evolution equation associated with the linear problem (2.1) is contained in Ref. 9, where the bi-Hamiltonian structure of this class is explicitly derived.

[1] See, for example, M. J. Ablowitz and H. Segur, SIAM Stud. Appl. Math. **4** (1981); F. Calogero and A. Degasperis *Spectral Transform and Solitons I* (North-Holland, Amsterdam, 1982).

[2] A. S. Fokas and M. J. Ablowitz, "Lectures on the inverse scattering transform for multidimensional (2 + 1) problems," in *Nonlinear Phenomena*, in *Lecture Notes in Physics*, Vol. 189, edited by K. Wolf (Springer, Berlin, 1983), pp. 137–183; M. J. Ablowitz, D. Bar Yaacov, and A. S. Fokas, Stud. Appl. Math. **69**, 135 (1983).

[3] S. V. Manakov, Physica D **3**, 420 (1981).

[4] R. Beals and R. R. Coifman, Comm. Pure Appl. Math. **37**, 39 (1984).

[5] S. V. Manakov, Zh. Eksp. Teor. Fiz. **67**, 543 (1974); M. Bruschi, S. V. Manakov, O. Ragnisco, and D. Levi, J. Math. Phys. **21**, 2749 (1980).

[6] D. Levi, L. Pilloni, and P. M. Santini, J. Phys. A: Math. Gen. **14**, 1567 (1981).

[7] M. Toda, Prog. Theor. Phys. **45**, 174 (1970); H. Flaschka, Prog. Theor. Phys. **51**, 703 (1974).

[8] V. Volterra, *Lecons sur la theorie mathematique de la lutte pour la vie* (Gauthier-Villars, Paris, 1931); R. Hirota and J. Satsuma, J. Phys. Soc. Jpn. **40**, 891 (1976).

[9] O. Ragnisco and P. M. Santini, "Recursion operator and bi-Hamiltonian structure of integrable multidimensional lattices," preprint, Dipartimento di Fisica, Università di Roma, 1987.

Eigenvalues and eigenvectors of the finite Fourier transform

M. L. Mehta[a)]
Service de Physique Théorique, CEN-Saclay, 91191 Gif-sur-Yvette, Cedex, France

(Received 27 March 1986; accepted for publication 3 December 1986)

The eigenvalues and eigenvectors of the $n\times n$ unitary matrix of finite Fourier transform whose j,k element is $(1/\sqrt{n})\exp[(2\pi i/n)jk]$, $i=\sqrt{-1}$, is determined. In doing so, a multitude of identities, some of which may be new, are encountered. A conjecture is advanced.

I. INTRODUCTION

The problem of diagonalizing the $n\times n$ unitary matrix

$$A_{jk}=(1/\sqrt{n})\exp[(2\pi i/n)jk],\quad i=\sqrt{-1},$$

arises in many contexts and has been studied extensively.[1-3] The eigenvalues of A were determined by Schur[1] to compute the trace of A or the quadratic Gauss sum.[1,2] We present in Sec. II a simple argument to recover the eigenvalues from the trace.

In spite of the abundant literature[3] concerning the eigenvectors of A, the situation is not equally satisfactory. We review briefly in Sec. III some known facts needed later. Section IV gives still another set of eigenvectors of A. This is the discrete analog of the continuum case where Hermite functions are their own Fourier transforms. One thus has an infinity of eigenvectors of A, only n of which are linearly independent, resulting in many nontrivial identities between infinite series involving Hermite functions over discrete points.

A conjecture is advanced.

II. EIGENVALUES OF A

To find the eigenvalues of A we note that

$$(A^2)_{jk}=\frac{1}{n}\sum_{l=0}^{n-1}\exp\left(\frac{2\pi i}{n}l(j+k)\right)=\begin{cases}1, & \text{if } j+k=0 \bmod n,\\ 0, & \text{otherwise,}\end{cases}\tag{2.1}$$

$$A^3=A^*,\quad A^4=1,\quad A^{4+j}=A^j,\quad j=1,2,\ldots,\tag{2.2}$$

so that

$$\operatorname{tr}A^{4j+2}=\tfrac{1}{2}(3+(-1)^n),\quad j=0,1,2,\ldots,\tag{2.3}$$

$$\operatorname{tr}A^{4j+3}=\operatorname{tr}A^*,\quad j=0,1,2,\ldots,\tag{2.4}$$

and according to Gauss (see Ref. 1),

$$\operatorname{tr}A=\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\exp\left(\frac{2\pi i}{n}j^2\right)=\frac{1}{2}(1+i)\{1+(-i)^n\}.\tag{2.5}$$

Consequently,

$$\operatorname{tr}A^s=1+\sum_{k=2}^{n}(-i)^{ks}.\tag{2.6}$$

Equation (2.6) is true for any integer s, but we need it only for $s=1,2,\ldots,n$ to conclude that the eigenvalues of A are

$$1,(-i)^2,(-i)^3,\ldots,(-i)^n.\tag{2.7}$$

Hence the multiplicities of the eigenvalues 1, -1, i, and $-i$ are, respectively, $[(n+4)/4]$, $[(n+2)/4]$, $[(n+1)/4]$, and $[(n-1)/4]$, where $[x]$ is the largest integer not greater than x.

III. EIGENVECTORS OF A

Let us write

$$A=\sum_{j=1}^{4}i^jB(j)=iB(1)-B(2)-iB(3)+B(4),\tag{3.1}$$

where

$$B(1)=\tfrac{1}{2}S+\tfrac{1}{4}(I-I'),\quad B(2)=\tfrac{1}{2}C+\tfrac{1}{4}(I+I'),$$
$$B(3)=-\tfrac{1}{2}S+\tfrac{1}{4}(I-I'),\quad B(4)=\tfrac{1}{2}C+\tfrac{1}{4}(I+I'),\tag{3.2}$$

$$C_{jk}=(1/\sqrt{n})\cos[(2\pi/n)jk],$$
$$S_{jk}=(1/\sqrt{n})\sin[(2\pi/n)jk],\tag{3.3}$$

$$I_{jk}=\delta_{j,k},\quad I'_{jk}=\delta_{j,-k}.\tag{3.4}$$

The Kronecker symbol $\delta_{j,-k}$ stands for 1 or 0 according to whether $j=-k \bmod n$ or $j\neq -k \bmod n$.

It is straightforward to verify that

$$C^3=C,\quad S^3=S,\tag{3.5}$$

and

$$B(j)B(k)=B(j)\delta_{j,k},\quad j,k=1,2,3,4.\tag{3.6}$$

Thus (3.1) is the decomposition of A into its principal idempotents[4] (projectors) and

$$AB(j)=i^jB(j),\tag{3.7}$$

so that the nonzero columns of $B(j)$ are the eigenvectors of A with the eigenvalue i^j.

The eigenvalues of C or S are ± 1 or 0 [Eq. (3.5)] and those of $B(j)$, any j, are 1 or 0 [Eq. (3.6)]. Denoting by $m(\lambda,M)$ the multiplicity of the eigenvalue λ of the matrix M, one has

$$m(1,C)=m(1,B(4))=m(1,A)=[(n+4)/4],\tag{3.8}$$

$$m(-1,C)=m(1,B(2))=m(-1,A)=[(n+2)/4],\tag{3.9}$$

$$m(1,S)=m(1,B(1))=m(i,A)=[(n+1)/4],\tag{3.10}$$

$$m(-1,S)=m(1,B(3))=m(-i,A)=[(n-1)/4].\tag{3.11}$$

[a)] Member of C. N. R. S.

 0022-2488/87/040781-05$02.50

The number of linearly independent column vectors of any matrix is the number of its nonzero eigenvalues; in particular that of $B(j)$ is $m(1,B(j))$, $j=1,2,3,4$, and its value is given above in each case.

To orthogonalize the columns of $B(j)$ we note the following simple facts. If a real symmetric matrix B satisfies $B^2=B$, then the eigenvalues of B are 1 or 0; B is positive semidefinite; its diagonal elements lie between 0 and 1, i.e., $0\leqslant B_{jj}\leqslant 1$; its trace, an integer, is the number of its linearly independent columns (or rows); if $B_{jj}=0$ or 1, then $B_{jk}=0$ for $j\neq k$. Moreover, if $B_{11}\neq 0$, then adding to the k th column, for $k=2,3,\ldots,n$, a constant multiple (depending on k) of the first column we can replace its $(1,k)$ element by zero. Thus

$$B\to\begin{bmatrix} B_{11} & 0 \\ B_{j1} & B' \end{bmatrix}, \tag{3.12}$$

where for $j,k=2,3,\ldots,n$, we have

$$B'_{jk}=B_{jk}-B_{1j}B_{1k}/B_{11}. \tag{3.13}$$

Using the relation

$$\sum_{j=1}^{n} B_{jk}B_{jl}=B_{kl}, \tag{3.14}$$

one easily verifies the following facts: (i) B' is real symmetric; (ii) $B'^2=B'$; (iii) $\operatorname{tr}B'=\operatorname{tr}B-1$; and (iv) the first column of B is orthogonal to any column of B', i.e.,

$$\sum_{j=2}^{n} B_{j1}B'_{jk}=0, \quad k=2,3,\ldots,n. \tag{3.15}$$

Since B' has the same properties as B, we can repeat the process with any nonzero diagonal element of B'. At each repetition we separate one more nonzero column, orthogonal to the earlier ones, and decrease the trace by unity. Ultimately we will be left with a zero matrix.

To summarize, the set of eigenvectors of A for the eigenvalue i^j is the set of $m(1,B(j))$ linearly independent columns of $B(j)$, which can be orthogonalized by a simple recursive procedure.

IV. EIGENVECTORS OF A; ANOTHER APPROACH

Let

$$F_{jk}\equiv F_{jk}(n)=\sum_{p=-\infty}^{\infty} e^{-(\pi/n)(pn+j)^2} \times H_k\left(\sqrt{\frac{2\pi}{n}}\,(pn+j)\right), \tag{4.1}$$

where

$$H_k(x)=e^{x^2}\left(-\frac{d}{dx}\right)^n e^{-x^2} =\sum_{m=0}^{[n/2]}\frac{n!(-1)^m}{m!(n-2m)!}(2x)^{n-2m} \tag{4.2}$$

is the k th Hermite polynomial. We will show that

$$\sum_{l=1}^{n} A_{jl}F_{lk}=i^kF_{jk}, \tag{4.3}$$

i.e., a nonzero F_{jk} is an eigenvector of A with the eigenvalue i^k.

Since F_{jk} considered as a function of j is periodic with period n, we have the Fourier expansion

$$F_{jk}=\sum_{l=-\infty}^{\infty} a_{lk}\exp\left(\frac{2\pi i}{n}lj\right),$$

with

$$a_{lk}=\frac{1}{n}\int_0^n e^{-(2\pi i/n)lx}\sum_{p=-\infty}^{\infty} e^{-(\pi/n)(pn+x)^2} \times H_k\left(\sqrt{\frac{2\pi}{n}}\,(pn+x)\right)dx.$$

Introducing the variable $y=\sqrt{2\pi/n}\,(pn+x)$ and noting that $\exp(2\pi ilp)=1$, one obtains

$$\begin{aligned} a_{lk}&=\frac{1}{\sqrt{2\pi n}}\sum_{p=-\infty}^{\infty}\int_{p\sqrt{2\pi n}}^{(p+1)\sqrt{2\pi n}} \exp\left(-\frac{1}{2}y^2 - ily\sqrt{\frac{2\pi}{n}}\right)H_k(y)\,dy \\ &=\frac{1}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}y^2\right) \times H_k(y)\exp\left(-ily\sqrt{\frac{2\pi}{n}}\right)dy \\ &=\frac{1}{\sqrt{n}}(-i)^k\exp\left(-\frac{\pi}{n}l^2\right)H_k\left(\sqrt{\frac{2\pi}{n}}\,l\right). \end{aligned}$$

In the last step we used the fact that $\exp(-\frac{1}{2}x^2)\,H_k(x)$ is its own Fourier transform. (See, for example, Ref. 5). Thus

$$\begin{aligned} F_{jk}&=(-i)^k\frac{1}{\sqrt{n}}\sum_{l=-\infty}^{\infty} e^{-(\pi/n)l^2+(2\pi i/n)lj}H_k\left(\sqrt{\frac{2\pi}{n}}\,l\right) \\ &=(-i)^k\frac{1}{\sqrt{n}}\sum_{l=1}^{n} e^{(2\pi i/n)lj}\sum_{p=-\infty}^{\infty} e^{-(\pi/n)(pn+l)^2} \times H_k\left(\sqrt{\frac{2\pi}{n}}\,(pn+l)\right), \end{aligned} \tag{4.4}$$

which is Eq. (4.3).

From (4.3) and the symmetry of A we deduce by the usual argument that

$$\sum_{j=1}^{n} F_{jk}F_{jl}=0, \quad \text{if } i^k\neq i^l, \tag{4.5}$$

i.e., if $k\neq l \bmod 4$. Eigenvectors corresponding to distinct eigenvalues are orthogonal.

Now Eq. (4.3) is valid for any $k\geqslant 0$, while the eigenvalues of A are finite in number, namely $1,(-i)^k, k=2,3,\ldots,n$. Thus we have to choose n values of k such that F_{jk} are linearly independent, and if possible express the other F_{jk} in terms of them. When n is even, the F_{jk} are linearly dependent for $k=0,1,2,\ldots,n-1$, as we will see below. Therefore we make the following conjecture.

Conjecture 4.1: The F_{jk} are linearly independent for the following values of k:

$$\begin{aligned} &k=0,1,2,\ldots,n-1, \quad \text{if } n \text{ is odd,} \\ &\text{and} \\ &k=0,1,2,\ldots,n-2,n, \quad \text{if } n \text{ is even.} \end{aligned} \tag{4.6}$$

To make sure of the linear independence of the F_{jk}, one should show that their determinant $\det[F_{jk}]$, $j=0,1,\ldots,n-1$, $k=0,1,\ldots,n-2$, $2[n/2]$, is nonzero. Replacing p by $-p-1$ in Eq. (4.1) one sees that

$$F_{jk}=(-)^k F_{n-j,k}. \tag{4.7}$$

Hence for k odd $F_{0k}=0$ and, in case n is even, $F_{n/2,k}=0$. In the determinant $[F_{jk}]$ add row $n-j$ to row j and then subtract $\frac{1}{2}\times$row j from row $n-j$ for $j=1,2,\ldots,[(n-1)/2]$. This factorizes $\det[F_{jk}]$ into two determinants, one with k even and the other with k odd.

Thus apart from a nonzero constant, $\det[F_{jk}]$ is the product of

$$D_1=\det\left[\sum_p \exp\left(-\frac{\pi}{n}p^2\right)\times H_{2k}\left(\sqrt{\frac{2\pi}{n}}\,p\right)\cos\left(\frac{2\pi}{n}jp\right)\right]_{j,k=0,1,\ldots,[(1/2)n]}$$

and

$$D_2=\det\left[\sum_p \exp\left(-\frac{\pi}{n}p^2\right)H_{2k-1}\left(\sqrt{\frac{2\pi}{n}}\,p\right)\times\sin\left(\frac{2\pi}{n}jp\right)\right]_{j,k=1,2,\ldots,[(1/2)(n-1)]}.$$

Had we taken vectors F_{jk} with $k=0,1,\ldots,n-1$ when n is even, the row and column corresponding to $j=n/2$, $k=n/2$ would appear not in D_1 but in D_2. However, $\sin[(2\pi/n)(n/2)p]=0$, and hence D_2 as well as $\det[F_{jk}]$ would have been zero.

As $H_k(x)$ is a polynomial in x of degree k and parity $(-1)^k$ and as H_{k-2}, H_{k-4},... occur in other columns, by adding appropriate multiples of these to the k th one, we may replace H_k in the above equations by any polynomial of the same degree and parity. In particular, we may replace $H_k(x)$ by x^k:

$$D_1=\text{const}\,\det[f_k((2\pi/n)j)]_{j,k=0,1,\ldots,[n/2]},$$

$$D_2=\text{const}\,\det[g_k((2\pi/n)j)]_{j,k=1,2,\ldots,[(n-1)/2]},$$

where

$$f_k(x)=\sum_{p=-\infty}^{\infty}\exp\left(-\frac{\pi}{n}p^2\right)p^{2k}\cos px$$

and

$$g_k(x)=\sum_{p=-\infty}^{\infty}\exp\left(-\frac{\pi}{n}p^2\right)p^{2k-1}\sin px.$$

Let us recall here the definition of a Chebyshev set.[6] A set of m functions $\{\varphi_1(x),\ldots,\varphi_m(x)\}$ is a Chebyshev set on $[\alpha,\beta]$ if and only if for any $\{a_1,a_2,\ldots,a_m\}$, not all zero, $\Sigma_{j=1}^m a_j\varphi_j\{x\}$ has at most $m-1$ zeros in $[\alpha,\beta]$.

Now, $\det[\varphi_j(x_k)]$, $j,k=1,2,\ldots,m$, considered as a function of x_1, say, is a linear combination of $\varphi_1(x_1),\ldots,\varphi_m(x_1)$, and is zero if x_1 is equal to any of the $m-1$ numbers $x_2,\ldots,x_m$. Thus $\det[\varphi_j(x_k)]$, $j,k=1,\ldots,m$, is not zero for $\alpha\leqslant x_1<x_2<\cdots<x_m\leqslant\beta$ if $\varphi_1(x),\ldots,\varphi_m(x)$ is a Chebyshev set on $[\alpha,\beta]$. Hence Conjecture 4.1 that D_1 and D_2 are nonzero is a consequence of the following conjectures.

Conjecture 4.2: For $n\geqslant 1$, $\{f_0(x),f_1(x),\ldots,f_{[n/2]}(x)\}$ is a Chebyshev set on $0\leqslant x\leqslant\pi$.

Conjecture 4.3: For $n\geqslant 3$,

$$\{g_1(x),g_2(x),\ldots,g_{[(n-1)/2]}(x)\}$$

is a Chebyshev set on $0<x<\pi$.

For small values of n, say $\lesssim 4$, it is not difficult to convince oneself that these conjectures are true. Since fast decreasing exponentials are present, terms beyond a certain value of p do not matter, and $f_k(x)$ and $g_k(x)$ contain only a finite number of effective terms. The question is how many? Intuitively about $n/2$, but this remains to be proved.

V. SOME IDENTITIES

In spite of, or maybe because of, the analogy with the continuum case, this approach is unsatisfactory. Eigenvectors come in a semi-infinite sequence and their hidden periodicity, if any, is difficult to discover. We do have the bilateral symmetry and periodicity in the components of each vector,

$$F_{j,k}=(-1)^k F_{n-j,k}=F_{j+n,k} \tag{5.1}$$

and numerical evidence is against a relation like

$$F_{j,k+4m}=\lambda_k F_{j,k}.$$

What one needs is a choice of n indices $k_1,k_2,\ldots,k_n$ such that the F_{jk} are linearly independent for these indices and to express any other F_{jk} as a linear combination of these.

Lacking this we may still deduce a few identities, some of which may be new. One of them, Eq. (4.5), may be written as

$$\sum_{p,q}\exp\left(-\frac{\pi}{n}(p^2+(qn+p)^2)\right)H_k\left(\sqrt{\frac{2\pi}{n}}\,p\right)\times H_l\left(\sqrt{\frac{2\pi}{n}}\,(qn+p)\right)=0 \tag{5.2}$$

if $k\neq l \bmod 4$. Here and below, sums over $p,q,\ldots$ are taken from $-\infty$ to ∞, unless explicitly stated otherwise. Equation (5.2) is nontrivial for $k+l$ even. A comparison of Eqs. (4.1) and (4.4) gives

$$\sum_p e^{-(\pi/n)p^2}H_k\left(\sqrt{\frac{2\pi}{n}}\,p\right)e^{(2\pi i/n)jp}=i^k\sqrt{n}\sum_p e^{-(\pi/n)(np+j)^2}H_k\left(\sqrt{\frac{2\pi}{n}}\,(pn+j)\right). \tag{5.3}$$

For example, with $n=2$, we have

$$\sum_p(-1)^p e^{-(\pi/2)p^2}H_{2k}(\sqrt{\pi}\,p)=(-1)^k\sqrt{2}\sum_p e^{-(\pi/2)(2p+1)^2}H_{2k}(\sqrt{\pi}(2p+1)). \tag{5.4}$$

Observing that for $l=4$ or 2,

$$\sum_{j=0}^{n-1}B_{jk}(l)=\frac{1}{2}(1\pm\sqrt{n})=\pm\sqrt{n}B_{0k}(l), \tag{5.5}$$

we deduce that for $k=0$ or 2 mod 4,

$$\sum_{j=0}^{n-1} F_{j,k} = \pm\sqrt{n}F_{0,k},$$

i.e.,

$$\sum_p \exp\left(-\frac{\pi}{n}p^2\right) H_{2k}\left(\sqrt{\frac{2\pi}{n}}\,p\right)$$
$$= (-1)^k\sqrt{n}\sum_p \exp(-\pi np^2)H_{2k}(\sqrt{2\pi n}p). \quad (5.6)$$

For $n = 3$ or 4, i^3 is not an eigenvalue of A, implying that $F_{j,4k+3} \equiv 0$ for $n = 3$ or 4, i.e.,

$$\sum_p \exp\left(-\frac{\pi}{3}(3p+1)^2\right) H_{4k+3}\left(\sqrt{\frac{2\pi}{3}}(3p+1)\right) = 0 \quad (5.7)$$

and

$$\sum_p \exp\left(-\frac{\pi}{4}(2p+1)^2\right)(-1)^p$$
$$\times H_{4k+3}\left(\sqrt{\frac{\pi}{2}}(2p+1)\right) = 0. \quad (5.8)$$

For $n = 4$ or 5, the eigenvalue -1 is nondegenerate, implying that $F_{0,4k+2} = (1-\sqrt{n})F_{j,4k+2}$, $0 < j < n$, $n = 4, 5$, i.e.,

$$\sum_p \exp\left(-\frac{\pi}{4}(4p+1)^2\right) H_{4k+2}\left(\sqrt{\frac{\pi}{2}}(4p+1)\right)$$
$$= \sum_p \exp(-\pi(2p+1)^2)H_{4k+2}(\sqrt{2\pi}(2p+1))$$
$$= -\sum_p \exp(-4\pi p^2)H_{4k+2}(\sqrt{8\pi}p) \quad (5.9)$$

and

$$\sum_p \exp\left(-\frac{\pi}{5}(5p+1)^2\right) H_{4k+2}\left(\sqrt{\frac{2\pi}{5}}(5p+1)\right)$$
$$= \sum_p \exp\left(-\frac{\pi}{5}(5p+2)^2\right) H_{4k+2}\left(\sqrt{\frac{2\pi}{5}}(5p+2)\right)$$
$$= (1-\sqrt{5})\sum_p \exp(-5\pi p^2)\, H_{4k+2}(\sqrt{10\pi}p). \quad (5.10)$$

Equations (5.9) and (5.10) are a little stronger than (5.6). For $n = 5$, 6, 7, or 8, the eigenvalue $-i$ is nondegenerate, implying that for these values of n,

$$F_{1,4k+3} : F_{2,4k+3} : F_{3,4k+3}$$
$$= \left(\sin\frac{2\pi}{n} - \frac{1}{2}\sqrt{n}\right) : \sin\frac{4\pi}{n} : \sin\frac{6\pi}{n},$$

i.e.,

$$\sin\frac{\pi}{5}\sum_p \exp\left(-\frac{\pi}{5}(5p+1)^2\right) H_{4k+3}\left(\sqrt{\frac{2\pi}{5}}(5p+1)\right)$$
$$= \left(\sin\frac{2\pi}{5} - \frac{\sqrt{5}}{2}\right)\sum_p \exp\left(-\frac{\pi}{5}(5p+2)^2\right)$$
$$\times H_{4k+3}\left(\sqrt{\frac{2\pi}{5}}(5p+2)\right), \quad (5.11)$$

$$\sum_p \exp\left(-\frac{\pi}{6}(6p+1)^2\right) H_{4k+3}\left(\sqrt{\frac{\pi}{3}}(6p+1)\right)$$
$$= (1-\sqrt{2})\sum_p \exp\left(-\frac{2}{3}\pi(3p+1)^2\right)$$
$$\times H_{4k+3}\left(\sqrt{\frac{4\pi}{3}}(3p+1)\right), \quad (5.12)$$

$$\left(\sin\frac{2\pi}{7} - \frac{\sqrt{7}}{2}\right)^{-1}\sum_p \exp\left(-\frac{\pi}{7}(7p+1)^2\right)$$
$$\times H_{4k+3}\left(\sqrt{\frac{2\pi}{7}}(7p+1)\right)$$
$$= \left(\sin\frac{4\pi}{7}\right)^{-1}\sum_p \exp\left(-\frac{\pi}{7}(7p+2)^2\right)$$
$$\times H_{4k+3}\left(\sqrt{\frac{2\pi}{7}}(7p+2)\right)$$
$$= \left(\sin\frac{6\pi}{7}\right)^{-1}\sum_p \exp\left(-\frac{\pi}{7}(7p+3)^2\right)$$
$$\times H_{4k+3}\left(\sqrt{\frac{2\pi}{7}}(7p+3)\right), \quad (5.13)$$

and a similar equation replacing 7 by 8.

For any n the multiplicity of the eigenvalue 1 is $[(n+4)/4]$. Therefore an $l \times l$ determinant $[F_{j,4k}]$ with $j = 0,1,\ldots,l-1$ and k taking l different values, is zero for $l > [(n+4)/4]$. Similar statements can be made concerning $\det[F_{j,4k+1}]$, $\det[F_{j,4k+2}]$, and $\det[F_{j,4k+3}]$. They will vanish as soon as their size becomes larger than $[(n+1)/4]$, $[(n+2)/4]$, and $[(n-1)/4]$, respectively.

Equation (5.6) for $n = 1$ seems to be equivalent to

$$\sum_p e^{-\pi p^2}H_{2k}(\sqrt{\pi}p) = (-1)^k\sum_p (2\sqrt{\pi}p)^{2k}e^{-\pi p^2} \quad (5.14)$$

of Grosjean[7] or to an identity of Schempp[7] involving Laguerre polynomials.

Many of the relations satisfied by oscillator wave functions have discrete analogs as identities. For this, one does the following replacements:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy \to \frac{1}{\sqrt{n}}\sum_{l=1}^{n}, \quad y \to \sqrt{\frac{2\pi}{n}}(pn+l),$$

$$x \to \sqrt{\frac{2\pi}{n}}(pn+j),$$

and adds a summation over p. For example,

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\, e^{ixy}e^{-(1/2)y^2}H_k(y) = i^k e^{-(1/2)x^2}H_k(x)$$

becomes the eigenvalue equation (4.3);

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy(2k+1-y^2)e^{-(1/2)y^2}H_k(y)e^{ixy}$$
$$= i^k x^2 e^{-(1/2)x^2}H_k(x)$$

becomes, after some simplification,

$$\frac{1}{\sqrt{n}}\sum_p \left(2k+1-\frac{2\pi}{n}p^2\right)e^{-(\pi/n)p^2}H_k\left(\sqrt{\frac{2\pi}{n}}\,p\right)e^{i(2\pi/n)pj}$$
$$= i^k\sum_p \frac{2\pi}{n}(pn+j)^2 e^{-(\pi/n)(pn+j)^2}$$
$$\times H_k\left(\sqrt{\frac{2\pi}{n}}\,(pn+j)\right),$$

and

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\{2yH_k(y)-H_{k+1}(y)\}e^{-(1/2)y^2}e^{ixy}$$
$$= 2k\cdot i^{k-1}e^{-(1/2)x^2}H_{k-1}(x)$$

becomes

$$\frac{1}{\sqrt{n}}\sum_p \left\{2\sqrt{\frac{2\pi}{n}}\,pH_k\left(\sqrt{\frac{2\pi}{n}}\,p\right)\right.$$
$$\left.-H_{k+1}\left(\sqrt{\frac{2\pi}{n}}\,p\right)\right\}e^{-(\pi/n)p^2}e^{i(2\pi/n)pj}$$
$$= 2k\cdot i^{k-1}\sum_p e^{-(\pi/n)(pn+j)^2}H_{k-1}\left(\sqrt{\frac{2\pi}{n}}\,(pn+j)\right).$$

To prove these relations, one observes that the right-hand sides considered as functions of j are periodic with period n and proceeds by expanding them in a Fourier series as at the beginning of Sec. IV.

ACKNOWLEDGMENTS

I am thankful to C. Itzykson for suggesting to me the problem and many illuminating explanations; to R. A. Askey and L. Michel; to the referee of *The Journal* for some recent references; and to my colleagues at Saclay for various discussions.

[1] I. Schur, Nachr. K. Gesell, Wiss. Göttingen, Math.-Phys. Kl. **1921**, 147; E. Landau, *Elementary Number Theory* (Chelsea, New York, 1966), pp. 207–212.

[2] B. C. Berndt and R. J. Evans, Bull. Am. Math. Soc. **5**, 107 (1981), containing about 200 references; see, also, Bull. Am. Math. Soc. **7**, 441 (1982).

[3] I. J. Good, Am. Math. Month. **69**, 259 (1962); J. H. McClellan and T. W. Parks, IEEE Trans. Audio Electroacoust. **AU-20**, 66 (1972); R. Tolimieri, Bull. Am. Math. Soc. **83**, 365 (1977); L. Auslander and R. Tolimieri, Bull. Am. Math. Soc. **1**, 847 (1979); P. Morton, J. Num. Th. **12**, 122 (1980); B. W. Dickinson and K. Steiglitz, IEEE Trans. Acoust. Speech Sig. Process. **ASSP-30**, 25 (1982); R. Tolimieri, Adv. Appl. Math. **5**, 56 (1984).

[4] See, for example, M. L. Mehta, *Elements of Matrix Theory* (Hindustan, Delhi, 1977), Sec. 4.6.

[5] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), formula 7.376.

[6] J. R. Rice, *The Approximation of Functions* (Addison–Wesley, Reading, MA, 1964), Vol. 1, pp. 54–66.

[7] W. Schempp and C. C. Grosjean, "Radar–sonar detection and Laguerre functions," in *Lecture Notes in Mathematics*, Vol. 1171 (Springer, New York, 1985), pp. 549 and 553–554.

Variational calculus on supermanifolds and invariance properties of superspace field theories

Ugo Bruzzo and Roberto Cianci
Dipartimento di Matematica, Universita' di Genova, Via L. B. Alberti 4, 16132 Genova, Italy

(Received 12 March 1986; accepted for publication 22 October 1986)

The foundations of a variational calculus on fibered supermanifolds are outlined, giving applications to the formulation of superspace field theories. Utiyama theorems and conservation laws related to local gauge and general invariance of the theory are proved.

I. INTRODUCTION

In view of the extensive use of superspace techniques made by physicists in the study of supersymmetric field theories, it is rather natural to wonder whether these techniques can be given a sound mathematical foundation in terms of a variational calculus on supermanifolds. This allows extending to superspace field theories typical features of ordinary Lagrangian theories, such as conservation laws, Utiyama theorem, etc.

As far as the definition of the action functional is concerned, one is faced with two major possibilities.

(i) One is to regard the action as an integral over a region of the supermanifold M. This approach should be viable since people working in supermanifold theory have recently succeeded in formulating a satisfactory theory of integration on supermanifolds.[1–4]

(ii) The other is to integrate on suitable local slices of M, which play the role of open sets in space-time M_0.

In this paper we follow the second approach. The main motivations for this choice are the following.

(i) Integrating on open sets in space-time allows a simple generalization of most of the concepts used in the variational calculus on ordinary fibered manifolds.[5,6]

(ii) In the group manifold approach[7,8] the action is constructed in a way that is easily recognized to be a particular case of our procedure for building up the action; as a consequence, the results obtained there are easily comparable with ours.

II. JET BUNDLES ON SUPERMANIFOLDS

Rather obviously, a field theory on a supermanifold M must contain, among its field variables, such objects as a connection form on a principal super fiber bundle P over M and the coframe field, perhaps in addition to matter or auxiliary fields. This entails that some of the field variables are differential one-forms. However, assuming that all the fields are zero-forms (functions) simplifies greatly the explicit writing of many equations, so that in this section we stick to that choice. Later on, we shall freely use generalizations of the equations given in this section which contain one-form fields.

In the language of variational calculus on fibered manifolds, the above assumption means that the fields are sections of a supervector bundle $\pi: E \to M$,[9] whose standard fiber is a supervector space $\mathbf{F}$.[10] In this paper we shall use Rogers' definition of supermanifold[11] as developed by Jadczyk and Pilch.[10] So, if $Q = Q_0 \oplus Q_1$ is a Banach–Grassmann algebra and $Q^{m,n}$ is the graded Banach Q_0-module $(Q_0)^m \times (Q_1)^n$, we say that M is an (m,n)-dimensional supermanifold if it is a Banach manifold[12] locally modeled on $Q^{m,n}$ by means of an atlas whose transition functions are supersmooth. A function $f: U \subset Q^{m,n} \to Q^{p,q}$ is said to be supersmooth if it is C^k $(k \geqslant 1)$ in the sense of Fréchet[13] and its first Fréchet differential is a Q_0-linear map $Df: Q^{m,n} \to Q^{p,q}$. Basic references for supermanifold theory are Refs. 10, 11, and 14–24.

Since we do not intend to build up a merely formal variational calculus, but rather to use the differential calculus on Banach spaces to define the action functional and discuss its extremization, we need to introduce suitable Banach spaces of fields. To this end we shall need the following lemma.[25]

Lemma 2.1: Let U be an open set in $Q^{m,n}$, and $E^k(U)$ the ring of C^k and supersmooth functions $f: U \to Q_0$ whose first k Fréchet differentials are bounded on U. Provided that $k \neq \infty, \omega$, $E^k(U)$ is a Banach space under the norm

$$\|f\| = \sum_{i=0,k} \sup_{x \in U} \|D^i f(x)\|,$$

where D^i is the ith Fréchet differential and the norms in the right-hand side (rhs) are taken in $L(\otimes^i Q^{m,n}, Q_0)$. ■

Let us now define jE, the bundle of first jets of sections s of E. Let U be an open set in M, with coordinates $\{x^A, A = 1,...,m+n\}$, and let $\{(x^A, z^r), r = 1,...,\dim \mathbf{F}\}$ be fibered coordinates in $\pi^{-1}(U)$. Given two sections $s_1, s_2 \in \Gamma E$ (in the following, ΓB will denote the space of local cross sections of any bundle B), we say that $s_1 \approx s_2$ in $x \in M$ if $s_1(x) = s_2(x)$ and $s_{1*}(x) = s_{2*}(x)$. The set $jE := E/\approx$ is a supervector bundle $j\pi: jE \to M$; in $j\pi^{-1}(U)$ we can put fibered coordinates $\{x^A, z^r, Z_A{}^r\}$ such that $Z_A{}^r = \partial z^r/\partial x^A$ on any section of E. Given $s \in \Gamma E$, the couple $js = (s, s_*)$ is a section of jE called the *jet extension* of s.

A couple of mappings $\mu: E \to E, \beta: M \to M$ is said to be an *automorphism* of E if it is a C^k automorphism of E as a vector bundle, and μ and β are supersmooth. In analogy with ordinary differential geometry,[5] it is easily proved that, for any automorphism (μ, β) of E, there exists a unique automorphism $(j\mu, \beta)$ of jE such that $j\mu \cdot js \cdot \beta^{-1} = j(\mu \cdot s \cdot \beta^{-1})$ $\forall s \in \Gamma E$. If $\{(\mu_t, \beta_t), t \in R\}$ is a group of local automorphisms of E, and $X \in \Gamma TE$ its generator, the generator $jX \in \Gamma(TjE)$ of $\{(j\mu_t, \beta_t)\}$ is called the *jet extension* of X.

Lemma 2.2: If, in local fibered coordinates,

$$x = b^A \frac{\partial}{\partial x^A} + h^r \frac{\partial}{\partial z^r},$$

 0022-2488/87/040786-06$02.50

then

$$jX = b^A \frac{\partial}{\partial x^A} + h^r \frac{\partial}{\partial z^r} + \left[\frac{dh^r}{dx^A} - Z_B{}^r \frac{\partial b^B}{\partial x^A}\right] \frac{\partial}{\partial Z_A{}^r},$$

where the "total derivative" d/dx^A maps functions on E into functions on jE and is defined as

$$\frac{d}{dx^A} = \frac{\partial}{\partial x^A} + Z_A{}^r \frac{\partial}{\partial z^r}. \quad \blacksquare$$

III. ACTION FUNCTIONAL AND FIELD EQUATIONS

According to our discussion of Sec. I, we intend to construct the action functional for the supermanifold field theory by integrating a suitable Lagrangian form over open sets in space-time. Now, under suitable conditions,[26] an (m,n)-dimensional supermanifold M defines intrinsically an m-dimensional ordinary differentiable manifold M_0, called the body manifold of M, together with a smooth projection Φ: $M \to M_0$. We assume that M admits a body manifold that we identify with space-time.

Let us recall some properties of supermanifolds that we shall need in order to define the action functional.

Lemma 3.1: Let M be a superanalytic supermanifold with body M_0. There exist covers $\mathbf{A} = \{U_\alpha, \alpha \in J\}$ of M and $\mathbf{A}_0 = \{V_\alpha, \alpha \in J\}$ of M_0 (where J is a suitable index set), such that the following conditions hold.

(i) $\Phi(U_\alpha) = V_\alpha \ \forall \alpha \in J$.

(ii) For each $\alpha \in J$ there exists an analytic imbedding i: $V_\alpha \to U_\alpha$ such that $\Phi \cdot i$ is surjective. In the following such imbeddings will be called local injections.

(iii) For any local injection i: $V_\alpha \to U_\alpha$ there exist functions f: $U_\alpha \to Q_0$ such that $i^* f | \partial V_\alpha = 0$.

(iv) For any couple of local injections i, k: $V_\alpha \to U_\alpha$ there exists a superanalytic diffeomorphism β of U_α such that $i = \beta \cdot k$.

(v) E is trivial over U_α.

(vi) If f is a p-form on $U \subset M$, $p \leqslant m$, and $i^* f = 0$ for all local injections i: $V \to U$, then $f = 0$.

Proof: Points (i)–(v) are proved in Ref. 25. As far as (vi) is concerned, choosing a coordinate patch (x, y) on U such that $i(r) = (r, 0)$, one has immediately $(\partial/\partial x^i) \lrcorner f = 0$. The condition $(\partial/\partial y^\alpha) \lrcorner f = 0$ is obtained by considering the collection of local injections $\{i_{k\beta}\}$, $k = 1,...,m$, $\beta = 1,...,n$, where

$$i_{k\beta}(r) = (r, \delta^\alpha_\beta r^k \delta \, b / \| b \|),$$

b being an arbitrary element of Q_1, and δ a suitably small real number. $\blacksquare$

Before defining the action functional, we need to introduce the following objects: (i) a Banach space of sections of E; (ii) a space of "admissible variations," i.e., of vertical vector fields in E that vanish on the boundary of the integration region; and (iii) a Lagrangian form. As far as point (i) is concerned, we note that the completed tensor product $\mathbf{B}_\alpha = E^k(U_\alpha) \otimes \mathbf{F}$ (the set of C^k, supersmooth sections of E over U_α whose first k Fréchet differentials are bounded) is a Banach space. The space of admissible variations is defined as

$$\mathbf{B}^{\sim}_{\alpha,i} = \{X \in \Gamma(\text{vert } TE) \text{ such that } i^* \ X^r | \partial V_\alpha = 0\}$$

for a given local injection i: $V_\alpha \to U_\alpha$, so that for each local injection there is a different set of admissible variations. $\mathbf{B}^{\sim}_{\alpha,i}$ is a Banach subspace of $\mathbf{B}_\alpha$.[25] Finally, a supermanifold Lagrangian Σ is a bounded, E^k horizontal m-form on jE with values in Q_0.

We define the action functional as the following i-dependent mapping of $\mathbf{B}_\alpha$ into Q_0:

$$A_\alpha(i,s) = \int_{V_\alpha} i^* \cdot js^* \Sigma \equiv \int_{V_\alpha} i^* \Sigma(s). \tag{3.1}$$

The expression in the rhs is basically the integral of a Banach space-valued function with domain in an open set in R^m. Such objects are well known to mathematicians.[13]

Definition: A section $s \in \Gamma E$ is critical if

$$DA_\alpha(i,s) \cdot (X) = 0 \tag{3.2}$$

for all local injections i, all $X \in \mathbf{B}^{\sim}_{\alpha,i}$, and all U_α.

Lemma 3.2: A section $s \in \Gamma E$ is critical if and only if

$$\int_{V_\alpha} i^* \cdot js^* \mathcal{L}_{jX} \Sigma = 0 \tag{3.3}$$

for all local injections i, all $X \in \mathbf{B}^{\sim}_{\alpha,i}$, and all U_α; $\mathcal{L}$ denotes the Lie derivative. $\blacksquare$

The next natural step is to translate the extremality condition (3.2) or (3.3) into a set of local field equations on M (Euler–Lagrange equations). However, in order that this can be done, some constraints have to be put on the supermanifold Lagrangian Σ. After setting, for notational convenience,

$$\frac{\partial \Sigma}{\partial z^r} = \frac{1}{m!} (-1)^{m + r(A_1 + \cdots + A_m)} dx^{A_m} \cdots dx^{A_1} \frac{\partial \Sigma_{A_1 \cdots A_m}}{\partial z^r},$$

$$\frac{\partial \Sigma}{\partial Z_B{}^r} = \frac{1}{m!} (-1)^{m + (B+r)(A_1 + \cdots + A_m)} \times dx^{A_m} \cdots dx^{A_1} \frac{\partial \Sigma_{A_1 \cdots A_m}}{\partial z_B{}^r},$$

Eq. (3.3) reads

$$\int_{V_\alpha} i^* X^r \left(js^* \frac{\partial \Sigma}{\partial z^r} - (-1)^{Br} \frac{\partial}{\partial x^B} js^* \frac{\partial \Sigma}{\partial Z^r} \right) + \int_{V_\alpha} i^* \frac{\partial}{\partial x^B} \left(X^r js^* \frac{\partial \Sigma}{\partial Z_B{}^r} \right) = 0,$$

where $\Sigma = (1/m!) \, dx^{A_m} \cdots dx^{A_1} \, \Sigma_{A_1 \cdots A_m}$ and $X^r(x) = s^* X^r$. Considering separately the terms with B even or odd, the last summand splits into two. The first term vanishes as a consequence of Stokes theorem and the boundary conditions, while the second reads

$$\int_{V_\alpha} i^* \frac{\partial}{\partial y^\mu} \left(X^r js^* \frac{\partial \Sigma^\sim}{\partial Z_\mu{}^r} \right) \tau, \quad \mu = m+1,...,m+n, \tag{3.4}$$

where we have denoted by τ the volume element in M_0 and

$$\Sigma^\sim = (1/m!) dx^{i_m} \cdots dx^{i_1} \, \Sigma_{i_1 \cdots i_m}.$$

Now, developing $X^r(x)$ and js^* $(\partial \Sigma^\sim / \partial Z_\mu{}^r)$ into superfields (namely, developing them into powers of the odd coordinates y^μ), one sees that the term (3.4) vanishes if

$$js^* \frac{\partial \Sigma}{\partial Z_\mu{}^r} = 0. \tag{3.5}$$

If Σ does not depend explicitly on the supermanifold coordinates, conditions (3.4) and (3.5) are equivalent.

Thus we have the following result.

Theorem 3.1 (local Euler–Lagrange equations on the supermanifold): Provided that Σ satisfies condition (3.5) for all $s\in\Gamma E$, a section $s\in\Gamma E$ is critical if and only if, on any coordinate patch $\{U,(x^i,y^\mu)\}$, it verifies

$$js^*\frac{\partial\Sigma}{\partial z^r}-(-1)^{Br}\frac{\partial}{\partial x^B}js^*\frac{\partial\Sigma}{\partial Z_B{}^r}=0. \tag{3.6}$$

■

The constraints (3.5) are similar to the "new" Euler–Lagrangian equations one finds in the calculus of variations on graded manifolds,[27] where they have been interpreted as the kinematical constraints arising, e.g., in superspace supergravity.[28] Our attitude here is somewhat different, since the constraints (3.5) are identically fulfilled by polynomial Lagrangians, while, on the other hand, the Wess–Zumino kinematical constraints can be obtained as field equations from a polynomial supermanifold Lagrangian.[29]

IV. REDUCTION TO SPACE-TIME

It is rather natural to ask whether, given a field theory formulated on a supermanifold, there exists an ordinary field theory on space-time in some sense naturally related to the former. In this section we examine this question. We shall see that requiring that the supermanifold Lagrangian Σ gives rise to a globally defined space-time Lagrangian L leads to constraints on the choice of Σ.

We denote by π_0: $E_0\to M_0$ the configuration bundle of the space-time theory, which is a supervector bundle over M_0 (the body manifold of M, identified with space-time) with standard fiber $\mathbf{F}$ (the same fiber of E). So the space-time fields are still Q valued (Grassmann valued), and therefore contain more information than the usual space-time fields. The possible physical significance of this extra information is still an open question.

In general, E_0 and E cannot be put in a global relation; however, given a local injection i: $V_\alpha\to U_\alpha$, one can construct in an obvious way a local bundle homomorphism τ_i: $\pi^{-1}(U_\alpha)\subset E\to\pi_0^{-1}(V_\alpha)\subset E_0$ such that, for any section $s\in\Gamma E$, the mapping $\sigma=\tau_i\cdot s\cdot i$ is a section of E_0. A similar homomorphism of $j\pi^{-1}(U_\alpha)\subset jE$ into $j\pi^{-1}(V_\alpha)\subset jE_0$, which we shall still denote by τ_i, can be constructed.

The Banach space $\mathbf{B}_{0\alpha}$ of C^k sections of E_0 over V_α is defined in analogy to $\mathbf{B}_\alpha$. The Lagrangian L of the space-time theory—a bounded, horizontal m form on jE_0 with values in Q_0—is defined tentatively by setting

$$j\sigma^*L=i^*\cdot js^*\Sigma \tag{4.1}$$

for each σ that can be written as $\sigma=\tau_i\cdot s\cdot i$ for some $s\in\Gamma E$. The next theorem states sufficient conditions in order that (4.1) makes sense.

Theorem 4.1: Assume that the supermanifold Lagrangian Σ satisfies the following conditions.

$$\text{(i)}\quad i^*\cdot js^*\frac{\partial\Sigma}{\partial Z_\mu{}^r}=0, \tag{4.2}$$

$\forall s\in\Gamma E$, $\forall$ local injection i; Eq. (4.2) has been written using coordinates $\{x,y\}$ in U_α such that $i(r)=(r,0)$, $r\in V_\alpha$.

$$\text{(ii)}\quad \beta^*\Sigma(s)=\Sigma(\beta^*s) \tag{4.3}$$

for any superdiffeomorphism β of M; then there exists on jE_0 an m-form L satisfying Eq. (4.1).

Remarks: (i) Conditions (4.2) have an obvious meaning: since the functions $i^*\Sigma_{i_1\cdots i_m}$ are the components of the space-time Lagrangian L, they must depend only on the coordinates of jE_0.

(ii) β^*s denotes the section obtained by pulling back s regarded as a collection of functions on M; it equals $\hat\beta\cdot s\cdot\beta^{-1}$, where $\hat\beta$ is the lift of β to E defined as β^{-1*}. Then condition (4.3) can be rewritten as

$$j\hat\beta^*\Sigma=\Sigma \tag{4.4}$$

and therefore is the *condition of invariance of* Σ *under superdiffeomorphisms*. This property of Σ will be further discussed in the next section.

Proof: First, we verify that L is independent of the choice of the section s such that $\sigma=\tau_i\cdot s\cdot i$. This amounts to requiring that

$$i^*\cdot js^*\mathcal{L}_{jX}\Sigma=0, \tag{4.5}$$

for any $X\in\Gamma(\text{vert } TE)$ satisfying $\tau_{i*}X=0$. Since this condition is equivalent to $i^*\cdot s^*X^r=0$ [where $X=X^r(\partial/\partial z^r)$], Eq. (4.5) can be turned into (4.2), so that the independence from s is assured.

Second, we verify whether L defined through (4.1) depends on the choice of the local injection i. This means checking whether

$$i^*\cdot js_1^*\Sigma=k^*\cdot js_2^*\Sigma, \tag{4.6}$$

whenever $\tau_i\cdot s\cdot i=\tau_k\cdot s\cdot k$, where i,k: $V_\alpha\to U_\alpha$ are two local injections. Since a superdiffeomorphism β of U_α exists such that $k=\beta\cdot i$ (see Lemma 3.1), we can choose s_1, s_2 so that $s_1=\beta^*s_2$ (i.e., $s_2=\hat\beta\cdot s_1\cdot\beta^{-1}$). Then (4.6) reads $i^*\cdot\beta^*\Sigma(s_2)=i^*\Sigma(\beta^*s_2)$; since this must hold $\forall i$, it is equivalent to (4.3). ■

To sum up, whenever the supermanifold Lagrangian Σ satisfies the requirements (4.2) and (4.3), it gives rise to a well-defined space-time Lagrangian L. In that case, we say that Σ is *projectable*. Then the action functional

$$a_\alpha(\sigma)=\int_{V_\alpha}j\sigma^*L$$

yields an ordinary field theory on space-time whose fields are Q valued (Grassmann valued). It is easily proved that any supermanifold Lagrangian which is a polynomial (in the horizontal exterior algebra over jE) of the forms z^r and $Z^r=dx^A\,Z_A{}^r$ is projectable.

V. INVARIANCE PROPERTIES

In this section we study the properties of invariance of a field theory on a supermanifold M under the action of gauge transformations and superdiffeomorphisms of M. The simplest class of nontrivial theory we can take into account are the pure gauge theories, where all the fields are gauge fields. So the configuration superbundle is $E=C(M)\times V(M)$, with $C(M)$ the bundle of connections on some principal super fiber bundle $P(M,G)$ over M, and $V(M)$ the superbundle of linear coframes on M.[9] The standard fiber G of P is a super-Lie group[14] playing the role of the gauge group. A

natural choice of local fibered coordinates in the tubes $\pi^{-1}(U)$ is $\{x^A, \omega_B{}^b, e_B{}^A\}$, where $\omega_B{}^b$ and $e_B{}^A$ are the components of the connection and coframe forms ω^b and e^A over the basis $\{dx^A\}$ (the index b runs over a basis of the superalgebra g of G). Concerning the further coordinates of jE, it is convenient not to choose the partial derivatives of the components of the connection and coframe, but other coordinates more directly related to the curvature and torsion forms Ω^b and T^A. Recalling the structure equations and Bianchi identities

$$\Omega^b = d\omega^b - \tfrac{1}{2}\omega^d\,\omega^a\,C_{ad}{}^b, \quad T^A = de^A - e^B\,\omega^b\,f_{bB}{}^A,$$
$$D\Omega^b = 0, \quad DT^A = e^B\,\Omega^b\,f_{bB}{}^A,$$

where the $C_{ad}{}^b$ are the structure constants of the superalgebra of G, and the $f_{bB}{}^A$ are the generators of a representation f of G on $Q^{m,n}$ we introduce in $j\pi^{-1}(U)$ local fibered coordinates $\{x^A, \omega_B{}^b, e_B{}^A, \hat{\Omega}_{AB}{}^b, \hat{T}_{AB}{}^C\}$ such that, for any section $s(x) = (\omega^b, e^A)$ of E, one has

$$js^*\hat{\Omega}_{AB}{}^a = -2\frac{\partial\omega_B{}^a}{\partial x^A} - (-1)^{A(B+b)}\omega_B{}^b\,\omega_A{}^c\,C_{cb}{}^a,$$
$$js^*\hat{T}_{AB}{}^C = -2\frac{\partial e_B{}^C}{\partial x^A} - 2(-1)^{A(B+D)}e_B{}^D\,\omega_A{}^a\,f_{aD}{}^C.$$

Here and in the following, we omit the wedge product symbol.

Now we give a proper definition of *gauge invariance* of the supermanifold Lagrangian Σ and deduce a generalization of the well-known Utiyama theorem[30] to the present situation. Let $A^* \in \Gamma(\mathrm{vert}\ TE)$ be the generator of a flow on E produced by a fiber-depending action of G on E defined through the representations f and Ad. We say that Σ is *locally G invariant* if

$$\mathscr{L}_{jA^*}\Sigma = 0 \tag{5.1}$$

for any choice of A^*.

Theorem 5.1: Σ is locally G invariant if and only if

$$\text{(i)}\ \frac{\partial\Sigma}{\partial\omega_B{}^a} = 0, \tag{5.2a}$$

$$\text{(ii)}\ \frac{\partial\Sigma}{\partial\hat{\Omega}_{AB}{}^a} + (-1)^{AB}\frac{\partial\Sigma}{\partial\hat{\Omega}_{BA}{}^a} = 0, \tag{5.2b}$$

$$\text{(iii)}\ \Big[(-1)^{b(B+C)}e_B{}^C f_{bC}{}^E\frac{\partial}{\partial e_B{}^E} + (-1)^{b(A+B+d)}\hat{\Omega}_{BA}{}^d\,C_{bd}{}^a\frac{\partial}{\partial\hat{\Omega}_{BA}{}^a} + (-1)^{b(A+B+C)}\hat{T}_{AB}{}^C f_{bC}{}^E\frac{\partial}{\partial\hat{T}_{AB}{}^E}\Big]\Sigma = 0. \tag{5.2c}$$

Proof: The proof is obtained by writing condition (5.1) explicitly. ■

So Σ is locally G invariant if and only if it does not depend explicitly on the connection, it depends on the derivatives of the connection only through the curvature form, and, finally, it is a G-invariant function of its variables.

Next, we wish to study the consequences of the supercovariance condition (4.4). If $\{\beta_t, t\in R\}$ is a group of superdiffeomorphisms of M with generator $Y\in\Gamma TM$, we denote by $Z\in\Gamma TE$ the generator of the lift $\{\hat{\beta}_t\}$ to E (note, for future use, that jZ contains the derivatives of order 0, 1, and 2 of Y). The infinitesimal version of the supercovariance condition (4.4) reads

$$\mathscr{L}_{jZ}\Sigma = 0. \tag{5.3}$$

Theorem 5.2 (Utiyama theorem for the invariance under superdiffeomorphisms): The supermanifold Lagrangian Σ is supercovariant if and only if

$$\text{(i)}\ \frac{\partial\Sigma}{\partial x^A} = 0, \tag{5.4a}$$

$$\text{(ii)}\ \frac{\partial\Sigma}{\partial\hat{T}_{AB}{}^C} + (-1)^{AB}\frac{\partial\Sigma}{\partial\hat{T}_{BA}{}^C} = 0, \tag{5.4b}$$

$$\text{(iii)}\ \omega_A{}^a\frac{\partial\Sigma}{\partial\omega_B{}^a} + \hat{\Omega}_{AD}{}^a\frac{\partial\Sigma}{\partial\hat{\Omega}_{BD}{}^a} + (-1)^{D(B+A)}\hat{\Omega}_{DA}{}^a \times\frac{\partial\Sigma}{\partial\hat{\Omega}_{DB}{}^a} + e_A{}^C\frac{\partial\Sigma}{\partial e_B{}^C} + \hat{T}_{AD}{}^C\frac{\partial\Sigma}{\partial\hat{T}_{BD}{}^C} + (-1)^{D(B+A)}\hat{T}_{DA}{}^C\frac{\partial\Sigma}{\partial\hat{T}_{DB}{}^C} - (-1)^{(A+B)B}\,dx^B\frac{\partial}{\partial x^A}\,\lrcorner\,\Sigma = 0. \tag{5.4c}$$

Proof: The theorem is proved simply by writing condition (5.3) explicitly and setting equal to zero the terms which multiply the derivatives of Y of order 0, 1, and 2, respectively. ■

So, Σ is supercovariant if and only if it does not depend on the supermanifold coordinates, it depends on the derivatives of the coframe only through the torsion form, and its "energy-momentum" form (5.4c) vanishes identically.

It is rather natural to wonder whether the conditions of supercovariance and G invariance entail some differential identities (conservation laws). As the following analysis will show, these identities do indeed exist, and can be written as differential equations on M. While, in general, they have a very complicated structure, whenever one assumes that Σ is a polynomial (in the horizontal exterior algebra over jE) of the forms ω, e, Ω, and T, the identities can be written rather nicely (the reader will realize that this is the same assumption made in the group manifold approach[7,8]). Under this condition we have the following.

Theorem 5.3: Σ admits differential field equations that can be written in the form

$$js^*G_b = 0, \quad js^*G_A = 0,$$

where

$$G_b = \frac{\partial\Sigma}{\partial\omega^b} + D\frac{\partial\Sigma}{\partial\Omega^b} + J_b,$$
$$G_A = \frac{\partial\Sigma}{\partial e^A} + D\frac{\partial\Sigma}{\partial T^A},$$
$$J_b = -e^A f_{bA}{}^B\frac{\partial\Sigma}{\partial T^B},$$

D being the ω-covariant exterior differential. Moreover, Σ is projectable (see Sec. IV), and therefore supercovariant. ■

In order to write down the differential identities, we need a general variation formula for evaluating $js^*\mathscr{L}_{jX}\Sigma$, where X is a projectable vector field on E.

Theorem 5.4: For all $s\in\Gamma E$ and all π projectable $X\in\Gamma TE$,

$$js^*\mathcal{L}_{jX}\Sigma = b^A\frac{\partial}{\partial x^A}\rfloor d\,js^*\Sigma - js^*(\hat{h}^a G_a + \hat{K}^A G_A) + d\,js^* f, \tag{5.5}$$

where

$$X = b^A(x)\frac{\partial}{\partial x^A} + h_A{}^a\frac{\partial}{\partial \omega_A{}^a} + K_A{}^B\frac{\partial}{\partial e_A{}^B},$$

$$\hat{h}^a = dx^A(h_A{}^a - b^B\omega_{AB}{}^a),$$

$$\hat{K}^A = dx^B(K_B{}^A - b^C e_{BC}{}^A),$$

$$f = jX\rfloor\Sigma + \hat{h}^a\frac{\partial\Sigma}{\partial\omega^a} + \hat{K}^A\frac{\partial\Sigma}{\partial e^A},$$

where $\omega_{AB}{}^a$ and $e_{AB}{}^C$ are the coordinates on jE corresponding to the partial derivatives of $\omega_B{}^a$ and $e_C{}^A$.

Outline of the proof: For any $u\in jE$ choose an $s\in\Gamma E$ such that $js(x) = u$ where $x = j\pi(u)$. Set $jX(u) = X_1 + X_2$ with $X_1 = js_*\cdot j\pi_* X$. Now, $(js^*\mathcal{L}_{jX}\Sigma(x)) = (j\pi_* jX\rfloor js^*\,d\Sigma) + d\,js^*(jX\rfloor\Sigma) + js^*(X_2\rfloor d\Sigma)$. Equation (5.5) is now proved by direct computation; in particular, since X_2 is vertical, the last term in the Lie derivative is computed straightforwardly. ■

Theorem 5.5 (Noether theorem for the gauge invariance): If Σ is locally G invariant, then, for all $s\in\Gamma E$,

$$js^*(DG_b - (-1)^{bA}e^A f_{bA}{}^B G_B) = 0. \tag{5.6}$$

Conversely, if the identity (5.6) holds, the action of G changes Σ only by an exact form, so that the field equations are unchanged.

Theorem 5.6 (Noether theorem for the invariance under superdiffeomorphisms): If Σ is locally G invariant and supercovariant then, for all $s\in\Gamma E$,

$$js^*DG_A = (\partial_A\rfloor\Omega^b)js^*G_b + (\partial_A\rfloor T^B)js^*G_B + \partial_A\rfloor js^*\,d\Sigma, \tag{5.7}$$

where $\{\partial_A\}$ is the frame dual to $\{e^A\}$. Conversely, if identity (5.7) holds, and Σ is locally G invariant, the action of a superdiffeomorphism changes Σ only by an exact form, so that the field equations are unchanged.

Proofs: The two Noether theorems are proved by inserting the variation formula (5.5) into the integral versions of the definitions of G invariance and supercovariance and then integrating by parts. ■

VI. AN APPLICATION

In this section we want to exemplify the contents of the previous sections by applying it to an example, namely, superspace $N = 1$ supergravity.[28,29] In this case we can assume that the gauge group G is the Lorentz group, and that the supermanifold M has dimension (4,4). So the fields are the spin connection ω^{ik} on M and the coframe field e^A, $i,k = 1,...,4$, $A = 1,...,8$. The "bosonic" torsion De^i will be denoted by T^i, and the spinorial torsion De^β, $\beta = 5,...,8$, by θ^β. The supermanifold Lagrangian is

$$\Sigma = \epsilon_{ijkh}e^i e^j\Omega^{kh} + 4\theta C\gamma_i\gamma_5 e e^i.$$

Spinor (odd) indices are omitted, and C is the charge conjugation matrix; for conventions concerning the algebra of the Dirac γ_i matrices, the reader is referred to Ref. 29. According to the discussion of the previous sections, we have the following.

(i) Σ satisfies the hypotheses of Theorem 5.3, so that it is projectable, supercovariant, and admits fields equations, that can be written as

$$G_i = 2\epsilon_{ijkh}e^j\Omega^{kh} - 4\theta C\gamma_i\gamma_5 e = 0,$$

$$G = 4C\gamma_i\gamma_5(2e^i\theta - eT^i) = 0,$$

$$G_{ik} = \epsilon_{ikhj}(2T^h + eC\gamma^h e)e^j = 0,$$

where we have split the coframe equation into its even and odd parts.

(ii) Since Σ is projectable, there exists a field theory on M_0, whose Lagrangian is

$$L = \epsilon_{ijhk}V^i V^j R^{hk} + 4\Psi C\gamma_i\gamma_5\psi V^i,$$

where V^i is the coframe on M_0, ψ is the gravitino field, $\Psi = D\psi$, and R^{ik} is the curvature on M_0. One can prove that, whenever $s\in\Gamma E$ is critical with respect to Σ, the section $\sigma\in\Gamma E_0$ defined by $\sigma = i^*s = \tau_i\cdot s\cdot i$, where i is a local injection, is critical with respect to L.

(iii) Σ is Lorentz invariant, so that Theorem 5.5 gives the identity

$$js^*DG_{ik} = \tfrac{1}{2}js^*(G_i e_k - G_k e_i).$$

Moreover, since Σ is supercovariant as well, we have the couple of identities

$$js^*DG_i = js^*[-4\theta C\gamma_i\gamma_5\theta + 2\epsilon_{ijhk}\Omega^{hk}(T^j + \tfrac{1}{2}eC\gamma^j e)],$$

$$js^*DG = js^*[4C\gamma_i\gamma_5\theta(T^i + \tfrac{1}{2}eC\gamma^i e) - C\gamma^i eG_i].$$

ACKNOWLEDGMENTS

This research was partly supported by the National Group for Mathematical Physics (GNFM) of the Italian Research Council (CNR) and by the Italian Ministry of Public Education through the research project "Geometria e Fisica."

[1] A. Rogers, J. Math. Phys. **26**, 385 (1985).
[2] A. Rogers, J. Math. Phys. **26**, 2749 (1985).
[3] A. Rogers, J. Math. Phys. **27**, 710 (1986).
[4] J. M. Rabin, Commun. Math. Phys. **103**, 431 (1985).
[5] A. Trautman, Commun. Math. Phys. **6**, 248 (1967).
[6] P. L. Garcia and A. Perez-Rendon, Commun. Math. Phys. **13**, 24 (1969).
[7] Y. Ne'eman and T. Regge, Riv. Nuovo Cim. **1**, 1 (1978).
[8] R. D'Auria, P. Fre', and T. Regge, in *Supergravity '81*, edited by S. Ferrara and J. C. Taylor (Cambridge U. P., Cambridge, 1982).
[9] U. Bruzzo and R. Cianci, Class. Quantum Grav. **1**, 213 (1984).
[10] A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).
[11] A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[12] S. Lang, *Differential Manifolds* (Addison–Wesley, Reading, MA, 1972).
[13] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (AMS Coll. Publ., Providence, RI, 1957), Vol. XXXI.
[14] A. Rogers, J. Math. Phys. **22**, 939 (1981).
[15] M. Rothstein, "On the resolution of some difficulties in the theory of G^∞ manifolds," preprint, University of Washington, Seattle, 1985.
[16] C. P. Boyer and S. Gitler, Trans. Am. Math. Soc. **285**, 241 (1984).
[17] V. S. Vladimirov and I. V. Volovich, Theor. Math. Phys. **59**, 317 (1984).
[18] V. S. Vladimirov and I. V. Volovich, Theor. Math. Phys. **60**, 743 (1985).
[19] J. Hoyos, M. Quiros, J. Ramirez Mittelbrunn, and F. J. De Urries, J.

Math. Phys. **25**, 833 (1984), and papers in the following pages.
[20]U. Bruzzo and R. Cianci, Lett. Math. Phys. **8**, 279 (1984).
[21]U. Bruzzo and R. Cianci, J. Phys. A **18**, 417 (1985).
[22]R. Cianci, J. Math. Phys. **25**, 451 (1984).
[23]B. DeWitt, *Supermanifolds* (Cambridge U. P., London, 1984).
[24]J. M. Rabin and L. Crane, Commun. Math. Phys. **100**, 141 (1985).
[25]U. Bruzzo and R. Cianci, Lett. Math. Phys. **11**, 21 (1986).
[26]R. Catenacci, C. Reina, and P. Teofilatto, J. Math. Phys. **26**, 671 (1985).
[27]D. Hernandez Ruiperez and J. Munoz Masqué, J. Math. Pures Appl. **64**, 87 (1985).
[28]J. Wess and B. Zumino, Phys. Lett. B **66**, 361 (1977).
[29]U. Bruzzo and R. Cianci, Commun. Math. Phys. **95**, 393 (1984).
[30]D. Bleecker, *Gauge Theories and Variational Principles* (Addison–Wesley, Reading, MA, 1981).

Classical particles with internal structure. II. Second-order internal spaces

M. V. Atre
Theoretical Physics Group, Tata Institute of Fundamental Research, Bombay 400005, India

N. Mukunda
Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012, India

(Received 8 August 1986; accepted for publication 22 October 1986)

A systematic study of classical relativistic particles with internal structure, initiated in a previous paper, is continued and a study of second-order internal spaces (SOS) is presented within the framework of the Lagrangian form of constrained dynamics. Such internal spaces Q are those for which a phase-space treatment must necessarily use the cotangent bundle T^*Q. The large variety of possible SOS's—ten discrete cases and two one-parameter families—is separated into those capable of a manifestly covariant description, and those for which special methods based on the transitive action of $SL(2,C)$ on a coset space are needed. The concept of the isotopy representation plays an important role in this context. Seven of the possible discrete SOS's are shown to be describable in a manifestly covariant way; two discrete and one one-parameter family of SOS's are shown to be unphysical in the sense that no Lagrangians can be written in which the internal and the space-time position variables are nontrivially coupled; and the remaining single discrete and one one-parameter family are shown to be physical though not describable in a manifestly covariant way. General phase-space methods uniformly applicable to all SOS's are developed; and as an illustrative example a Lagrangian model for an SOS in which the internal space is spanned by two orthonormal spacelike unit vectors is presented.

I. INTRODUCTION

The subject of relativistic particles with internal structure has received a great deal of attention over a long period of time. A variety of approaches has been used in discussing different aspects of this problem. In a previous paper, hereafter referred to as I, we have briefly reviewed the history of this subject, trying to highlight the different physical ideas and points of view used by various authors.[1] We also developed a new approach based on group theoretic and differential geometric methods, intended to systematically classify and investigate all possible physical systems of this type, within the framework of the Lagrangian formalism.

A relativistic point particle with internal structure possesses, in addition to space-time position variables x^μ, internal coordinates q^r describing an internal space Q (Ref. 2). An important physical condition imposed in I was that the space Q must admit a transitive action by the group $G = SL(2,C)$, the twofold universal covering group of the homogeneous Lorentz group $SO(3,1)$ of special relativity. As a result, each conceivable Q is the coset space G/H for some subgroup H in G. The further restriction was then made to subgroups H, which are connected Lie subgroups of G. Since all such subgroups are known up to conjugation,[3] there results a complete catalog of all possible internal spaces Q. The possible subgroups H consist of two continuous one-parameter families and 13 other distinct cases, leading to a similar variety of possible Q's. Among these H's are included both the entire group G and the trivial case consisting of the identity alone, corresponding, respectively, to Q being trivial or G itself. In the notation of Patera *et al.*[3] the one-parameter families of subgroups are written as F_5^φ and F_{11}^φ with φ the concerned parameter; and the 13 other subgroups are F_n for $n = 1,...,4,6,...,10,12,...,15$. At the two ends of this list, we have $F_1 = G$ and $F_{15} = e$. The infinitesimal generators of each of these F's have been listed in I (Ref. 4).

The set of all internal spaces Q was separated in I into two types, which were, respectively, called first-order spaces (FOS's) and second-order spaces (SOS's). The distinguishing property of the former is that in the phase-space description of a relativistic object with such an internal space, one can work at the level of Q itself (as far as the internal degrees of freedom are concerned) without passing to the cotangent bundle T^*Q at all; at least, the nature of Q permits this provided the Lagrangian is properly chosen. All other Q's are, by definition, SOS's. It was possible to determine the possible FOS's by using the Kostant–Kirillov–Souriau theorem, which relates coset spaces G/H carrying a G-invariant symplectic structure to orbits in the Lie algebra $\mathfrak{g}$ of G under the adjoint action by G (Ref. 5). In this way we found that there are three possible FOS's, namely G/F_9, G/F_{10}, and a third one which can be exhibited as $SO(3,1)/F_{10}$ or G/N_0, where N_0 is a suitable subgroup in G. (Actually N_0 does not obey the condition of connectedness which was imposed on H above, so there is a departure here from the general pattern which, however, causes no difficulties.) The important geometrical and phase-space properties of FOS's, and of Lagrangian models for relativistic objects with such internal structures, were discussed in detail in I (Ref. 6). The purpose of the present paper is to take up the study of SOS's and of objects with such internal variables in a systematic way.

The number of possible coset spaces $Q = G/H$ that qualify in principle as SOS's is quite large, as compared to the

 0022-2488/87/040792-15$02.50

case of FOS's. For convenience we introduce the following notation to keep track of them:

$$Q_n = G/F_n, \quad \text{for } n = 2,3,4,6,7,8,12,13,14,15,$$
$$Q_n^{\varphi} = G/F_n^{\varphi}, \quad \text{for } n = 5,11. \tag{1.1}$$

(The value $n = 1$ is omitted since then Q is trivial; the values $n = 9, 10$ leads to FOS's.) Given any one of these Q's as an internal space, the total configuration space is the product $\mathscr{M} \times Q$, where $\mathscr{M}$ is Minkowski space-time with coordinates x^μ. The complete set of Lagrangian variables describing $T(\mathscr{M} \times Q) \simeq T\mathscr{M} \times TQ$ is $x^\mu, q^r, \dot{x}^\mu$, and $\dot{q}^r$; the velocities are defined with respect to an unspecified evolution parameter s. (It is understood of course that the q^r may only be local coordinates on Q.) Now it was shown in I that, in case Q is an FOS, the most general Lagrangian splits into the sum of two parts, the first of which is a function $\mathscr{L}_0(q,\dot{q})$ linear in $\dot{q}^r$ and the second of which is $\dot{q}$ independent. With Q chosen to be an SOS, however, there is no such natural breakup of the Lagrangian into separate parts distinguished by different $\dot{q}$ dependences. One must construct the most general function of all the Lagrangian variables subject to the conditions of Poincaré and reparametrization invariances described in I. The important physical question which comes up for a given Q is whether it is possible to couple the internal variables $q,\dot{q}$ and the space-time variables x, $\dot{x}$ in a nontrivial way when constructing the Lagrangian. This question has an easy answer—in the affirmative—if Q can be described in a manifestly covariant way. In that case in place of the minimal set of internal coordinates q^r consisting of independent variables, one can use an overcomplete system of coordinates for Q, made up of (sets of) quantities transforming linearly under G, with suitable G-invariant restrictions imposed upon them. The possibility of coupling the internal variables and velocities to $\dot{x}^\mu$ is then immediate, and one can easily form Lorentz scalar combinations that can enter a Lagrangian. For other Q's, however, the answer to the basic question is not immediately evident, and special methods must be devised to handle them and to determine whether each of them is physical or unphysical in this sense.

The material of this paper is arranged as follows. In Sec. II all those Q's that can be given manifestly covariant descriptions are discussed, and such a description is developed for each of them. In every case the description is via configurations of one, two, or at most three vectors with respect to SO(3,1), and the concerned SOS's are Q_n for $n = 3,4,6,12,13,14,15$. For handling the remaining SOS's, Sec. III develops techniques based on what is called the isotopy representation associated with the action of G on a coset space G/H. This is a linear matrix representation of H, of the same dimension as G/H. We give systematic procedures for determining whether Lorentz scalars, four-vectors, second-rank symmetric tensors,..., can be constructed on TQ, with various degrees of dependence on $\dot{q}$: these objects are candidates for coupling to $\dot{x}^\mu, \dot{x}^\mu\dot{x}^\nu, \ldots$. By applying these methods, it is shown that the SOS's Q_2, Q_3^{φ}, Q_7 are unphysical—that is, it is impossible to construct Lagrangians for point particles with such internal spaces, with a nontrivial coupling between $\dot{x}^\mu$ and $q,\dot{q}$. The two remaining candidates for SOS's, namely Q_8 and Q_{11}^{φ}, are taken up in Sec. IV. (Actually these are an infinity of SOS's since Q_{11}^{φ} is a one-parameter family.) For Q_8, it is possible to show quite directly that it is physically admissible, though a full treatment would require the methods of Sec. III. On the other hand, by application of those methods it is shown that in any Q_{11}^{φ} as well coupling of internal and space-time variables is definitely possible. The general features of a phase-space treatment for an object with a physical SOS as internal space are developed in Sec. V, with particular attention paid to the manifestly covariant cases. It is possible in all cases to regard the internal contribution $S_{\mu\nu}$ to the total four-dimensional angular momentum as the canonical momentum conjugate to the internal space Q. As an illustration of the general ideas, a manifestly covariant Lagrangian model based on Q_{13} as an internal space is presented in Sec. VI. This is intended to capture some of the features of the Hanson–Regge model, but with useful simplifications.[7] Section VII contains some remarks and a table presenting the main results so that the overall picture concerning the large number of conceivable SOS's can be easily grasped. We emphasize that the aim of this paper is to survey the set of all possible SOS's, classify them in a useful way, and develop suitable methods to deal with them, and not to discuss in complete detail the dynamics of any particular model. Thus we carry the development far enough so that one can in principle construct the most general Lagrangian in any specific case, provided the concerned SOS is a physically admissible one; an example is provided by Sec. VI. Further study of the constraint structures, dynamics and external field couplings in interesting cases will be taken up elsewhere. The Appendix contains a useful lemma relating to finite-dimensional representations of SL(2,C).

II. MANIFESTLY COVARIANT SOS'S

Let Q be any one of the possible SOS's, with q^r a system of (local) coordinates for it. On account of space-time translation invariance, the Lagrangian for an isolated point object with internal space Q must be a function only of $\dot{x}$, q, and $\dot{q}$, and so it is written $\mathscr{L}(\dot{x},q,\dot{q})$. While $\dot{x}^\mu$ transforms as a translation invariant four-vector under the Poincaré group $\mathscr{P}$, the transitive G action on Q appears as a set of point transformations on the internal variables q^r. (Suitable notation to deal with these transformations will be set up in Sec. III.) The internal velocities $\dot{q}^r$ then transform linearly among themselves, possibly with functions of q for coefficients. Given these behaviors of the arguments of $\mathscr{L}$, the two conditions to be obeyed by $\mathscr{L}$ are (i) it must be G invariant, and (ii) it must be homogeneous of degree 1 in the velocities $\dot{x}^\mu$ and $\dot{q}^r$. The former condition guarantees the Poincaré invariance of the system, hence the validity of the conservation laws; the latter leads to reparametrization invariance of the action. One can then see that the general form of the Lagrangian must be[8]

$$\mathscr{L}(\dot{x},q,\dot{q}) = (-\dot{x}^2)^{1/2} f(\cdots\zeta\cdots), \tag{2.1}$$

where the arguments ζ of the real function f form a complete independent set of Lorentz scalars formed out of $\dot{x}$, q, and $\dot{q}$; with each ζ being homogeneous of degree zero in the velocities. We now consider in general terms the way in which to build up the ζ's.

On account of the transitive action of G on Q, there can be no nontrivial Lorentz scalar functions of the q^r alone; some dependence on $\dot{q}$ and/or $\dot{x}$ must be present in every ζ. The possibility that there is no $\dot{x}$ dependence is a particular case of the following more general procedure, so we consider how $\dot{x}$ can be combined with q and $\dot{q}$ to form ζ's. Now $\dot{x}^\mu$ belongs to the four-vector or $D^{(1/2,1/2)}$ representation of SO(3,1). The homogeneous monomials $\dot{x}^\mu\dot{x}^\nu$, $\dot{x}^\mu\dot{x}^\nu\dot{x}^\lambda$,... (on removal of traces) give quantities belonging to the various symmetric traceless tensor representations $D^{(1,1)}$, $D^{(3/2,3/2)}$,..., generally $D^{(j,j)}$, of SO(3,1). First let us consider ζ's with no $\dot{q}$ dependence. We attempt to construct as many independent sets of functions $V_\mu(q)$, $t_{\mu\nu}(q)$,..., as possible on Q, transforming according to the linear SO(3,1) representations $D^{(1/2,1/2)}, D^{(1,1)}$,..., of SO(3,1) when the q^r are subjected to the point transformations realizing G. A general "multiplet" of such functions belonging to the representation $D^{(j,j)}$ will be called a quantity of type j on Q. By contracting $V_\mu(q), t_{\mu\nu}(q)$,..., with $\dot{x}^\mu, \dot{x}^\mu\dot{x}^\nu$,..., and then dividing through by suitable powers of $(-\dot{x}^2)$, we produce possible ζ's with no $\dot{q}$ dependence. To include $\dot{q}$ dependence, we follow a similar procedure with the following difference: we look for quantities $V_\mu(q,\dot{q}), t_{\mu\nu}(q,\dot{q})$,..., generally quantities of type j on TQ, at first linear in $\dot{q}$, then quadratic in $\dot{q}$, and so on. (While in the $\dot{q}$ independent case we are interested in $j = \frac{1}{2},1,...$, now we include the possibility $j = 0$ as well.) These can then be contracted among themselves (which would reproduce $j = 0$) or with $\dot{x}^\mu, \dot{x}^\mu\dot{x}^\nu$,..., to again give us, after division by suitable powers of $(-\dot{x}^2)$, possible ζ's.

While this is the systematic procedure for building up the set of ζ's for any given SOS, there are several SOS's that allow a more direct treatment. These are the ones which can be given a manifestly covariant description in terms of suitable configurations of sets of space-time four-vectors.[9] Such descriptions invariably use an overcomplete set of coordinates in the interest of manifest covariance, subject to suitable Lorentz invariant restrictions. These manifestly covariant SOS's are Q_n, for $n = 3,4,6,12,13,14$, and 15. We shall describe them directly in geometrical terms and then verify that they are indeed realizations of the coset spaces G/F_n for the above mentioned values of n.

As in I we denote by e_0, e_1, e_2, e_3 a set of orthonormal vectors in space-time in the indicated directions, and by the letters t,s, and l a general unit positive timelike, a unit spacelike, and a positive lightlike vector, respectively. Looking first at manifolds formed by one vector at a time, the set $Q(t)$ of all unit positive timelike vectors,

$$Q(t) = \{a^\mu | a^2 = -1,\ a^0 > 0\} \tag{2.2}$$

forms the unit timelike hyperboloid on which SO(3,1) acts transitively. A representative point on $Q(t)$ is e_0 or in component form (1,0,0,0), with stability group SO(3) in SO(3,1) or SU(2) in SL(2,C). Since SU(2) is generated by J_1, J_2, J_3, it follows, as is well known, that $Q(t)$ is a realization of the coset space G/F_3:

$$Q(t) = Q_3 = G/F_3 = \mathrm{SO}(3,1)/\mathrm{SO}(3). \tag{2.3}$$

Similarly, $Q(s)$ defined by

$$Q(s) = \{a^\mu | a^2 = 1\}, \tag{2.4}$$

is a realization of Q_4 because the representative point $e_3 = (0,0,0,1)$ has a stability group SO(2,1) $\subset$ SO(3,1) generated by J_3, K_1, and K_2:

$$Q(s) = Q_4 = G/F_4 = \mathrm{SO}(3,1)/\mathrm{SO}(2,1). \tag{2.5}$$

Lastly,

$$Q(l) = \{a^\mu | a^2 = 0,\ a^0 > 0\} \tag{2.6}$$

defines the positive light cone; a representative point on it is $e_0 + e_3 = (1,0,0,1)$ with stability group $E(2) \subset$ SO(3,1) generated by J_3, $N_1 = J_1 - K_2$, $N_2 = J_2 + K_1$. Thus this is a realization of Q_6:

$$Q(l) = Q_6 = G/F_6 = \mathrm{SO}(3,1)/E(2). \tag{2.7}$$

Thus the configurations of one vector at a time, $Q(t)$, $Q(s)$, and $Q(l)$, respectively, realize the SOS's Q_3, Q_4, and Q_6; and for an object with such an internal space the vector a^μ with suitable conditions on it can be used as the internal variable.

We next look at configurations formed by two vectors at a time, the members of the pair being independent and mutually orthogonal.[10] Because of the signature of the space-time metric, the only possible choices are *ts*, *ss*, and *sl*. Let us then define $Q(ts)$ by

$$Q(ts) = \{(a^\mu, b^\mu) | a^2 = -1,\ b^2 = 1,\ a^0 > 0,\ a.b = 0\}. \tag{2.8}$$

Since any pair $(a^\mu, b^\mu) \in Q(ts)$ can be carried by a suitable SO(3,1) transformation to the standard pair (e_0, e_3), it follows that SO(3,1) acts transitively on $Q(ts)$. Now the subgroup of SO(3,1) leaving this representative pair invariant (i.e., leaving each of e_0 and e_3 invariant) is clearly SO(2) generated by J_3 [U(1) within SL(2,C)], hence it follows that

$$Q(ts) = Q_{12} = G/F_{12} = \mathrm{SO}(3,1)/\mathrm{SO}(2). \tag{2.9}$$

In a similar way we define

$$Q(ss) = \{(a^\mu, b^\mu) | a^2 = b^2 = 1,\ a.b = 0\}. \tag{2.10}$$

This also carries a transitive action by SO(3,1), and the pair (e_1, e_2) can be taken as a representative point. [In this case it must be understood that "points" of $Q(ss)$ are ordered pairs, and (a^μ, b^μ) must be distinguished from (b^μ, a^μ).] The transformations of SO(3,1) that leave both e_1 and e_2 invariant are pure Lorentz transformations in the three-direction, i.e., SO(1,1) with K_3 as its generator. This is the subgroup F_{13} of the general classification, so

$$Q(ss) = Q_{13} = G/F_{13} = \mathrm{SO}(3,1)/\mathrm{SO}(1,1). \tag{2.11}$$

Lastly in this category we have the manifold

$$Q(sl) = \{(a^\mu, b^\mu) | a^2 = 1,\ b^2 = 0,\ b^0 > 0,\ a.b = 0\}, \tag{2.12}$$

for which a representative point is $(e_2, e_0 + e_3)$. The stability group of this pair is generated by $N_2 = J_2 + K_1$, i.e., it is the subgroup F_{14}, so

$$Q(sl) = Q_{14} = G/F_{14}. \tag{2.13}$$

For an object with any one of the SOS's Q_{12}, Q_{13}, or Q_{14} as internal space, we can use a suitable pair of vectors (a^μ, b^μ) as internal coordinates.

The third and final type of configuration we consider is that provided by three mutually orthogonal unit vectors, say

tss, of which the first is positive timelike and the other two spacelike. Once again the sequence in which the last two are given is important. In four-dimensional space-time such a triad (a^μ, b^μ, c^μ) can be completed to a tetrad by adjoining another unit spacelike vector orthogonal to each member of the triad, and with an unambiguous direction. A standard tetrad to which any other tetrad can be taken by a *unique* element of SO(3,1) is of course (e_0, e_1, e_2, e_3). This suffices to show that

$$Q(tss) = \{(a^\mu, b^\mu, c^\mu) | a^2 = -1, \quad b^2 = c^2 = 1,\ a^0 > 0,\ a.b = b.c = c.a = 0\} \quad (2.14)$$

is a realization of the group SO(3,1) itself, i.e.,

$$Q(tss) = Q_{15} = \mathrm{SO}(3,1) = \mathrm{SL}(2,c)/z_2. \quad (2.15)$$

From the discussion above it is clear that in the case of any manifestly covariant SOS the construction of a complete set of ζ's is quite straightforward, since it is reduced to a geometrical problem that is easy to visualize. Once the ζ's are in hand, the most general Lagrangian can be constructed as indicated in Eq. (2.1), and the constraints and dynamics can be studied. It goes without saying that in these cases it is always possible to couple $\dot{x}^\mu$ and the internal variables q, $\dot{q}$ nontrivially.

The use of Q_{15} as the internal space is exemplified by the classic work of Hanson and Regge,[7] as well as in the work of Halbwachs.[11] In a sense Q_{13} appears in the work of Cognola *et al.*,[12] except that the condition of transitive G action is not imposed. The vector model of Mukunda *et al.*[6] is based on Q_4 as the internal space. In Sec. VI, we shall use Q_{13} as an illustrative example to develop a particular Lagrangian model with an interesting internal symmetry.

The remaining SOS's are Q_2, Q_5^φ, Q_7, Q_8, and Q_{11}^φ, of which the second and fifth are one-parameter families. Though Q_2, Q_7, and Q_8 can be visualized geometrically and group theoretically, none of them nor Q_5^φ, Q_{11}^φ can be given a manifestly covariant description, and suitable methods are needed to handle them. To their development we now turn.

III. METHOD OF ISOTOPY REPRESENTATION—DETERMINATION OF UNPHYSICAL SOS's

For any given SOS, the main kinematic or geometric problem to be solved first before the general Lagrangian (2.1) can be written is the construction of a complete set of ζ variables. In the process one would discover whether, in a given case, the $\dot{x}^\mu$ combine nontrivially with the internal q's and $\dot{q}$'s in some of the ζ's. The successive steps one must go through in building up the ζ's were briefly described in the previous section. For those SOS's that can be given a manifestly covariant description, this procedure can be essentially sidestepped in favor of direct geometrical arguments. However, for the remaining three discrete and two one-parameter families of possible Q's we have no alternative but to develop in detail the steps outlined in the previous section.

In order to keep track of the arguments and the algebra we shall name the steps of the procedure of Sec. II in this way. We call step A_0 the attempt to construct quantities of type j on Q for $j = \frac{1}{2}, 1, \frac{3}{2}, \ldots$. Next we call step A_1 the search for quantities of type j on TQ, for $j = 0, \frac{1}{2}, 1, \ldots$, which are linear in the internal velocities $\dot{q}$. This is followed by step A_2 which is the search for quantities of type j on TQ, again for $j = 0, \frac{1}{2}, 1, \ldots$, which are quadratic in the $\dot{q}$. And then follow steps $A_3, \ldots$. All these steps must be gone through for each choice Q of the SOS. To set up the necessary machinery required for these steps, we begin with some well-known properties of the transitive G action on a coset space $Q = G/H$ (Ref. 13.).

It will be adequate to describe both Q and TQ with local coordinates q^r and $q^r, \dot{q}^r$, respectively; for convenience we sometimes write v^r in place of $\dot{q}^r$. There is a distinguished "origin" in Q which is denoted q_0: it is the identity coset consisting of H itself. An element $g \in G$ acting on Q maps the point q^r to an image point q'^r: the dependence of q' on g and q is denoted by a set of functions φ:

$$q'^r = \varphi^r(g;q), \quad q', q \in Q, \quad g \in G. \quad (3.1)$$

Sometimes to save on symbols we write gq for $\varphi(g;q)$. Two immediate properties of φ are

$$\varphi^r(e;q) = q^r, \quad (3.2a)$$

$$\varphi^r(h;q_0) = q_0^r, \quad h \in H. \quad (3.2b)$$

The fact that the point transformations φ follow the composition law of G leads to the functional relations

$$\varphi(g_2;\varphi(g_1;q)) = \varphi(g_2 g_1;q),$$

i.e.,

$$\varphi(g_2;g_1 q) = \varphi(g_2 g_1;q). \quad (3.3)$$

The transformation law for the velocities $\dot{q}^r$ is linear, with coefficients generally dependent on q. It involves the Jacobian matrix $M(g;q)$ of the point transformation φ, with elements

$$M^r{}_s(g;q) = \frac{\partial \varphi^r(g;q)}{\partial q^s}. \quad (3.4)$$

The superscript (subscript) is the row (column) index. Then a point $(q,v) \in TQ$ consisting of $q \in Q$ and a velocity vector $v \in T_q Q$ is carried by $g \in G$ to $(q',v') \in TQ$, where q' is given by Eq. (3.1) and

$$v'^r = M^r{}_s(g;q) v^s. \quad (3.5)$$

In a compact way the action of G on TQ can be written as

$$(q,v) \overset{g \in G}{\rightarrow} (\varphi(g;q), M(g;q)v). \quad (3.6)$$

The properties of φ^r imply corresponding properties for M which help reduce the latter to its essential elements. Thus by differentiating Eqs. (3.3) and (3.2a), respectively, with respect to q we get a functional equation for M, and its value at the identity of G:

$$M(g_2 g_1;q) = M(g_2;g_1 q) M(g_1;q), \quad (3.7a)$$

$$M(e;q) = 1. \quad (3.7b)$$

Two important consequences follow from here on suitable choices of arguments:

$$M(g;q)^{-1} = M(g^{-1};gq), \quad (3.8a)$$

$$M(h_2;q_0) M(h_1;q_0) = M(h_2 h_1;q_0). \quad (3.8b)$$

We see from the second of these that the matrices $M(h;q_0)$

provide us with a linear representation of the subgroup $H \subset G$ which leaves the distinguished point $q_0 \in Q$ invariant. This is called the isotopy representation of H associated with the transitive action of G on G/H; it is of the same dimension as G/H and will be written $D(h)$:

$$M(h;q_0) = D(h), \quad h \in H. \tag{3.9}$$

It will play an important role in the sequel. By setting $q \to q_0$, $g \to h \in H$ in Eq. (3.6) we see that the isotopy representation is the one by which the velocities v^r at q_0 transform when one applies an element $h \in H$:

$$(q_0, v) \overset{h \in H}{\to} (q_0, D(h)v). \tag{3.10}$$

The functional equation (3.7a) can be exploited to express $M(g;q)$ for general arguments entirely in terms of $D(h)$ and a matrix function of q alone.[14] For this, some choice of an element $l(q) \in G$ must be made for each $q \in Q$ such that

$$\varphi(l(q);q_0) = q. \tag{3.11}$$

We agree to set $l(q_0) = e$. It is then easy to see that for any $g \in G$ and $q \in Q$,

$$h(g,q) \equiv l(gq)^{-1} g l(q) \in H. \tag{3.12}$$

Then by manipulating the arguments of the M's in Eq. (3.7a) and also using Eq. (3.8a) we obtain

$$M(g;q) = M(l(gq);q_0) D(h(g,q)) M(l(q);q_0)^{-1}. \tag{3.13}$$

Therefore $M(g;q)$ is determined for all g and q by the isotopy representation matrices $D(h)$ for all $h \in H$, and one Jacobian matrix $M(l(q);q_0)$ for each $q \in Q$. This breakup of $M(g;q)$ suggests, in conjunction with the transformation law for the velocities $\dot{q} = v$ appearing in Eqs. (3.5) and (3.6), that we define auxiliary quantities $w \in T_q Q$ in this way:

$$w = M(l(q);q_0)^{-1} v. \tag{3.14}$$

These quantities w^r are like angular velocity components, being linear combinations of the true velocities v^r with q-dependent coefficients. When q is mapped to $q' = gq$ by $g \in G$, then $w \in T_q Q$ is taken to $w' \in T_{q'} Q$ by a matrix of the isotopy representation:

$$w' = D(h(g,q))w. \tag{3.15}$$

With this machinery relating to the action of G on G/H, we can tackle the problem of constructing the ζ's.

We consider step A_0 first. Let us denote by $(\mathscr{D}^{(j)}_{AB}(g))$ the matrices of the representation (j,j) of G—this notation is hereafter used in place of the previous $D^{(j,j)}$ to avoid overuse of the letter D already earmarked for the isotopy representation—and by $\mathscr{V}^{(j)}$ the vector space carrying this representation. The relevant values of j are $\frac{1}{2}, 1, \frac{3}{2}, \ldots$; and for given j we ask if we can find $(2j+1)^2$ functions $V_A(q)$ on Q (arranged into a column vector) obeying the functional equation

$$V(gq) = \mathscr{D}^{(j)}(g) V(q). \tag{3.16}$$

The necessary and sufficient conditions for existence of such $V(q)$ can be derived by a three stage argument. First we set $q \to q_0$, $g \to h \in H$ in (3.16) to get a condition on $V(q_0) = f \in \mathscr{V}^{(j)}$:

$$\mathscr{D}^{(j)}(h) f = f, \quad \text{all } h \in H. \tag{3.17}$$

Suppose there is a nonvanishing solution f. It means that $\mathscr{V}^{(j)}$ contains a one-dimensional subspace spanned by f, invariant under the representation of H obtained by restricting the representation $\mathscr{D}^{(j)}$ of G to H. (Of course there may be several independent f's.) Next we take $q \to q_0$, $g \to l(q)$ in (3.16) to get a determination of $V(q)$ at all $q \in Q$ in terms of f:

$$V(q) = \mathscr{D}^{(j)}(l(q)) f. \tag{3.18}$$

Because of Eq. (3.17), $V(q)$ so determined is independent of the precise choice of $l(q)$ as long as the requirement (3.11) is met. The third and final stage is to confirm that $V(q)$ so determined does obey condition (3.16); this is easily done, and once again it uses the property (3.17) of f. The result for step A_0 of this three stage analysis can be expressed in two equivalent ways: (a) a nonvanishing quantity of type j can be found on Q if and only if the representation $\mathscr{D}^{(j)}(h)$ of H contains the identity or trivial representation of H; and (b) $\mathscr{V}^{(j)}$ must contain a vector f whose stability group is larger than or equal to H.

Next we turn to step A_1, the attempt to construct quantities of type j linear in $\dot{q}$, for $j = 0, \frac{1}{2}, 1, \ldots$. It is convenient to express such a collection of $(2j+1)^2$ functions on TQ, $V_A(q,\dot{q})$, as linear combinations of the w^r:

$$V_A(q,\dot{q}) = F_{Ar}(q) w^r. \tag{3.19}$$

The property of $V(q,\dot{q})$ is that when q and w are transformed to q' and w' by an element $g \in G$, following Eqs. (3.1) and (3.15), then $V = Fw$ transforms by $\mathscr{D}^{(j)}(g)$:

$$F_{Ar}(q') w'^r = \mathscr{D}^{(j)}_{AB}(g) F_{Br}(q) w^r. \tag{3.20}$$

This means that $F_{Ar}(q)$ must obey the functional equation

$$D^r{}_{r'}(h(g,q)) F_{Ar'}(gq) = \mathscr{D}^{(j)}_{AB}(g) F_{Br}(q), \tag{3.21}$$

for all $q \in Q$, $g \in G$. The general solution to this equation can be developed by a three-stage argument in the same pattern as with step A_0. First set $q \to q_0$, $g \to h \in H$ in (3.21); $F(q_0)$ is then required to obey

$$\mathscr{D}^{(j)}_{AB}(h) F_{Br}(q_0) = D^r{}_{r'}(h) F_{Ar'}(q_0), \quad \text{all } h \in H. \tag{3.22}$$

The invariant interpretation of this requirement will be given later; for the present assume a nonzero $F(q_0)$ can be found. Next set $q \to q_0$, $g \to l(q)$ in (3.21), to determine $F(q)$ at all $q \in Q$ in terms of $F(q_0)$:

$$F_{Ar}(q) = \mathscr{D}^{(j)}_{AB}(l(q)) F_{Br}(q_0). \tag{3.23}$$

Here in contrast to Eq. (3.18) of step A_0, the choice of $l(q)$ is relevant in determining $F(q)$ (Ref. 15). The third stage is to verify that this $F(q)$ does obey the condition (3.21) for all g. It is easily verified that this is so, thanks to Eq. (3.22). The necessary and sufficient condition to be able to find a quantity of type j linear in $\dot{q}$ is thus that there should be nonzero quantities $F_{Ar}(q_0)$ connecting $\mathscr{D}^{(j)}(h)$ and $D(h)$ in the manner of Eq. (3.22). The interpretation of this intertwining relation is as follows. Regard each column $F_{Ar}(q_0)$, r fixed, as a vector with $(2j+1)^2$ components in $\mathscr{V}^{(j)}$: the rth vector f_r has $F_{Ar}(q_0)$ for its A th component. Then writing Eq. (3.22) as

$$\mathscr{D}^{(j)}(h) f_r = D^{r'}{}_r(h) f_{r'}, \tag{3.24}$$

we read it as requiring that the action of $\mathscr{D}^{(j)}(h)$ on the rth vector $f_r \in \mathscr{V}^{(j)}$ must be to produce a linear combination of

all the f's with the matrix elements $D^r{}_r(h)$ of the isotopy representation as coefficients. There is no condition here that these vectors f_r must be linearly independent.

Step A_2 can be handled in a similar manner. The $(2j+1)^2$ quantities $V_A(q,\dot{q})$ of type j, quadratic in $\dot{q}$, are written as

$$V_A(q,\dot{q}) = F_{Ars}(q)w^r w^s, \tag{3.25}$$

with coefficient functions $F_{Ars}(q)$ symmetric in rs. It is now easy to see that Eq. (3.20) is here replaced by

$$F_{Ars}(q')w'^r w'^s = \mathscr{D}^{(j)}_{AB}(g)F_{Brs}(q)w^r w^s, \tag{3.26}$$

which in turn leads to the basic functional equation

$$\mathscr{D}^{(j)}_{AB}(g)F_{Brs}(q) = D^{r'}{}_r(h(g,q))D^{s'}{}_s(h(g,q))F_{Ar's'}(gq). \tag{3.27}$$

The first stage of the three stage analysis is to take $q\to q_0$, $g\to h\in H$: then $F_{Ars}(q_0)$ must obey

$$\mathscr{D}^{(j)}_{AB}(h)F_{Brs}(q_0) = D^{r'}{}_r(h)D^{s'}{}_s(h)F_{Ar's'}(q_0), \quad \text{all } h\in H. \tag{3.28}$$

The second stage is to set $q\to q_0$, $g\to l(q)$ in (3.27), which determines $F(q)$ all over Q:

$$F_{Ars}(q) = \mathscr{D}^{(j)}_{AB}(l(q))F_{Brs}(q_0). \tag{3.29}$$

Here again, as with step A_1, the choice of $l(q)$ does affect the value of $F(q)$ (Ref. 15). The third and final stage is to confirm that the $F(q)$ so determined does obey (3.27), and this is straightforward. The necessary and sufficient condition to be able to find a quantity of type j on TQ quadratic in $\dot{q}$ is thus that there should be a nonzero $F(q_0)$ obeying Eq. (3.28). We can again interpret this in an intrinsic manner: We require a set of vectors $f_{rs} = f_{sr} \in \mathscr{V}^{(j)}$, the A th component of f_{rs} being $F_{Ars}(q_0)$, which behave under the action of $\mathscr{D}^{(j)}(h)$ in the following manner:

$$\mathscr{D}^{(j)}(h)f_{rs} = D^{r'}{}_r(h)D^{s'}{}_s(h)f_{r's'}. \tag{3.30}$$

Generalizing the requirements (3.17), (3.24), (3.30) in steps A_0, A_1, A_2, the pattern for A_3,... is clear.

In order to solve the above requirements in any particular case, it is more convenient to deal with their infinitesimal forms corresponding to the Lie algebra $\mathfrak{h}$ of H. Let H_a, $a = 1,\ldots,6 - \dim Q$, be a basis for $\mathfrak{h}$. They are linear combinations of the generators J_j, K_j of $\mathfrak{g}$. The symbols J_j, K_j, H_a will be used both for the abstract generators and for their representatives in any of the representations $\mathscr{D}^{(j)}(g)$ of G of interest to us. On the other hand, the generator matrices of the isotopy representation $D(h)$ of H will be written $(h_a)^r{}_s$, consistent with the arrangement of indices on $M(g;q)$ and $D(h)$. These matrices h_a are easy to obtain from the Lie bracket relations of $\mathfrak{g}$. For, suppose we extend H_a to a basis for $\mathfrak{g}$ by adjoining additional elements P_r, $r = 1,\ldots,\dim Q$. In a suitable neighborhood of the identity in G, each $g\in G$ can be uniquely expressed as a product $\exp(q^r P_r)\exp(\xi^a H_a)$, so (q^r,ξ^a) are (local) coordinates for G; while $\exp(q^r P_r)$ and q^r are (local) coset representatives and coordinates respectively for $Q = G/H$. By premultiplying $\exp(q^r P_r)$ by an (infinitesimal) element of H and following through the changes caused in the q^r, we find that the matrices $(h_a)^r{}_s$ can be read off from the bracket relations involving an H_a and a P_r:

$$[H_a,P_r] = (h_a)^s{}_r P_s + \text{ terms in } \mathfrak{h}. \tag{3.31}$$

The infinitesimal versions of Eqs. (3.17), (3.24), and (3.30) can now be written. They are

$$\text{step } A_0\text{:}\quad H_a f = 0; \tag{3.32a}$$

$$\text{step } A_1\text{:}\quad H_a f_r = (h_a)^{r'}{}_r f_{r'}; \tag{3.32b}$$

$$\text{step } A_2\text{:}\quad H_a f_{rs} = (h_a)^{r'}{}_r f_{r's} + (h_a)^{s'}{}_s f_{rs'},\quad f_{rs} = f_{sr}; \tag{3.32c}$$

and so on for the later steps A_3,... .

We now apply the methods developed above to show that for the second-order internal spaces Q_2, Q^q_ξ, and Q_7 there are no ξ's at all and hence no possibility of coupling space-time and internal variables in these cases. We shall treat these Q's in the sequence Q_2, Q_7, and Q^q_ξ. The following notations for combinations of the generators J_j, K_j of $\mathfrak{g}$ will be useful (the N's have been already used):

$$\begin{aligned} &N_1 = J_1 - K_2,\quad N_2 = J_2 + K_1;\\ &J_\varphi = \sin\varphi\, J_3 + \cos\varphi\, K_3,\quad 0<\varphi<\pi/2 \text{ or } \pi/2<\varphi<\pi;\\ &P_1 = J_1 + K_2,\quad P_2 = J_2 - K_1;\\ &P_\varphi = \cos\varphi\, J_3 - \sin\varphi\, K_3. \end{aligned} \tag{3.33}$$

Subscripts $a,b,\ldots$ will be used to go over the two "transverse" values 1,2. An elementary lemma relating to the representations $\mathscr{D}^{(j)}(g)$ of G for $j\geq\frac{1}{2}$, proved in the Appendix, will be of great use in the sequel. It states that if a vector $f\in\mathscr{V}^{(j)}$ is annihilated by N_a and is an eigenvector of K_3 with a real nonnegative eigenvalue, then f vanishes identically

$$\left.\begin{aligned} &N_a f = 0,\quad a = 1,2\\ &K_3 f = \lambda f,\quad \lambda\geq 0 \end{aligned}\right\} \Rightarrow f = 0. \tag{3.34}$$

The case $Q_2 = G/F_2$: The four generators of F_2 are J_3, K_3, and N_a, $a = 1,2$. To extend these to a basis for $\mathfrak{g}$ we choose the two additional elements P_a, $a = 1,2$. The commutators of the form $[H,P]$ appearing in Eq. (3.31) have the values

$$\begin{aligned} &[J_3,P_a] = \epsilon_{ab}P_b,\\ &[K_3,P_a] = P_a,\\ &[N_a,P_b] = 2\delta_{ab}K_3 + 2\epsilon_{ab}J_3. \end{aligned} \tag{3.35}$$

Therefore the 2×2 matrices that generate the isotopy representation of F_2 can be read off:

$$J_3\to\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},\quad K_3\to\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad N_a\to\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}. \tag{3.36}$$

(This is therefore a nonfaithful representation of F_2.) Let us now examine step A_0 for this case. We seek a vector $f\in\mathscr{V}^{(j)}$, for $j\geq\frac{1}{2}$, annihilated by all four generators J_3, K_3, and N_a: by the lemma, f vanishes and so there are no quantities of type j on Q_2 for any $j\geq\frac{1}{2}$. Next we examine steps A_1, A_2,... . We can quite easily deal with the general step A_n for any $n\geq 1$. We are required to find vectors $f_{a_1a_2\cdots a_n}\in\mathscr{V}^{(j)}$, symmetric in all the a's and with each $a = 1,2$, obeying on account of (3.36) the conditions

$$\begin{aligned} &N_a f_{a_1a_2\cdots a_n} = 0,\\ &K_3 f_{a_1a_2\cdots a_n} = n f_{a_1a_2\cdots a_n}. \end{aligned} \tag{3.37}$$

The values of j of interest here are $j = 0, \frac{1}{2}, 1, \ldots$. If $j = 0$, the left-hand sides vanish, then so does $f_{a_1 a_2 \cdots a_n}$ since $n \geqslant 1$. If $j \geqslant \frac{1}{2}$, the lemma immediately tells us again that $f_{a_1 \cdots a_n} = 0$. Therefore no quantities of type j on TQ_2, for any $j = 0, \frac{1}{2}, \ldots$, and of any degree $n = 1,2,\ldots$ in the $\dot{q}$, can be constructed. This completes the proof for this case that no ζ's exist.

The case $Q_7 = G/F_7$: The three generators of F_7 are K_3 and N_a. We extend them to a basis for $\mathfrak{g}$ by taking P_1, P_2 as defined in (3.33) and $P_3 = J_3$. From the subset of commutation relations of $\mathfrak{g}$,

$$\begin{aligned} &[K_3, P_a] = P_a, \quad [K_3, P_3] = 0, \\ &[N_a, P_b] = 2\delta_{ab} K_3 + 2\epsilon_{ab} P_3, \\ &[N_a, P_3] = -\epsilon_{ab} N_b, \end{aligned} \tag{3.38}$$

we infer that the 3×3 matrices generating the isotopy representation of F_7 are

$$\begin{aligned} K_3 &\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ N_1 &\to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}, \\ N_2 &\to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -2 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{3.39}$$

Step A_0 for Q_7 can be handled as in the case of Q_2: we need a vector $f \in \mathscr{V}^{(j)}$, $j \geqslant \frac{1}{2}$, annihilated by N_a and K_3, and by the lemma $f = 0$. Therefore there are no quantities of type j on Q_7 for any $j \geqslant \frac{1}{2}$. Turning next to step A_n for general $n \geqslant 1$: we seek a system of vectors $f_{r_1 r_2 \cdots r_n} \in \mathscr{V}^{(j)}$, symmetric in the r's and with each $r = 1,2,3$, behaving in a suitable way under action by N_a and K_3. The symmetry in the indices allows us to write f_{n_1,n_2,n_3} for these vectors where $n_1 + n_2 + n_3 = n$ and n_1 is the number of 1's, n_2 the number of 2's, n_3 the number of 3's in the collection $r_1 r_2 \cdots r_n$. Then by using the isotopy generator matrices (3.39) and extending the pattern of Eqs. (3.32), we see that $f_{n_1,n_2,n_3} \in \mathscr{V}^{(j)}$ must obey

$$N_1 f_{n_1,n_2,n_3} = 2n_2 f_{n_1,n_2-1,n_3+1}, \tag{3.40a}$$

$$N_2 f_{n_1,n_2,n_3} = -2n_1 f_{n_1-1,n_2,n_3+1}, \tag{3.40b}$$

$$K_3 f_{n_1,n_2,n_3} = (n_1 + n_2) f_{n_1,n_2,n_3}, \quad (n_1 + n_2) \geqslant 0. \tag{3.40c}$$

The range of j values here is $0, \frac{1}{2}, 1, \ldots$. For $j = 0$, the left-hand sides vanish, so the only possibly nonvanishing f is $f_{0,0,n}$. But by taking, say, $n_1 = 0$, $n_2 = 1$, $n_3 = n - 1$ in (3.40a) we see that $f_{0,0,n} = 0$ as well. For $j \geqslant \frac{1}{2}$ we use a stepwise argument. The lemma shows to begin with that $f_{0,0,n} = 0$. Using this we see next that the two vectors $f_{1,0,n-1}$ and $f_{0,1,n-1}$ must be annihilated by N_a and are eigenvectors of K_3 with eigenvalue $+1$. By the lemma they must both vanish. At the next step the three vectors $f_{2,0,n-2}, f_{1,1,n-2}, f_{0,2,n-2}$ are seen to vanish; and so on for all f_{n_1,n_2,n_3}. Thus no quantities of type j on TQ_7, for any $j = 0, \frac{1}{2}, \ldots$, and of any degree $n = 1,2,\ldots$ in the $\dot{q}$, can be constructed. In this case too, there are no ζ's.

The case $Q_5^\varphi = G/F_5^\varphi$: The three generators of F_5^φ are J_φ, N_1, and N_2. These are extended to a basis for $\mathfrak{g}$ by adjoining the three elements P_1, P_2, and $P_3 = P_\varphi$ of Eq. (3.33). The commutators among the two sets are

$$\begin{aligned} &[J_\varphi, P_a] = \cos\varphi\, P_a + \sin\varphi\, \epsilon_{ab} P_b, \\ &[J_\varphi, P_3] = 0, \\ &[N_a, P_b] = 2\delta_{ab}(\cos\varphi\, J_\varphi - \sin\varphi\, P_\varphi) \\ &\qquad + 2\epsilon_{ab}(\sin\varphi\, J_\varphi + \cos\varphi\, P_\varphi), \\ &[N_a, P_3] = -\sin\varphi\, N_a - \cos\varphi\, \epsilon_{ab} N_b. \end{aligned} \tag{3.41}$$

Therefore the 3×3 matrices generating the isotopy representation of F_5^φ are

$$\begin{aligned} J_\varphi &\to \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ N_1 &\to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -2\sin\varphi & +2\cos\varphi & 0 \end{pmatrix}, \\ N_2 &\to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -2\cos\varphi & -2\sin\varphi & 0 \end{pmatrix}. \end{aligned} \tag{3.42}$$

We shall examine steps A_0, A_1, and A_2 in sequence in this case since unlike the two previous cases it is now a bit cumbersome to deal directly with a general A_n. Let us first take A_0. For some $j \geqslant \frac{1}{2}$ we seek a vector f annihilated by all the generators of F_5^φ:

$$N_a f = J_\varphi f = 0, \quad f \in \mathscr{V}^{(j)}. \tag{3.43}$$

In the notation of the Appendix, the first two conditions determine f to be a multiple of the basis vector $\psi^{(j)}_{j,-j}$ of the space $\mathscr{V}^{(j)}$:

$$N_a f = 0 \Rightarrow f = c.\psi^{(j)}_{j,-j}. \tag{3.44}$$

Now on this basis vector we have [Eq. (A7)]

$$\begin{aligned} &J_3 \psi^{(j)}_{j,-j} = 0, \\ &K_3 \psi^{(j)}_{j,-j} = -2j\psi^{(j)}_{j,-j}, \end{aligned} \tag{3.45}$$

so the third condition on f,

$$\begin{aligned} J_\varphi f &= c.(\sin\varphi\, J_3 + \cos\varphi\, K_3)\psi^{(j)}_{j,-j} \\ &= -2j\, c.\cos\varphi\, \psi^{(j)}_{j,-j} = 0, \end{aligned} \tag{3.46}$$

determines $c = 0$ since both j and $\cos\varphi$ are nonzero. Thus we see fairly easily that there are no quantities of type j for $j \geqslant \frac{1}{2}$ on Q_5^φ.

Next we take up step A_1. Here for some $j \geqslant 0$, we seek three vectors $f_r \in \mathscr{V}^{(j)}$, $r = 1,2,3$, obeying

$$\begin{aligned} N_1 \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} &= \begin{pmatrix} 0 & 0 & -2\sin\varphi \\ 0 & 0 & 2\cos\varphi \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}, \\ N_2 \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} &= \begin{pmatrix} 0 & 0 & -2\cos\varphi \\ 0 & 0 & -2\sin\varphi \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}, \\ J_\varphi \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} &= \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}. \end{aligned} \tag{3.47}$$

For $j=0$, since the left-hand sides vanish, and both $\cos\varphi$ and $\sin\varphi$ are nonzero, the vanishing of all f_r follows. Thus there are no Lorentz scalars on TQ_5^{φ} linear in the $\dot{q}$'s. For $j\geqslant\frac{1}{2}$: we see that f_3 must be annihilated by N_a and J_φ, so by the argument just used for step A_0 it follows that $f_3=0$. Using this we get for f_1 and f_2:

$$N_a f_b=0\Rightarrow f_a=c_a\psi^{(j)}_{j,-j}$$
$$\Rightarrow J_\varphi f_a=-2j\cos\varphi f_a. \tag{3.48}$$

This combined with the third equation of (3.47) demands that f_a fulfill

$$\begin{pmatrix}\cos\varphi & \sin\varphi\\ -\sin\varphi & \cos\varphi\end{pmatrix}\begin{pmatrix}f_1\\ f_2\end{pmatrix}=-2j\cos\varphi\begin{pmatrix}f_1\\ f_2\end{pmatrix}. \tag{3.49}$$

But this forces $f_a=0$ since the 2×2 matrix on the left has eigenvalues $e^{\pm i\varphi}$ that have nonvanishing imaginary parts. Thus the answer to step A_1 for Q_5^{φ} is that there are no quantities of type j linear in $\dot{q}$ for any $j=0,\frac{1}{2},1,\ldots$.

For step A_2, we need six vectors $f_{rs}=f_{sr}\in\mathscr{V}^{(j)}$, $j=0,\frac{1}{2},\ldots$, obeying

$$N_1\begin{pmatrix}f_{11}\\ f_{12}\\ f_{13}\\ f_{22}\\ f_{23}\\ f_{33}\end{pmatrix}=\begin{pmatrix}-4\sin\varphi f_{13}\\ 2\cos\varphi f_{13}-2\sin\varphi f_{23}\\ -2\sin\varphi f_{33}\\ 4\cos\varphi f_{23}\\ 2\cos\varphi f_{33}\\ 0\end{pmatrix},$$

$$N_2\begin{pmatrix}f_{11}\\ f_{12}\\ f_{13}\\ f_{22}\\ f_{23}\\ f_{33}\end{pmatrix}=\begin{pmatrix}-4\cos\varphi f_{13}\\ -2\sin\varphi f_{13}-2\cos\varphi f_{23}\\ -2\cos\varphi f_{33}\\ -4\sin\varphi f_{23}\\ -2\sin\varphi f_{33}\\ 0\end{pmatrix}, \tag{3.50}$$

$$J_\varphi\begin{pmatrix}f_{11}\\ f_{12}\\ f_{13}\\ f_{22}\\ f_{23}\\ f_{33}\end{pmatrix}=\begin{pmatrix}2\cos\varphi f_{11}+2\sin\varphi f_{12}\\ -\sin\varphi f_{11}+2\cos\varphi f_{12}+\sin\varphi f_{22}\\ \cos\varphi f_{13}+\sin\varphi f_{23}\\ -2\sin\varphi f_{12}+2\cos\varphi f_{22}\\ -\sin\varphi f_{13}+\cos\varphi f_{23}\\ 0\end{pmatrix}.$$

These effects of N_a and J_φ are obtained by iterating Eqs. (3.47) on each index on f_{rs}. For $j=0$, since both $\cos\varphi$ and $\sin\varphi$ are nonzero, the vanishing of N_a and J_φ forces $f_{rs}=0$. (While the vanishing of f_{13}, f_{23}, and f_{33} is immediate, for f_{11}, f_{12}, and f_{22} a little work is needed.) For $j\geqslant\frac{1}{2}$, the conditions on f_{33} are the same as on f in step A_0, therefore $f_{33}=0$. Using this, the conditions on f_{13} and f_{23} are the same as on f_1 and f_2 in step A_1 (after $f_3=0$ had been established there), so $f_{13}=f_{23}=0$. It remains to examine the conditions on f_{ab}, a and $b=1,2$, taking into account $f_{r3}=0$. We have to solve $N_a f_{bc}=0$,

$$J_\varphi\begin{pmatrix}f_{11}\\ f_{12}\\ f_{22}\end{pmatrix}=\begin{pmatrix}2\cos\varphi & 2\sin\varphi & 0\\ -\sin\varphi & 2\cos\varphi & \sin\varphi\\ 0 & -2\sin\varphi & 2\cos\varphi\end{pmatrix}\begin{pmatrix}f_{11}\\ f_{12}\\ f_{22}\end{pmatrix}. \tag{3.51}$$

By a now familiar argument, the first two conditions on f_{ab} force each f_{ab} to be a multiple of $\psi^{(j)}_{j,-j}$ and therefore an eigenvector of J_φ with eigenvalue $-2j\cos\varphi$. However the eigenvalues of the 3×3 matrix appearing in the J_φ condition are $2e^{i\varphi}$, $2e^{-i\varphi}$, and $2\cos\varphi$ corresponding to the combinations $f_{11}-f_{22}-2if_{12}$, $f_{11}-f_{22}+2if_{12}$, and $f_{11}+f_{22}$, respectively; and each of these eigenvalues is definitely distinct from $-2j\cos\varphi$. It is by this delicate argument that we conclude that f_{ab} must also vanish, so the answer to step A_2 is that there are no quantities of type $j=0,\frac{1}{2},1,\ldots$ on TQ_5^{φ}, quadratic in the $\dot{q}$'s. The analysis of steps $A_3,\ldots$ follows a similar pattern.

As we shall mention in the concluding section, both Q_2 and Q_7 can be visualized geometrically even though they are not manifestly covariant, nevertheless the detailed analysis of this section is needed to see that there are no ζ's in these cases as well as in Q_5^{φ}.

IV. PHYSICAL NONMANIFESTLY COVARIANT SOS'S

The SOS's that have still to be examined, to see if nontrivial ζ's can be constructed, are $Q_8=G/F_8$ and $Q_{11}^{\varphi}=G/F_{11}^{\varphi}$. We shall tackle them in this section. For Q_8 a geometrical argument shows that nontrivial ζ's certainly exist, while a complete treatment (here omitted) must use the methods of the previous section. For Q_{11}^{φ} those methods are essential to show that ζ's do exist.

We take up $Q_8=G/F_8$ first. The subgroup F_8 is generated by K_3 and N_2. These generators belong to the su(1,1) subalgebra of g spanned by J_2, K_3, and K_1. This corresponds to the SO(2,1) subgroup of SO(3,1) acting on the dimensions 013, which is the stability group of the (unit) spacelike vector $e_2=(0,0,1,0)$. Of course, K_3, K_1, and J_2 separately annihilate e_2. On the other hand, N_2 is one of the three generators belonging to the stability group of the positive lightlike vector $e_0+e_3=(1,0,0,1)$, the others being N_1 and J_3. Of course, K_3 does not annihilate e_0+e_3; rather, the corresponding pure Lorentz transformations scale e_0+e_3 by a positive factor. All these facts suggest that Q_8 can be obtained by a simple procedure from the space $Q(sl)=Q_{14}=G/F_{14}$ defined in Eqs. (2.12) and (2.13). The representative point $(a^\mu,b^\mu)=(e_2,e_0+e_3)$ of $Q(sl)$ is invariant under the one-parameter subgroup generated by N_2. To accommodate the additional generator K_3, we define an SO(3,1) invariant equivalence relation on $Q(sl)$ as follows:

$$(a^\mu,\kappa b^\mu)\sim(a^\mu,b^\mu),\quad \kappa>0. \tag{4.1}$$

This freedom can be used to reduce the time component of b^μ to unity, when the space part of b^μ becomes a unit vector $\hat{b}$. The quotient space $Q(sl)/\sim$ can then be described as follows:

$$Q(sl)/\sim=\{(a^\mu,\hat{b})\,|\,a^2=1,\ \hat{b}\in S^2,\ \mathbf{a}\cdot\hat{\mathbf{b}}=a^0\}. \tag{4.2}$$

From the linear four-vector SO(3,1) transformation law for the lightlike vector b^μ we get a nonlinear action by SO(3,1) on $\hat{b}\in S^2$; this combined with the SO(3,1) action on a^μ as a four-vector fixes the action of SO(3,1) on $Q(sl)/\sim$. Any $(a^\mu,\hat{b})$ can be carried via SO(3,1) to the standard configuration $a^\mu=(0,0,1,0)$, $\hat{b}=(0,0,1)$; and the stability group of

this configuration is easily seen to be generated by N_2 and K_3. Thus we have the result

$$Q_8 = G/F_8 = Q(sl)/\sim. \tag{4.3}$$

This suffices to show that with Q_8 as an internal space, nontrivial ζ's and coupling between internal and space-time variables certainly exist, since, for example, both a^μ and $\dot{a}^\mu$ could be contracted with $\dot{x}^\mu$. We shall be content with this remark, realizing that a complete analysis must use the methods based on the isotopy representation.

In the case of the SOS $Q^{\varphi}_{11} = G/F^{\varphi}_{11}$, not even as much of a geometrical picture as with Q_8 is easily available, and one has to rely totally on the methods of the previous section to demonstrate that nontrivial ζ's do exist. We shall now indicate how this happens. The subgroup F^{φ}_{11} is one dimensional, with the single generator J_φ. Thus Q^{φ}_{11} is a five-dimensional manifold, and the isotopy representation of F^{φ}_{11} is likewise five dimensional. The generator matrix for this representation is obtained by extending J_φ to a basis for $\mathfrak{g}$ with the help of five additional elements P_r, $r = 1,...,5$: P_1 and P_2 as in Eq. (3.33), $P_3 = N_1$, $P_4 = N_2$, and $P_5 = P_\varphi$ of Eq. (3.33). The commutators of J_φ with the P_r are

$$\begin{aligned}
&[J_\varphi, P_a] = \cos\varphi\, P_a + \sin\varphi\, \epsilon_{ab} P_b, \\
&[J_\varphi, P_{a+2}] = -\cos\varphi\, P_{a+2} + \sin\varphi\, \epsilon_{ab} P_{b+2}, \\
&[J_\varphi, P_5] = 0, \quad a,b = 1,2.
\end{aligned} \tag{4.4}$$

Therefore the isotopy representation is generated by

$$J_\varphi \to \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 & 0 & 0 \\ \sin\varphi & \cos\varphi & 0 & 0 & 0 \\ 0 & 0 & -\cos\varphi & -\sin\varphi & 0 \\ 0 & 0 & \sin\varphi & -\cos\varphi & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.5}$$

We now analyze step A_0 for Q^{φ}_{11}: for some $j \geq \frac{1}{2}$ we ask if there is a vector $f \in \mathscr{V}^{(j)}$ annihilated by J_φ in the representation $\mathscr{D}^{(j)}$ of G:

$$J_\varphi f = 0. \tag{4.6}$$

This question is easy to answer because in the canonical basis for $\mathscr{V}^{(j)}$, described in the Appendix, J_φ is diagonal:

$$J_\varphi\, \psi^{(j)}_{m_1,m_2} = (m_2 e^{-i\varphi} - m_1 e^{i\varphi}) \psi^{(j)}_{m_1,m_2}. \tag{4.7}$$

Since both $\cos\varphi$ and $\sin\varphi$ are nonzero, the only way this eigenvalue can vanish is when $m_1 = m_2 = 0$, which restricts j to integer values; and in that case f is unique up to a factor:

$$f = c \cdot \psi^{(j)}_{0,0}, \quad j = 1,2,.... \tag{4.8}$$

We conclude that it is possible to construct precisely one quantity of type j (up to a factor) for each integer value of j, on Q^{φ}_{11}. This already shows the existence of ζ variables, even without any $\dot{q}$ dependence. We can examine the structure of these quantities of type j on Q^{φ}_{11} a little further. The value $j = 1$ corresponds to symmetric traceless rank-2 tensors, written in Cartesian form as $t_{\mu\nu}$; for $j = 2$ we have symmetric traceless rank-4 tensors $t_{\mu\nu\rho\sigma}$, and so on. Now from Eq. (4.8) we see that within any representation space $\mathscr{V}^{(j)}$ a null eigenvector of J_φ is always and only obtained by J_3 and K_3 individually annihilating this vector

$$J_\varphi f = 0 \Leftrightarrow J_3 f = K_3 f = 0. \tag{4.9}$$

Using this, the Cartesian form of $\psi^{(1)}_{0,0}$ is easily calculated to have only diagonal nonzero components

$$t_{00}(q_0) = t_{11}(q_0) = t_{22}(q_0) = -t_{33}(q_0) = 1,$$

i.e.,

$$t_{\mu\nu}(q_0) = \delta_{\mu\nu} - 2\delta_{\mu 3}\delta_{\nu 3}. \tag{4.10}$$

As explained in the previous section, $t_{\mu\nu}(q_0)$ determines $t_{\mu\nu}(q)$ elsewhere on Q^{φ}_{11}. Now the case $j = 2$ can be built up from $j = 1$ in a Lorentz invariant way:

$$t_{\mu\nu\rho\sigma}(q) = t_{\mu\nu}(q) t_{\rho\sigma}(q) + t_{\mu\rho}(q) t_{\sigma\nu}(q) + t_{\mu\sigma}(q) t_{\nu\rho}(q) - \text{traces}. \tag{4.11}$$

The removal of traces is easy since from Eq. (4.10) we have for all q:

$$t_{\mu\nu}(q) t^{\mu}{}_{\lambda}(q) = g_{\nu\lambda}. \tag{4.12}$$

At q_0, since $t_{\mu\nu}(q_0)$ is annihilated by J_3 and K_3, the method of construction guarantees that $t_{\mu\nu\rho\sigma}(q_0)$ will also be annihilated by J_3 and K_3, hence by J_φ. Therefore $t_{\mu\nu\rho\sigma}(q)$ of Eq. (4.11) must be (apart from a factor) just the quantity of type $j = 2$ that would be obtained by starting with $\psi^{(2)}_{0,0}$ at q_0 and then using Eq. (3.18) to move to other points of Q^{φ}_{11}. Similar arguments evidently work for $j = 3,4,...$. We conclude that the basic algebraically independent symmetric traceless tensor on Q^{φ}_{11} is the second-rank $t_{\mu\nu}(q)$ for $j = 1$, all others being obtained from this one by Lorentz invariant methods. Moreover it can be easily checked that the stability group of $t_{\mu\nu}(q_0)$ is exactly $SO(2) \times SO(1,1)$ generated by J_3 and K_3, no larger. The only algebraically independent ζ variable independent of $\dot{q}$ is therefore

$$\zeta = \dot{x}^\mu \dot{x}^\nu t_{\mu\nu}(q)/(-\dot{x}^2). \tag{4.13}$$

These results concerning the algebraically independent quantities of type j we have been able to form on Q^{φ}_{11} tell us something about the manifold structure of Q^{φ}_{11} itself, but we shall come back to this after looking briefly at step A_1.

The values of j relevant for step A_1 are $j = 0, \frac{1}{2}, 1, ...$. For any such j, we seek vectors $f_r \in \mathscr{V}^{(j)}$, $r = 1,...,5$, which behave as follows under action by J_φ:

$$J_\varphi \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 & 0 & 0 \\ -\sin\varphi & \cos\varphi & 0 & 0 & 0 \\ 0 & 0 & -\cos\varphi & \sin\varphi & 0 \\ 0 & 0 & -\sin\varphi & -\cos\varphi & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix}. \tag{4.14}$$

These requirements are more conveniently written as

$$J_\varphi(f_1 \pm if_2) = e^{\mp i\varphi}(f_1 \pm if_2),$$
$$J_\varphi(f_3 \pm if_4) = -e^{\pm i\varphi}(f_3 \pm if_4), \tag{4.15}$$
$$J_\varphi f_5 = 0.$$

Consider first $j=0$, when $J_\varphi = 0$. We see immediately that of the five f's, which are now real numbers, only f_5 can be nonzero. Setting $f_5 = 1$, we see that we have succeeded in finding a Lorentz scalar $s(q,\dot{q})$ linear in the internal velocities

$$s(q,\dot{q}) = w^5 = M(l(q)^{-1};q)^5{}_r\dot{q}^r. \tag{4.16}$$

Despite appearances, $s(q,\dot{q})$ is in fact independent of the precise choice of $l(q)$. Next let us consider values of $j \geqslant \frac{1}{2}$. We know that in the representation $\mathscr{D}^{(j)}$ of G, J_φ is diagonal in the canonical basis, with eigenvalues given in Eq. (4.7), while we require f_r obey Eqs. (4.15). Comparing the two, and subject to reality conditions, the f_r can only be

$$f_1 + if_2 \sim \psi^{(j)}_{0,1}, \quad f_1 - if_2 \sim \psi^{(j)}_{-1,0},$$
$$f_3 + if_4 \sim \psi^{(j)}_{1,0}, \quad f_3 - if_4 \sim \psi^{(j)}_{0,-1}, \tag{4.17}$$
$$f_5 \sim \psi^{(j)}_{0,0}.$$

This restricts j to integer values, so quantities of type j on $TQ^{\varphi}_{\mathrm{fl}}$ linear in $\dot{q}$ exist only for $j = 1,2,\ldots$ [leaving aside the scalar $s(q,\dot{q})$ already constructed]. Moreover, for each such j we find three linearly independent $V_A(q,\dot{q})$: one starts with only f_1 and f_2 nonzero, another with only f_3 and f_4 nonzero, and the third with only f_5 nonzero. The last of these is immediately recognized to be simply related to the scalar $s(q,\dot{q})$ and to the quantities of type j on $Q^{\varphi}_{\mathrm{fl}}$ already formed in step A_0. Indeed, comparison of the last of Eqs. (4.17) with Eq. (4.8) shows that if we take $f_1 = \cdots = f_4 = 0$ in step A_1, we simply reproduce

$$t_{\mu\nu}(q)s(q,\dot{q}),\ t_{\mu\nu\rho\sigma}(q)s(q,\dot{q}),\ldots \tag{4.18}$$

as we would in any case expect. The other two possible choices for f_r, however, lead to new tensors linear in $\dot{q}$.

It is obvious that quantities of type j quadratic, cubic,... in $\dot{q}$ will also exist and nontrivial Lagrangians involving coupling between $\dot{x}$ and the internal variables can definitely be constructed. However, a systematic attack on this problem is hampered by the fact that a simple geometrical description of $Q^{\varphi}_{\mathrm{fl}}$ is not easily obtained. We would like to make the following observations concerning the nature of $Q^{\varphi}_{\mathrm{fl}}$. We have seen in step A_0 that a symmetric traceless second-rank tensor $t_{\mu\nu}(q)$ exists on $Q^{\varphi}_{\mathrm{fl}}$. However, at each point q, the stability group of $t_{\mu\nu}(q)$ is larger than, and contains, that of q. For example, at q_0, the former is generated by J_3 and K_3 and so is $\mathrm{SO}(2)\times\mathrm{SO}(1,1)$ with the topological structure $S^1 \times R$; while the latter is $F^{\varphi}_{\mathrm{fl}}$ generated by J_φ and with the structure of R. One might ask if any other linearly transforming quantity could be set up on $Q^{\varphi}_{\mathrm{fl}}$ such that its value together with that of $t_{\mu\nu}(q)$ uniquely determine the point q, because by the previous remark the value of $t_{\mu\nu}(q)$ certainly does not fix q. However, Eq. (4.9) shows that this is impossible: *any* linearly transforming quantity invariant under J_φ is also invariant under J_3 and K_3 separately. While $Q^{\varphi}_{\mathrm{fl}}$ is of dimension 5, $t_{\mu\nu}(q)$ takes values in the coset space $G/\mathrm{SO}(2)\times\mathrm{SO}(1,1)$ and so can supply us with only four independent coordinates. The problem is to understand the nature of the fifth coordinate. In other words, we need to know how q is related to q_0 (more generally q_1 to q_2) if $t_{\mu\nu}(q) = t_{\mu\nu}(q_0)$ [or $t_{\mu\nu}(q_1) = t_{\mu\nu}(q_2)$]. Comparing $F^{\varphi}_{\mathrm{fl}}$ generated by J_φ and $\mathrm{SO}(2)\times\mathrm{SO}(1,1)$ generated by J_3 and K_3, we deduce that the fifth coordinate is of the nature of an angle.

We leave the study of the SOS's Q_8 and $Q^{\varphi}_{\mathrm{fl}}$ at this stage, having satisfied ourselves that nontrivial Lagrangians can be constructed for particles with such internal structures.

V. GENERAL PHASE-SPACE ASPECTS

We have found many SOS's that permit nontrivial coupling of internal and space-time variables and construction of interesting Lagrangians. To set up a phase-space formalism for the internal spaces one may use a local approach varying from one SOS to another; or one may set up a formalism that is as uniform as possible for all SOS's. In the former approach, one has (local) coordinates q^r for Q accompanied by canonical momenta p_r, and the Legendre map involving the Lagrangian (2.1) is

$$p_r = \frac{\partial \mathscr{L}}{\partial \dot{q}^r}. \tag{5.1}$$

The internal contributions $S_{\mu\nu}$ to the SO(3,1) generators would be derived objects formed out of q^r and p_r. Alternatively, one can regard the $S_{\mu\nu}$ themselves as a kind of generalized canonical momenta conjugate to the q^r, with a suitable number of independent relations reducing the independent components of $S_{\mu\nu}$ from 6 to dim Q. The advantage of this approach is that the $S_{\mu\nu}$ are always globally defined quantities on T^*Q, while this may not be the case for the p_r. We shall develop the necessary formulas to achieve this, and at the end specialize to the manifestly covariant cases.

We begin with the action of infinitesimal SO(3,1) transformations on Q. An infinitesimal element of SO(3,1) can be written as

$$\Lambda^\mu{}_\nu \simeq \delta^\mu{}_\nu + \omega^\mu{}_\nu,$$
$$\omega_{\mu\nu} = -\omega_{\nu\mu}, \quad |\omega_{\mu\nu}| \ll |. \tag{5.2}$$

Let the corresponding point transformation $\varphi(\Lambda;q)$ be written to first order in the ω's as

$$q^r \to q'^r \simeq q^r + \tfrac{1}{2}\omega^{\mu\nu}\varphi^r{}_{\mu\nu}(q). \tag{5.3}$$

The $6\times$dim Q functions $\varphi^r{}_{\mu\nu}(q)$ determine, by integration, the transitive action of G on Q. Associated with them are the six vector fields

$$X_{\mu\nu} = \varphi^r{}_{\mu\nu}(q)\frac{\partial}{\partial q^r} \tag{5.4}$$

which obey the commutation relations of so(3,1) or $\mathfrak{g}$:

$$[X_{\mu\nu}, X_{\rho\sigma}] = g_{\mu\rho}X_{\nu\sigma} - g_{\nu\rho}X_{\mu\sigma} + g_{\mu\sigma}X_{\rho\nu} - g_{\nu\sigma}X_{\rho\mu}. \tag{5.5}$$

The G action on Q naturally lifts to a canonical action on T^*G with generators $S_{\mu\nu}$ which are functions on T^*Q linear in p_r,

$$S_{\mu\nu} = -\varphi^r{}_{\mu\nu}(q)p_r. \tag{5.6}$$

The usual fundamental PB's among the q's and p's,

$$\{q^r, p_s\} = \delta^r_s, \quad \{q^r, q^s\} = \{p_r, p_s\} = 0, \tag{5.7}$$

combined with Eq. (5.5) lead to the complete system of PB's

$$\{q^r,q^s\}=0, \tag{5.8a}$$

$$\{q^r,S_{\mu\nu}\}=-\varphi^r_{\mu\nu}(q), \tag{5.8b}$$

$$\{S_{\mu\nu},S_{\rho\sigma}\}=g_{\mu\rho}S_{\nu\sigma}-g_{\nu\rho}S_{\mu\sigma}+g_{\mu\sigma}S_{\rho\nu}-g_{\nu\sigma}S_{\rho\mu}. \tag{5.8c}$$

If we wish, however, we can regard these equations rather than (5.7) as the basic PB relations for T^*Q. In that case however the proper number of restrictions on the $S_{\mu\nu}$ must be included. These are easy to obtain. For a general point $q\in Q$, let $H_q\subset G$ be the corresponding stability group. The infinitesimal generators of H_q have parameters $\omega^{\mu\nu}$ obeying

$$\omega^{\mu\nu}\varphi^r_{\mu\nu}(q)=0. \tag{5.9}$$

These must be regarded as a system of linear conditions on $\omega^{\mu\nu}$ for given q. The general solution for $\omega^{\mu\nu}$ will involve 6-dim Q independent parameters ϵ^α (typically a subset of components of $\omega^{\mu\nu}$) and can be written

$$\omega^{\mu\nu}=\epsilon^\alpha\psi^{\mu\nu}_\alpha(q). \tag{5.10}$$

Putting these expressions into (5.9) and remembering that the ϵ^α are independent, we get the identities

$$\psi^{\mu\nu}_\alpha(q)\varphi^r_{\mu\nu}(q)=0 \tag{5.11}$$

characterizing the action of G on Q. These identities combine with the original definition of $S_{\mu\nu}$ in Eq. (5.6) to give the expected number of (linear) restrictions on $S_{\mu\nu}$:

$$\psi^{\mu\nu}_\alpha(q)S_{\mu\nu}=0,\quad \alpha=1,\ldots,6-\dim Q. \tag{5.12}$$

Once these restrictions are in hand, we can treat the $S_{\mu\nu}$ as the basic internal momentum variables. Then the PB's (5.8a) and (5.8c) have a uniform appearance for all cases, while only (5.8b) is related to the action of G on each particular Q.

If Q is one of those SOS's possessing a manifestly covariant description, as given in Sec. II, useful simplifications occur. The space Q can be described by an overcomplete system of variables ξ^R, say, transforming by some generally reducible linear representation $\mathscr{D}^R{}_S(g)$ of SO(3,1):

$$\xi^R\to\xi'^R=\mathscr{D}^R{}_S(g)\xi^S. \tag{5.13}$$

Typically, as shown in Sec. II, ξ may be a single four-vector, possibly normalized, a pair of mutually orthogonal four-vectors, etc. Correspondingly the ξ's obey a set of Lorentz invariant relations which may be written as

$$\chi(\xi)=0. \tag{5.14}$$

[See, for instance, the restrictions appearing in Eqs. (2.2), (2.4), (2.6), (2.8), (2.10), and (2.12).] If these relations were to be used to eliminate some of the ξ's in favor of the others, it would amount to making some local choice of independent q^r. However, our aim is to avoid such elimination of ξ's and instead carry them all along subject to the restrictions (5.14). Let a Lagrangian given as a function, $\mathscr{L}$, of $\dot{x}$, q, $\dot{q}$ in Eq. (2.1), be viewed as a function $\mathscr{L}'$ of $\dot{x},\xi,\dot{\xi}$:

$$\mathscr{L}(\dot{x},q,\dot{q})=\mathscr{L}'(\dot{x},\xi,\dot{\xi}). \tag{5.15}$$

It is understood that even if $\mathscr{L}$ is given the functional form of $\mathscr{L}'$ is ambiguous. We must develop a way by which the Legendre map for $S_{\mu\nu}$ can be easily written using $\mathscr{L}'$ in a manifestly covariant way. Let the infinitesimal generators of the representation $\mathscr{D}(g)$ of SO(3,1) be $\mathscr{s}_{\mu\nu}$. If q^r were a local coordinate system for Q, their changes under an infinitesimal SO(3,1) transformation must produce the expected linear changes in ξ^R:

$$\delta q^r=\tfrac{1}{2}\omega^{\mu\nu}\varphi^r_{\mu\nu}(q)$$

$$\Rightarrow\delta\xi^R=\tfrac{1}{2}\omega^{\mu\nu}(\mathscr{s}_{\mu\nu})^R{}_S\xi^S. \tag{5.16}$$

We will now exploit the fact that the expression of the ξ's in terms of the q's is, by definition, independent of the $\dot{q}$'s. Thus even if the transformation parameters $\omega^{\mu\nu}$ were to be made functions of the evolution parameter s, the implication (5.16) would be valid. If we make the variations (5.16) on the two sides of (5.15) and equate the coefficients of $\dot{\omega}^{\mu\nu}$ we get the identities

$$\varphi^r_{\mu\nu}(q)\frac{\partial\mathscr{L}}{\partial\dot{q}^r}=\xi^S(\mathscr{s}_{\mu\nu})^R{}_S\frac{\partial\mathscr{L}'}{\partial\dot{\xi}^R}. \tag{5.17}$$

Here we can treat the various components ξ^R as though they were algebraically independent, as this causes no error in the calculation of $S_{\mu\nu}$. Putting together Eqs. (5.1), (5.6), (5.17) we get a manifestly covariant form for the Legendre map relating to $S_{\mu\nu}$:

$$S_{\mu\nu}=-\xi^S(\mathscr{s}_{\mu\nu})^R{}_S\frac{\partial\mathscr{L}'}{\partial\dot{\xi}^R}. \tag{5.18}$$

One may convince oneself that the PB's among $S_{\mu\nu}$ and ξ^R cause the appropriate linear transformations on ξ:

$$\{S_{\mu\nu},\xi^R\}=(\mathscr{s}_{\mu\nu})^R{}_S\xi^S. \tag{5.19}$$

This manifestly covariant phase space formalism makes it unnecessary, for the SOS's listed in Sec. II, to eliminate redundant variables and to restrict oneself to independent q's. At the same time, the restrictions (5.12) on $S_{\mu\nu}$ in any manifestly covariant case can also be expressed using the ξ variables quite easily, so at all stages the manifest covariance can be preserved.

VI. MODEL BASED ON Q_{13}

As an illustration of the general methods developed in the preceding sections, we discuss in this section the construction of the Lagrangian for a point particle with the SOS Q_{13} as the internal space. The Lagrangian must obey the conditions coming from Poincaré and reparametrization invariances. In addition we shall require that it possess a certain natural internal symmetry suggested by the nature of Q_{13}, and that it also leads to two reasonable primary constraints. Thus we shall determine the form of the most general Lagrangian having all these four properties. It will then appear that this model is something like a scaled-down version of the original Hanson–Regge model, which had an internal SO(3,1) symmetry; the advantage will be that the restrictions on the Lagrangian arising from the requirement of the primary constraints can be explicitly solved.

The space $Q_{13}=G/F_{13}=Q(ss)$ can be visualized as consisting of pairs of spacelike orthonormal vectors (a^μ,b^μ), Eq. (2.10). The eight variables a^μ, b^μ are an example of the ξ^R of the previous section, and give an overcomplete coordi-

nate system for Q_{13}. The Lorentz invariant restrictions (5.14) are in this case

$$a^2 = b^2 = 1, \quad a.b = 0. \tag{6.1}$$

From these follow three kinematic restrictions on the internal velocities:

$$a.\dot{a} = b.\dot{b} = 0, \quad a.\dot{b} = -b.\dot{a}. \tag{6.2}$$

We can see that in the generic case the vectors a^μ, b^μ, $\dot{a}^\mu$, $\dot{b}^\mu$ are a system of four linearly independent vectors; indeed one can exhibit such configurations consistent with (6.2), for example,

$$\begin{aligned} &a^\mu = (0,0,0,1), \quad b^\mu = (0,0,1,0); \\ &\dot{a}^\mu = (\beta,0,\alpha,0), \quad \dot{b}^\mu = (\gamma,\delta,0,-\alpha). \end{aligned} \tag{6.3}$$

Configurations where this linear independence breaks down, while they certainly can occur, will be singular configurations and will be explicitly excluded in the sequel. We now define an internal SO(2) group acting on Q_{13}, and commuting with the physical SO(3,1) action, as follows:

$$\begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \tag{6.4}$$

This group is naturally suggested by the manifestly covariant description of Q_{13} and is analogous to the internal SO(3,1) group occurring in the original Hanson–Regge model. We shall now require that the Lagrangian be invariant under this internal SO(2) group. In the spirit of Eq. (2.1), we must then build up a complete set of ζ variables which are moreover individually SO(2) invariant: the Lagrangian can depend only on such ζ's. Our approach will be to first find a complete set of ζ's in the sense of Sec. II, then isolate those that are SO(2) invariant.

Let us first look at the ζ's which do not involve $\dot{x}$. Subject to later making them homogeneous of degree zero in the velocities, the restrictions (6.2) limit us to four possible Lorentz scalars:

$$a.\dot{b}, \ \dot{a}^2, \ \dot{a}.\dot{b}, \ \dot{b}^2. \tag{6.5}$$

We rearrange these into three combinations that are SO(2) invariant and one that changes under SO(2):

$$\begin{aligned} &a.\dot{b}, \ \dot{a}^2 + \dot{b}^2, \ \dot{a}^2\dot{b}^2 - (\dot{a}.\dot{b})^2 \to \text{SO(2) invariant;} \\ &\tan^{-1}(2\dot{a}.\dot{b}/(\dot{a}^2 - \dot{b}^2)) \to \text{SO(2) variant.} \end{aligned} \tag{6.6}$$

We now remark that it is algebraically simpler to divide through by powers of $a.\dot{b}$ rather than of $\dot{x}^2$ to obtain expressions homogeneous of degree zero in the velocities. Thus we obtain two SO(2) invariant ζ variables not involving $\dot{x}$ at all,

$$\begin{aligned} &\zeta_1 = (\dot{a}^2 + \dot{b}^2)/(a.\dot{b})^2, \\ &\zeta_2 = (\dot{a}^2\dot{b}^2 - (\dot{a}.\dot{b})^2)/(a.\dot{b})^4; \end{aligned} \tag{6.7}$$

and mention that the SO(2) transformation property of the last variable listed in (6.6) is

$$\tan^{-1}(2\dot{a}.\dot{b}/(\dot{a}^2 - \dot{b}^2)) \to \tan^{-1}(2\dot{a}.\dot{b}/(\dot{a}^2 - \dot{b}^2)) - 2\theta. \tag{6.8}$$

Let us now include $\dot{x}^\mu$ in the construction of ζ's. Since a^μ, b^μ, $\dot{a}^\mu$, and $\dot{b}^\mu$ are generically linearly independent, we have exactly four independent Lorentz scalars involving $\dot{x}^\mu$,

$$\dot{x}.a, \ \dot{x}.b, \ \dot{x}.\dot{a}, \ \dot{x}.\dot{b}. \tag{6.9}$$

In principle $\dot{x}^2$ can be expressed as a function of these, though we shall soon arrange for it to be one of the independent variables. The above four Lorentz scalars can be rearranged into three independent SO(2) invariant combinations and a fourth SO(2) variant one:

$$\begin{aligned} &(\dot{x}.a)^2 + (\dot{x}.b)^2, \ (\dot{x}.\dot{a})^2 + (\dot{x}.\dot{b})^2, \\ &\dot{x}.a\,\dot{x}.\dot{b} - \dot{x}.b\,\dot{x}.\dot{a} \to \text{SO(2) invariant,} \\ &\tan^{-1}(\dot{x}.b/\dot{x}.a) \to \text{SO(2) variant.} \end{aligned} \tag{6.10}$$

This last combination changes by $-\theta$ under the transformation (6.4). Dividing through by suitable powers of $a.\dot{b}$, we define three more SO(2) invariant ζ's as

$$\begin{aligned} &\zeta_4 = \{(\dot{x}.a)^2 + (\dot{x}.b)^2\}/(a.\dot{b})^2, \\ &\zeta_5 = \{(\dot{x}.\dot{a})^2 + (\dot{x}.\dot{b})^2\}/(a.\dot{b})^4, \\ &\zeta_6 = \{\dot{x}.a\,\dot{x}.\dot{b} - \dot{x}.b\,\dot{x}.\dot{a}\}/(a.\dot{b})^3. \end{aligned} \tag{6.11}$$

One more and the last such ζ is evidently the combination

$$\tan^{-1}(2\dot{a}.\dot{b}/(\dot{a}^2 - \dot{b}^2)) - 2\tan^{-1}(\dot{x}.b/\dot{x}.a). \tag{6.12}$$

However, this is an awkward variable and in its place we wish to use

$$\zeta_3 = \dot{x}^2/(a.\dot{b})^2. \tag{6.13}$$

That this is possible is shown by the following argument. Treating $\dot{x}^2$ as a dependent variable, we can always express it in the form

$$\begin{aligned} \dot{x}^2 = (a.\dot{b})^2 \Phi(&\zeta_1,\zeta_2,\zeta_4,\zeta_5,\zeta_6, \\ &\tan^{-1}(2\dot{a}.\dot{b}/(\dot{a}^2 - \dot{b}^2)) - 2\tan^{-1}(\dot{x}.b/\dot{x}.a)). \end{aligned} \tag{6.14}$$

Now since there are four independent components to $\dot{x}^\mu$, and $\dot{x}^\mu$ appears only in the last four arguments of Φ, this function must depend in an essential way on these four arguments, in particular on the last of them. Therefore it should be possible in principle to turn Eq. (6.14) inside out and express the combination (6.12) in terms of $\zeta_1,\ldots,\zeta_6$. This establishes that a complete set of SO(2) invariant ζ's is given by Eqs. (6.7), (6.11), (6.13) and contains six variables. At this stage the most general SO(2) invariant Lagrangian is

$$\mathscr{L}(\dot{x},a,b,\dot{a},\dot{b}) = a.\dot{b}\tilde{f}(\zeta_1,\ldots,\zeta_6) \tag{6.15}$$

and involves one real function f of six real arguments.

At this point we restrict the Lagrangian further by demanding that it leads to two primary constraints that we shall specify. They involve the momentum p_μ canonically conjugate to x^μ, given in terms of Lagrangian variables by

$$\begin{aligned} p_\mu = \frac{\partial \mathscr{L}}{\partial \dot{x}^\mu} = (a.\dot{b})^{-1}\{&2\dot{x}_\mu \tilde{f}_3 + 2(\dot{x}.a\,a_\mu + \dot{x}.b\,b_\mu)\tilde{f}_4 \\ &+ 2(a.\dot{b})^{-2}(\dot{x}.\dot{a}\,\dot{a}_\mu + \dot{x}.\dot{b}\,\dot{b}_\mu)\tilde{f}_5 \\ &+ (a.\dot{b})^{-1}(\dot{x}.a\,\dot{b}_\mu + \dot{x}.\dot{b}\,a_\mu \\ &- \dot{x}.b\dot{a}_\mu - \dot{x}.\dot{a}b_\mu)\tilde{f}_6\}. \end{aligned} \tag{6.16}$$

Here $\tilde{f}_j$ stands for $\partial\tilde{f}/\partial\zeta_j$, $j = 1,\ldots,6$. The desired primary constraints are

$$p.a \approx 0, \quad p.b \approx 0. \tag{6.17}$$

It is convenient to use the complex combination $z = a + ib$

with a simple SO(2) transformation property. The demand on $\mathcal{L}$ can then be expressed as

$$p.z = 2(a.\dot{b})^{-1}\dot{x}.z(\tilde{f}_3 + \tilde{f}_4 + \tfrac{1}{2}\tilde{f}_6) - 2i(a.\dot{b})^{-2}\dot{x}.\dot{z}(\tilde{f}_5 + \tfrac{1}{2}\tilde{f}_6) \approx 0. \tag{6.18}$$

Now by straightforward algebra one can show that $\dot{x}.\dot{z}$ and $\dot{x}.z$ are related by combinations of ζ's:

$$\dot{x}.\dot{z} = a.\dot{b}\,\dot{x}.z(i\zeta_6 + (\zeta_4\zeta_5 - \zeta_6^2)^{1/2})/\zeta_4. \tag{6.19}$$

Using this in (6.18) we get the condition[16]

$$\tilde{f}_3 + \tilde{f}_4 + \tfrac{1}{2}\tilde{f}_6 + (\zeta_6 - i(\zeta_4\zeta_5 - \zeta_6^2)^{1/2})(\tilde{f}_5 + \tfrac{1}{2}\tilde{f}_6)/\zeta_4 = 0, \tag{6.20}$$

that is, $\tilde{f}$ is subject to two first-order partial differential equations

$$\left(\frac{\partial}{\partial\zeta_3} + \frac{\partial}{\partial\zeta_4} + \frac{1}{2}\frac{\partial}{\partial\zeta_6}\right)\tilde{f} = 0, \qquad \left(\frac{\partial}{\partial\zeta_5} + \frac{1}{2}\frac{\partial}{\partial\zeta_6}\right)\tilde{f} = 0. \tag{6.21}$$

These conditions make $\tilde{f}$ a function of four particular combinations of the six ζ's. Retaining the symbol f for this function, the most general Lagrangian possessing all the properties demanded is

$$\mathcal{L}(\dot{x},a,b,\dot{a},\dot{b}) = a.\dot{b}\,f(\zeta_1,\zeta_2,\zeta_3 - \zeta_4,\zeta_4 + \zeta_5 - 2\zeta_6), \tag{6.22}$$

and involves one real function f of four real arguments.

For the phase-space treatment of this model we can use the manifestly covariant formalism set up in the previous section. In place of individual momenta conjugate to a^μ and b^μ, we have a tensor $S_{\mu\nu}$ jointly conjugate to both a^μ and b^μ. Taken together, the fundamental PB relations for the internal phase space, consistent with Eq. (6.1), are

$$\begin{aligned} &\{a_\mu,a_\nu\} = \{a_\mu,b_\nu\} = \{b_\mu,b_\nu\} = 0; \\ &\{S_{\mu\nu},a_\rho\} = g_{\mu\rho}a_\nu - g_{\nu\rho}a_\mu, \\ &\{S_{\mu\nu},b_\rho\} = g_{\mu\rho}b_\nu - g_{\nu\rho}b_\mu; \\ &\{S_{\mu\nu},S_{\rho\sigma}\} = g_{\mu\rho}S_{\nu\sigma} - g_{\nu\rho}S_{\mu\sigma} + g_{\mu\sigma}S_{\rho\nu} - g_{\nu\sigma}S_{\rho\mu}. \end{aligned} \tag{6.23}$$

In addition of course we have the usual PB's among x^μ and p_μ. The single expected condition on $S_{\mu\nu}$, Eq. (5.12) in the present model, is also consistent with the PB's (6.23) and is

$$\epsilon^{\mu\nu\rho\sigma}a_\rho b_\sigma S_{\mu\nu} = 0. \tag{6.24}$$

This arises from the fact that an infinitesimal Lorentz transformation that does not alter a given pair (a^μ, b^μ) has parameters $\omega_{\mu\nu}$ proportional to $\epsilon_{\mu\nu\rho\sigma}a^\rho b^\sigma$. This condition on $S_{\mu\nu}$ reduces the number of its independent components to 5, the dimension of Q_{13}; it is also consistent with the Legendre map [analog of Eq. (5.1)]:

$$S_{\mu\nu} = a_\mu\frac{\partial\mathcal{L}}{\partial\dot{a}^\nu} - a_\nu\frac{\partial\mathcal{L}}{\partial\dot{a}^\mu} + b_\mu\frac{\partial\mathcal{L}}{\partial\dot{b}^\nu} - b_\nu\frac{\partial\mathcal{L}}{\partial\dot{b}^\mu}. \tag{6.25}$$

Concerning the full phase-space treatment of this model (which will be taken up elsewhere) the following remarks may be made. An overcomplete set of phase-space coordinates is made up of x^μ, p^μ, a^μ, b^μ, and $S_{\mu\nu}$. The generator for the internal SO(2) group introduced in Eq. (6.4) is $a^\mu b^\nu$ since it has the following PB's with a and b:

$$\{a^\mu b^\nu S_{\mu\nu},a_\lambda\} = -b_\lambda, \qquad \{a^\mu b^\nu S_{\mu\nu},b_\lambda\} = a_\lambda. \tag{6.26}$$

Therefore it will turn out to be a constant of motion. It is of course both Lorentz and SO(2) invariant. More generally, one expects that there will be five algebraically independent Lorentz and SO(2) invariant variables on phase space, made up out of p,a,b, and S. Among these would be included $-p^2$, the square of the invariant mass, and the square of the Pauli–Lubanski vector. This counting can be seen to be correct by going to the special Lorentz frame in which $p^\mu = \sqrt{-p^2}e_0^\mu$, $a = e_1$, and $b = e_2$, after which there is no more freedom of Lorentz transformation. In this frame, Eq. (6.24) makes S_{03} vanish, so each of the other five components of $S_{\mu\nu}$ is a possible candidate for a Lorentz scalar combination of variables. Adding p^2 to this set, we get a total of six independent Lorentz scalars that can be built out of p, a, b, and S. [Note that this is after the primary constraints (6.17) have been taken into account.] Since SO(2) is a one-parameter group, it then follows that there will be five quantities in the model that are both Lorentz and SO(2) invariant. One then expects that the reparametrization invariance will lead to a Lorentz and SO(2) invariant primary constraint, independent of the already specified primary constraints (6.17), which will be the generator of reparametrizations. This will be a relationship among the five variables mentioned above, which should arise by expressing each of them in terms of the four Lagrangian quantities $\zeta_1, \zeta_2, \zeta_3 - \zeta_4, \zeta_4 + \zeta_5 - 2\zeta_6$, and then eliminating the latter.

VII. CONCLUDING REMARKS

In this paper we have continued the study of classical relativistic particles with internal structure initiated in I, and developed methods for a thorough analysis of second-order internal spaces. The large number of possible SOS's has been separated into those permitting a manifestly covariant description, and others; and in order to deal with the latter, special methods based on the isotopy representation have been devised. The Lagrangian formalism has been used throughout, and in the manifestly covariant cases it is obvious that nontrivial coupling between space-time and internal variables (and velocities) is always possible. A particularly interesting result has been our finding that the internal spaces Q_2, Q_5^φ, and Q_7 are unphysical, though perfectly admissible from the point of view of four-dimensional kinematics, in that it is impossible to couple the space-time variables and the internal spaces in these cases. This is a result that one might not have anticipated in advance. While the spaces Q_5^φ are not easily visualized, both Q_2 and Q_7 do admit of geometrical descriptions. Thus Q_2 is just the unit sphere S^2 in three dimensions, and it can be easily obtained from $Q_6 = Q(l)$ by the same kind of quotienting procedure that was used in Sec. IV to develop a picture of Q_8. [Starting with the description of $Q_6 = Q(l)$ given in Eq. (2.6), one defines the equivalence relation $a^\mu \sim \kappa a^\mu$ for all $\kappa > 0$, then passes to the quotient $Q_6/\sim$ which is Q_2.] Another way to realize that $Q_2 = S^2$ is via

the Iwasawa decomposition $G = \mathrm{SU}(2)AN$, A being generated by K_3 and N by N_1 and N_2. Since F_2 has J_3, K_3, and N_a as generators, we see that $Q_2 = G/F_2 = \mathrm{SU}(2)AN/\mathrm{U}(1)AN \simeq \mathrm{SU}(2)/\mathrm{U}(1) = S^2$. Again by the Iwasawa decomposition, the space $Q_7 = G/F_7$ is seen to be the SU(2) group manifold, since $F_7 = AN$. In spite of the fact that both Q_2 and Q_7 are visualizable in these ways, though of course not manifestly covariantly, they turn out to be incapable of coupling to the space-time variable $\dot{x}^\mu$.

The impossibility of coupling in the case of Q_7 is understandable and reassuring form the point of view of the Hanson–Regge model.[7] In that model, the starting internal space is SO(3,1); and by imposing several primary constraints and invariant relations one arranges things so that at the end only SO(3) survives as the physical internal space. One can naturally ask whether the procedure of Hanson and Regge could be significantly simplified by using from the very beginning the manifold of SO(3) as the internal space. Since, again by the Iwasawa theorem, SO(3) does admit a transitive action by SO(3,1), at first sight one would suppose that this should definitely be possible, and only an element of SO(3) and its "velocity" would occur as internal variables in the starting Lagrangian. However, our analysis showing that no coupling between $\dot{x}^\mu$ and TQ_7 is possible demonstrates that there is no way to bypass the route taken in the Hanson–Regge model to achieve the desired result that the final surviving internal space be SO(3).

A possible coupling scheme which, if it had occurred, might have been termed trivial, would be this[17]: one is able to construct Lorentz scalars $s(q,\dot{q})$ homogeneous of some degree in the $\dot{q}$, leading to corresponding ζ's by division by suitable powers of $(-\dot{x}^2)$, but there are no quantities of type $j \geqslant \frac{1}{2}$ on Q or TQ capable of coupling to $\dot{x}^\mu$, $\dot{x}^\mu\dot{x}^\nu$,... . In such a case, the only possible dependence of the Lagrangian on $\dot{x}^\mu$ would have been through $\dot{x}^2$, and one would have had separate invariances under the action of SO(3,1) on x^μ and on Q. Of course with a manifestly covariant Q this situation does not arise, and only Q_8 and Q^{φ}_{11} remain to be examined from this point of view. However, the results of Sec. IV show that with Q_8 a vector a^μ is available for direct coupling to $\dot{x}^\mu$, and with Q^{φ}_{11}, even though the scalar $s(q,\dot{q})$ has been found [Eq. (4.16)], there does exist a tensor $t_{\mu\nu}(q)$ which could be contracted with $\dot{x}^\mu\dot{x}^\nu$. Thus the kind of trivial coupling visualized in the opening sentence of this paragraph does not actually arise.

TABLE I. Pattern of SOS's.

	Coupling permitted	No coupling
Manifestly covariant Q's	Q_3, Q_4, Q_6, Q_{12} Q_{13}, Q_{14}, Q_{15}	...
Not manifestly covariant, but visualizable, Q's	Q_8	Q_2, Q_7
Not easily visualizable Q's	Q^{φ}_{11}	Q^{φ}_5

Our analysis of the SOS's Q_8, Q^{φ}_{11} has been carried only far enough to show that they are physically admissible. A more complete treatment is definitely worthwhile and remains to be given. This involves among other things developing a "picture" for Q^{φ}_{11}, allowing us to handle it to some extent geometrically. The model based on Q_{13} that we have offered in Sec. VI as a "scaled-down" version of the Hanson–Regge model has been chosen so that the most general form of the Lagrangian could be explicitly displayed. This was a motivation for the choice of primary constraints (6.17). One could also consider as an alternative the primary constraints chosen by Hanson and Regge, namely

$$p^\mu S_{\mu\nu} \approx 0,$$

which are quadratic in the conjugate momenta. To ensure them, one has to subject the function $\tilde{f}$ to nonlinear partial differential equations, as in the Hanson–Regge model, so that an explicit form for the Lagrangian cannot be obtained. Even within our model, a complete analysis of the constraints and phase-space structure remains to be given.

Finally, since the number of SOS's is so large in variety and complexity as compared to the FOS's studied in I, we present Table I containing the principal results of this paper, so that one can get a feeling for the situation concerning SOS's in a compact way. The various SOS's are separated on the one hand into those possessing a manifestly covariant description, those that are otherwise visualizable, and those that cannot be easily pictured at all; and on the other hand into those permitting coupling to $\dot{x}$, and those which are unphysical.

We plan to come back to some of the questions raised above, as well as to the general problems of quantization and of couplings to external fields, elsewhere.

APPENDIX: A PROPERTY OF SL(2,C) REPRESENTATIONS

In the conventions adopted in this paper, the SL(2,C) generators J_j, K_j obey the commutation rules

$$\begin{aligned} &[J_j,J_k] = -[K_j,K_k] = \epsilon_{jkl}J_l, \\ &[J_j,K_k] = \epsilon_{jkl}K_l. \end{aligned} \tag{A1}$$

These differ from the conventional relations by a factor i: stated precisely, the present generators are $-i$ times the usual ones. If we define the two combinations

$$S_j = \tfrac{1}{2}(J_j + iK_j), \quad S'_j = \tfrac{1}{2}(J_j - iK_j), \tag{A2}$$

then as is well known S_j and S'_j independently obey the angular momentum commutation relations:

$$\begin{aligned} &[S_j,S_k] = \epsilon_{jkl}S_l, \\ &[S'_j,S'_k] = \epsilon_{jkl}S'_l, \\ &[S_j,S'_k] = 0. \end{aligned} \tag{A3}$$

The expressions for N_1, N_2, J_3, and K_3 are important:

$$\begin{aligned} &N_1 = J_1 - K_2 = S_+ + S'_-, \\ &N_2 = J_2 + K_1 = -i(S_+ - S'_-), \\ &J_3 = S_3 + S'_3, \\ &K_3 = -i(S_3 - S'_3). \end{aligned} \tag{A4}$$

Here we have used the familiar raising and lowering combinations

$$S_{\pm} = S_1 \pm iS_2, \quad S'_{\pm} = S'_1 \pm iS'_2. \tag{A5}$$

The finite-dimensional representation $\mathscr{D}^{(j_1,j_2)}$ of SL(2,C) is obtained by choosing the spin j_1 representation for S_j and the spin j_2 representation for S'_j. When $j_1 = j_2$, we have denoted $\mathscr{D}^{(j,j)}$ by $\mathscr{D}^{(j)}$. The canonical basis for $\mathscr{D}^{(j_1,j_2)}$ is made up of simultaneous eigenvectors of S_3 and S'_3:

$$\begin{aligned} S_3\psi^{(j_1,j_2)}_{m_1,m_2} &= -im_1\psi^{(j_1,j_2)}_{m_1,m_2}, \\ S'_3\psi^{(j_1,j_2)}_{m_1,m_2} &= -im_2\psi^{(j_1,j_2)}_{m_1,m_2}, \\ -j_1 \leqslant m_1 \leqslant j_1, &\quad -j_2 \leqslant m_2 \leqslant j_2. \end{aligned} \tag{A6}$$

In this basis both J_3 and K_3 are diagonal:

$$\begin{aligned} J_3\psi^{(j_1,j_2)}_{m_1,m_2} &= -i(m_1+m_2)\psi^{(j_1,j_2)}_{m_1,m_2}, \\ K_3\psi^{(j_1,j_2)}_{m_1,m_2} &= (m_2-m_1)\psi^{(j_1,j_2)}_{m_1,m_2}. \end{aligned} \tag{A7}$$

The only vector annihilated by N_1 and N_2 is the one with maximum m_1 and minimum m_2, since it must be annihilated by the raising operator S_+ and the lowering operator S'_-:

$$\begin{aligned} N_a\psi = 0 &\Rightarrow S_+\psi = S'_-\psi = 0 \\ &\Rightarrow \psi = c\psi^{(j_1,j_2)}_{j_1,-j_2}. \end{aligned} \tag{A8}$$

This vector is an eigenvector of K_3 with eigenvalue $-(j_1+j_2)$ which cannot be positive. This proves the lemma of Sec. III. In the text, where only the case $j_1 = j_2 = j$ is used, the basis vectors $\psi^{(j_1,j_2)}_{m_1,m_2}$ are written more simply as $\psi^{(j)}_{m_1,m_2}$.

[1] M. V. Atre and N. Mukunda, J. Math. Phys. **27**, 2908 (1986). A general discussion of physical features of such systems and references to many important and early papers on this subject are given there.

[2] The term "internal" only means that these coordinates are unaffected by space-time translations, though they are certainly acted upon by homogeneous Lorentz transformations. They must therefore not be confused with internal degrees of freedom such as isospin and hypercharge in particle physics.

[3] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597 (1975).

[4] See the Appendix in Ref. 1 above.

[5] An account of this is given in, for instance, B. Y. Chu, Trans. Am. Math. Soc. **197**, 45 (1974).

[6] The case G/F_{10} has been studied earlier in N. Mukunda, H. van Dam, and L. C. Biedenharn, Phys. Rev. D **22**, 1938 (1980); D **23**, 1451 (1981); *Relativistic Models of Extended Hadrons Obeying a Mass-Spin Trajectory Constraint, Lecture Notes in Physics,* Vol. 165 (Springer, Berlin, 1982).

[7] A. J. Hanson and T. Regge, Ann. Phys. (NY) **87**, 498 (1974).

[8] We use the metric $g_{00} = -1$.

[9] See in this connection the discussion given by D. Finkelstein, Phys. Rev. **100**, 924 (1955).

[10] This latter property causes no loss of generality.

[11] F. Halbwachs, Prog. Theor. Phys. **24**, 291 (1960).

[12] G. Cognola, R. Soldati, L. Vanzo, and S. Zerbini, Nuovo Cimento B **76**, 109 (1983).

[13] The material to follow is familiar from work related to nonlinear realizations of groups in the context of the σ models.

[14] This analysis is quite similar in structure to the way in which Wigner rotations are derived in the representation theory of the Poincaré group.

[15] However, it should be emphasized that the quantities $V_A(q,\dot{q})$ written in terms of the indicated arguments are independent of the choice of the element $l(q)$.

[16] The combination $\zeta_4\zeta_5 - \zeta_6^2$ is positive definite, hence the radical is real.

[17] M. V. Atre and N. Mukunda, Curr. Sci. **54**, 610 (1985).

Hamiltonization for singular and nonsingular mechanics

Maria L. Espindola, Nelson L. Teixeira, and Oslim Espindola
Departamento de Física-CCEN, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil

(Received 25 November 1985; accepted for publication 12 November 1986)

A Hamiltonization procedure valid for both singular and nonsingular mechanics is proposed. A comparison with Dirac's theory (for singular systems) is developed.

I. INTRODUCTION

A centenary tradition prescribes the use of Legendre transformations in the passage from the Lagrangian to the Hamiltonian description of classical mechanics.[1] We will prove that this prescription must not be some *a priori* postulate. It will appear as a consequence in nonsingular problems, and it will give inadequate results in singular ones.[2]

In Sec. II we define the problem in its generality. Sections III and IV will be devoted to solving two extreme cases of mechanics: the nonsingular and the pure singular cases, respectively. The intermediate (mixed) case is discussed in Sec. V. In Sec. VI our results are compared with those obtained by Dirac's theory.

II. ANALYSIS OF THE HAMILTONIZATION PROCEDURE

Consider a Lagrangian function

$$L(q_1,...,q_N,\dot{q}_1,...,\dot{q}_N,t)\equiv L(q,\dot{q},t),$$

with the usual equations of motion [given by Hamilton's principle $\delta\int L(q,\dot{q},t)dt=0$]

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}_i}\right]-\frac{\partial L}{\partial q_i}=0 \quad (i=1,...,N). \tag{1}$$

Irrespective of the fact that Eqs. (1) are a first- or second-order differential system, it is possible to define a new function $H(p,q,t)\equiv H(p_1,...,p_N,q_1,...,q_N,t)$ and new equations of motion

$$H(p,q,t)\equiv\sum p_i\dot{q}_i - L(q,\dot{q},t), \tag{2}$$

$$\dot{q}_i\equiv\frac{\partial H}{\partial p_i}, \tag{3a}$$

$$\dot{p}_i\equiv -\frac{\partial H}{\partial q_i}. \tag{3b}$$

Equations (2) and (3a) define a partial differential equation for H:

$$H=\sum p_i\frac{\partial H}{\partial p_i}-L\left(q,\frac{\partial H}{\partial p},t\right), \tag{4}$$

which has a complete solution given by

$$H'=\sum A_i p_i - L(q,A,t), \tag{5}$$

the A_i being arbitrary functions of the q_i variables.

Under certain circumstances Eq. (4) has an additional solution: the envelope (singular) solution. This solution exists whenever some functions A_i $(i=1,...,k)$ are defined as functions of p and q through

$$\frac{\partial H'}{\partial A_i}=0 \quad (i=1,...,N) \tag{6}$$

(which define the envelope solution).

The case where $k=N$ is called the pure nonsingular mechanics and will be studied in Sec. III. The case where Eq. (6) does not define any function A_i is discussed in Sec. IV, and will be called the pure singular case. The mixed problem will be studied in Sec. V.

III. THE PURE NONSINGULAR MECHANICS

Consider the Lagrangian function with N degrees of freedom,

$$L(q,\dot{q},t)=\sum a_{ij}(q,t)\dot{q}_i\dot{q}_j - V(q,t) \quad (i,j=1,...,N), \tag{7}$$

with the Hessian matrix $(\partial^2 L/\partial\dot{q}_i\,\partial\dot{q}_j)=(a_{ij})\neq 0$.

The partial differential equation for H obtained from Eqs. (2) and (3a) now reads

$$H=\sum p_i\frac{\partial H}{\partial p_i}-\sum a_{ij}\frac{\partial H}{\partial p_i}\frac{\partial H}{\partial p_j}+V. \tag{8}$$

The corresponding complete solution is

$$H'=\sum p_i A_i-\sum a_{ij}A_iA_j+V, \tag{9}$$

where the A_i are arbitrary functions of the q_i variables. The envelope solution is obtained from

$$\frac{\partial H'}{\partial A_i}=0=p_i-2\sum a_{ij}A_j\rightarrow p_i=2\sum a_{ij}A_j. \tag{10}$$

This algebraic system is easily solved (since $\det a_{ij}\neq 0$):

$$A_l=\frac{1}{2}\sum b_{lm}p_m$$

(where b is the inverse of the matrix a; i.e., $\Sigma a_{ij}b_{jm}=\delta_{ij}$).

In consequence it is easy to construct the envelope solution

$$H_E=\frac{1}{4}\sum b_{lm}p_lp_m+V,$$

which is the usual Hamiltonian. As it should be, since condition (10) and Eq. (3a) imply $p_i=2\Sigma a_{ij}\dot{q}_j=\partial L/\partial\dot{q}_i$. This (usual) definition of the momenta cannot be maintained in the subsequent cases.

As a consequence $\partial H/\partial q_i=-\partial L/\partial q_i$ since

$$\frac{\partial H}{\partial q_i}=+\sum p_l\frac{\partial A_l}{\partial q_i}-\frac{\partial L}{\partial q_i}-\sum\frac{\partial L}{\partial A_l}\frac{\partial A_l}{\partial q_i}$$

$$=-\frac{\partial L}{\partial q_i}-\sum\frac{\partial A_l}{\partial q_i}\left(\frac{\partial L}{\partial A_l}-p_l\right).$$

And the compatibility among the set of Eqs. (3) and the Lagrangian description is automatically achieved here, since

 0022-2488/87/040807-03$02.50

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} = \frac{\partial L}{\partial q_i}.$$

IV. THE PURE SINGULAR CASE

Consider the most general Lagrangian for the pure singular case with N degrees of freedom:

$$L(q,\dot{q},t) = \sum a_i(q,t)\dot{q}_i - V(q,t). \tag{11}$$

The corresponding Hamiltonian must obey the linear partial differential equation obtained from Eqs. (2) and (3a),

$$H = \sum (p_i - a_i)\frac{\partial H}{\partial p_i} + V, \tag{12}$$

whose general solution is

$$H(p,q,t) = \sum G_i(q,t)(p_i - a_i) + V(q,t), \tag{13}$$

with the G_i being arbitrary functions.

The Hamiltonian equations of motion [Eqs. (3)] read

$$\dot{q}_i = \frac{\partial H}{\partial p_i} = G_i, \tag{14a}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} = -\sum (p_j - a_j)\frac{\partial G_j}{\partial q_i} - \frac{\partial V}{\partial q_i} + \sum G_j \frac{\partial a_j}{\partial q_i}. \tag{14b}$$

Due to the arbitrary character of the functions G_i, compatibility of the above equations with Lagrangian ones is not automatically achieved. It must be imposed.

From Lagrangian equations we have

$$\dot{a}_i = -\frac{\partial V}{\partial q_i} + \sum \frac{\partial a_j}{\partial q_i}\dot{q}_j \quad (i,j = 1,...,N). \tag{15}$$

These equations together with Eq. (14a) gives the functions G_i as solutions of the algebraic system:

$$\frac{\partial V}{\partial q_i} + \frac{\partial a_i}{\partial t} = \sum \left[\frac{\partial a_j}{\partial q_i} - \frac{\partial a_i}{\partial q_j}\right] G_j \quad (i,j = 1,...,N). \tag{16}$$

The second set of Hamiltonian equations [Eq. (14b)] furnishes [using also Eq. (15)]

$$\dot{p}_j - \dot{a}_j = -\sum \frac{\partial G_i}{\partial q_j}(p_i - a_i) \quad (i,j = 1,...,N). \tag{17}$$

These equations are, in fact, definitions of the momenta p_j. (The solutions $p_j = a_j$, i.e., $p_j = \partial L/\partial \dot{q}_j$, imply that $H = V$ and, in consequence $\dot{q}_j = 0$, which are incompatible with Lagrangian equations.)

In fact, Eq. (17) is a linear first-order partial differential equation for the variables $C_i(q) \equiv p_i - a_i$. Its general solution is given in terms of an arbitrary function. Each particularization of this arbitrary function furnishes a definition for the variable p_i (different from $\partial L/\partial \dot{q}_i$).

As an example consider the Lagrangian

$$L = q_2\dot{q}_1 - 2q_1\dot{q}_2 + q_1q_2. \tag{18}$$

Following the steps previously described [Eqs. (13)–(16)] we construct the Hamiltonian

$$H = G_1(p_1 - q_2) + G_2(p_2 + 2q_1) - q_1q_2,$$

where

$$G_1 = -q_1/3 \quad \text{and} \quad G_2 = q_2/3. \tag{19}$$

Consequently,

$$H = (q_2p_2 - q_1p_1)/3. \tag{20}$$

The functions p_i are defined as solutions of the following system [Eqs. (17) and (19)]:

$$\dot{p}_1 - \dot{q}_2 = (p_1 - q_2)/3,$$

$$\dot{p}_2 + 2\dot{q}_1 = -(p_2 + 2q_1)/3.$$

These solutions are given by

$$p_1 = q_2 f_1(q_1q_2), \quad p_2 = q_1 f_2(q_1q_2),$$

the f_i being arbitrary functions of the product q_1q_2 (as remarked before, we must have $f_1 \neq 1$ and $f_2 \neq -2$).

V. THE INTERMEDIATE (MIXED) CASE

Consider the solution given by Eq. (5) for the partial differential equation for the Hamiltonian H,

$$H' = \sum A_ip_i - L(q,A) \quad (i = 1,...,N). \tag{5}$$

In the mixed case some relations $A_i = A_i(q,p)$ $(i = 1,...,k)$ will be determined by the envelope condition:

$$\frac{\partial H'}{\partial A_i} = 0 \quad (i = 1,...,k). \tag{6}$$

These conditions imply that $p_i = \partial L/\partial \dot{q}_i$ $(i = 1,...,k)$ (as discussed in Sec. III), and can be substituted in Eq. (5) resulting in a Hamiltonian that depends only on q, p_i, and A_j, where $i = 1,...,k$ and $j = k,...,N$.

The following steps of the Hamiltonization procedure are the same as in the pure singular case.

As an example consider the Lagrangian function

$$L = \tfrac{1}{2}a(q)\dot{q}_1^2 + b(q)\dot{q}_2 - V(q). \tag{21}$$

The corresponding complete solution of the partial differential equation for the Hamiltonian is [Eq. (5)]

$$H' = A_1p_1 + A_2p_2 - \tfrac{1}{2}aA_1^2 - bA_2 + V.$$

Equations (3a) furnish

$$\dot{q}_i = \frac{\partial H'}{\partial p_i} = A_i \quad (i = 1,2).$$

The envelope condition [Eq. (6)] gives

$$p_1 = aA_1 = a\dot{q}_1 = \frac{\partial L}{\partial \dot{q}_1}.$$

Then the resulting Hamiltonian is

$$H' = (1/2a)p_1^2 + (p_2 - b)A_2 + V. \tag{22}$$

The corresponding Euler–Lagrange equations for the above Lagrangian [Eq. (21)] are

$$\frac{d}{dt}(a\dot{q}_1) = \frac{1}{2}\frac{\partial a}{\partial q_1}\dot{q}_1^2 + \frac{\partial b}{\partial q_1}\dot{q}_2 - \frac{\partial V}{\partial q_1}$$

and

$$\frac{db}{dt} = \frac{\partial b}{\partial q_1}\dot{q}_1 + \frac{\partial b}{\partial q_2}\dot{q}_2 = \frac{1}{2}\frac{\partial a}{\partial q_2}\dot{q}_1^2 + \frac{\partial b}{\partial q_2}\dot{q}_2 - \frac{\partial V}{\partial q_2}. \tag{23}$$

As $\dot{q}_2 = \partial H/\partial q_2 = A_2$, then

$$-\frac{1}{2}\frac{\partial a}{\partial q_1}\dot{q}_1^2+\frac{\partial b}{\partial q_1}A_2-\frac{\partial V}{\partial q_1}=\frac{\partial a}{\partial q_2}\dot{q}_1A_2+a\ddot{q}_1$$

and

$$\frac{\partial b}{\partial q_1}\dot{q}_1=\frac{1}{2}\frac{\partial a}{\partial q_2}\dot{q}_1^2-\frac{\partial V}{\partial q_2}.$$

This system has the solution

$$A_2=\left(\frac{1}{2}\frac{\partial a}{\partial q_1}\dot{q}_1^2+a\ddot{q}_1+\frac{\partial V}{\partial q_1}\right)\left(\frac{\partial b}{\partial q_1}-\frac{\partial a}{\partial q_2}\dot{q}_1\right)^{-1},$$

with

$$\dot{q}_1=\left(\frac{\partial b}{\partial q_1}\pm\left[\left(\frac{\partial b}{\partial q_1}\right)^2+2\frac{\partial a}{\partial q_2}\frac{\partial V}{\partial q_2}\right]^{1/2}\right)\left(\frac{\partial a}{\partial q_2}\right)^{-1}.$$

To finalize the momentum p_2 is defined as the solution of the equation

$$\dot{p}_2=-\frac{\partial H}{\partial q_2}=\frac{1}{2a^2}\frac{\partial a}{\partial q_2}p_1^2-(p_2-b)\frac{\partial A_2}{\partial q_2}-\frac{\partial V}{\partial q_2}+\frac{\partial b}{\partial q_2}A_2,$$

which may be written as [using Eq. (23) and $p_1=a\dot{q}_1$]

$$\frac{d}{dt}(p_2-b)=\frac{\partial A_2}{\partial q_2}(p_2-b).$$

Defining $C(q)\equiv p_2-b$, we have the following system:

$$\frac{dq_1}{\dot{q}_1}=\frac{dq_2}{A_2}=\frac{dC}{\partial A_2/\partial q_2}. \tag{24}$$

To clarify the procedure let us consider the following Lagrangian:

$$L=\tfrac{1}{2}q_2\dot{q}_1^2+q_1\dot{q}_2-q_1q_2,$$

i.e., $a=q_2$, $b=q_1$, and $V=q_1q_2$.

Then the Hamiltonian is [Eq. (22)]

$$H'=(1/2q_2)p_1^2+(p_2-q_1)A_2+q_1q_2,$$

where

$$A_2=-(1\pm2\sqrt{1+2q_1})q_2/(1+2q_1).$$

Then Eq. (24) furnishes

$$C(q)=q_2F([1-2q_1\pm\sqrt{1+2q_1}]q_2)$$

and

$$p_2=q_1+C(q),$$

where F is an arbitrary function and the momentum p_2 is defined by the choice of F.

VI. FINAL REMARKS

Our approach unifies, into the same formalism, singular and nonsingular problems. It is worthwhile to compare our approach to the well-known Dirac theory for singular mechanics.[2]

Starting from the same Lagrangian [Eq. (11)] Dirac defines the momentum p as a "weak equality"

$$p_i\approx\frac{\partial L}{\partial\dot{q}_i}. \tag{25}$$

His "canonical Hamiltonian" is

$$H_c\approx V.$$

And the "total Hamiltonian" is given by

$$H\approx V+v_m\phi_m, \tag{26}$$

where the ϕ_m are the constraints of Dirac's theory,

$$\phi_m=p_m-\frac{\partial L}{\partial\dot{q}_m}=p_m-a_m\approx0. \tag{27}$$

These constraints must obey the consistency condition of Dirac's theory,

$$\dot{\phi}_m=[\phi_m,H]\approx0,\quad\text{for }\phi_m\approx0, \tag{28}$$

which results in a so-called[2] "relation of type 1." These relations are given by

$$v_m\left(\frac{\partial a_m}{\partial q_{m'}}-\frac{\partial a_{m'}}{\partial q_m}\right)\approx\frac{\partial V}{\partial q_{m'}}+\frac{\partial a_{m'}}{\partial t}. \tag{29}$$

If we compare this result with that obtained in the previous section we observe that this equation is identical to Eq. (16), where v_m plays the role of G_i.

To clarify, consider as an example the Lagrangian used in Sec. IV [Eq. (18)].

Equations (29) imply that

$$v_1\approx-q_1/3\quad\text{and}\quad v_2\approx q_2/3,$$

and, finally Eqs. (26) and (27) give the Hamiltonian

$$H\approx(q_2p_2-q_1p_1)/3,$$

which is identical to the one given in Eq. (20).

If we now consider the Lagrangian given by Eq. (21) (Sec. V), then the Hamiltonian given by Dirac's theory is

$$H\approx H_c+v_m\phi_m.$$

As $p_1=a\dot{q}_1$ and $p_2\approx b$, then

$$H_c\approx p_1^2/2a+V$$

and the constraint of the Dirac's theory is

$$\phi=p_2-b\approx0.$$

The "total Hamiltonian" is

$$H\approx p_1^2/2a+V+v(p_2-b),$$

where $v=\dot{q}_2$.

This Hamiltonian is identical with that given by Eq. (22) (Sec. V) if we consider that there $A_2=\dot{q}_2$.

Finally we observe that the Hamiltonization procedure for pure singular and mixed cases developed in this paper and in Dirac's theory have identical results. But our approach is a more general Hamiltonization procedure, and does not need "weak equalities," a variational procedure involving p's, q's, and $\dot{q}$'s, a "super phase space," etc.

ACKNOWLEDGMENTS

We would like to thank Professor L. J. Negri for fruitful discussions.

[1] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1971), Chap. 7.

[2] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964); E. C. G. Sudarshan and N. Mukunda, *Quantum Mechanics* (Wiley, New York, 1974), Chap. 4; S. Shanmugadhasan, J. Math. Phys. **14**, 677 (1973); K. Kamimura, Nuovo Cimento B **68**, 33 (1982).

On the strength of Maxwell's equations

N. F. J. Matthews
North Carolina State University, Raleigh, North Carolina 27695-7911

(Received 6 September 1985; accepted for publication 26 November 1986)

The "strength" of a set of field equations (first defined by Einstein as the number of Taylor coefficients of field variables that could be chosen arbitrarily) is used to show that the amount of initial data required by the electromagnetic formulation of Maxwell's theory in free space is equal, without approximation, to that required by the potential formulation. In each formulation, the strength is interpreted in terms of the amount of initial data required to provide a solution of the Cauchy initial-value problem and in terms of the invariance properties of the formulation. Equality of the strengths of the two formulations of Maxwell's theory is used to support the assertion that knowledge of the strengths of other established field theories provides a means for predicting the possible existence of unknown formulations of the theories.

I. INTRODUCTION

In his investigations of field theories, Einstein[1] used the concept of "strength" of a system of field equations for comparing the worth of one system with that of another. He assumed that each field variable could be written as a Taylor series in the neighborhood of a point P. By repeated differentiation and evaluation of the field equations at P, he could obtain relations among the Taylor coefficients. In effect, the field equations would specify some coefficients in terms of others so that, in principle, many coefficients could be eliminated from the series. The remaining Taylor coefficients would be free data; they could be chosen freely provided they remained consistent with the field equations. Einstein called the amount of this free data the strength of the system of field equations.

Several authors have examined strengths of field theories. Penney,[2] Hoenselaers,[3] Burman,[4] and Murphy[5] presented strengths of various field theories including those of Einstein, Maxwell, Weyl, Klein–Gordon, and Dirac. Mariwalla[6] showed that approximations of the strengths of Maxwell's potential and electromagnetic formulations were equal. Schutz[7] suggested that the strength of a field theory was related in a well-defined manner to the number and character of arbitrary functions required to specify a solution locally.

As a supplement to the works of Mariwalla[6] and of Schutz,[7] this paper presents verification that the strength of each formulation of Maxwell's theory can be interpreted as initial data required to yield a solution to the field equations of that formulation in free space. Also, the amount of initial data required for a solution of the electromagnetic formulation is shown to be equal, without approximation, to that required for a solution of the potential formulation.

Exact equivalence of the strengths of the two formulations leads to a useful concept. An investigator familiar with the second-order potential formulation, but ignorant of the first-order electromagnetic one, could reasonably decide to examine the possibility of expressing the content of Maxwell's theory by a set of first-order field equations simply because the character of initial data predicted by the strength of the potential formulation is precisely that of a set of first-order differential equations. And, of course, the investigation would meet with success. This suggests that *the concept of strength of a field theory may be useful in finding unknown formulations of established field theories.*

II. COUNTING TAYLOR COEFFICIENTS

A function $V(x^\mu)$ expressed as a Taylor series in the neighborhood of a point $P(x_p^0,x_p^1,x_p^2,x_p^3)$ has the form[8]

$$V(x^\mu)=\sum_{n=0}^{\infty}\frac{1}{n!}\left[(x^\alpha-x_p^\alpha)\frac{\partial}{\partial x^\alpha}\right]^n V_p, \tag{2.1}$$

where the subscript on V indicates that derivatives of V are evaluated at P. Taking $V(x^\mu)$ as a function of order[9] zero, we identify an nth-order Taylor coefficient as the coefficient of

$$(x^0-x_p^0)^H(x^1-x_p^1)^I(x^2-x_p^2)^J(x^3-x_p^3)^K,$$

where H, I, J, and K are non-negative integers and

$$n=H+I+J+K.$$

In particular, V_p is the only zeroth-order coefficient, $(V_{,\alpha})_p$ are the four first-order coefficients, and $\frac{1}{2}(V_{,\alpha\beta})_p$ are the ten second-order coefficients.[10]

The number of nth-order Taylor coefficients appearing in the description of V is

$$N_n^4=(n+3)!/3!n! \quad (n\geqslant 0), \tag{2.2a}$$

$$N_n^4=0 \quad (n<0). \tag{2.2b}$$

By induction the number of nth-order coefficients for a zeroth-order function of m variables is[11]

$$N_n^m=(n+m-1)!/(m-1)!n! \quad (n\geqslant 0,\ m\geqslant 1), \tag{2.3a}$$

$$N_n^m=0 \quad (n<0 \text{ or } m<1). \tag{2.3b}$$

An easily derived relation exists among numbers of Taylor coefficients of different orders:

$$N_n^m=N_n^{m-1}+N_{n-1}^m. \tag{2.4}$$

This relation is used below repeatedly in simplifying expressions for the strength of field equations.

 0022-2488/87/040810-05$02.50

In following sections, the total number of free nth-order Taylor coefficients of one set of functions is compared with the total number of free nth-order Taylor coefficients of another set of functions for the case in which the functions of one set are linear combinations of first derivatives of the functions of the second set. Assigning each function an order permits accurate comparison of Taylor coefficients having the same dimensions. Since $V(x^\mu)$ is taken here to be a function of order zero, its nth-order coefficients contain nth derivatives of V, and V has N^4_n nth-order coefficients. If V were an rth-order function, its nth-order Taylor coefficients would contain $(n-r)$th derivatives of V, and V would have N^4_{n-r} nth-order coefficients. In Sec. III, Maxwell's potentials A^α are shown to be functions of first order. Maxwell's electromagnetic components $F^{\alpha\beta}$ are then necessarily of second order.

III. STRENGTH OF MAXWELL'S EQUATIONS

If the concept of strength of a field theory is to be of value, then the strength should be independent of the manner in which the theory is presented. One would expect the theory to have the same strength when considered from different points of view.

In this section, strengths of the electromagnetic and potential formulations of Maxwell's theory in free space are derived and shown to be precisely equivalent. Conformal invariance of the theory is included in each case. Accounting for conformal invariance in the analysis lays the groundwork for expanding the concept of strength beyond its previous treatment in the literature. For simplicity, both formulations are written in Minkowski coordinates.

A. Electromagnetic formulation

Maxwell's equations in free space are

$$C^\alpha \equiv F^{\alpha\beta}{}_{,\beta} = 0, \tag{3.1a}$$

$$D_{\alpha\beta\sigma} \equiv F_{\alpha\beta,\sigma} + F_{\sigma\alpha,\beta} + F_{\beta\sigma,\alpha} = 0. \tag{3.1b}$$

The electromagnetic contravariant and covariant components, $F^{\alpha\beta}$ and $F_{\alpha\beta}$, are antisymmetric in α and β. They are related by

$$F^{\alpha\beta} = \eta^{\alpha\sigma}\eta^{\beta\mu}F_{\sigma\mu}, \tag{3.2}$$

where $\eta^{00} = 1$, $\eta^{11} = \eta^{22} = \eta^{33} = -1$, and $\eta^{\alpha\beta} = 0$ for $\alpha \neq \beta$.

The $4+4$ field equations contain six linearly independent field variables, F_{01}, F_{02}, F_{03}, F_{12}, F_{23}, and F_{31}, where the first three are electric field components and the last three are magnetic field components. Each is taken to be of order 2 and each is a function of four independent variables x^μ. Each function has its own Taylor series so that the six functions together contain $6N^4_{n-2}$ nth-order Taylor coefficients.

Many relations derivable from the field equations exist among these nth-order coefficients. Repeatedly differentiating each of the eight field equations $n-3$ times yields a total of $8N^4_{n-3}$ relations among the nth-order coefficients of $F^{\alpha\beta}$.

These $8N^4_{n-3}$ relations are not independent because the field equations themselves are not independent. Two identities exist among the field equations:

$$C^\alpha{}_{,\alpha} = 0, \tag{3.3a}$$

$$D_{\alpha\beta\sigma,\mu} - D_{\beta\sigma\mu,\alpha} + D_{\sigma\mu\alpha,\beta} - D_{\mu\alpha\beta,\sigma} = 0, \tag{3.3b}$$

where no two indices in (3.3b) have the same value. These identities exist because of the functional forms of the field equations; they would be valid even if the field equations did not state that C^α and $D_{\alpha\beta\sigma}$ were zero. Repeatedly differentiating the identities $n-4$ times yields a total of $2N^4_{n-4}$ identities among the $8N^4_{n-3}$ relations. Therefore, there are just $8N^4_{n-3} - 2N^4_{n-4}$ independent relations among the nth-order coefficients resulting from the field equations. This number should be subtracted from the total number of nth-order coefficients of $F^{\alpha\beta}$.

The field equations are invariant with respect to a conformal transformation,[12] which consists of the space-time translation

$$x^\mu = x'^\mu + \epsilon^\mu, \tag{3.4a}$$

the proper homogeneous Lorentz transformation

$$x^\mu = \Lambda^\mu{}_\sigma x'^\sigma, \tag{3.4b}$$

the scale transformation

$$x^\mu = s x'^\mu, \tag{3.4c}$$

and the acceleration transformation

$$x^\mu = (1 + 2a^\alpha x'_\alpha + a^\alpha a_\alpha x'^\beta x'_\beta)^{-1}(x'^\mu + a^\mu x'^\sigma x'_\sigma). \tag{3.4d}$$

The four parameters ϵ^μ of order -1, the six parameters $\Lambda^\mu{}_\sigma$ of order zero, the single parameter s of order zero, and the four parameters a^μ of first order play no part in the character of the fields themselves. These 15 parameters can be chosen to eliminate four Taylor coefficients of order -1, seven coefficients of order zero, and four coefficients of first order from those of $F^{\alpha\beta}$. Since from (2.3) $N^1_n = 0$ for $n < 0$ and $N^1_n = 1$ for $n \geqslant 0$, this indicates that

$$4N^1_{n+1} + 3N^1_n - 3N^1_{n-1} - 4N^1_{n-2}$$

should be subtracted from the number of nth-order Taylor coefficients of $F^{\alpha\beta}$.

The number Z^E_n of free nth-order coefficients[13] is then

$$Z^E_n = 6N^4_{n-2} - (8N^4_{n-3} - 2N^4_{n-4}) - (4N^1_{n+1} + 3N^1_n - 3N^1_{n-1} - 4N^1_{n-2}). \tag{3.5}$$

The first three terms on the right-hand side can be eliminated by repeated use of (2.4). The result is

$$Z^E_n = 4N^3_{n-2} + 2N^2_{n-2} - 4N^1_{n+1} - 3N^1_n + 3N^1_{n-1} + 4N^1_{n-2}. \tag{3.6}$$

This can be made more amenable to interpretation if it is noted from (2.3a) and (2.3b) that $N^1_n = 0$ for $n < 0$ and $N^1_n = 1$ for $n \geqslant 0$ and if the result in (3.6) is expressed in the form

$$Z^E_{-1} = -4, \quad Z^E_0 = -7, \quad Z^E_1 = -4, \tag{3.7a}$$

$$Z^E_n = 4N^3_{n-2} + 2N^2_{n-2} \quad (n \geqslant 2). \tag{3.7b}$$

Interpretation of Z^E_n is straightforward. The negative numbers for Z^E_{-1}, Z^E_0, and Z^E_1 appear because of the conformal invariance of the system. The quantity N^3_{n-2} is the number of nth-order Taylor coefficients for a second-order function of three variables, and N^2_{n-2} is that for a second-

order function of two variables. Thus Z_n^E for $n \geqslant 2$ corresponds to four second-order functions of three variables and two second-order functions of two variables. These functions can be chosen freely provided they are consistent with the field equations. The selected functions can be viewed as initial data, which can be correlated with the Cauchy initial-value problem as applied to the field equations,[7] an interpretation that is verified by an example in Appendix A.

This development complements the work of Mariwalla[6] in several respects. Conformal invariance is included for the first time here in the derivation of Z_n^E, resulting in (3.7a). The result for Z_n^E in (3.7b), together with its detailed interpretation, is original here.

Mariwalla[6] expressed Z_n^E and the strength Z_n^P for the potential formulation in the infinite series

$$Z_n^E = N_{n-2}^4 (E_0 + E_1/n + E_2/n^2 + \cdots), \tag{3.8a}$$

$$Z_n^P = N_{n-2}^4 (P_0 + P_1/n + P_2/n^2 + \cdots), \tag{3.8b}$$

where all E_m and P_m were constants. Mariwalla[6] showed that $E_0 = P_0 = 0$ and $E_1 = P_1$. The coefficients E_m and P_m for $m > 1$ were not considered. In effect, this meant that Z_n^E and Z_n^P were shown to be equal only for large n. Exact equivalence of the free data for the two formulations did not follow from this result.

In Sec. III B, Z_n^P is derived and shown to be equal to Z_n^E for all n. Next, Z_n^P is shown to represent initial data required to yield a solution of the potential formulation of Maxwell's equations in free space. Finally, the suggestion that the strength concept may be useful in finding unknown formulations of established field theories is discussed.

B. Potential formulation

In free space, the potentials A^σ satisfy the field equations

$$K^\alpha \equiv \eta^{\mu\beta} A^{\alpha}{}_{,\mu,\beta} - \eta^{\mu\alpha} A^{\beta}{}_{,\mu,\beta} = 0. \tag{3.9}$$

Since each of the functions A^σ is of first order, the four functions together contain $4N_{n-1}^4$ nth-order Taylor coefficients.

The system is conformally invariant so that, as in the electromagnetic formulation,

$$4N_{n+1}^1 + 3N_n^1 - 3N_{n-1}^1 - 4N_{n-2}^1$$

should be subtracted from the number of nth-order coefficients of A^σ to account for conformal invariance of the system.

Many relations, derivable from the four field equations, exist among these nth-order coefficients. Repeatedly differentiating (3.9) $n-3$ times yields a total of $4N_{n-3}^4$ relations among the nth-order coefficients. The $4N_{n-3}^4$ relations are not independent, however, because the four field equations are not independent. An identity exists among them:

$$K^{\alpha}{}_{,\alpha} = 0. \tag{3.10}$$

This identity comes about because of the form of the field equations. It would be valid even if K^α were not zero. Repeatedly differentiating this identity $n-4$ times yields N_{n-4}^4 identities among the nth-order relations. Therefore, there are just $4N_{n-3}^4 - N_{n-4}^4$ independent relations among the nth-order coefficients. This number must be subtracted from the total number of nth-order coefficients describing A^σ to account for the field equations.

The field equations (3.9) exhibit gauge invariance. They remain unchanged by a gauge transformation

$$A_\alpha = A'_\alpha + \Omega_{,\alpha}, \tag{3.11}$$

where $\Omega(x^\mu)$ is an arbitrary function of order zero. Many of the nth-order coefficients of A^σ can be removed by performing a gauge transformation. Such coefficients should not be included in the number of free nth-order coefficients because they are not necessary for describing the field; they can be transformed away.

Since the gauge function Ω is of order zero, N_n^4 nth-order coefficients are required to describe it. However, the linear transformation

$$\Omega = \Omega' + \delta, \tag{3.12}$$

where δ is any constant of zero order, leaves the gauge transformation (3.11) itself unchanged. Since δ plays no role in describing the gauge transformation, the quantity $N_n^1 - N_{n-1}^1$ should be subtracted from the number of nth-order coefficients describing Ω. This means that $N_n^4 - (N_n^1 - N_{n-1}^1)$ must be subtracted from the number of nth-order coefficients of A^σ in order to account for the gauge invariance of the field equations.

The constant δ can be interpreted as a quantity corresponding to the parameter s associated with the scale transformation (3.4c) in a conformal transformation. The quantity s is dimensionless. Its order is arbitrarily taken here to be zero. Taking the order of δ to be equal to the order of s requires both δ and Ω to have order 0, A^α to have order 1, and $F^{\alpha\beta}$ to have order 2.

The number Z_n^P of free nth-order coefficients[14] for the potential formulation is then

$$\begin{aligned} Z_n^P = {} & 4N_{n-1}^4 - (4N_{n-3}^4 - N_{n-4}^4) \\ & - (N_n^4 - N_n^1 + N_{n-1}^1) \\ & - (4N_{n+1}^1 + 3N_n^1 - 3N_{n-1}^1 - 4N_{n-2}^1). \end{aligned} \tag{3.13}$$

It should be noted that the gauge function Ω and the constant δ play essential roles in the development of (3.13); if they were not considered, the calculated number of free nth-order coefficients would be incorrect.

The expression on the right-hand side of (3.13) can be simplified by repeated use of (2.4). The result is

$$\begin{aligned} Z_n^P = {} & 4N_{n-2}^3 + 2N_{n-2}^2 - 4N_{n+1}^1 \\ & - 3N_n^1 + 3N_{n-1}^1 + 4N_{n-2}^1. \end{aligned} \tag{3.14}$$

Comparing (3.14) with (3.6) yields

$$Z_n^E = Z_n^P \quad (\text{all } n). \tag{3.15}$$

This agreement between the strengths of the electromagnetic and potential formulations means that each formulation requires exactly the same amount of initial data for a solution to the respective set of field equations. Of course, the form of the initial data in each formulation is expected to be different, but this does not detract from the significance of (3.15).

Verification that Z_n^P of (3.14) represents the amount of initial data required for a solution of (3.9) begins with an

examination of (3.9) in expanded form:

$$\eta^{jk}A^{\,0}_{j,k} - \eta^{00}A^{\,j}_{j,0} = 0, \tag{3.16a}$$

$$\eta^{00}A^{\,i}_{,0,0} = -\eta^{jk}A^{\,i}_{j,k} + \eta^{ij}A^{\,\alpha}_{,\alpha,j}. \tag{3.16b}$$

The form of (3.16) precludes the determination of $A^{\,0}$ at time $t + \Delta t$ even if $A^{\,\alpha}$ and $A^{\,\alpha}_{,0}$ are known everywhere at time t. This impasse is overcome by imposition of a gauge condition, which here is taken to be that of Lorentz:

$$A^{\,\alpha}_{,\alpha} = 0. \tag{3.17}$$

As Mariwalla[6] showed, a gauge condition does not change the strength $Z^{\,P}_{n}$ of the potential formulation if an appropriate constraint is imposed on the gauge function Ω used in any gauge transformation. Equation (3.17) remains unchanged under any gauge transformation for which the gauge function Ω satisfies[15]

$$\eta^{\alpha\beta}\Omega_{,\alpha,\beta} = 0. \tag{3.18}$$

Including (3.17) in the strength calculation adds $N^{\,4}_{n-2}$ to $Z^{\,P}_{n}$, while including (3.18) in the strength calculation subtracts $N^{\,4}_{n-2}$ from $Z^{\,P}_{n}$.

Equation (3.14) can be written, with the aid of (2.4), in the form

$$Z^{\,P}_{n} = (4N^{\,3}_{n-1} + 2N^{\,3}_{n-2} + N^{\,2}_{n-2}) - (N^{\,3}_{n} + N^{\,3}_{n-1}) - (4N^{\,1}_{n+1} + 2N^{\,1}_{n} - 2N^{\,1}_{n-1} - 4N^{\,1}_{n-2}). \tag{3.19}$$

Terms in the first set of parentheses correspond to initial data required to yield solutions of the field equations (3.9) and the Lorentz condition (3.17) together. Terms in the second set of parentheses correspond to initial data required to yield solutions of the constraint (3.18) on Ω. Thus initial data in the form of four first-order functions of three variables, two second-order functions of three variables, and one second-order function of two variables are required to yield solutions of (3.9) and (3.17) for $A^{\,\alpha}$. Initial data in the form of one zeroth-order function of three variables and one first-order function of three variables are required to yield a solution of (3.18) for Ω. Since Ω plays no physical role in the description of the field, the solution for Ω can be used in a gauge transformation (3.11) to eliminate irrelevant data in $A^{\,\alpha}$.

It is not obvious that the amount of initial data of the first set of parentheses in (3.19) is sufficient to yield solutions of the four field equations (3.9) and the Lorentz condition (3.17) for $A^{\,\sigma}$. An example demonstrating that they are sufficient is given in Appendix B.

Terms in the third set of parentheses of (3.19) correspond to the reduction in initial data permitted by a space-time translation (3.4a), a homogeneous Lorentz transformation (3.4b), and an acceleration transformation (3.4d). The difference between the coefficients of $N^{\,1}_{n} - N^{\,1}_{n-1}$ in (3.19) and (3.14) indicates that the system is no longer invariant with respect to a scale transformation (3.4c).

Equation (3.14) shows that the amount of initial data required for the set of second-order field equations (3.9), relating functions of first order, is precisely the same as the amount of initial data required for a set of first-order differential equations relating functions of second order. If an investigator, unaware of the existence of the electromagnetic formulation, obtained the result of (3.14) for the potential formulation, he might feel compelled to seek a reformulation of the theory in the form of eight first-order differential equations having two identities of second order. He would expect the equations to relate six dependent variables, each of which could be written as a linear combination of first derivatives of $A^{\,\sigma}$. He would also expect the equations to exhibit conformal invariance. He could not be certain that the first-order equations existed, but he could be sure that they would have the predicted form if they did exist. And, of course, the investigator would meet with success in this case.

These results for Maxwell's theory suggest that the concept of strength of a field theory as described here may be useful in obtaining unknown formulations, together with their invariance properties, of established field theories.

APPENDIX A: INITIAL DATA FOR THE ELECTROMAGNETIC FORMULATION

Presented here is an example demonstrating that the number of Taylor coefficients of (3.7b) corresponds to initial data required for a solution to the electromagnetic formulation of Maxwell's theory. Suppose F_{01}, F_{02}, F_{23}, and F_{31} are given at all points in space at time $x^0 = 0$, and F_{03} and F_{12} are given at all points on the $x^3 = 0$ plane at $x^0 = 0$. These six initial conditions constitute an example of four second-order functions of three variables and two second-order functions of two variables that can be chosen freely.

With these initial conditions given, Maxwell's equations provide solutions at all points in space for all time. The components F_{03} and F_{12} are found at all points in space at time $x^0 = 0$ from (3.1a) with $\alpha = 0$ and from (3.1b) with $\alpha = 1$, $\beta = 2$, and $\sigma = 3$:

$$F_{03}(0,x^1,x^2,(n+1)\Delta x^3) = F_{03} - (F_{01,1} + F_{02,2})\Delta x^3, \tag{A1}$$

$$F_{12}(0,x^1,x^2,(n+1)\Delta x^3) = F_{12} - (F_{23,1} + F_{31,2})\Delta x^3, \tag{A2}$$

where the right-hand sides are evaluated at $(0,x^1,x^2,n\Delta x^3)$ and where the relations $F^{0j} = -F_{0j}$ from (3.3) have been employed. The quantity Δx^3 is incremental length and n is an integer running from 0 to $\pm\infty$.

The electric field components F_{0j} and the magnetic field components F_{jk} are found for all time, respectively, from (3.1a) with $\alpha = j$ and from (3.1b) with $\alpha = j$, $\beta = k$, and $\sigma = 0$:

$$F_{0j}((n+1)\Delta x^0,x^m) = F_{0j} - (F_{j1,1} + F_{j2,2} + F_{j3,3})\Delta x^0, \tag{A3}$$

$$F_{jk}((n+1)\Delta x^0,x^m) = F_{jk} + (F_{0k,j} - F_{0j,k})\Delta x^0, \tag{A4}$$

where the right-hand sides are evaluated at $(n\Delta x^0,x^m)$, where Δx^0 is the speed of light times incremental time, and where n is an integer running from 0 to $\pm\infty$. The relations $F^{0j} = -F_{0j}$ and $F^{jk} = F_{jk}$ from (3.2) have been employed in arriving at (A3).

APPENDIX B: INITIAL DATA FOR THE POTENTIAL FORMULATION

Presented here is an example demonstrating that the terms in the first set of parentheses of (3.19) correspond to

the amount of initial data required for a complete solution of the potential field equations (3.9) and the Lorentz condition (3.17).

If initial data for the determination of A^{σ} are taken to be $A^{\sigma}(0,x^m)$, $A^{1}_{,0}(0,x^m)$, $A^{2}_{,0}(0,x^m)$, and $A^{3}_{,0}(0,x^1,x^2,0)$, then (3.17) gives $A^{0}_{,0}(0,x^m)$:

$$A^{0}_{,0}(0,x^m) = -A^{j}_{,j}(0,x^m). \tag{B1}$$

The remaining initial condition $A^{3}_{,0}(0,x^m)$ is found first by combining (3.9) and (3.17) and writing the result for $\alpha = 0$:

$$A^{0}_{,0,0} + \eta^{jk}A^{0}_{j,k} = 0. \tag{B2}$$

Next, differentiating (3.17) with respect to x^0 gives

$$A^{0}_{,0,0} + A^{j}_{j,0} = 0. \tag{B3}$$

Subtracting (B3) from (B2) and solving for $(A^{3}_{,0})_{,3}$ yield

$$(A^{3}_{,0})_{,3} = \eta^{jk}A^{0}_{j,k} - (A^{1}_{,0})_{,1} - (A^{2}_{,0})_{,2}. \tag{B4}$$

The right-hand side is known at all points in space at $x^0 = 0$. Since $A^{3}_{,0}(0,x^1,x^2,0)$ is known, $A^{3}_{,0}(0,x^m)$ can be found from (B4) by stepping the x^3 coordinate.

Since the components $A^{\sigma}(0,x^m)$ are given and since $A^{\sigma}_{,0}(0,x^m)$ are now known, the components $A^{\sigma}(x^{\alpha})$ can be found from (3.16b) and (3.17) everywhere for all time by stepping the time coordinate.

Finally, if initial data for Ω is given in the form of functions for $\Omega(0,x^m)$ and $\Omega_{,0}(0,x^m)$, a complete solution for Ω at all points in space-time can be found from (3.18) by stepping the time coordinate. In effect, the amount of initial data required for finding a complete solution for Ω can be removed from the amount of total initial data required for a solution for A^{σ} by performing a gauge transformation.

[1] A. Einstein, *The Meaning of Relativity* (Princeton U. P., Princeton, NJ, 1955), 4th ed., p. 133.
[2] R. Penney, J. Math. Phys. **6**, 1607 (1965).
[3] C. Hoenselaers, Prog. Theor. Phys. **58**, 1185 (1977).
[4] R. Burman, Czech. J. Phys. B **27**, 113 (1977).
[5] G. L. Murphy, Int. J. Theor. Phys. **18**, 323 (1979).
[6] K. H. Mariwalla, J. Math. Phys. **15**, 468 (1974).
[7] B. F. Schutz, J. Math. Phys. **16**, 855 (1975).

[8] Greek indices take on values 0,1,2,3, while Latin indices take on values 1,2,3. Repeated indices are summed.
[9] The order of a function here is a number describing the dimension of the function relative to a reference function whose order is a number chosen arbitrarily. The derivative with respect to x^{α} of a function of order r has order $r+1$. Multiplying a function of order r by x^{α} yields a function of order $r-1$.
[10] Partial differentiation of a variable Q^{μ} with respect to x^{σ} is denoted by $Q^{\mu}_{,\sigma}$.
[11] The quantity on the right-hand side of (2.3a) was designated $(^{m}_{n})$ by Einstein,[1] and later $[^{m}_{n}]$ by Mariwalla.[6] It is designated here by N^{m}_{n}, a symbol more easily displayed, read, and manipulated algebraically.

[12] T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962).

[13] Equation (13) of Mariwalla[6] contains a minor error; the upper limit on the summation should be $d-2$ rather than d. With this change, with $d=4$, and with n replaced by $n-2$, Eq. (13) of Mariwalla's paper reduces to the first three terms on the right-hand side of (3.5). The last four terms on the right-hand side of (3.5) were not obtained by Mariwalla since he ignored conformal invariance of the system.
[14] Mariwalla's[6] equation just below his Eq. (16) reduces for $d=4$ and for n replaced by $n-1$ to the first four terms on the right-hand side of (3.13). Mariwalla did not obtain the fifth and sixth terms of (3.13) because he did not consider the transformation (3.12), and he did not obtain the last four terms on the right-hand side of (3.13) because he did not consider conformal invariance of the system.

[15] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1965), p. 181.

The equivariant inverse problem and the Maxwell equations

R. J. Noriega and C. G. Schifini
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

(Received 24 June 1986; accepted for publication 26 November 1986)

In affirmative, the equivariant inverse problem for Maxwell-type Euler–Lagrange expressions is solved. This allows the proof of the uniqueness of the Maxwell equations.

I. INTRODUCTION

It is very well known that Maxwell equations can be written in covariant form as[1]

$$F^{ij}{}_{|j} = J^i, \tag{1.1}$$

$$*F^{ij}{}_{|j} = 0, \tag{1.2}$$

where F^{ij} is a skew-symmetric tensor and $*F^{ij} = \eta^{ijhk}F_{hk}$. Equation (1.2) is equivalent to the existence of a covector field ψ_i such that

$$F_{ij} = \psi_{i,j} - \psi_{j,i}. \tag{1.3}$$

The Maxwell equations can be deduced from a variational principle as follows. If

$$L = L(g_{ij};\psi_i;\psi_{i,j}), \tag{1.4}$$

then from a variation of ψ_i we obtain as Euler–Lagrange equations,

$$E^i(L) = \frac{\partial L}{\partial \psi_i} - \frac{\partial}{\partial x^j}\left(\frac{\partial L}{\partial \psi_{i,j}}\right) = 0; \tag{1.5}$$

on choosing

$$L = \sqrt{g}F^{ij}F_{ij}, \tag{1.6}$$

Eqs. (1.5) become (1.1).

The left-hand side of (1.1) has two properties of covariance: (1) by a transformation of coordinates it changes as a vector; and (2) by a change of gauge, i.e., by a transformation of the type $\psi_i \to \psi_i + \phi_{,i}$, where ϕ is a scalar, it is invariant. These properties are also possessed by the Lagrangian (1.6). However, the last assertion is not mandatory since the Lagrangian does not have, in general, any physical meaning, although the Euler–Lagrange expressions do have a meaning. The main purpose of this article is to prove that the situation already encountered with the Maxwell equations is found always in spaces of dimension 4, i.e., the assumption of the two covariance properties for the Euler–Lagrange expressions implies that the Lagrange is equivalent to (it has the same Euler–Lagrange expressions as) a Lagrangian with the same properties.[2] This solves for the affirmative the equivariant inverse problem[3] for Maxwell-type Euler–Lagrange expressions.

Precisely, we consider a quantity

$$B^i = B^i(g_{ij};g_{ij,h};\psi_i;\psi_{i,j};\psi_{i,jh}) \tag{1.7}$$

such that

(i) B^i is a vector,

(ii) B^i is gauge invariant, (1.8)

(iii) $B^i = E^i(L_1)$ for L_1 of the type (1.4).

Here we do not assume any covariance property for L_1 with respect to transformations of coordinates or changes of gauge. We will prove that (1.7) and (1.8) imply the existence of $L = L(g_{ij};\psi_i;\psi_{i,j})$, which is a gauge-invariant scalar density such that $E^i(L) = B^i$.

II. THE EQUIVARIANT INVERSE PROBLEM

The condition (iii) in (1.8) written in full is

$$B^i = L^i - L^{i,j;hk}g_{hk,j} - L^{i,j;h}\psi_{h,j} - \tfrac{1}{2}(L^{i,j;h,k} + L^{i,k;h,j})\psi_{h,kj}, \tag{2.1}$$

where $L^i = \partial L/\partial \psi_i$, $L^{i,j} = \partial L/\partial \psi_{i,j}$, and $L^{hk} = \partial L/\partial g_{hk}$. The coefficients of (2.1) depend only on g_{ij}, ψ_i, and $\psi_{i,j}$.

Since B^i is gauge invariant, by the replacement theorem[4] we have

$$B^i = H^i - \tfrac{1}{2}H^{ijh}F_{hj} - \tfrac{2}{3}H^{ijhk}F_{hk|j}, \tag{2.2}$$

where

$$H^i = L^i(g_{ij};0;\tfrac{1}{2}F_{ij}), \quad H^{ijh} = L^{i,j;h}(g_{ij};0;\tfrac{1}{2}F_{ij}),$$
$$H^{ijhk} = \tfrac{1}{2}(L^{i,j;h,k} + L^{i,k;h,j})(g_{ij};0;\tfrac{1}{2}F_{ij}). \tag{2.3}$$

Since B^i is a tensorial density, the same is true for $\partial B^i/\partial \psi_{h,kj}$. From (2.1) this derivative is $\frac{1}{2}(L^{i,j;h,k} + L^{i,k;h,j})$. Then H^{ijhk} defined by (2.3) is also a tensorial density. Hence, by (2.2), $H^i - \frac{1}{2}H^{ijh}F_{hj}$ is a vector density and depends only on g_{ij} and F_{ij}. It is known[5] that such vector densities are zero, and so we obtain

$$B^i = -\tfrac{2}{3}H^{ijhk}F_{hk|j}. \tag{2.4}$$

Denoting $W^{ijhk} = \frac{1}{2}(L^{i,j;h,k} + L^{i,k;h,j})$, from (2.1) and (2.2) we deduce

$$W^{ijhk}\psi_{h,kj} = H^{ijhk}(\psi_{h,kj} - \psi_{k,hj}) + (\cdots), \tag{2.5}$$

where in $(\cdots)$ we have gathered the terms that do not depend on $\psi_{h,kj}$. From the equality of the cross derivatives it is easily obtained that $W^{ijhk} = W^{hkij}$, and by (2.3), $H^{ijhk} = H^{hkij}$. Also, $W^{ijhk} = W^{ikhj}$ is valid and so $H^{ijhk} = H^{ikhj}$.

Differentiating (2.5) with respect to $\psi_{a,bc}$ and taking into account the mentioned symmetries, we have

$$2W^{icab} = 2H^{icab} - H^{icba} - H^{ibca}. \tag{2.6}$$

Changing i with a and c with b in (2.6) and subtracting the resulting equation to (2.6), it follows that

$$H^{abci} + H^{acbi} - H^{icba} - H^{ibca} = 0. \tag{2.7}$$

Changing a with c and b with i in (2.7) and subtracting the resulting equation to (2.7), we have $H^{acbi} = H^{ibca}$, or else

$$H^{acbi} = H^{caib}. \tag{2.8}$$

 0022-2488/87/040815-03$02.50

Denoting

$$T^{ijhk} = -\tfrac{1}{3}(H^{ijhk} - H^{ijkh}), \tag{2.9}$$

it follows that T^{ijhk} is a tensor, such that

$$B^i = \sqrt{g} T^{ijhk} F_{hk|j}, \tag{2.10}$$

and besides,

$$T^{ijhk} = -T^{jihk} = T^{hkij}. \tag{2.11}$$

Using Theorem 3 in Kerrighan,[5] we deduce

$$\begin{aligned} T^{ijhk} = {} & A\,{}^*F^{ij}\,{}^*F^{hk} - B(F^{ik}\,{}^*F^{jh} - F^{ih}\,{}^*F^{jk} \\ & - F^{jk}\,{}^*F^{ih} + F^{jh}\,{}^*F^{ik}) + CF^{ij}F^{hk} \\ & + 2D(g^{ih}g^{jk} - g^{ik}g^{jh}) + E\eta^{ijhk}, \end{aligned} \tag{2.12}$$

where A, B, C, D, and E are scalar concomitants of g_{ij} and F_{ij}. Then

$$\begin{aligned} B^i = \sqrt{g}\{ & (A\,{}^*F^{ij}\,{}^*F^{hk} + B(F^{ij}\,{}^*F^{hk} + {}^*F^{ij}F^{hk}) \\ & + CF^{ij}F^{hk})F_{hk|j} + DF^{ij}{}_{|j}\}. \end{aligned} \tag{2.13}$$

Since B^i is a Euler–Lagrange expression, it satisfies certain identities[6]:

$$B^{i;j,hk} = B^{j;i,hk}, \tag{2.14}$$

$$B^{i;j,h} = -B^{j;i,h} + 2\frac{\partial}{\partial x^k}(B^{j;i,hk}), \tag{2.15}$$

$$B^{i;j} = B^{j;i} - \frac{\partial}{\partial x^h}(B^{j;i,h}) + \frac{\partial^2}{\partial x^h \partial x^k}(B^{j;i,hk}). \tag{2.16}$$

It is easy to see that (2.14) is satisfied identically. Let us see what restrictions (2.15) imposes on A, B, C, and D. It is known[7] that there are functions $\lambda_1, \lambda_2, \lambda_3$, and λ_4 of two real variables such that

$$\begin{aligned} & A = \lambda_1(\phi,\psi), \quad B = \lambda_2(\phi,\psi), \\ & C = \lambda_3(\phi,\psi), \quad D = \lambda_4(\phi,\psi); \end{aligned} \tag{2.17}$$

where

$$\phi = {}^*F^{ij}F_{ij}, \quad \psi = F^{ij}F_{ij}. \tag{2.18}$$

Equation (2.15) gives rise, in view of (2.13), to the following:

$$\begin{aligned} & \Big\{ 2\frac{\partial\lambda_1}{\partial\psi}(F^{hk}\,{}^*F^{ac}\,{}^*F^{ib} - F^{ab}\,{}^*F^{ic}\,{}^*F^{hk}) + 2\frac{\partial\lambda_2}{\partial\psi}(F^{hk}F^{ac}\,{}^*F^{ib} - F^{ab}F^{ic}\,{}^*F^{hk}) + 2\frac{\partial\lambda_2}{\partial\phi}({}^*F^{hk}F^{ab}\,{}^*F^{ic} - {}^*F_{ab}\,{}^*F^{ic}F^{hk}) \\ & + 2\frac{\partial\lambda_2}{\partial\phi}({}^*F^{hk}F^{ac}F^{ib} - {}^*F^{ab}F^{ic}F^{hk}) + \frac{\partial\lambda_4}{\partial\phi}({}^*F^{hk}g^{ai}g^{cb} - {}^*F^{hk}g^{ab}g^{ci} - 2\,{}^*F^{ab}g^{hi}g^{kc}) \\ & + \frac{\partial\lambda_4}{\partial\psi}(F^{hk}g^{ai}g^{cb} - F^{hk}g^{ab}g^{ci} - 2F^{ab}g^{hi}g^{kc}) + \lambda_2({}^*F^{ib}g^{ha}g^{kc} + \tfrac{1}{2}\,{}^*F^{hk}g^{ib}g^{ca} - \tfrac{1}{2}\,{}^*F^{hk}g^{ia}g^{cb}) \\ & + \lambda_3(F^{ib}g^{ha}g^{kc} + \tfrac{1}{2}F^{hk}g^{ib}g^{ca} - \tfrac{1}{2}F^{hk}g^{ia}g^{cb})\Big\}(\psi_{h,kc} - \psi_{k,hc}) + (i \leftrightarrow a) = (\cdots), \end{aligned} \tag{2.19}$$

where $(i \leftrightarrow a)$ denotes symmetrization of the previous expression with respect to i and a, and where we have gathered in $(\cdots)$ all the terms that do not depend on the second derivatives of ψ_i.

It is known[8] that given a point on the underlying manifold there is a coordinate system such that, at the point

$$(g_{ij}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$(F_{ij}) = \begin{pmatrix} 0 & \alpha & 0 & 0 \\ -\alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \\ 0 & 0 & -\beta & 0 \end{pmatrix}, \tag{2.20}$$

from where we deduce

$$(F^{ij}) = \begin{pmatrix} 0 & -\alpha & 0 & 0 \\ \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \\ 0 & 0 & -\beta & 0 \end{pmatrix},$$

$$({}^*F^{ij}) = \begin{pmatrix} 0 & 2\beta & 0 & 0 \\ -2\beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\alpha \\ 0 & 0 & -2\alpha & 0 \end{pmatrix}. \tag{2.21}$$

Differentiating (2.19) with respect to $\psi_{a,bc}$ and choosing $i = a = 1$, $c = b = s = 4$, and $r = 3$, we obtain

$$4\alpha\frac{\partial\lambda_4}{\partial\phi} + 2\beta\frac{\partial\lambda_4}{\partial\psi} = 2\alpha\lambda_2 + \beta\lambda_3. \tag{2.22}$$

Differentiating (2.19) with respect to $\psi_{a,bc}$ and choosing $i = a = 4$, $c = b = s = 2$, and $r = 1$, we have

$$4\beta\frac{\partial\lambda_4}{\partial\phi} - 2\alpha\frac{\partial\lambda_4}{\partial\psi} = 2\beta\lambda_2 - \alpha\lambda_3. \tag{2.23}$$

Let us suppose for the moment that $\det(F_{ij}) \neq 0$. Then from (2.22) and (2.23) it follows easily that

$$\lambda_2 = 2\frac{\partial\lambda_4}{\partial\phi}, \quad \lambda_3 = 2\frac{\partial\lambda_4}{\partial\psi}. \tag{2.24}$$

Differentiating (2.19) with respect to $\psi_{a,bc}$ and choosing now $a = b = 1$, $c = i = 2$, $r = 3$, and $s = 4$, we obtain

$$2\beta\frac{\partial\lambda_2}{\partial\phi} + \alpha\frac{\partial\lambda_2}{\partial\psi} = 2\beta\frac{\partial\lambda_1}{\partial\psi} + \alpha\frac{\partial\lambda_3}{\partial\phi}. \tag{2.25}$$

Finally, differentiating (2.19) with respect to $\psi_{a,bc}$ and choosing $a = b = 4$, $c = i = 3$, $r = 1$, and $s = 2$, we have

$$2\alpha\frac{\partial\lambda_1}{\partial\psi} + \beta\frac{\partial\lambda_2}{\partial\psi} = 2\alpha\frac{\partial\lambda_2}{\partial\phi} + \beta\frac{\partial\lambda_3}{\partial\phi}. \tag{2.26}$$

From (2.25) and (2.26) it follows that

$$\frac{\partial\lambda_2}{\partial\phi} = \frac{\partial\lambda_1}{\partial\psi}, \quad \frac{\partial\lambda_2}{\partial\psi} = \frac{\partial\lambda_3}{\partial\phi}. \tag{2.27}$$

Now, (2.24) and (2.27) are the integrability conditions we

need to establish the existence of a scalar $T = T(\phi,\psi)$ such that

$$\lambda_1 = 2\frac{\partial^2 T}{\partial\phi^2}, \quad \lambda_2 = 2\frac{\partial^2 T}{\partial\phi\,\partial\psi}, \qquad \lambda_3 = 2\frac{\partial^2 T}{\partial\psi^2}, \quad \lambda_4 = \frac{\partial T}{\partial\psi}. \tag{2.28}$$

In this case, a straightforward computation proves that

$$B^i = E^i(L), \tag{2.29}$$

where $L = \sqrt{g}T$ is a gauge-invariant scalar density. This solves the equivariant inverse problem since $\det(F_{ij}) \neq 0$ is a dense subset of the space of variables g_{ij}, F_{ij} [and then (2.29) is valid everywhere by a continuity argument].

III. THE MAXWELL EQUATIONS

We have proved in Sec. II that if B^i is of the type (1.7) and it satisfies (1.8), then

$$B^i = L^{ij}{}_{|j}, \tag{3.1}$$

where $L^{ij} = \partial L/\partial F_{ij}$ and $L = \sqrt{g}T(\phi,\psi)$, with ϕ and ψ given by (2.18). By similarity, with the Maxwell equations (1.1) and (1.2), we could claim the field equations to be

$$L^{ij}{}_{|j} = J^i, \tag{3.2}$$

$$^*L^{ij}{}_{|j} = 0. \tag{3.3}$$

Now, Eqs. (3.3) corresponding to the Maxwell internal equations do not depend on the charge and current distribution, and so they should be satisfied identically. Written out in full, they are

$$\begin{aligned} ^*L^{ij}{}_{|j} = \sqrt{g}\Big[& 16\frac{\partial^2 L}{\partial\phi^2}\,{}^*F^{hk}F^{ij}F_{hk|j} + 16\frac{\partial^2 L}{\partial\phi\,\partial\psi}F^{hk}F^{ij}F_{hk|j} \\ & + 8\frac{\partial L}{\partial\phi}F^{ij}{}_{|j} + 4\frac{\partial^2 L}{\partial\psi\,\partial\phi}\,{}^*F^{hk}\,{}^*F^{ij}F_{hk|j} \\ & + 4\frac{\partial^2 L}{\partial\psi^2}\,{}^*F^{ij}F^{hk}F_{hk|j}\Big] = 0. \end{aligned} \tag{3.4}$$

Differentiating (3.4) with respect to $\psi_{a,bc}$, we obtain

$$\begin{aligned} & 4\frac{\partial^2 L}{\partial\phi^2}({}^*F^{ab}F^{ic} + {}^*F^{ac}F^{ib}) + 4\frac{\partial^2 L}{\partial\phi\,\partial\psi}(F^{ab}F^{ic} + F^{ac}F^{ib}) \\ & + \frac{\partial^2 L}{\partial\psi\,\partial\phi}({}^*F^{ab}\,{}^*F^{ic} + {}^*F^{ac}\,{}^*F^{ib}) \\ & + \frac{\partial^2 L}{\partial\psi^2}(F^{ab}\,{}^*F^{ic} + F^{ac}\,{}^*F^{ib}) \\ & + \frac{\partial L}{\partial\phi}(2g^{ia}g^{bc} - g^{ib}g^{ac} - g^{ic}g^{ab}) = 0. \end{aligned} \tag{3.5}$$

In the coordinate system in which (2.20) is valid and choosing $a = b = 1, c = i = 3$, we have $\partial L/\partial\phi = 0$ if we assume $\det(F_{ij}) \neq 0$. Choosing now $a = i = 1$, $b = c = 2$, it follows that $\partial^2 L/\partial\psi^2 = 0$. By a continuity argument we deduce that everywhere,

$$L = \sqrt{g}(C\psi + A), \tag{3.6}$$

where A and C are real numbers, and so

$$L^{ij}{}_{|j} = C\sqrt{g}F^{ij}{}_{|j},$$

which makes (3.2) the usual Maxwell equations.

ACKNOWLEDGMENT

We thank the referee for helpful suggestions.

[1]The tensor g_{ij} is the metric tensor of a four-space-time manifold and $g = |\det(g_{ij})|$. We use the summation convention and we raise and lower indices with g_{ij} and g^{ij}. We denote $\eta^{ijhk} = (\sqrt{g})^{-1}\epsilon^{ijhk}$, where ϵ^{ijhk} are the Levi-Civita permutation symbols. The vertical bar stands for covariant derivative and the comma stands for the usual partial derivative. Here, J^i is the charge–current vector.

[2]This is not true for all dimensions. If it is 3, then $B^i = \epsilon^{ijh}F_{jh}$ is of the form (1.7) and satisfies (1.8) with $L = \frac{1}{2}\epsilon^{ijh}\psi_i F_{jh}$, but L cannot be replaced by a gauge invariant Lagrangian. See, for example, M. C. Calvo, C. Lopez, R. J. Noriega, and C. G. Schifini, "Gauge invariance of Euler–Lagrange expressions in Einstein–Yang–Mills field theories," submitted to Gen. Relativ. Gravit.

[3]I. M. Anderson, Ann. Math. **120**, 329 (1984).

[4]G. W. Horndeski, Tensor (N.S.) **32** (2), 131 (1978).

[5]B. Kerrighan, Gen. Relativ. Gravit. **13** (1), 19 (1981).

[6]I. M. Anderson and T. Duchamp, Am. J. Math. **102** (5), 781 (1980).

[7]R. J. Noriega and C. G. Schifini, Gen. Relativ. Gravit. **16** (3), 292 (1984).

[8]V. Hlavatý, J. Rat. Mech. Anal. **1**, 539 (1952).

Markovian limit for a reduced operation-valued stochastic process

Alberto Barchielli
Dipartimento di Fisica dell'Università di Milano, Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Via Celoria 16, 20133 Milano, Italy

(Received 31 October 1984, accepted for publication 18 November 1986)

Operation-valued stochastic processes give a formalization of the concept of continuous (in time) measurements in quantum mechanics. In this article, a first stage M of a measuring apparatus coupled to the system S is explicitly introduced, and continuous measurement of some observables of M is considered (one can speak of an indirect continuous measurement on S). When the degrees of freedom of the measuring apparatus M are eliminated and the weak coupling limit is taken, it is shown that an operation-valued stochastic process describing a direct continuous observation of the system S is obtained.

I. INTRODUCTION

In the framework of the operational approach to quantum mechanics,[1-3] a formalism for treating continuous (in time) measurements has been developed[4-8] and some applications worked out.[9-11] This formalism allows us to describe a situation in which one or more observables are measured with continuity in a certain time interval. These continuous measurements can be realized by means of a set of operation-valued measures (for this terminology see, for instance, Refs. 1–3) satisfying a certain composition law which is a kind of Markov property. We have called this set of measures an *operation-valued stochastic process* (OVSP).[4]

In quantum mechanics there are two ways for describing measurements. A first one is to consider the system of interest alone and to describe measurements on it in an abstract way by means of the so-called operations; this situation can be called a direct measurement. A second, more concrete way for describing measurements is to consider explicitly a first stage of the measuring apparatus, interacting with the system of interest; from a measurement on this apparatus one can gain information on the system itself. This situation can be called an indirect measurement. These two descriptions are essentially equivalent in the sense that, if in the case of indirect measurement one eliminates the degrees of freedom of the apparatus by taking the partial trace over the Hilbert space of the apparatus, then one recovers a direct measurement, described by operations.[3,12] A similar consistency property does not hold in the usual formulation of quantum mechanics (in which observables are represented by self-adjoint operators and state changes are described by von Neumann reduction postulate). This is one of the reasons for generalizing quantum mechanics.

In the case of continuous measurements we have a similar consistency problem. Consider explicitly a first stage of the measuring apparatus, coupled to the system of interest, and make a continuous measurement of some observable of the apparatus. From this indirect continuous measurement we can obtain information about the behavior of the system. Then, we can eliminate the degrees of freedom of the apparatus and in this way we obtain some kind of direct measurement as in the general case. However, in general, we do not obtain an OVSP because the Markov property does no longer hold for the reduced system. Therefore to include in the description more and more pieces of the measuring apparatus brings us to different results and the question arises of where to cut this chain. The problem is to find suitable conditions under which the Markov property is (approximately) recovered also for the reduced description. Then to include or not to include a stage of the measuring apparatus for which such conditions hold gives essentially the same results and, therefore, it is at this stage that the cut can be consistently taken.

Now, "Markovian" limits for a reduced quantum dynamics are extensively studied in the literature.[13-21] Since in our theory of continuous observations measurements and dynamics are strictly linked, we expect that, under these limits, also the Markov property characterizing the OVSP's will be preserved. Indeed, we shall show that, under a suitable weak coupling limit, the Markov property for a reduced OVSP is recovered. Therefore, when this limit gives a good approximation, one can consistently ignore the details of the measuring apparatus and take into account only the system.

In this paper we do not try to attain full mathematical rigor. This is because too many technical details could obscure the main physical ideas, but, above all, because the results of this paper open more general mathematical problems. In Sec. VI we shall discuss how the topics treated here deserve a more general mathematical frame. Only in this more general perspective, mathematical rigor becomes essential.

The plan of the paper is as follows: after a short review of the continuous measurement formalism (Sec. II), we consider a quantum system coupled to an apparatus undergoing continuous measurement and eliminate the degrees of freedom of the apparatus (Sec. III); then, we take the weak coupling limit, in the form introduced in Ref. 18 (Sec. IV) and show that in this way an OVSP is reobtained (Sec. V). Some open problems are pointed out in the conclusions (Sec. VI).

II. OPERATION-VALUED STOCHASTIC PROCESSES

This section is devoted to a sketch of the formalism of continuous measurements. We limit ourselves to some definitions and general results; the physical justifications of the concepts introduced below can be found in the literature quoted.

Let D be the nuclear space of the n-component, real C^∞

 0022-2488/87/040818-09$02.50

functions on $\mathbb{R}$ with compact support and D' its topological dual space. We denote by x_k the distribution x ($x\in D'$) applied to the test function k ($k\in D$). We equip D' with the family of σ algebras $\{\Sigma_{(t_1,t_2)}, t_1<t_2, t_1,t_2\in\mathbb{R}\}$, where $\Sigma_{(t_1,t_2)}$ is the σ algebra generated by the sets (cylinder sets) of the form $\{x\in D': (x_{k^1},\ldots,x_{k^s})\in B\}$ (B a Borel subset of $\mathbb{R}^s$) with the condition that the supports of the test functions $\mathbf{k}^j$ are contained in the time interval (t_1,t_2). This is the framework of classical *generalized stochastic processes* (GSP's).[22] The elements of D' play the role of random variables.

Let now $\mathcal{H}$ be the Hilbert space of the system and $T(\mathcal{H})$ the space of trace-class operators on $\mathcal{H}$. We call (time-translation invariant) OVSP a family

$$\{\mathcal{F}(t_2,t_1;N),\ N\in\Sigma_{(t_1,t_2)},\ t_1<t_2,\ t_1,t_2\in\mathbb{R}\}$$

of linear maps from $T(\mathcal{H})$ into itself with the following properties.

(i) $\mathcal{F}(t_2,t_1;N)$ is completely[3] positive.

(ii) $\mathcal{F}(t_2,t_1;\cdot)$ is strongly σ additive on $\Sigma_{(t_1,t_2)}$.

(iii) $\mathcal{F}(t_2,t_1;D')$ is trace preserving (normalization).

(iv) The following composition law holds:

$$\mathcal{F}(t_3,t_2;M)\mathcal{F}(t_2,t_1;N)=\mathcal{F}(t_3,t_1;N\cap M)\,,$$
$$\forall N\in\Sigma_{(t_1,t_2)},\quad \forall M\in\Sigma_{(t_2,t_3)},\quad t_1<t_2<t_3\,. \tag{2.1}$$

(v) The time-translation invariance holds, i.e.,

$$\mathcal{F}(t_2,t_1;N)=\mathcal{F}(t_2+\bar t,t_1+\bar t;N_{\bar t}),\quad \forall \bar t\in\mathbb{R},\quad \forall N\in\Sigma_{(t_1,t_2)}\,, \tag{2.2}$$

where $N_{\bar t}\in\Sigma_{(t_1+\bar t,t_2+\bar t)}$ is obtained from N by the time translation in D', defined by duality from the time shift in D

$$\mathbf{k}\xrightarrow{\bar t}\mathbf{k}_{\bar t},\quad \mathbf{k}_{\bar t}(t)=\mathbf{k}(t-\bar t)\,. \tag{2.3}$$

When the system is prepared in the state W at time t_1, probabilities are given (for $N\in\Sigma_{(t_1,t_2)}$) by

$$P(N\,|\,W,t_1):=\mathrm{Tr}\{\mathcal{F}(t_2,t_1;N)W\}\,. \tag{2.4}$$

The triple $\{D',\Sigma_{(t_1,t_2)},P(\cdot\,|\,W,t_1)\}$ for given W,t_1,t_2, is a GSP in the sense of Gel'fand.[22] Equation (2.1) is the Markov property mentioned in the Introduction.

In the theory of GSP's it is useful to introduce the notion of *characteristic functional* (the functional Fourier transform of the probability measure).[22] Analogously, for the OVSP's it is useful to introduce the notion of *characteristic operator*,[4,6] defined by

$$\mathcal{G}(t_2,t_1;[\mathbf{k}]):=\int_{D'}e^{ix_k}\mathcal{F}(t_2,t_1;dx)\,, \tag{2.5}$$

where $\mathbf{k}(t)$ is a test function with support contained in (t_1,t_2). The quantity

$$L([\mathbf{k}]\,|\,W,t_0):=\mathrm{Tr}\{\mathcal{G}(\bar t,t_0;[\mathbf{k}])W\} \tag{2.6}$$

gives the characteristic functional of the associated GSP [in $(t_0,\bar t)$]; from Eq. (2.6) all the probabilities and, so, all the physically interesting quantities can be computed (see Ref. 4). Moreover, given the characteristic operator $\mathcal{G}(\cdots)$, it is possible to reconstruct the OVSP $\mathcal{F}(\cdots)$.

In Ref. 8 we have shown that a family

$$\{\mathcal{G}(t_2,t_1;[\mathbf{k}]),\ \mathbf{k}\in D,\ \mathrm{supp}(\mathbf{k})\subset(t_1,t_2),\ t_1<t_2,\ t_1,t_2\in\mathbb{R}\}$$

of bounded operators acting on $T(\mathcal{H})$ is the characteristic operator of an OVSP if and only if the following properties hold.

(a) $\mathcal{G}(t_2,t_1;[\mathbf{0}])$ is trace preserving (normalization).

(b) $\mathcal{G}(t_2,t_1;[\mathbf{k}])$ is strongly continuous in $\mathbf{k}$.

(c) $\mathcal{G}(\cdots)$ is completely positive definite, which means that the map of $T(\mathcal{H})$ into itself

$$\sum_{i,j=1}^{m}\alpha_i^*\mathcal{G}(t_2,t_1;[\mathbf{k}_i-\mathbf{k}_j])\alpha_j$$

is completely positive for any choice of the integer m, of the complex numbers α_i and of the test functions $\mathbf{k}_i$.

(d) The following composition law holds:

$$\mathcal{G}(t_3,t_2;[\mathbf{k}_2])\mathcal{G}(t_2,t_1;[\mathbf{k}_1])=\mathcal{G}(t_3,t_1;[\mathbf{k}_1+\mathbf{k}_2])\,, \tag{2.7}$$

for

$$\mathrm{supp}(\mathbf{k}_1)\subset(t_1,t_2),\quad \mathrm{supp}(\mathbf{k}_2)\subset(t_2,t_3),\quad t_1<t_2<t_3\,.$$

(e) Time-translation invariance holds, i.e.,

$$\mathcal{G}(t_2,t_1;[\mathbf{k}])=\mathcal{G}(t_2+\bar t,t_1+\bar t;[\mathbf{k}_{\bar t}]),\quad \forall\bar t\in\mathbb{R}\,. \tag{2.8}$$

Now, we want to construct a simple class of OVSP's. Consider a family of operators acting on $T(\mathcal{H})$ defined by the differential equation (derivatives are taken in the strong sense)

$$\frac{\partial}{\partial t}\mathcal{G}(t,t_0;[\mathbf{k}])=\mathcal{K}(\mathbf{k}(t))\mathcal{G}(t,t_0;[\mathbf{k}]),\quad k\in D\,, \tag{2.9}$$

with the initial condition

$$\mathcal{G}(t_0,t_0;[\mathbf{k}])=1\,, \tag{2.10}$$

where, for any $\mathbf{k}\in\mathbb{R}^n$, $\mathcal{K}(\mathbf{k})$ is a suitable operator, densely defined in $T(\mathcal{H})$; we call it the generator of $\mathcal{G}(\cdots)$. Then, properties (d) and (e) are satisfied and one can give sufficient conditions on $\mathcal{K}$ in order that also properties (a), (b), (c) hold.[23] However, it is difficult to find out the explicit form of the most general generator satisfying these conditions.

Note that the family of operators $\{\mathcal{G}(t),\ t\geqslant 0\}$, where

$$\mathcal{G}(t-t_0):=\mathcal{G}(t,t_0;[\mathbf{0}])\equiv\mathcal{F}(t,t_0;D')\,, \tag{2.11}$$

turns out to be a quantum dynamical semigroup.[13–19] Its meaning is to give the dynamics of the system, including the perturbation due to the measuring apparatus.[4] The operator $\mathcal{K}(\mathbf{0})$ is the generator of this semigroup. Only in the case of uniform continuity is the most general form of $\mathcal{K}(\mathbf{0})$ known.[24]

In Ref. 7, by computations involving quantum stochastic calculus,[25] a form for $\mathcal{K}(\mathbf{k})$ has been found, which is not the most general one, but which allows us to construct a physically interesting class of OVSP's. The structure of the generator $\mathcal{K}(\mathbf{k})$ given in Ref. 7 can be written in the following way:

$$\mathcal{K}(\mathbf{k})=\mathcal{L}+i\sum_{j=1}^{n}k_j\mathcal{R}_j-\frac{1}{4}\sum_{i,j=1}^{n}k_i\Gamma_{ij}k_j+\mathcal{K}_p(\mathbf{k})\,,$$
$$\Gamma_{ij}\in\mathbb{R}\,, \tag{2.12}$$

$$\mathcal{K}_p(\mathbf{k})X=\sum_{j=1}^{N}\left(S_j^\dagger XS_j\exp(i\mathbf{k}\cdot\boldsymbol{\alpha}_j)-\frac{1}{2}\{S_jS_j^\dagger,X\}\right),\quad \boldsymbol{\alpha}_j\in\mathbb{R}^n\,, \tag{2.13}$$

$$\mathscr{L}X = -\frac{i}{\hbar}[H,X] - \frac{1}{4}\sum_{r,s=1}^{m} D_{rs}[Q_r,[Q_s,X]] - \frac{i}{4}\sum_{r,s=1}^{m} E_{rs}[Q_r,\{Q_s,X\}], \quad D_{rs},E_{rs}\in\mathbf{R}, \tag{2.14}$$

$$\mathscr{R}_j X = \tfrac{1}{2}(R_j X + X R_j^\dagger), \tag{2.15}$$

$$R_j = \sum_{s=1}^{m} C_{js}Q_s + c_j I, \quad j=1,2,\ldots,n, \quad C_{js},c_j\in\mathbf{C}, \tag{2.16}$$

where $[A,B]$ denotes the commutator and $\{A,B\}$ the anticommutator, H, Q_s, and S_j are operators in $\mathscr{h}$, $H^\dagger = H$, $Q_s^\dagger = Q_s$, I is the identity operator on $\mathscr{h}$, and $\{I, Q_s, s=1,\ldots,m\}$ is a set of linearly independent operators. In Eqs. (2.12)–(2.16) three real matrices (Γ,D,E) and a complex one (C) are involved; Γ has dimensions $n\times n$, D and E have $m\times m$, and C has $n\times m$. Moreover, in order that property (c) of the characteristic operator holds,[4] we must have

$$\Gamma\geqslant 0, \quad \det\Gamma\neq 0, \tag{2.17}$$

$$D_{rs} = D_{sr}, \quad E_{rs} = -E_{sr}, \quad D+iE\geqslant 0, \tag{2.18}$$

$$D + iE - C^\dagger\Gamma^{-1}C\geqslant 0. \tag{2.19}$$

The operators involved in Eqs. (2.13)–(2.16) can be bounded or unbounded. Reference 7 deals with the case of bounded operators. Some rigorous results in the case of unbounded operators are given in Ref. 5, where Feynman integral techniques are used, and in Refs. 9 and 11, where linear systems are treated.

In Eq. (2.12) we have written the generator as the sum of a "Gaussian" (the first three terms) and a "Poissonian" part $[\mathscr{K}_p(\mathbf{k})]$. These names are due to analogies with the classical case. From a physical point of view a pure Gaussian OVSP is particularly suited for describing a continuous measurement of quantities like position and momentum of a particle[9–11] and a pure Poissonian OVSP for describing counting processes.[2,7]

The moments of the measured quantities can be obtained, when they exist, by functional differentiation of the characteristic functional (2.6) (Ref. 4). In the following we shall need the expressions for mean values and two-time correlation functions in the case of a pure Gaussian OVSP ($S_j\equiv 0$, $\forall j$). For the mean values (taking $t=0$ as initial time) we have

$$\langle x_k\rangle := \int_{y\in\mathbf{R}} yP\left(y\leqslant\int_0^{+\infty} dt\,\mathbf{x}(t)\cdot\mathbf{k}(t) < y+dy\,|\,W,0\right) = \sum_{i=1}^{n}\int_0^{+\infty} dt\,k_i(t)\langle x_i(t)\rangle, \tag{2.20}$$

where

$$\langle x_i(t)\rangle = -i\frac{\delta}{\delta k_i(t)}L([\mathbf{k}]\,|\,W,0)|_{\mathbf{k}=0} = \mathrm{Tr}\{\tfrac{1}{2}(R_i+R_i^\dagger)\,\mathscr{G}(t)W\}; \tag{2.21}$$

similarly, for the two-time correlation functions we obtain

$$\begin{aligned}\langle(x_i(t_1)-\langle x_i(t_1)\rangle)(x_j(t_2)-\langle x_j(t_2)\rangle)\rangle &= \tfrac{1}{2}\Gamma_{ij}\delta(t_1-t_2)+\tfrac{1}{4}\theta(t_1-t_2)\\ &\times\mathrm{Tr}((R_i+R_i^\dagger)\mathscr{G}(t_1-t_2)\mathscr{R}_j\mathscr{G}(t_2)W)\\ &+\tfrac{1}{4}\theta(t_2-t_1)\,\mathrm{Tr}((R_j+R_j^\dagger)\,\mathscr{G}(t_2-t_1)\mathscr{R}_i\mathscr{G}(t_1)W)\\ &-\langle x_i(t_1)\rangle\langle x_j(t_2)\rangle = :\Delta_{ij}(t_1,t_2) = \Delta_{ji}(t_2,t_1),\end{aligned} \tag{2.22}$$

where $\theta(t)$ is the usual step function. Note that, as the correlations come from a true probability measure, $\Delta_{ij}(t_1,t_2)$ is positive definite, i.e., for any choice of the complex functions $k_j(t)$, we have

$$\sum_{i,j=1}^{n}\int_0^{+\infty} dt_1\,dt_2\,k_i^*(t_1)\Delta_{ij}(t_1,t_2)k_j(t_2)\geqslant 0. \tag{2.23}$$

III. A SYSTEM COUPLED TO A CONTINUOUSLY OBSERVED MEASURING APPARATUS

Consider now a system S, in which we are interested (with Hilbert space $\mathscr{h}_S$), coupled to a continuously observed system M (with Hilbert space $\mathscr{h}_M$). We call M the measuring apparatus; in reality we can include in M, in addition to (a first stage of) the apparatus itself, also other things as thermal reservoirs and so on.

Let $\mathscr{L}_S$ be the generator of a quantum dynamical semigroup on $T(\mathscr{h}_S)$, giving the dynamics of S by itself. In general this dynamics can be dissipative or purely Hamiltonian. The system S interacts with M via a purely Hamiltonian coupling $\mathscr{L}_I$, which we write as

$$\mathscr{L}_I W = -(i/\hbar)[H_I,W], \tag{3.1a}$$

$$H_I = \sum_{q=1}^{N} A_q\otimes B_q, \tag{3.1b}$$

where $I_{\mathscr{h}_S}, A_q$, $q=1,\ldots,N$, are linearly independent self-adjoint operators in $\mathscr{h}_S$ and B_q, $q=1,\ldots,N$, self-adjoint operators in $\mathscr{h}_M$. System M undergoes a continuous measurement, which we assume, for simplicity, to be of pure Gaussian type. In the last section we shall discuss what happens when also a Poissonian type of measurement is considered. Therefore the continuous measurement on M (and its dynamics) is described by equations of the type of Eqs. (2.9)–(2.19) (with $\mathscr{K}_p\equiv 0$).

Summarizing, the continuous observation (and the dynamics of $S+M$) is described by an OVSP on $T(\mathscr{h}_S\otimes\mathscr{h}_M)$, whose characteristic operator satisfies the equations

$$\frac{\partial}{\partial t}\tilde{\mathscr{G}}(t,t_0;[\phi]) = \tilde{\mathscr{K}}(\phi(t))\tilde{\mathscr{G}}(t,t_0;[\phi]), \tag{3.2a}$$

$$\tilde{\mathscr{G}}(t_0,t_0;[\phi]) = 1. \tag{3.2b}$$

The generator $\tilde{\mathscr{K}}(\phi)$ is given by

$$\tilde{\mathscr{K}}(\phi) = \mathscr{L}_S\otimes\mathscr{J}_M + \mathscr{J}_S\otimes\mathscr{L}_M + \mathscr{L}_I + i\sum_{j=1}^{n}\phi_j\mathscr{J}_S\otimes\mathscr{R}_j - \tfrac{1}{4}\sum_{i,j=1}^{n}\phi_i\Gamma_{ij}\phi_j, \tag{3.3}$$

where (1) $\mathscr{J}_S$ and $\mathscr{J}_M$ are the identity operators on $T(\mathscr{h}_S)$ and $T(\mathscr{h}_M)$, respectively; (2) $\mathscr{L}_M$ is the generator of a quantum dynamical semigroup on $T(\mathscr{h}_M)$ with the struc-

ture given by Eqs. (2.14) and (2.18); (3) $\mathscr{R}_j$ is an operator in $T(\mathcal{h}_M)$ given by

$$\mathscr{R}_j X = \tfrac{1}{2}(R_j X + X R_j^\dagger) \tag{3.4}$$

[R_j is an operator in $\mathcal{h}_M$ with structure (2.16)]; and (4) Γ is an invertible positive real $n\times n$ matrix [see Eq. (2.17)] and Γ, C, D, E satisfy Eq. (2.19).

We assume that the dynamical semigroup $\exp(\mathscr{L}_M t)$ admits a stationary state ρ_M, i.e.,

$$\mathscr{L}_M \rho_M = 0, \quad \rho_M \in T(\mathcal{h}_M), \quad \rho_M \geqslant 0, \quad \mathrm{Tr}_M\{\rho_M\} = 1, \tag{3.5}$$

and, moreover, that

$$\mathrm{Tr}_M\{R_j \rho_M\} = 0, \quad j = 1,\ldots,n, \tag{3.6}$$

$$\mathrm{Tr}_M\{B_q \rho_M\} = 0, \quad q = 1,\ldots,N. \tag{3.7}$$

Equations (3.6) and (3.7) are not restrictions. If Eq. (3.6) does not hold, we can make the replacement $R_j \to R_j' = R_j - \mathrm{Tr}_M\{R_j\rho_M\}$; this amounts to shifting the stochastic variables by a constant. If Eq. (3.7) does not hold, without changing $\mathscr{K}(\phi)$, we can make the replacements

$$B_q \to B_q' = B_q - \mathrm{Tr}_M\{B_q \rho_M\}$$

and

$$\mathscr{L}_S \to \mathscr{L}_S' = \mathscr{L}_S - \frac{i}{\hbar}\left[\sum_{q=1}^{N} (\mathrm{Tr}_M\{B_q\rho_M\}) A_q, \cdot\right].$$

Let us prepare at time t_0 the measuring apparatus in the stationary state ρ_M (not correlated with the state of system S), so that at time t_0 we have the initial state

$$W = \rho \otimes \rho_M, \quad \rho \in T(\mathcal{h}_S), \quad \rho \geqslant 0, \quad \mathrm{Tr}_S\{\rho\} = 1. \tag{3.8}$$

Then, we introduce the projection operators $\mathscr{P}_i$, $i = 0,1$, on $T(\mathcal{h}_S \otimes \mathcal{h}_M)$ defined by

$$\mathscr{P}_0 X = (\mathrm{Tr}_M X) \otimes \rho_M, \quad \forall X \in T(\mathcal{h}_S \otimes \mathcal{h}_M), \tag{3.9a}$$

and

$$\mathscr{P}_1 = 1 - \mathscr{P}_0. \tag{3.9b}$$

Note that the initial state (3.8) satisfies the equation

$$\mathscr{P}_0 W = W. \tag{3.10}$$

Because we are interested only in the trace of $\mathscr{G}(\cdots)$ applied to W [cf. Eq. (2.6) and the following comments], we can first eliminate the degrees of freedom of the measuring apparatus by taking the partial trace over $\mathcal{h}_M$ and then we can take the trace over $\mathcal{h}_S$. Let us define for any operator $\mathscr{A}$ acting on $T(\mathcal{h}_S \otimes \mathcal{h}_M)$,

$$\mathscr{A}_{(ij)} := \mathscr{P}_i \mathscr{A} \mathscr{P}_j. \tag{3.11}$$

Taking into account that the initial state satisfies Eq. (3.10), we have that all the probabilities can be obtained from the *reduced characteristic operator* $\mathscr{G}_{(00)}(\cdots)$.

Using the well-known projection operator technique, we derive from Eq. (3.2a) a system of two coupled differential equations for $\mathscr{G}_{(00)}(\cdots)$ and $\mathscr{G}_{(10)}(\cdots)$; by eliminating $\mathscr{G}_{(10)}(\cdots)$, we obtain for $\mathscr{G}_{(00)}(\cdots)$ a kind of "generalized master equation,"

$$\begin{aligned}\frac{\partial}{\partial t}\mathscr{G}_{(00)}(t,t_0;[\phi]) &= \mathscr{K}_{(00)}(\phi(t))\mathscr{G}_{(00)}(t,t_0;[\phi]) \\ &+ \int_{t_0}^{t} dt'\, \mathscr{K}_{(01)}(\phi(t))\mathscr{U}(t,t';[\phi]) \\ &\times \mathscr{K}_{(10)}(\phi(t'))\mathscr{G}_{(00)}(t',t_0;[\phi]),\end{aligned} \tag{3.12}$$

where $\mathscr{U}(t,t';[\phi])$ is the solution of the differential equation

$$\frac{\partial}{\partial t}\mathscr{U}(t,t';[\phi]) = \mathscr{K}_{(11)}(\phi(t))\mathscr{U}(t,t';[\phi]), \tag{3.13a}$$

with the initial condition

$$\mathscr{U}(t',t';[\phi]) = 1. \tag{3.13b}$$

Now, Eq. (3.12) does not imply Eq. (2.7) for $\mathscr{G}_{(00)}(\cdots)$ and, therefore, the reduced characteristic operator does not give rise to an OVSP for system S alone: in the general case the structure of the measuring apparatus and the details of the interaction between S and M have an essential role and M cannot be eliminated in an effective way. As we shall see in the following section, Eq. (3.12) reduces, under a suitable limit, to a Markovian master equation and, so, Eq. (2.7) holds. Only in this case we obtain a reduced OVSP for system S itself.

By Eqs. (3.5)–(3.7) we have

$$(\mathscr{L}_S \otimes \mathscr{J}_M)\mathscr{P}_i = \mathscr{P}_i(\mathscr{L}_S \otimes \mathscr{J}_M), \tag{3.14a}$$

$$\begin{aligned}&(\mathscr{J}_S \otimes \mathscr{L}_M)\mathscr{P}_i = \mathscr{P}_i(\mathscr{J}_S \oplus \mathscr{L}_M), \\ &(\mathscr{J}_S \otimes \mathscr{L}_M)\mathscr{P}_0 = 0,\end{aligned} \tag{3.14b}$$

$$(\mathscr{J}_S \otimes \mathscr{R}_j)_{(00)} = 0, \quad \mathscr{L}_{I(00)} = 0; \tag{3.14c}$$

therefore the components of the generator $\mathscr{K}(\phi)$ are given by

$$\mathscr{K}_{(00)}(\phi) = (\mathscr{L}_S \otimes \mathscr{J}_M)\mathscr{P}_0 - \frac{1}{4}\sum_{i,j=1}^{n} \phi_i \Gamma_{ij} \phi_j \mathscr{P}_0, \tag{3.15a}$$

$$\mathscr{K}_{(01)}(\phi) = \mathscr{P}_0\left(\mathscr{L}_I + i\sum_{j=1}^{n}\phi_j(\mathscr{J}_S \otimes \mathscr{R}_j)\right), \tag{3.15b}$$

$$\mathscr{K}_{(10)}(\phi) = \left(\mathscr{L}_I + i\sum_{j=1}^{n}\phi_j(\mathscr{J}_s \otimes \mathscr{R}_j)\right)\mathscr{P}_0, \tag{3.15c}$$

$$\begin{aligned}\mathscr{K}_{(11)}(\phi) &= (\mathscr{L}_S \otimes \mathscr{J}_M)\mathscr{P}_1 + \mathscr{J}_S \otimes \mathscr{L}_M + \mathscr{L}_{I(11)} \\ &+ i\sum_{j=1}^{n}\phi_j(\mathscr{J}_S \otimes \mathscr{R}_j)_{(11)} \\ &- \frac{1}{4}\sum_{i,j=1}^{n}\phi_i\Gamma_{ij}\phi_j\mathscr{P}_1.\end{aligned} \tag{3.15d}$$

Finally, let us rewrite Eq. (3.12) in the equivalent integral form (that will be useful in the following):

$$\tilde{\mathcal{G}}_{(00)}(t,t_0;[\phi]) - \mathcal{P}_0$$
$$= \int_{t_0}^{t} dt' \int_{t'}^{t} dt'' \exp\left\{(t-t'')\mathcal{L}_S \otimes \mathcal{J}_M - \frac{1}{4}\sum_{i,j=1}^{n}\int_{t'}^{t} dt_1\, \phi_i(t_1)\Gamma_{ij}\phi_j(t_1)\right\}\mathcal{P}_0\left(\mathcal{L}_I + i\sum_{j=1}^{n}\phi_j(t'')\mathcal{J}_S\otimes\mathcal{R}_j\right)$$
$$\times T\left\{\exp\int_{t'}^{t''} dt_2\left(\mathcal{J}_S\otimes\mathcal{L}_M + \mathcal{L}_S\otimes\mathcal{J}_M + \mathcal{L}_{I(11)} + i\sum_{j=1}^{n}\phi_j(t_2)(\mathcal{J}_S\otimes\mathcal{R}_j)_{(11)}\right)\right\}$$
$$\times\left(\mathcal{L}_I + i\sum_{j=1}^{n}\phi_j(t')\mathcal{J}_S\otimes\mathcal{R}_j\right)\mathcal{P}_0\,\tilde{\mathcal{G}}_{(00)}(t',t_0;[\phi])\,, \tag{3.16}$$

where T means the chronologically ordered product and we have used expressions (3.15).

IV. THE WEAK COUPLING LIMIT

In this section we want to see whether, under a suitable limit, Eq. (3.12) reduces to a Markovian equation of the type (2.9), so that a true OVSP is recovered for system S alone.

Note that $\tilde{\mathcal{G}}(t,t_0;[\mathbf{0}])$ gives the dynamics of the total system $S+M$ and $\tilde{\mathcal{G}}_{(00)}(t,t_0;[\mathbf{0}])$ the reduced dynamics of S alone [cf. Eq. (2.11)]. Therefore our first problem is to find a limiting situation in which $\tilde{\mathcal{G}}_{(00)}(t,t_0;[\mathbf{0}])$ becomes a Markovian dynamics.

Consider the case of weak coupling between S and M. We know that the usual second-order perturbation theory applied to a reduced dynamics [as that given by Eq. (3.16) for $\phi(t)=\mathbf{0}$] in general does not preserve positivity and, so, does not give rise to a *bona fide* limiting dynamics (see, for instance, Ref. 19). Only if a scaling parameter λ is introduced and the limit $\lambda\to 0$ taken, complete positivity is preserved. A scaling procedure is needed also in our case in order that the positivity property (c) of Sec. II remains valid and a true reduced OVSP is obtained in the limit.

Two schemes for the weak coupling limit are known and both can be adapted to our case: that of Davies[13] and that of Palmer.[18] The first one treats the free motion of S on the order one and the dissipation on the order λ^2; in this case the existence of the limit $\lambda\to 0$ needs very restrictive hypotheses on $\mathcal{L}_S$ ($\mathcal{L}_S = -(i/\hbar)[H_S,\cdot]$, H_S with only discrete spectrum). The second one uses a different scaling for which the free motion of S and the dissipation are of the same order of magnitude; no special hypothesis on $\mathcal{L}_S$ is necessary and more general forms for the limiting dynamics are obtained (as in the case of the singular coupling limit[14–16]). For these reasons we choose the second type of scaling, which is characterized by

$$\mathcal{L}_I \to \lambda\mathcal{L}_I\,, \tag{4.1}$$
$$\mathcal{L}_S \to \lambda^2\mathcal{L}_S\,, \tag{4.2}$$
$$\tau = \lambda^2 t\,. \tag{4.3}$$

The introduction of a rescaled time τ is necessary for compensating the vanishing of the coupling (4.1).

In our case, in which also measurements are considered, we have to rescale also the continuously measured quantities, otherwise in the limit $\lambda\to 0$ all information on S is lost. If we denote by $\mathbf{x}(t)$ the old stochastic variables associated with the test functions $\phi(t)$, we have to introduce new stochastic variables $\mathbf{z}(\tau)$ and test functions $\mathbf{k}(\tau)$ in such a way that

$$\int d\tau\,\mathbf{z}(\tau)\cdot\mathbf{k}(\tau) = \int dt\,\mathbf{x}(t)\cdot\boldsymbol{\phi}(t)\,. \tag{4.4}$$

Therefore we must take

$$\mathbf{z}(\tau) = \lambda^{\alpha-1}\mathbf{x}(\tau/\lambda^2)\,, \tag{4.5a}$$
$$\mathbf{k}(\tau) = \lambda^{-(\alpha+1)}\boldsymbol{\phi}(\tau/\lambda^2)\,, \tag{4.5b}$$

where α is a real scaling exponent to be determined.

If we introduce Eqs. (4.1)–(4.3) and (4.5b) into Eq. (3.16) and set

$$G^{(\lambda)}_{(00)}(\tau,\tau_0;[\mathbf{k}]) := \tilde{\mathcal{G}}^{(\lambda)}_{(00)}(\tau/\lambda^2,\tau_0/\lambda^2;[\lambda^{\alpha+1}\mathbf{k}(\lambda^2 t)])\,, \tag{4.6}$$

after some simple calculations we obtain

$$G^{(\lambda)}_{(00)}(\tau,\tau_0;[\mathbf{k}]) - \mathcal{P}_0$$
$$= \int_{\tau_0}^{\tau} d\tau' \exp\left\{(\mathcal{L}_S\otimes\mathcal{J}_M)(\tau-\tau') - \frac{\lambda^{2\alpha}}{4}\sum_{i,j=1}^{n}\int_{\tau'}^{\tau} d\tau''\, k_i(\tau'')\Gamma_{ij}k_j(\tau'')\right\}$$
$$\times\int_0^{(\tau-\tau')/\lambda^2} du\,\mathcal{N}_{(\lambda)}(u;\tau';[\mathbf{k}])G^{(\lambda)}_{(00)}(\tau',\tau_0;[\mathbf{k}])\,, \tag{4.7}$$

where

$$\mathcal{N}_{(\lambda)}(u;\tau';[\mathbf{k}]) = \mathcal{P}_0\left(\mathcal{L}_I(\lambda^2 u) + i\lambda^{\alpha}\sum_{j=1}^{n}k_j(\tau'+\lambda^2 u)\,(\mathcal{J}_S\otimes\mathcal{R}_j)\right)$$
$$\times T\left\{\exp\int_0^u du'\left(\mathcal{J}_S\otimes\mathcal{L}_M + \lambda\mathcal{L}_{I(11)}(\lambda^2 u') + i\lambda^{\alpha+1}\sum_{j=1}^{n}k_j(\tau'+\lambda^2 u')(\mathcal{J}_S\otimes\mathcal{R}_j)_{(11)}\right)\right\}$$
$$\times\left(\mathcal{L}_I + i\lambda^{\alpha}\sum_{j=1}^{n}k_j(\tau')(\mathcal{J}_S\otimes\mathcal{R}_j)\right)\mathcal{P}_0 \tag{4.8}$$

and

$$\mathcal{L}_I(t) = \exp\{-(\mathcal{L}_S\otimes\mathcal{J}_M)t\}\mathcal{L}_I\exp\{(\mathcal{L}_S\otimes\mathcal{J}_M)t\}\,. \tag{4.9}$$

From these equations it is apparent that, if we want the limit $\lambda\to 0$ to exist and the continuous measurement on M to give

information about S, we must take

$$\alpha = 0 . \tag{4.10}$$

Consider now the formal expansion of the operator $\mathcal{N}_{(\lambda)}(\cdots)$,

$$\begin{aligned}\mathcal{N}_{(\lambda)}(u;\tau';[\mathbf{k}]) = \mathcal{P}_0\Big[\mathcal{L}_I(\lambda^2 u) + i\sum_{j=1}^{n} k_j(\tau'+\lambda^2 u)\mathcal{J}_S\otimes\mathcal{R}_j\Big]\\ \times\Big\{e^{(\mathcal{J}_S\otimes\mathcal{L}_M)u} + \sum_{r=1}^{\infty}\lambda^r\int_0^u du_1\int_0^{u_1} du_2\cdots\int_0^{u_{r-1}} du_r\, e^{(\mathcal{J}_S\otimes\mathcal{L}_M)(u-u_1)}\\ \times\Big[\mathcal{L}_{I(11)}(\lambda^2 u_1) + i\sum_{j=1}^{n} k_j(\tau'+\lambda^2 u_1)(\mathcal{J}_S\otimes\mathcal{R}_j)_{(11)}\Big]\times\cdots\\ \times e^{(\mathcal{J}_S\otimes\mathcal{L}_M)(u_{r-1}-u_r)}\Big[\mathcal{L}_{I(11)}(\lambda^2 u_r) + i\sum_{j=1}^{n} k_j(\tau'+\lambda^2 u_r)(\mathcal{J}_S\otimes\mathcal{R}_j)_{(11)}\Big]e^{(\mathcal{J}_S\otimes\mathcal{L}_M)u_r}\Big\}\\ \times\Big[\mathcal{L}_I + i\sum_{j=1}^{n} k_j(\tau')\,\mathcal{J}_S\otimes\mathcal{R}_j\Big]\mathcal{P}_0 .\end{aligned} \tag{4.11}$$

If one takes into account the explicit expressions for $\mathcal{L}_I$, $\mathcal{R}_j$, $\mathcal{P}_i$, one sees that in the integrals they appear some kind of multitime correlation functions of M, which can be written as

$$\mathrm{Tr}_M\{\mathcal{C}_{j_1}\exp[\mathcal{L}_M u_1]\mathcal{C}_{j_2}\exp[\mathcal{L}_M u_2]\cdots\mathcal{C}_{j_s}\exp[\mathcal{L}_M u_s]\rho_M\},$$

where $\mathcal{C}_j := [B_j,\cdot]$, for $j = 1,\ldots,N$, $\mathcal{C}_j := \{B_{j-N},\cdot\}$ for $j\neq N+1,\ldots,2N$, $\mathcal{C}_j := \mathcal{R}_{j-2N}$, for $j = 2N+1,\ldots,2N+n$. If these quantities decay "sufficiently fast" in time, all the terms but the first in $\mathcal{N}_{(\lambda)}$ vanish for $\lambda\to 0$ (see Ref. 16 for more precise statements in a similar situation). In this case we obtain the Markovian master equation (in integral form)

$$\begin{aligned}G_{(00)}^{(\lambda\to 0)}(\tau,\tau_0;[\mathbf{k}]) - \mathcal{P}_0\\ = \int_{\tau_0}^{\tau} d\tau'\exp\Big\{(\mathcal{L}_S\otimes\mathcal{J}_M)(\tau-\tau') - \frac{1}{4}\sum_{i,j=1}^{n}\int_{\tau'}^{\tau} d\tau''\, k_i(\tau'')\Gamma_{ij}k_j(\tau'')\Big\}\int_0^{+\infty} du\,\mathcal{P}_0\Big[\mathcal{L}_I + i\sum_{j=1}^{n} k_j(\tau')\mathcal{J}_S\otimes\mathcal{R}_j\Big]\\ \times e^{(\mathcal{J}_S\otimes\mathcal{L}_M)u}\Big[\mathcal{L}_I + i\sum_{j=1}^{n} k_j(\tau')\mathcal{J}_s\otimes\mathcal{R}_j\Big]\mathcal{P}_0 G_{(00)}^{(\lambda\to 0)}(\tau',\tau_0;[\mathbf{k}]) .\end{aligned} \tag{4.12}$$

Finally, if we set

$$\begin{aligned}\mathcal{G}(\tau,\tau_0;[\mathbf{k}])\rho &:= \lim_{\lambda\to 0}\mathrm{Tr}_M(G_{(00)}^{(\lambda)}(\tau,\tau_0;[\mathbf{k}])\rho\otimes\rho_M)\\ &= \lim_{\lambda\to 0}\mathrm{Tr}_M\{\tilde{\mathcal{G}}^{(\lambda)}(\tau/\lambda^2,\tau_0/\lambda^2;[\lambda\,\mathbf{k}(\lambda^2 t)])\rho\otimes\rho_M\},\quad \forall\rho\in T(\mathcal{H}_S) ,\end{aligned} \tag{4.13}$$

from Eq. (4.12) we have (changing τ in t) that the reduced characteristic operator satisfies the differential equation

$$\frac{\partial}{\partial t}\mathcal{G}(t,t_0;[\mathbf{k}]) = \mathcal{K}(\mathbf{k}(t))\mathcal{G}(t,t_0;[\mathbf{k}]) , \tag{4.14}$$

with the initial condition

$$\mathcal{G}(t_0,t_0;[\mathbf{k}]) = 1 ; \tag{4.15}$$

the generator $\mathcal{K}(\mathbf{k})$ is given by

$$\begin{aligned}\mathcal{K}(\mathbf{k})\rho = \mathcal{L}_S\rho - \frac{1}{4}\sum_{i,j=1}^{n} k_i\Gamma_{ij}k_j\rho\\ + \int_0^{+\infty} du\,\mathrm{Tr}_M\Big\{\Big[\mathcal{L}_I + i\sum_{j=1}^{n} k_j\mathcal{J}_S\otimes\mathcal{R}_j\Big]\\ \times e^{(\mathcal{J}_S\otimes\mathcal{L}_M)u}\Big[\mathcal{L}_I + i\sum_{i=1}^{n} k_i\mathcal{J}_S\otimes\mathcal{R}_i\Big]\rho\otimes\rho_M\Big\} .\end{aligned} \tag{4.16}$$

V. THE STRUCTURE OF THE REDUCED CHARACTERISTIC OPERATOR

In this section we want to prove that the generator (4.16) has the structure (2.12)–(2.19) with $\mathcal{K}_p\equiv 0$, i.e., it is of pure Gaussian type.

In order to write down the explicit expression of $\mathcal{K}(\mathbf{k})$ in a compact form, let us introduce some more definitions. First, for any operator $\mathcal{B}$ acting on $T(\mathcal{H}_M)$, we define its transpose $\mathcal{B}^T$ acting on $B(\mathcal{H}_M)$ (Banach space of bounded operators on $\mathcal{H}_M$) by

$$\mathrm{Tr}_M[(\mathcal{B}^T Y)X] = \mathrm{Tr}_M[Y(\mathcal{B}X)] ,$$

$$\forall X\in T(\mathcal{H}_M),\quad \forall Y\in B(\mathcal{H}_M) . \tag{5.1}$$

Then, we introduce the pre-inner product in $B(\mathcal{H}_M)$

$$\langle A,B\rangle := \mathrm{Tr}_M(A^\dagger B\rho_M) . \tag{5.2}$$

The completion of $B(\mathcal{H}_M)$ with respect to this inner product is a Hilbert space, denoted by $L^2(\mathcal{H}_M;\rho_M)$ (see Ref. 26, pp. 81–87, and Ref. 17, p. 165). In $L^2(\mathcal{H}_M;\rho_M)$ we define a time-translation operator by

$$\mathcal{T}(t) := \exp(\mathcal{L}_M^T t),\quad t\geq 0 . \tag{5.3}$$

It can be shown that $\mathscr{T}(t)$ is a contraction and, therefore, it can be extended to the whole $L^2(\mathcal{h}_M;\rho_M)$; then, $\{\mathscr{T}(t), t\geqslant 0\}$ is a strongly continuous contraction semigroup on $L^2(\mathcal{h}_M;\rho_M)$ (Ref. 17, pp. 165–166). For negative times we define

$$\mathscr{T}(t) := \mathscr{T}(-t)^*, \quad t<0, \tag{5.4}$$

where the star denotes the adjoint with respect to the inner product (5.2). Note that we have

$$(\langle A,\mathscr{T}(t)B\rangle)^* = \langle B,\mathscr{T}(-t)A\rangle, \quad \forall t\in\mathbb{R}. \tag{5.5}$$

Recalling Eqs. (3.1), (3.4), (3.9) and using the above definitions, we can write the generator $\mathscr{K}(\mathbf{k})$ of the limiting characteristic operator $\mathscr{G}(\cdots)$ [see Eqs. (4.14)–(4.16)] in the following way:

$$\mathscr{K}(\mathbf{k})\rho = \mathscr{L}\rho + \frac{i}{2}\sum_{j=1}^{n}\sum_{q=1}^{N} k_j (C_{jq}^{(S)} A_q\rho + C_{jq}^{(S)*}\rho A_q) - \frac{1}{4}\sum_{i,j=1}^{n} k_i \Gamma_{ij}^{(S)} k_j \rho, \tag{5.6}$$

where

$$\mathscr{L}\rho = \mathscr{L}_S\rho - \frac{i}{\hbar}[H_1,\rho] - \frac{1}{4}\sum_{q,p=1}^{N} D_{qp}^{(S)}[A_q,[A_p,\rho]] - \frac{i}{4}\sum_{q,p=1}^{N} E_{qp}^{(S)}[A_q,\{A_p,\rho\}], \tag{5.7}$$

$$H_1 = \frac{i}{4\hbar}\sum_{q,p=1}^{N}\left(\int_0^{+\infty} dt - \int_{-\infty}^{0} dt\right) \times \langle B_q,\mathscr{T}(t)B_p\rangle A_q A_p + \text{H.c.}, \tag{5.8}$$

$$D_{qp}^{(S)} = \frac{1}{\hbar^2}\int_{-\infty}^{+\infty} dt\,(\langle B_p,\mathscr{T}(t)B_q\rangle + \langle B_q,\mathscr{T}(t)B_p\rangle), \tag{5.9a}$$

$$E_{qp}^{(S)} = \frac{i}{\hbar^2}\int_{-\infty}^{+\infty} dt\,(\langle B_p,\mathscr{T}(t)B_q\rangle - \langle B_q,\mathscr{T}(t)B_p\rangle), \tag{5.9b}$$

$$C_{jq}^{(S)} = -\frac{i}{\hbar}\left\{\int_{-\infty}^{+\infty} dt\,\langle R_j,\mathscr{T}(t)B_q\rangle + \int_{-\infty}^{0} dt\,(\langle R_j^\dagger,\mathscr{T}(t)B_q\rangle + \langle B_q,\mathscr{T}(t)R_j\rangle)\right\}, \tag{5.10}$$

$$\Gamma_{ij}^{(S)} = \Gamma_{ij} + \frac{1}{2}\int_{-\infty}^{+\infty} dt\,(\langle R_i,\mathscr{T}(t)R_j\rangle + \langle R_j,\mathscr{T}(t)R_i\rangle) + \frac{1}{2}\int_0^{+\infty} dt\,(\langle R_i,\mathscr{T}(t)R_j^\dagger\rangle + \langle R_j,\mathscr{T}(t)R_i^\dagger\rangle) + \frac{1}{2}\int_{-\infty}^{0} dt\,(\langle R_j^\dagger,\mathscr{T}(t)R_i\rangle + \langle R_i^\dagger,\mathscr{T}(t)R_j\rangle). \tag{5.11}$$

Note that the operator $\mathscr{G}(\cdots)$ satisfies Eqs. (2.9), (2.10) and its generator equations (2.12) (with $\mathscr{K}_p\equiv 0$), (2.15), and (2.16). In order that $\mathscr{G}(\cdots)$ be the characteristic operator of a true OVSP, we must also prove that $\Gamma^{(S)}$ is an invertible positive real $n\times n$ matrix, $\mathscr{L}$ is the generator of a quantum dynamical semigroup and

$$D^{(S)} + iE^{(S)} - C^{(S)\dagger}\Gamma^{(S)-1}C^{(S)} \geqslant 0. \tag{5.12}$$

In writing Eq. (5.12) we have not taken into account the contribution of the (possible) dissipative part of $\mathscr{L}_S$, because $\mathscr{G}(\cdots)$ is a characteristic operator for any choice of $\mathscr{L}_S$ and in particular when one has a purely Hamiltonian generator $\mathscr{L}_S = -(i/\hbar)[H_S,\cdot]$.

From definitions (5.9a), (5.9b), and (5.11) and from Eq. (5.5), we see immediately that we have

$$\Gamma_{ij}^{(S)} = \Gamma_{ji}^{(S)}\in\mathbb{R}, \quad D_{qp}^{(S)} = D_{pq}^{(S)}\in\mathbb{R}, \quad E_{qp}^{(S)} = -E_{pq}^{(S)}\in\mathbb{R}. \tag{5.13}$$

Then, consider the correlation functions (2.22) for the continuous observation on system M, when $\mathscr{L}_I\equiv 0$ and $W\equiv\rho_M$ [so that we have $\mathscr{G}(t) = \exp(\mathscr{L}_M t)$, $\exp(\mathscr{L}_M t)\rho_M = \rho_M$, $\langle x_j(t)\rangle = 0$]. We obtain

$$\begin{aligned}\Delta_{ij}(t_1,t_2) = {}& \tfrac{1}{2}\Gamma_{ij}\delta(t_1-t_2) + \tfrac{1}{4}\theta(t_1-t_2)\{\langle R_i^\dagger,\mathscr{T}(t_2-t_1)R_j\rangle \\ & + \langle R_j,\mathscr{T}(t_1-t_2)R_i\rangle + \langle R_i,\mathscr{T}(t_2-t_1)R_j\rangle + \langle R_j,\mathscr{T}(t_1-t_2)R_i^\dagger\rangle\} \\ & + \tfrac{1}{4}\theta(t_2-t_1)\{\langle R_j^\dagger,\mathscr{T}(t_1-t_2)R_i\rangle + \langle R_i,\mathscr{T}(t_2-t_1)R_j\rangle \\ & + \langle R_j,\mathscr{T}(t_1-t_2)R_i\rangle + \langle R_i,\mathscr{T}(t_2-t_1)R_j^\dagger\rangle\}.\end{aligned} \tag{5.14}$$

By taking in Eq. (2.23)

$$k_j(t) = c_j\sqrt{2\alpha}\,e^{-\alpha t}, \quad c_j\in\mathbb{C}, \quad \alpha>0, \tag{5.15}$$

we have

$$0 \leqslant \sum_{i,j=1}^{n} c_i^* c_j\left\{\frac{1}{2}\Gamma_{ij} + \frac{1}{4}\int_{-\infty}^{+\infty} dt\,e^{-\alpha|t|}\,[\langle R_i,\mathscr{T}(t)R_j\rangle + \langle R_j,\mathscr{T}(t)R_i\rangle] + \frac{1}{4}\int_0^{+\infty} dt\,e^{-\alpha t}\,[\langle R_j,\mathscr{T}(t)R_i^\dagger\rangle + \langle R_i,\mathscr{T}(t)R_j^\dagger\rangle] + \frac{1}{4}\int_{-\infty}^{0} dt\,e^{-\alpha|t|}\,[\langle R_i^\dagger,\mathscr{T}(t)R_j\rangle + \langle R_j^\dagger,\mathscr{T}(t)R_i\rangle]\right\}; \tag{5.16}$$

and taking the limit $\alpha \to 0+$, we obtain

$$\Gamma^{(S)} \geqslant 0 . \tag{5.17}$$

Equation (5.12), as it stands, requires an invertible $\Gamma^{(S)}$, but, if it is not so, we can easily modify this equation in order to take into account this fact. For simplicity, we do not treat this case here and assume that $\Gamma^{(S)}$ is invertible.

Now we have only to prove the positivity property (c) of Sec. II. But the quantity

$$\sum_{i,j=1}^{m} \alpha_i^* \alpha_j \mathcal{G}(\tau_2,\tau_1;[\mathbf{k}_i - \mathbf{k}_j]) = \sum_{i,j=1}^{M} \alpha_i^* \alpha_j \lim_{\lambda \to 0} \mathrm{Tr}_M(\mathcal{G}^{(\lambda)}(\tau_2/\lambda^2,\tau_1/\lambda^2; [\lambda(\mathbf{k}_i(\lambda^2 t) - \mathbf{k}_j(\lambda^2 t))](\cdot) \otimes \rho_M)) \tag{5.18}$$

is completely positive, because complete positivity is not destroyed under any limiting procedure.[17,24] In particular, by taking $\mathbf{k}_i = \mathbf{0}$, we obtain that $\mathcal{L}$ is the generator of a quantum dynamical semigroup and this is true, for any choice of $\mathcal{L}_S$, if and only if

$$D^{(S)} + iE^{(S)} \geqslant 0 . \tag{5.19}$$

Finally, from the complete positivity of (5.19), for $\mathbf{k}_i \neq \mathbf{0}$, have also that Eq. (5.12) holds [in Ref. 4 we have proved that Eq. (2.19) is a necessary condition for the positivity of an OVSP with a generator of pure Gaussian structure].

It would be interesting to have independent direct proofs of Eqs. (5.12) and (5.19). However, I am able to give this proof only for Eq. (5.19). A very formal proof of this equation goes as follows. Since $\{\mathcal{T}(t),\ t \geqslant 0\}$ is a strongly continuous contraction semigroup on a Hilbert space, extended to negative times by Eq. (5.4), we can write

$$\mathcal{T}(t) = \exp\{-i\mathcal{A}t - \mathcal{B}|t|\}, \tag{5.20}$$

with

$$\mathcal{A} = \mathcal{A}^*, \quad \mathcal{B} = \mathcal{B}^* \geqslant 0 . \tag{5.21}$$

Then we have

$$D_{qp}^{(S)} + iE_{qp}^{(S)} \equiv \frac{1}{\hbar^2} \int_{-\infty}^{+\infty} dt \langle B_q, \mathcal{T}(t) B_p \rangle = (2/\hbar^2) \langle (-i\mathcal{A} + \mathcal{B})^{-1} B_q , \mathcal{B}(-i\mathcal{A} + \mathcal{B})^{-1} B_p \rangle ; \tag{5.22}$$

from the positivity of $\mathcal{B}$, Eq. (5.19) follows. More rigorously, Eq. (5.19) is a direct consequence of Theorem 3.8 of Ref. 27.

VI. FINAL REMARKS

As we have seen, if we consider a continuous measurement on M, of pure Gaussian type, the reduced OVSP gives, in the weak coupling limit, a continuous measurement on S of Gaussian type. The same result holds even for more general OVSP's. Consider a continuous measurement on M of the type discussed in Sec. II, with a generator with nonvanishing Poisson part. On eliminating the degrees of freedom of M and under the scaling discussed in Sec. IV, it could be shown that also in this case a pure Gaussian reduced OVSP is obtained. This result is not surprising if one thinks of the classical analog. Let us try to explain this point.

If we consider the definitions of Sec. II in the classical case ($\mathcal{K} \equiv \mathbb{C}$), we see that our OVSP's reduce to GSP's with independent values at any time.[22] Among these processes Eq. (2.9) selects those GSP's which are the derivative of ordinary Markov processes with independent increments. As is well known, these processes are strictly related to infinitely divisible probability distributions. The classical Lévy–Kintchine formula gives the most general distribution of this type.[28] An important subclass of infinitely divisible probabilities is that of *stable distributions*, which possess a domain of attraction.[28] Under suitable scalings, the probability distribution of the sum of identically distributed quantities goes to one of the stable distributions.

Also in the quantum case it is possible to give a definition of infinite divisibility. In Ref. 23 the notion of *infinitely divisible instrument* is introduced, and it is shown that the problem of finding the most general infinitely divisible instrument is equivalent to the problem of finding the most general OVSP, whose characteristic operator satisfies Eq. (2.9). Moreover, the scaling procedure used in the present paper is very reminiscent of that used in the usual central limit theorem (probability distributions with finite first and second moments are in the domain of attraction of the normal distribution). Therefore it is not surprising that only OVSP's of Gaussian type be obtained in the limit. Let us stress that the limiting procedure discussed in Sec. IV is more complicated than the classical one, because OVSP's involve not only the probabilities, but also the quantum dynamics.

The above considerations suggest the following open problems in the quantum case: (1) to find some quantum analog of the Lévy–Khintchine formula for the infinitely divisible instruments; (2) to give a physically motivated and mathematically well-posed definition of stable instrument and to find the most general one; (3) to introduce the notion of domain of attraction for stable instruments; and (4) to formulate suitable central limit theorems in the quantum case, in such a way that the results of this paper could be regarded as a particular case of these theorems. None of these problems is trivial. Only in the case of point (1) some partial results are known.[8,23]

[1] G. Ludwing, *Foundations of Quantum Mechanics* (Springer, Berlin, 1982).
[2] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[3] K. Kraus, *States, Effects, and Operations, Lecture Notes in Physics*, Vol. 190 (Springer, Berlin, 1983).
[4] A. Barchielli, L. Lanz, and G. M. Prosperi, Found. Phys. **13**, 779 (1983).
[5] G. Lupieri, J. Math. Phys. **24**, 2329 (1983).
[6] A. Barchielli, L. Lanz, and G. M. Prosperi, in *Proceedings of the International Symposium on Foundations of Quantum Mechanics*, edited by S. Kamefuchi, H. Ezawa, Y. Murayama, M. Namiki, S. Namura, Y. Ohnuki, and T. Yajima (Phys. Soc. Jpn., Tokyo, 1984), p. 165.
[7] A. Barchielli and G. Lupieri, J. Math. Phys. **26**, 2222 (1985).
[8] A. Barchielli, in *Stochastic Processes in Classical and Quantum Systems*, edited by S. Albeverio, G. Casati, and D. Merlini, *Lecture Notes in Physics*, Vol. 262 (Springer, Berlin, 1986), p. 14.
[9] A. Barchielli, Nuovo Cimento B **74**, 113 (1983).
[10] A. S. Holevo, Teor. Mat. Fiz. **57**, 424 (1983) [Theor. Math. Phys. **57**, 1234 (1984)].
[11] A. Barchielli, Phys. Rev. D **32**, 347 (1985).
[12] M. Ozawa, J. Math. Phys. **25**, 79 (1984).
[13] E. B. Davies, Commun. Math. Phys. **39**, 91 (1974).

[14]V. Gorini and A. Kossakowski, J. Math. Phys. **17**, 1298 (1976).
[15]A. Frigerio and V. Gorini, J. Math. Phys. **17**, 2123 (1976).
[16]A. Frigerio, C. Novellone, and M. Verri, Rep. Math. Phys. **12**, 279 (1977).
[17]V. Gorini, A. Frigerio, M. Verri, A. Kossakowski, and E. C. G. Sudarshan, Rep. Math. Phys. **13**, 149 (1978).
[18]P. F. Palmer, J. Math. Phys. **18**, 527 (1977).
[19]H. Spohn, Rev. Mod. Phys. **53**, 569 (1980).
[20]R. Dümcke, J. Math. Phys. **24**, 311 (1983).
[21]R. Dümcke, in *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes*, edited by L. Accardi, A. Frigerio, and V. Gorini, *Lecture Notes in Mathematics*, Vol. 1055 (Springer, Berlin, 1984), p. 113.
[22]I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. 4, *Applications of Harmonic Analysis* (Academic, New York, 1964).
[23]A. S. Holevo, Theory Probab. Its Appl. (USSR) **31**, 560 (1986); **32**, 142 (1987).
[24]G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).
[25]R. L. Hudson and K. R. Parthasarathy, Commun. Math. Phys. **93**, 301 (1984).
[26]A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[27]D. E. Evans and J. T. Lewis, Commun. Dublin Inst. Adv. Stud. A **24**, 1 (1977).
[28]W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1971), Vol. II.

Algebra representations on eigenfunctions of the Rosen–Morse potential

M. J. Englefield
Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia

(Received 26 August 1986; accepted for publication 3 December 1986)

The bound state eigenfunctions of the Rosen–Morse [N. Rosen and P. M. Morse, Phys. Rev. **42**, 210 (1932)] potential are investigated using ladder operators that give a representation of the algebra A_1. The representations are generally infinite-dimensional and indecomposable, the representation space containing unbounded functions as well as the normalizable eigenfunctions. Operators giving eigenfunctions of a potential with different strength are also found, giving a representation of D_2. The A_1 (or su_2) representations are identified in terms of a classification proposed by Sannikov (S. S. Sannikov, Yad. Fiz. **6**, 1294 (1967) [Sov. J. Nucl. Phys. **6**, 939 (1968)]), and the connection with representations obtained by Gruber and Klimyk [B. Gruber and A. U. Klimyk, J. Math Phys. **19**, 2009 (1978); **25**, 755 (1984)] is given. The evaluation of matrix elements is considered.

I. INTRODUCTION

In the early years of quantum mechanics several examples of the Schrödinger equation, including the Rosen–Morse[1] potential $-c\,\mathrm{sech}^2 x$, were solved using known properties of special functions. The harmonic oscillator problem and the hydrogen atom problem[2] were also solved by algebraic methods (matrix mechanics) which were equivalent to determining the properties of a representation of a Lie algebra associated with the system. More recently, relevant algebras have been discovered for several other cases where the Schrödinger equation is analytically soluble, allowing unsuspected algebraic treatments.[3–10] In general the required eigenfunctions are a basis for a representation of the Lie algebra, which contains (ladder) operators which transform one eigenfunction into another. In some cases it is easiest to use a potential algebra where the operators generate an eigenfunction belonging to a system with a different potential strength. For the Rosen–Morse potential such operators were given by Alhassid, Gursey, and Iachello[8] and these results have been extended by Frank and Wolf[10] to the modified Pöschl–Teller potential, which contains the Rosen–Morse potential as a special case.

Interest in the potential $-c\,\mathrm{sech}^2 x$ is partly due to the fact that when the strength constant $c = n^2 + n$, with n integer, the reflection coefficient is zero at all energies.[11] This quantum-mechanical curiosity became famous as an exactly soluble inverse scattering problem because the reflectionless potentials were the soliton solutions[12] of the Korteweg–de Vries equation. The potential algebra has been used[8] to relate the reflection coefficients of different potentials; if $c = \nu(\nu+1)$, then ν changes in integer steps.

This paper considers first the algebra where the ladder operators generate eigenfunctions of the same potential.[8,13,14] For $c = n^2 + n$, these eigenfunctions are associated Legendre functions[15] or spherical harmonics with a change of variable, so the algebra is evidently su_2. The usual angular momentum operators can be used[8] provided the eigenfunctions are multiplied by a phase factor $e^{im\phi}$, introducing a redundant variable ϕ exactly as in the work of Armstrong[3] and Berrondo and Palmer.[7] The operators retain their commutation properties when the integer n is replaced by an arbitrary ν.

The su_2 algebra is extended in Sec. III by considering not only the operators corresponding to an angular momentum $\mathbf{L} = \mathbf{r}\times\mathbf{p}$, but also those corresponding to $\mathbf{p}$. This gives a new algebra based on six operators with the commutation relations of $so_{3,1}$ or D_2; the potential algebra[8] can be constructed as quadratic functions of these D_2 operators.

The main result of this paper, in Sec. IV, is the identification of the representations of the su_2 algebra, by obtaining the functions annihilated by the ladder operators. If 2ν is not an integer, the ladder operators generate an infinite sequence of unbounded functions; the square-integrable eigenfunctions span only a small part of the representation space. When 2ν is an odd integer, the infinite-dimensional representation is reducible but indecomposable. A quotient space may be used to get the familiar finite-dimensional representation, but half of this space consists of functions that are not normalizable. For integer ν there are finite-dimensional representation spaces which include a zero-energy wave function which is not normalizable; however, it is bounded, and adjoint operators may be defined by using an inner product integral with the weight function $\mathrm{sech}^2 x$, as discussed in Sec. V. Infinite-dimensional, indecomposable representations of su_2 were classified and labeled by Sannikov[16]; each representation found here is identified with one given by Sannikov; in some cases they are isomorphic to those given more recently by Gruber and Klimyk.[17,18] The representations of D_2 have not been identified with any given previously.

The concluding section of the paper obtains some matrix elements with the aid of algebraic methods.

II. SHIFT OPERATORS CHANGING THE ENERGY

Consider the Schrödinger equation for a Rosen–Morse[1] potential:

$$H\psi = -\frac{d^2\psi}{dx^2} - \frac{\nu(\nu+1)}{\cosh^2 x}\psi = -k^2\psi \quad (\nu>0). \tag{1}$$

When $k = \nu$ the function

$$\mathrm{sech}^{\nu} x \tag{2}$$

 0022-2488/87/040827-06$02.50

is a square-integrable solution which (having no zeros) represents the ground state. The potential constant $\nu(\nu+1)$ in (1) is invariant under the substitution $\nu \to -\nu-1$. The same substitution in (2) shows that

$$\chi(x) = \cosh^{\nu+1} x \tag{3}$$

is a solution of (1) when $k = \nu+1$. This solution is not square-integrable.

The possibility of a group-theoretical approach to (1) is indicated by the[15] fact that the substitution $\tanh x = \cos\theta$ gives the associated Legendre equation; conversely a solution of (1), at least for integer values of ν, may be obtained from the spherical harmonic $Y_\nu^m(\theta,\phi)$ after deleting the factor $e^{im\phi}$. The usual angular momentum operators become

$$\begin{aligned} L_0 &= -i\frac{\partial}{\partial\phi}, \\ L_\pm &= e^{\pm i\phi}\left(\mp\cosh x\frac{\partial}{\partial x} + i\sinh x\frac{\partial}{\partial\phi}\right), \end{aligned} \tag{4}$$

which satisfy the commutation relations

$$[L_0, L_\pm] = \pm L_\pm, \quad [L_+, L_-] = 2L_0. \tag{5}$$

The Casimir operator is

$$L^2 = -\cosh^2 x\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial\phi^2}\right) \tag{6}$$

and it is easy to verify that $\psi(x)$ satisfying (1) is equivalent to

$$\begin{aligned} L_0\psi e^{\pm ik\phi} &= \pm k\psi e^{\pm ik\phi}, \\ L^2\psi e^{\pm ik\phi} &= \nu(\nu+1)\psi e^{\pm ik\phi}. \end{aligned} \tag{7}$$

For example, for any $\nu > 0$,

$$f^{-\nu} = e^{-i\nu\phi}\operatorname{sech}^\nu x \text{ and } f^{\nu+1} = e^{i(\nu+1)\phi}\cosh^{\nu+1} x \tag{8}$$

are solutions of (7) corresponding to (2) and (3).

Thus, if a solution of (1) is multiplied by $e^{-ik\phi}$, further solutions may be generated using the shift operators $L_\pm$. For example, since

$$L_+ f^{-\nu} = 2\nu e^{-i(\nu-1)\phi}\sinh x\operatorname{sech}^\nu x \tag{9}$$

the function $\sinh x\operatorname{sech}^\nu x$ satisfies (1) when $k = \nu-1$. The functions in (8) give zero on applying L_-. Since $L_\pm\psi(x)e^{ik\phi} = 0$ reduces to a first-order differential equation for ψ, for a given k the functions which are annihilated by the shift operators are unique:

$$L_- f^k = 0, \quad L_+ f_k = L_+(e^{ik\phi}\operatorname{sech}^k x) = 0. \tag{10}$$

[Note: alternative equations could be obtained by interchanging ϕ with $-\phi$ and L_+ with L_-, but this effectively only duplicates the results. One consequence of the choice of $e^{\pm ik\phi}$ factors made in (8) is that when ν is noninteger, square-integrable functions of x always appear with a factor $e^{-ik\phi}$ with $k > 0$.]

One difference from angular momentum theory is that in general k (in $e^{ik\phi}$) is not an integer, as $\nu-k$ is an integer. Also functions of x generated by $L_\pm$ may be unbounded; for example, the function obtained on the right-hand side of (9) is square-integrable only if $\nu > 1$, whereas $f^{-\nu}$ is square integrable for $\nu > 0$. More generally, as $x \to \infty$,

$$\begin{aligned} L_+^n f^{-\nu} &\sim \left[\frac{1}{2}e^{x+i\phi}\left(-\frac{\partial}{\partial x} + i\frac{\partial}{\partial\phi}\right)\right]^n 2^\nu e^{-\nu x - i\nu\phi} \\ &= 2^\nu\nu(\nu-1)\cdots(\nu-n+1)e^{kx+ik\phi} \\ &\quad (k = n-\nu). \end{aligned} \tag{11}$$

Since $f^{-\nu}$ is even and L_+ changes the parity, a factor $(-1)^n$ in (11) gives the behavior as $x \to -\infty$. Thus $L_+^n f^{-\nu}$ is square-integrable when $k < 0$, i.e., $n < \nu$, unbounded if $n > \nu$ when ν is nonintegral, and bounded but not square-integrable when $n = \nu$ in the integer case. This gives the following square-integrable eigenfunctions of (1): $L_+^n f^{-\nu}$ $(n = 0,1,2,\ldots; n < \nu)$ belonging to the eigenvalues $-(\nu-n)^2$. The eigenfunctions have the same parity as n. Later, in Eq. (24), a normalization factor will be given.

When ν is not an integer, further $(n > \nu)$ applications of L_+ generate unbounded solutions of (1). When ν is an integer, $L_+^\nu f^{-\nu}$ is a bounded zero-energy wave function having the same parity as ν. The functions $L_+^{\nu+m}f^{-\nu}$ and $L_+^{\nu-m}f^{-\nu}$ have the same parity, and are both solutions of (1) with $k^2 = m^2$ $(k = \pm m)$. So (apart from the factors $e^{\pm im\phi}$) one function must be a multiple of the other, and the $L_+^{\nu+m}f^{-\nu}$ $(m = 1,2,\ldots,\nu)$ just give the square-integrable eigenfunctions again, the sequence terminating as in (10).

Example $(\nu = 1)$:

$$\begin{aligned} &f^{-1} = e^{-i\phi}\operatorname{sech} x, \\ &L_+ f^{-1} = 2\tanh x \quad \text{(zero-energy wave function)}, \\ &L_+^2 f^{-1} = -2e^{i\phi}\operatorname{sech} x = -2f_1 \quad [\text{cf. (10)}], \\ &L_+ f_1 = 0. \end{aligned}$$

When ν is a half-integer, $L_+^m f^{-\nu}$ and $L_+^{2\nu-m}f^{-\nu}$ $(m < \nu)$ are also both solutions of the same equation (1) with $k^2 = (m-\nu)^2$, but they are opposite parity (as $L_+^{2\nu-m} = L_+^{2\nu-2m}L_+^m$ and $2\nu-2m$ is odd). Thus for $m < \nu$ the $L_+^{2\nu-m}f^{-\nu}$ are the second (unbounded) solutions of (1).

Example $(\nu = \frac{1}{2})$:

$$\begin{aligned} f^{-1/2} &= e^{-i\phi/2}\operatorname{sech}^{1/2} x \\ &\to e^{i\phi/2}\operatorname{sech}^{1/2} x\sinh x \quad \text{(second solution)} \\ &\to -e^{3i\phi/2}\cosh^{3/2} x = -f^{3/2} \end{aligned}$$

(and further applications of L_+ produce more unbounded functions.)

Although this adaptation of angular momentum theory has yielded the eigenvalues of H, other familiar techniques require the shift operators $L_\pm$ to be adjoint. This can be achieved only by using an unusual inner product, as discussed later in Sec. V.

III. OPERATORS CHANGING THE POTENTIAL

The operators (4) were obtained by applying the transformation $\tanh x = \cos\theta$ to the (angular momentum) operators which are the components of $\mathbf{L}(\theta,\phi) = -\mathbf{r}\times\nabla$. The same transformation can be applied to the components of $\nabla = \mathbf{i}\,\partial/\partial X + \mathbf{j}\,\partial/\partial Y + \mathbf{k}\,\partial/\partial Z$, using the polar coordinates θ and ϕ, but eliminating the radial coordinate r by setting $r = 1$ and $\partial/\partial r = 0$. For example,

$$\frac{\partial}{\partial Z} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta} \to -\sin\theta\frac{\partial}{\partial\theta} = \frac{\partial}{\partial x}.$$

This produces (with a factor $-i$ to get standard commutators)

$$U_0 = -i\frac{\partial}{\partial x}, \quad U_\pm = e^{\pm i\phi}\left(i\sinh x\frac{\partial}{\partial x} \pm \cosh x\frac{\partial}{\partial \phi}\right) \tag{12}$$

giving

$$[L_\pm, U_\pm] = 0, \quad [L_\pm, U_0] = \mp U_\pm,$$
$$[L_\pm, U_\mp] = \pm 2U_0, \quad [L_0, U_0] = 0, \tag{13}$$

$$[L_0, U_\pm] = \pm U_\pm . \tag{14}$$

These equations are the commutation relations of a vector operator, so selection rules on matrix elements are expected. Those rules which arise because factors $e^{ik\phi}$ produce $\delta_{k,k'}$ terms are irrelevant for matrix elements (of momentum U_0, for example) between solutions of (1). However, one can conclude that $U_\pm$ may change ν by ± 1, shifting to a different potential. Also (14) and (12) show that $U_\pm$ raise and lower the energy eigenvalue and do not change parity. If ψ_k^ν denotes a square-integrable solution of (1), then ($k > 0$)

$$U_\pm \psi_k^\nu e^{-ik\phi} = (a_\pm \psi_{k\pm 1}^{\nu} + b_\pm \psi_{k\pm 1}^{\nu\pm 1})e^{-i(k\mp 1)\phi}. \tag{15}$$

The first term does not appear when U_- acts on the lowest two states:

$$U_- f^{-\nu} = i\nu f^{-\nu-1}, \tag{16}$$
$$U_-(L_+ f^{-\nu}) = [i\nu^2/(\nu+1)](L_+ f^{-\nu-1}).$$

On a lattice of points representing solutions of (7), the action of the operators $L_\pm$ and $U_\pm$ can be indicated by arrows as in Fig. 1.

It is well known[15] that the operators

$$B = \frac{d}{dx} + \nu\tanh x, \quad C = -\frac{d}{dx} + (\nu+1)\tanh x, \tag{17}$$

obtained from the Darboux transformation, generate eigenfunctions of a different potential without changing the energy:

$$B\psi_k^\nu = b\psi_k^{\nu-1}, \quad C\psi_k^\nu = c\psi_k^{\nu+1}. \tag{18}$$

Figure 1 suggests that $L_\mp U_\pm$ should be some combinations of B and C. From (4), (6), and (12) one gets

$$\pm iL_\pm U_\mp = -(\tanh x)L^2 + (1\mp L_0)\frac{\partial}{\partial x}. \tag{19}$$

Applying each side of (19) to $\psi_{\nu-n}^\nu e^{i(n-\nu)\phi} = L_+^n f^{-\nu}$, the combinations of B and C required to give $L_\mp U_\pm$ can be determined by inspection:

$$i(2\nu+1)L_+U_- = (\nu+1)(1-n)B - \nu(2\nu+2-n)C,$$
$$i(2\nu+1)L_-U_+ = (\nu+1)(2\nu-n-1)B + \nu(n+2)C. \tag{20}$$

A normalization constant for the function $L_+^n f^{-\nu}$ may be obtained by considering the construction of the same function by successive applications of C to $f^{-\nu+n}$. Note that C in (17) depends on ν, which increases by 1 with each application, so at this stage C is rewritten $C(\nu)$. The normalization integral for $f^{-\nu+n}$ can be evaluated using the β function to get

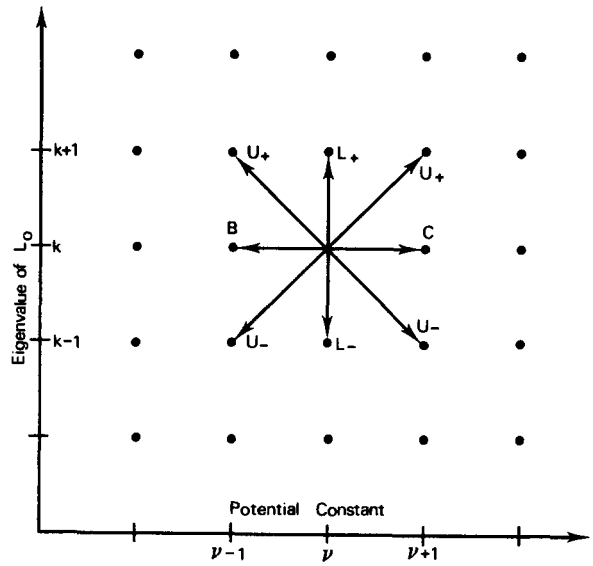


FIG. 1. The energy eigenvalue $-k^2$ is lowered by L_- when $k<0$.

$$\int_{-\infty}^{\infty} \mathrm{sech}^{2\lambda} x\,dx = \frac{\Gamma(\lambda)\sqrt{\pi}}{\Gamma(\frac{1}{2}+\lambda)} = \frac{[2^\lambda\Gamma(\lambda)]^2}{2\Gamma(2\lambda)}. \tag{21}$$

Hence the normalized function for the ground state is

$$[\Gamma(\tfrac{1}{2}+\nu-n)/\Gamma(\tfrac{1}{2})\Gamma(\nu-n)]^{1/2}\,\mathrm{sech}^{\nu-n}x. \tag{22}$$

The theory of the Darboux transformation gives[19] both the change in normalization when C is applied, and the change in the asymptotic form of the function. Applying to (22)

$$C(\nu-n-1-j)/[(2\nu-2n+j)j]^{1/2}$$

with $j = 1,2,\ldots,n$, retains the normalization, and gives a function with asymptotic behavior (as $x\to\infty$)

$$\left[\frac{\Gamma(2\nu-n+1)}{n!(\nu-n)}\right]^{1/2}\frac{1}{\Gamma(\nu-n)}e^{-(\nu-n)x}. \tag{23}$$

Comparing (23) with (11) shows that, provided $n<\nu$ ($k = n-\nu < 0$)

$$|\nu\, n-\nu> = \left[\frac{(\nu-n)\Gamma(2\nu-n+1)}{n!}\right]^{1/2} \times\frac{1}{2^\nu\Gamma(\nu+1)}L_+^n f^{-\nu} \tag{24}$$

is normalized.

The commutators between components of U are

$$[U_0, U_\pm] = \mp L_\pm, \quad [U_+, U_-] = -2L_0. \tag{25}$$

Equations (5), (13), (14), and (25) exhibit a basis for an so(3,1) algebra. Its invariants are

$$L^2 - U^2 = 0, \quad \mathbf{U}\cdot\mathbf{L} = 0. \tag{26}$$

Since neither $L_\pm$ nor $U_\pm$ are complex conjugates, the algebra should be regarded as the complex algebra D_2 rather than the real algebra so(3,1), and the "angular momentum" subalgebra is the complex algebra A_1.

IV. IDENTIFICATION OF REPRESENTATIONS

The nonunitary representations of the complex algebra A_1 were classified by Sannikov[16] according to the number of

basis functions which are annihilated by the shift operators $L_\pm$. These functions have been called extremal vectors by Gruber.[17] Using (4), the partial differential equations $(L_\pm)u=0$ have solutions of the form $u(x,\phi)=f(e^{\mp i\phi}\cosh x)$. Assuming that the Casimir operator

$$L^2=L_-L_+ +L_0^2+L_0=L_+L_- +L_0^2-L_0 \quad (27)$$

has the eigenvalue $\nu^2+\nu$ then determines f, giving two two-dimensional subspaces of extremal vectors:

$$L_+(\alpha f_\nu+\beta f_{-\nu-1})=0 \quad \text{[cf. (10)]}, \quad (28)$$

$$L_-(\alpha f^{-\nu}+\beta f^{\nu+1})=0 \quad \text{[cf. (8)]}. \quad (29)$$

The notation shows the eigenvalue of L_0 as a superscript or subscript.

Now consider the representation space containing $f^{-\nu}$. If

$$g_\nu=-e^{i\nu\phi}\,\mathrm{sech}^\nu x\int\cosh^{2\nu}x\,dx \quad (30)$$

then $L_+g_{-\nu-1}=f^{-\nu}$, and the representation space is spanned by the eigenfunctions (of L_0) $L_+^n f^{-\nu}$, $L_-^n g_{-\nu-1}$. The latter functions belong to the eigenvalues $-\nu-1-n$ of L_0 and so cannot be in the subspaces (28) or (29). The function $L_+^n f^{-\nu}$ belongs to the eigenvalue $n-\nu$; it can be in (28) if $2\nu=n$ (an integer), and it can be in (29) if $2\nu+1=n$. However, the extremal vectors are all even parity, and L_+ changes the parity, so the integer n must be even.

The extremal vectors are therefore as follows.

(i) If ν is neither integer nor half-integer, then $f^{-\nu}$ is the only extremal vector in the representation space.

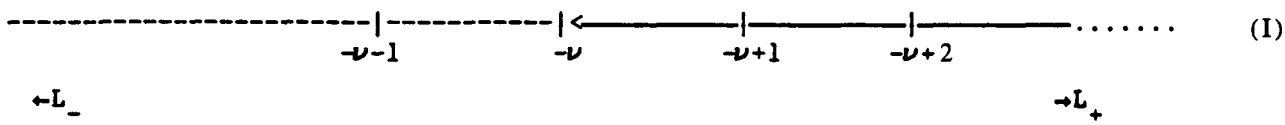
(ii) If ν is an integer, there is one other extremal vector $L_+^{2\nu}f^{-\nu}=(\text{const})f_\nu$, annihilated by L_+.

(iii) If ν is half-integer (2ν is an odd integer), there is one other extremal vector (in addition to $f^{-\nu}$)

$$L_+^{2\nu+1}f^{-\nu}=(\text{const})f^{\nu+1} \quad (31)$$

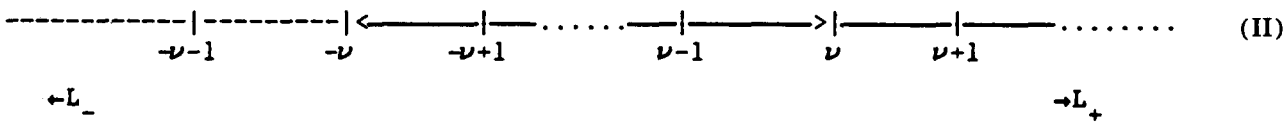
also annihilated by L_-. Note that (31) follows from the fact that the left-hand side is known to satisfy (7) with $k=\nu+1$, and so has the form $\psi(x)e^{ik\phi}$, where ψ is even and satisfies (1) with $k=\nu+1$. Thus ψ is a multiple of (3). A similar argument applies in (ii).

The representations can be illustrated by diagrams on which L_+ shifts to the right, L_- shifts to the left, and arrows indicate the extremal vectors. The numbers show the eigenvalues of L_0, i.e., the weights.



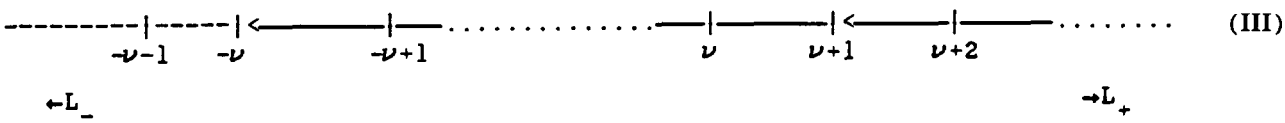
(I)

Diagram (I) illustrates the case where 2ν is not integer, the complete space realizing the representation $D^+(-\nu-1,-\nu)$, using Sannikov's notation.[16] That part of the space spanned by the $L_-^n g_{-\nu-1}$ [dashed line in Diagram (I)] can be omitted in favor of the invariant subspace spanned by the $L_+^n f^{-\nu}$, realizing the representation $D^+(\nu)$. The representation $D^+(-\nu-1,-\nu)$ is thus reducible but indecomposable. There is a $[\nu]$-dimensional subspace of square-integrable functions, but this is not invariant.



(II)

Diagram (II) illustrates the case where ν is an integer, the complete space realizing Sannikov's $D_2(\nu,0)$. The invariant subspace spanned by the $L_+^n f^{-\nu}$ realizes Sannikov's $D_2^+(\nu)$. This is also reducible but indecomposable, and contains the (usual) finite-dimensional representation $D(\nu)$. This representation space has a 2ν-dimensional subspace of square-integrable functions [each bound state of (1) appears twice with different factors $e^{\pm ik\phi}$], but also contains the zero-energy wave function $L_+^\nu f^{-\nu}$ (see Sec. II). One could also realize Sannikov's $D_2^-(\nu)$ by omitting the $L_+^n f^{-\nu}$ with $n>2\nu$, and retaining the $L_-^n g_{-\nu-1}$. (This is an undesirable alternative if $e^{-ik\phi}$ with $k>0$ is going to indicate a square-integrable function.)



(III)

Diagram (III) illustrates the case where 2ν is an odd integer, the complete space realizing Sannikov's $D^+(\nu,\frac{1}{2})$. Omitting the $L_-^n g_{-\nu-1}$ gives an invariant subspace spanned by the $L_+^n f^{-\nu}$, realizing $D_1^+(\nu)$. This is also reducible but indecomposable: the invariant subspace spanned by the $L_+^{2\nu+n}f^{-\nu}$ ($n=1,2,\ldots$) realizes $D^+(-\nu-1)$. The quotient space

$$\{L_+^n f^{-\nu}:\ n=0,1,\ldots\}/\{L_+^n f^{-\nu}:\ n=2\nu+1,2\nu+2,\ldots\}$$

realizes the (usual) finite-dimensional representation $D(\nu)$.

This is formally obtained by taking only the functions $L_+^n f^{-\nu}$ $(n=0,1,...,2\nu)$ as a basis, and including in L_+ a projection operator P off the denominator space, so that $PL_+(L_+^{2\nu} f^{-\nu}) = Pf^{\nu+1} = 0$. The representation space has a $(\nu+\frac{1}{2})$-dimensional noninvariant subspace of normalizable functions. For every half-integer $k \leqslant \nu$, both the even and the odd solution of (1) appear in the basis for $D^+(\nu,\frac{1}{2})$.

When 2ν is not an odd integer, representation spaces are also obtained by taking the second solutions of (1) of the opposite parity as basis functions. Then all functions in the representation space are unbounded. The details can be obtained from the work given above by the substitutions

$$L_+ \leftrightarrow L_-, \quad \phi \to -\phi, \quad \nu \to -\nu - 1. \tag{32}$$

For integer ν Sannikov's $D_1(\nu,0) \supset D^{\pm}(-\nu-1)$ is realized, while $D_1^{\pm}(\nu)$ and $D(\nu)$ can be obtained from quotient spaces. If (32) is applied to the $D^+(\nu,\frac{1}{2})$ representation shown in Diagram (III), then Sannikov's $D^-(\nu,\frac{1}{2}) \supset D_1^-(\nu) \supset D^-(-\nu-1)$ is realized.

Another discussion of indecomposable representations was given by Gruber and Klimyk.[17] The representation in Diagram (II) realizes the τ_Λ given in their Fig. 6, with $\Lambda = -\nu - 1$; the representation in Diagram (III) realizes the τ_Λ given in their Fig. 7, with $\Lambda = \nu$. Similarly the representations using unbounded second solutions of (1) realize their σ_Λ. In their later paper[18] realizations of $D_2^+(-\Lambda-1)$ and $D^+(-\Lambda-1)$ are given in Eq. (10.1), and a realization of $D_2(\Lambda,h)$ given in Eq. (10.4) ($h=0$ or $\frac{1}{2}$).

Gruber[20] has also given indecomposable representations of the algebra D_2, defining extremal vectors as those annihilated either by both U_+ and L_+ or by both U_- and L_-. For a given k, the unique functions annihilated by $U_\pm$ are

$$U_\pm (\sinh^{\mp k} x) e^{ik\phi} = 0. \tag{33}$$

Equations (10) and (33) show that there are no such vectors satisfying Gruber's definition in the representation spaces used here, which are therefore inequivalent to any of the representations considered by Gruber.

V. MATRIX ELEMENTS

When ν is an integer, the functions (24) can be expressed in terms of spherical harmonics by comparing (24) with the analogous equation from angular momentum theory:

$$Y_l^{-m} = \left[\frac{(l+m)!}{(2l)!(l-m)!}\right]^{1/2} L_+^{l-m} Y_l^{-l}.$$

Apart from the factor $e^{i(n-l)\phi}$, and with $\cos\theta = \tanh x$,

$$[(2l-2n)/(2l+1)]^{1/2} Y_l^{-l+n}(\theta,\phi) \quad (n=0,1,...,l-1) \tag{34}$$

are thus the normalized eigenfunctions ψ_{l-n}^l for the potential $-l(l+1)\operatorname{sech}^2 x$. Known matrix elements for spherical harmonics should therefore translate into results for the Rosen–Morse system, subject to two provisos. First, there is no result from a matrix element that vanishes from the integration over ϕ. Second, because $-\sin\theta\, d\theta = \operatorname{sech}^2 x\, dx$, the angular momentum matrix elements of any operator V will correspond to the Rosen–Morse matrix elements of the operator $(\operatorname{sech}^2 x)V$.

The simplest example uses the normalization condition for the spherical harmonics:

$$1 = \int_0^\pi |Y_l^{n-l}|^2 \sin\theta\, d\theta$$

$$= \frac{2l+1}{2l-2n} \int_{-\infty}^{\infty} |\psi_{l-n}^l|^2 \operatorname{sech}^2 x\, dx. \tag{35}$$

The expectation value for the potential energy $-l(l+1)\operatorname{sech}^2 x$ is therefore $-l(l+1)(2l-2n)/(2l+1)$; since the energy eigenvalue is $-(l-n)^2$, the expectation value for the kinetic energy is $(l-n)(2nl+n+l)/(2l+1)$. The orthogonality condition for the spherical harmonics Y_l^m and Y_L^M gives no result for the Rosen–Morse system unless $m=M$, when it shows that a potential matrix element is zero if evaluated between eigenstates with the same energy but for potentials differing by an integer change in ν:

$$\int_{-\infty}^{\infty} \bar{\psi}_{l-n}^l \psi_{l-n}^{l+k} \operatorname{sech}^2 x\, dx = 0 \quad (k>0, \quad n<l).$$

Although this argument only demonstrates results when ν is an integer l, a proof of a spherical harmonics property for integer l will usually apply to any Rosen–Morse states by using the integral in (35) to define their inner product. Denote this by parentheses, and use an asterisk for the corresponding complex conjugation of operators:

$$(\nu\, n-\nu|\mu\, m-\mu)$$

$$= (2\pi)^{-1} \int_{-\pi}^{\pi} e^{i\phi(m-\mu-n+\nu)} d\phi$$

$$\times \int_{-\infty}^{\infty} \bar{\psi}_{\nu-n}^{\nu} \psi_{\mu-m}^{\mu} \operatorname{sech}^2 x\, dx, \tag{36}$$

$$\left(\frac{\partial}{\partial\phi}\right)^* = -\frac{\partial}{\partial\phi}, \quad L_+^* = L_-, \quad L_-^* = L_+. \tag{37}$$

To get (37) from (36) one must assume that $\nu-\mu$ is an integer.

Using (22) and (21), the expectation value of $\operatorname{sech}^2 x$ in the ground state is

$$(\nu\ -\nu|\nu\ -\nu) = \nu/(\nu+\tfrac{1}{2}). \tag{38}$$

A recurrence formula for the expectation value in other states can be derived using (36), (37), and (27),

$$v_n = (L_+^n f^{-\nu}, L_+^n f^{-\nu})$$

$$= (L_+^{n-1} f^{-\nu}, L_+^* L_+ L_+^{n-1} f^{-\nu})$$

$$= n(2\nu - n + 1) v_{n-1}.$$

Incorporating the normalization factor given in (24) shows that $(\nu\, n-\nu|\nu\, n-\nu)/(\nu-n)$ is independent of n, and (38) gives the value $1/(\nu+\frac{1}{2})$. Thus the above results on expectation values of potential energy and kinetic energy extend from integer l to arbitrary ν.

Results can also be deduced from the matrix elements of $L_\pm$. From (24), if $n<\nu$,

$$L_+|\nu\, n-\nu\rangle$$

$$= [(\nu-n)(n+1)(2\nu-n)/(\nu-n-1)]^{1/2}$$

$$\times |\nu\, n-\nu+1\rangle, \tag{39}$$

and using also (27) gives

$$L_-|\nu\, n-\nu\rangle = [n(\nu-n)(2\nu-n+1)/(\nu-n+1)]^{1/2} \times|\nu\, n-\nu-1\rangle. \quad (40)$$

These equations show that $L_\pm$ are equivalent to the n-dependent operators A_n and C_n used by Bauhain[13] or the operators $A_n^\pm$ used by Nieto and Simmons.[14] Our operators $L_\pm$ avoid the n dependence by introducing the extra variable ϕ, as in the treatment of the hydrogenic radial functions by Armstrong.[3] The basis functions used by Sannikov[16] satisfy different equations owing to different normalization.

From (39), (40), and (4) one obtains the Rosen–Morse matrix elements of $\sinh x$ and of $\cosh x(d/dx)$; these operators play a basic role in Nieto and Simmons' definition of coherent states.

[1]N. Rosen and P. M. Morse, Phys. Rev. **42**, 210 (1932).
[2]W. Pauli, Z. Phys. **36**, 336 (1926).
[3]L. Armstrong, Phys. Rev. A **3**, 1546 (1971).
[4]P. Cordero and G. C. Ghirardi, Fortschr. Phys. **20**, 105 (1972).
[5]B. Beers and H. H. Nickle, J. Math. Phys. **13**, 1592 (1972); H. H. Nickle and B. L. Beers, J. Phys. A **5**, 1658 (1972).
[6]R. D. Levine and C. E. Wulfman, Chem. Phys. Lett. **60**, 372 (1979); C. E. Wulfman and R. D. Levine, *ibid.* **97**, 361 (1983).
[7]H. Berrondo and A. Palmer, J. Phys. A **13**, 773 (1980).
[8]Y. Alhassid, F. Gursey, and F. Iachello, Ann. Phys. (NY) **148**, 367 (1983); **167**, 181 (1986).
[9]B. Schuh, J. Phys. A **18**, 803 (1985).
[10]A. Frank and K. B. Wolf, J. Math. Phys. **26**, 973 (1985).
[11]I. Kay and H. E. Moses, J. Appl. Phys. **27**, 1503 (1956).
[12]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[13]M. Bauhain, Lett. Nuovo Cimento **14**, 475 (1975).
[14]M. M. Nieto and L. M. Simmons, Phys. Rev. D **20**, 1342 (1979).
[15]G. L. Lamb, *Elements of Soliton Theory* (Wiley, New York, 1980), p. 38.
[16]S. S. Sannikov, Yad. Fiz. **6**, 1294 (1967) [Sov. J. Nucl. Phys. **6**, 939 (1968)].
[17]B. Gruber and A. U. Klimyk, J. Math. Phys. **19**, 2009 (1978).
[18]B. Gruber and A. U. Klimyk, J. Math. Phys. **25**, 755 (1984).
[19]H. Flaschka and D. W. McLaughlin, in *Backlund Transformations*, edited by R. M. Miura (Springer, New York, 1976), p. 258; P. Deift and E. Trubowitz, Commun. Pure Appl. Math **32**, 165 (1979).
[20]B. Gruber, Proc. R. Ir. Acad. A **82**, 13 (1982); B. Gruber and R. Lenczewski, J. Phys. A **16**, 3703 (1983).

On the problem of local hidden variables in algebraic quantum mechanics

Miklós Rédei
Loránd Eötvös University, H-1088 Budapest, Rákóczi ut 5, Hungary

(Received 20 March 1986; accepted for publication 29 October 1986)

Given two Bose-type quasilocal C^*-algebras $\mathscr{A}$, $\mathscr{B}$, their state spaces $E(\mathscr{A})$, $E(\mathscr{B})$, and a positive, unit preserving map $L: \mathscr{B} \to \mathscr{A}$, respecting the local structure of $\mathscr{A}$ and $\mathscr{B}$ $(\mathscr{A}, E(\mathscr{A}))$ is said to have $(\mathscr{B}, E(\mathscr{B}))$ as a local hidden theory via L if for all states φ in $E(\mathscr{A})$, $L^*\varphi$ can be decomposed in $E(\mathscr{B})$ via a subcentral measure into states with pointwise strictly less dispersion than the dispersion of φ. After motivating this definition of local hidden theory it is shown that if, in addition, L factorizes on disjoint local algebras, then $(\mathscr{B}, E(\mathscr{B}))$ is not a local hidden theory of $(\mathscr{A}, E(\mathscr{A}))$ via L.

I. INTRODUCTION

In this paper we prove a negative result on the following problem: Let $\mathscr{A}, \mathscr{B}$ and $E(\mathscr{A})$, $E(\mathscr{B})$ be two unital C^*-algebras and their state spaces, respectively,[1] and let L: $\mathscr{B} \to \mathscr{A}$ be a linear onto map that maps the unit of $\mathscr{B}$ into the unit of $\mathscr{A}$ and has the property $x \geqslant 0$ implies $Lx \geqslant 0$. Such a map L will be called a positive, unit preserving map and the set of all positive, unit preserving maps from $\mathscr{B}$ onto $\mathscr{A}$ will be denoted by $\mathscr{D}(\mathscr{B}, \mathscr{A})$. If $L \in \mathscr{D}(\mathscr{B}, \mathscr{A})$, then $(L^*\varphi)(x) = \varphi(Lx)$, $x \in \mathscr{B}$, defines a state $L^*\varphi$ over $\mathscr{B}$ for all states φ in $E(\mathscr{A})$. The problem is under what conditions on $\mathscr{A}, \mathscr{B}$ and a positive map $L \in \mathscr{D}(\mathscr{B}, \mathscr{A})$ is it true that for every φ in $E(\mathscr{A})$ we can find a positive, normalized (regular Borel) measure μ on the state space $E(\mathscr{B})$, with the help of which $L^*\varphi$ can be obtained in the integral form

$$L^*\varphi(x) = \int \omega(x) d\mu(\omega), \tag{1}$$

such that the dispersions of all the states occurring in the integration are strictly less than the dispersion of φ; more precisely: For all $\omega \in \operatorname{supp} \mu$ and all $x \in \mathscr{B}$ we have

$$\sigma_\varphi(Lx) > \sigma_\omega(x), \quad \text{if } \sigma_\varphi(Lx) > 0, \tag{2}$$

$$\sigma_\varphi(Lx) = \sigma_\omega(x), \quad \text{if } \sigma_\varphi(Lx) = 0, \tag{3}$$

where $\sigma_\varphi(Lx) = \varphi((Lx)^2) - (\varphi(Lx))^2$ and $\sigma_\omega(x) = \omega(x^2) - (\omega(x))^2$ are the dispersions defined in the usual way and $\operatorname{supp} \mu$ denotes the support set of μ, which is the smallest closed set in $E(\mathscr{B})$ with μ-measure equal to 1.

This problem is a very natural generalization in operator algebraic framework of the problem known for a long time as the problem of hidden variables in quantum mechanics. Thus in the case when the decomposition described above can be given, we call the pair $(\mathscr{B}, E(\mathscr{B}))$ a hidden theory of $(\mathscr{A}, E(\mathscr{A}))$ via the hiding map L.

The original formulation of the hidden variable problem, which was given by von Neumann, can be phrased as follows: $\mathscr{A}$ equals to the C^*-algebra of the bounded linear operators $\mathscr{B}(\mathscr{H})$ on some Hilbert space $\mathscr{H}$, $\mathscr{A} = \mathscr{B}$, and L is the identity map. Furthermore, as is well known, von Neumann demanded of the hidden states that they be dispersion-free, which was shown by him to be impossible under weak assumptions on $\mathscr{H}$.[2]

Later on, the problem was reformulated by Misra, who investigated the problem without assuming $\mathscr{A}$ to be $\mathscr{B}(\mathscr{H})$ on some $\mathscr{H}$ but still keeping the requirement of the hidden states that they have zero dispersion on all observables, and also in his definition $\mathscr{A} = \mathscr{B}$, L = identity. Misra proved that there is a nontrivial dispersion-free state in $E(\mathscr{A})$ if and only if $\mathscr{A}$ contains a nontrivial (closed) two-sided ideal Z such that the quotient algebra $\mathscr{A}/Z$ is commutative, which implies that if $\mathscr{A}$ is simple (i.e., $\mathscr{A}$ does not contain a nontrivial two-sided ideal) then no state over $\mathscr{A}$ can be dispersion-free.[3]

However, it is natural to generalize the hidden variable problem in the way done above for two reasons at least. First, one may argue that it can never be decided on a physical basis whether a state is strictly dispersion-free or not because of the finite, nonzero, unavoidable error occurring in every real physical measurement; therefore all the hidden variable definitions formulated in terms of dispersion-free states are physically too strong and inaccessible for measurements. As one seeks for hidden variables in the hope of becoming able to decrease the uncertainty—expressed by the nonzero dispersions—of the quantum mechanical states by finding the hidden states, the requirement that the latter have strictly less dispersions seems to be the weakest requirement one should insist on. Second, the assumption that both the quantum mechanical states to be decomposed into hidden states and also the hidden states themselves are defined over the same algebra is a very strong restriction, which is mathematically unnecessary and which is hard—if possible at all—to justify physically.

Some negative results on the problem were proved for rather general $\mathscr{A}$, $\mathscr{B}$, and L in a previous paper.[4] However, it is the problem of *local* hidden variables that has become the main subject of research since Bell's work.[5] As the so-called quasilocal algebras were introduced into the field theory and statistical physics just in order to express the physical locality concept in a mathematically precise and treatable way,[6] it is very natural to investigate the problem in the case where $\mathscr{A}, \mathscr{B}$ are quasilocal algebras and both L and the averaging process [that is, the measure μ in (1) considered as a mathematical representative of some physical averaging process] respect the local structure of $\mathscr{A}$ and $\mathscr{B}$ in some appropriate sense (see the definition of local hidden theory below). Before giving and motivating these locality

 0022-2488/87/040833-03$02.50

definitions we first describe briefly those quasilocal algebras that our result applies to. For more general quasilocal algebras the reader is referred to our standard reference.[1]

II. QUASILOCAL ALGEBRAS AND LOCAL HIDDEN THEORIES

Let $(I,\leqslant)$ be a directed set endowed with a relation $\perp$ (called orthogonality) between pairs of elements of I such that (1) if $\alpha_1\in I$, then there is an $\alpha_2\in I$ with $\alpha_1\perp\alpha_2$; (2) if $\alpha\leqslant\alpha_1$ and $\alpha_1\perp\alpha_2$, then $\alpha\perp\alpha_2$; and (3) if $\alpha\perp\alpha_1$ and $\alpha\perp\alpha_2$, then there is an $\alpha_3\in I$ such that $\alpha\perp\alpha_3$ and α_1, $\alpha_2\leqslant\alpha_3$. Assume, furthermore, that $(I,\leqslant,\perp)$ has the following properties: For each pair $\alpha_1,\alpha_2\in I$ there is an element (the least upper bound of α_1 and α_2) denoted by $\alpha_1\vee\alpha_2$ such that (4) $\alpha_1\leqslant\alpha_1\vee\alpha_2$, $\alpha_2\leqslant\alpha_1\vee\alpha_2$; and (5) if $\alpha_1\leqslant\alpha$ and $\alpha_2\leqslant\alpha$, then $\alpha_1\vee\alpha_2\leqslant\alpha$. For each pair α_1, α_2 there exists an $\alpha\perp\alpha_1$ such that $\alpha_2\leqslant\alpha_1\vee\alpha$.

A set I having all these properties is called an index set. Typically I consists of the open, bounded subsets of the configuration space R^d ordered by inclusion and the orthogonality relation $\alpha_1\perp\alpha_2$ corresponds to the disjointness of α_1 and α_2.

The pair $(\mathscr{A},(\mathscr{A}_\alpha,\alpha\in I))$ with an index set I is called a (Bose-type) quasilocal algebra if $\mathscr{A}$, $\mathscr{A}_\alpha$ are unital C^*-algebras and the following four conditions are fulfilled: (1) If $\alpha_1\leqslant\alpha_2$, then $\mathscr{A}_{\alpha_1}\subset\mathscr{A}_{\alpha_2}$; (2) $\mathscr{A}$ equals to the norm closure of the union $\cup_{\alpha\in I}\mathscr{A}_\alpha$; (3) the local algebras $\mathscr{A}_\alpha$ and $\mathscr{A}$ have a common unit; and (4) $[A_{\alpha_1},B_{\alpha_2}]=0$ for all $A_{\alpha_1}\in\mathscr{A}_{\alpha_1}$ and $B_{\alpha_2}\in\mathscr{A}_{\alpha_2}$ if $\alpha_1\perp\alpha_2$. We assume, in addition, that for each $\alpha\in$I there is a Hilbert space $\mathscr{H}_\alpha$ such that $\mathscr{A}_\alpha$ is isomorphic to $\mathscr{B}(\mathscr{H}_\alpha)$, the C^*-algebra of all bounded linear operators on $\mathscr{H}_\alpha$, and that $\mathscr{A}_{\alpha_1}\cup\mathscr{A}_{\alpha_2}$ generate $\mathscr{A}_{\alpha_1\vee\alpha_2}$ in the weak operator topology.

Definition: Let $(\mathscr{A},(\mathscr{A}_\alpha,\alpha\in\text{I}))$ and $(\mathscr{B},(\mathscr{B}_\beta,\beta\in J))$ be two quasilocal algebras. A positive map L from $\mathscr{B}$ onto $\mathscr{A}$ is called *local* if (1) the local algebras are mapped onto local algebras in such a way that (2) disjoint algebras are mapped onto disjoint algebras[7] and (3) the restriction of L to each local algebra is continuous in the ultraweak operator topology.

Definition: A local map L is said to be *locally factorizable* if $L(A_{\beta_1}B_{\beta_2})=LA_{\beta_1}LB_{\beta_2}$ whenever A_{β_1} and B_{β_2} belong to disjoint local algebras.

Besides the locality of the hiding map it is also necessary to request of μ some kind of locality property if the decomposition (1) via μ is thought to be more than a mere mathematical expression. To motivate the property μ should possess, let us first define the restriction maps $r_\beta: E(\mathscr{B})\to E(\mathscr{B}_\beta)$, $\beta\in J$, by $(r_\beta\varphi)(x)=\varphi(x)$ $(x\in\mathscr{B}_\beta)$. Here r is known to be w^*-continuous,[8] thus given a Borel subset E_β of $E(\mathscr{B}_\beta)$, $r^{-1}(E_\beta)\subset E(\mathscr{B})$ is the Borel set of states extending the states φ in E_β to the whole quasilocal algebra. Let $\delta\perp\beta$ and $E_\delta\subset E(\mathscr{B}_\delta)$ also be a Borel set. Then the Borel subset G_β of $E(\mathscr{B})$ defined by

$$G_\beta=r_\beta^{-1}(E_\beta)\setminus r_\delta^{-1}(E_\delta)$$

consists of states over the quasilocal algebra $\mathscr{B}$ that extend the states in E_β without extending any of the states in E_δ. This means that no information about the local algebra $\mathscr{B}_\delta$ contained in E_δ can be obtained from G_β and, similarly, the states in the accordingly defined G_δ "do not feel" those properties of the local algebra $\mathscr{B}_\beta$ (disjoint from $\mathscr{B}_\delta$) that are described by the states in E_β. The idea of physical locality demands of μ that it does not mix up the states in G_β and G_δ, in other words, the two positive functionals ω_β, ω_δ obtained by averaging via μ over G_β and G_δ, respectively,

$$\omega_\beta=\int_{G_\beta}\omega\,d\mu(\omega),\quad \omega_\delta=\int_{G_\delta}\omega\,d\mu(\omega),$$

should be independent.

A natural independence condition of positive linear functionals over C^*-algebras is their disjointness: Two positive functionals ω_1, ω_2 are disjoint if the two representations π_{ω_1}, π_{ω_2} associated with ω_1 and ω_2 are disjoint, i.e., π_{ω_1}, π_{ω_2} have no unitary equivalent subrepresentations.[9]

All this motivates the following definition: μ is said to be local if for two arbitrary disjoint Borel sets E_1, E_2 in $E(\mathscr{B})$ the two functionals $\omega_1=\int_{E_1}\omega\,d\mu(\omega)$ and $\omega_2=\int_{E_2}\omega\,d\mu(\omega)$ are disjoint. Recall[10] that a measure μ on $E(\mathscr{B})$ is called a subcentral measure if for arbitrary Borel set E in $E(\mathscr{B})$, $\int_E\omega\,d\mu(\omega)$ and $\int_{E(\mathscr{B})\setminus E}\omega\,d\mu(\omega)$, are disjoint. It is easily seen that if μ is subcentral then it also is local in the sense above (the converse is obvious). We sum up with the following definition: Given $(\mathscr{A},E(\mathscr{A}))$ with a quasilocal algebra $\mathscr{A}$, a hidden theory $(\mathscr{B},E(\mathscr{B}))$ of $(\mathscr{A},E(\mathscr{A}))$ via L is called a local hidden theory if $\mathscr{B}$ is a quasilocal algebra, L is local, and for each $\varphi\in E(\mathscr{A})$ we can find a subcentral measure μ having the properties (1)–(3).

III. A NEGATIVE RESULT

Before stating our result we recall a few facts that will be used in the proof of it.

A positive map $L\in\mathscr{D}(\mathscr{B},\mathscr{A})$ is a contraction, i.e., $\|Lx\|\leqslant\|x\|$ for all x and for a positive $L\in\mathscr{D}(\mathscr{B},\mathscr{A})$ the generalized Cauchy–Schwarz inequality is valid[11]: $L(x^*x)\geqslant(Lx)^*Lx$ for all normal (and, therefore, also for all self-adjoint) x. A state φ over the quasilocal algebra $\mathscr{A}$ is called locally normal if φ is given by a density matrix ρ_α on each local algebra $\mathscr{A}_\alpha$. An arbitrary state φ is called factor state if the von Neumann algebra $\pi_\varphi(\mathscr{A})''$ generated in the Gel'fand–Naimark–Segal (GNS) representation π_φ induced by φ is a factor, i.e., there are no nontrivial projectors in the center $\pi_\varphi(\mathscr{A})''\cap\pi_\varphi(\mathscr{A})'$. Note that the set of locally normal factor states is nonempty by a straightforward argument,[12] and a cluster-type characterization of locally normal factor states is known.

Theorem[13]**:** The following two conditions are equivalent if φ is a locally normal state: (1) φ is a factor state; (2) given $\alpha_1\in$I and $\epsilon>0$ there exists an α' such that

$$|\varphi(A_{\alpha_1}B_\alpha)-\varphi(A_{\alpha_1})\varphi(B_\alpha)|\leqslant\epsilon\|A_{\alpha_1}\|\,\|B_\alpha\|,$$

for all $A_{\alpha_1}\in\mathscr{A}_{\alpha_1}$, all $B_\alpha\in\mathscr{A}_\alpha$, and all $\alpha\perp\alpha'$.

We can now formulate the result.

Proposition: Let $\mathscr{A}$ and $\mathscr{B}$ be quasilocal algebras with $\mathscr{A}$ simple.[14] Then $(\mathscr{B},E(\mathscr{B}))$ is not a local hidden theory of $(\mathscr{A},E(\mathscr{A}))$ via a locally factorizable map L.

Proof: Assume that $(\mathscr{B},E(\mathscr{B}))$ is a local hidden theory of $(\mathscr{A},E(\mathscr{A}))$ via a locally factorizable map L and let $\varphi\in E(\mathscr{A})$ be an arbitrary state. With $\mathscr{A}$ being simple, no

state over $\mathscr{A}$ can be dispersion-free by the cited result of Misra; thus there is a self-adjoint $x_1 \in \mathscr{B}$ such that $\sigma_\varphi(Lx_1) > 0$. Integrating (2) with respect to μ we get

$$\varphi((Lx_1)^2 - Lx_1^2) > \varphi(Lx_1)^2 - \int_{E(\mathscr{B})} \omega(x_1)^2 \, d\mu(\omega). \tag{4}$$

The left-hand side of (4) is not greater than zero by the generalized Cauchy–Schwarz inequality for L; therefore if μ is the Dirac—or point measure concentrated at $L^*\varphi$—then (4) is a contradiction since the right-hand side of (4) equals zero in this case.

Let φ now be a locally normal factor state over $\mathscr{A}$. By the locality of L, $L^*\varphi$ is a locally normal state over $\mathscr{B}$, which is also a factor state by the following argument. Assume that $L^*\varphi$ is not a factor state. Then by the theorem characterizing locally normal factor states there is a $\beta_1 \in J$ and an $\epsilon > 0$ such that for all $\beta \in J$ there is a $\beta_2 \in J$, $\beta_2 \perp \beta$, and there are $A_{\beta_1} \in \mathscr{B}_{\beta_1}$ and $B_{\beta_2} \in \mathscr{B}_{\beta_2}$ for which the following inequality holds:

$$|\varphi(LA_{\beta_1}B_{\beta_2}) - \varphi(LA_{\beta_1})\varphi(LB_{\beta_2})| > \epsilon \|A_{\beta_1}\| \, \|B_{\beta_2}\|. \tag{5}$$

We may assume that $\beta_1 \perp \beta_2$, for if this were not the case, then choosing $\beta_3 = \beta_1 \vee \beta_2$ and applying the theorem again with β_3 as β we can find a $\beta_2' \perp \beta_3$ such that (5) holds with some $B_{\beta_2'} \in \mathscr{B}_{\beta_2'}$. But then $\beta_2' \perp \beta_1$ holds, too, by the properties of the index set J. Thus A_{β_1} and B_{β_2} can be chosen from disjoint local algebras, therefore, by the local factorizability of L we get from (5)

$$|\varphi(LA_{\beta_1}LB_{\beta_2}) - \varphi(LA_{\beta_1})\varphi(LB_{\beta_2})| > \epsilon \|A_{\beta_1}\| \, \|B_{\beta_2}\| > \epsilon' \|LA_{\beta_1}\| \, \|LB_{\beta_2}\|,$$

with some $\epsilon' > 0$. This implies by the theorem and by the locality properties of L that φ is not a factor state.

Thus $L^*\varphi$ is a factor state, that is, $\pi_{L^*\varphi}$ is a factor representation. Now let $\pi_\mu = \int_{E(\mathscr{B})}^{\oplus} \pi_\omega \, d\mu(\omega)$ be the direct integral representation of $\mathscr{B}$ defined by μ in the usual way.[15] Since μ is subcentral it also is orthogonal,[16] which implies that $\pi_\mu = \pi_{L^*\varphi}$ by the theorem of Effros,[17] i.e., π_μ is a factor representation, too. The proof is complete by noting that the only subcentral measure that decomposes a factor representation is the Dirac measure. This can explicitly be shown as follows: The map $K_\mu : L^\infty(E(\mathscr{B}), \mu) \to \pi_{L^*\varphi}(\mathscr{B})'$ defined by

$$\langle \Omega_{L^*\varphi}, K_\mu(f)\pi_{L^*\varphi}(x)\Omega_{L^*\varphi} \rangle = \int_{E(\mathscr{B})} f(\omega)\omega(x) d\mu(\omega)$$

is a $*$-isomorphism by Tomita's theorem.[18] If μ is not the Dirac measure, then supp μ contains at least two different points ω_1, ω_2, which can be separated by two Borel sets E_1, E_2, such that $0 < \mu(E_1) < 1$. Let χ_{E_1} be the characteristic function of E_1. Then $K_\mu(\chi_{E_1})$ is a nontrivial projector in $\pi_{L^*\varphi}(\mathscr{B})'$. Since μ is subcentral, $K_\mu(\chi_{E_1})$ lies not only in $\pi_{L^*\varphi}(\mathscr{B})'$ but also in $\pi_{L^*\varphi}(\mathscr{B})''$ by Proposition 4.2.9 of Ref. 1, i.e., $K_\mu(\chi_{E_1})$ is a nontrivial central projector of the von Neumann algebra $\pi_{L^*\varphi}(\mathscr{B})''$, which is impossible since $\pi_{L^*\varphi}$ is a factor representation.

IV. CLOSING REMARKS

It is well known that one of the assumptions, the "locality" or "factorizability" as it was called later,[19] which the derivation of Bell's inequalities for stochastic hidden variables was based on, has been criticized in the sense that it has been questioned whether it is a sufficient or even necessary condition of expressing physical locality for stochastic hidden variables.[20] More recently strong arguments have been given in favor of independence of the factorizability from a very reasonable physical locality condition[21]; thus the problem of existence of stochastic local hidden variables has become in some sense open again and the proposition above may be a step ahead in clarifying the role of locality in the problem of hidden variables. The locality conditions in the definition of local hidden theory are quite natural; however, it would be desirable to know whether a proposition similar to the one proved here can be obtained without the assumption of local factorizability of L, which is less natural and which was motivated by the original factorizability property.

It would also be interesting to know whether $(\mathscr{A}, E(\mathscr{A}))$ has a local hidden theory $(\mathscr{B}, E(\mathscr{B}))$ via some L in the case where $\mathscr{A}, \mathscr{B}$ are the more general quasilocal algebras of relativistic quantum field theory.

[1] For all the definitions and elementary facts in connection with the C^*-algebra theory we refer to O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, Berlin, 1979), Vol. I, which will be used here as a standard reference.

[2] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).

[3] B. Misra, Nuovo Cimento A **47**, 841 (1968).

[4] M. Rédei, "Non-existence of hidden variables in algebraic approach," Found. Phys. (to be published).

[5] J. S. Bell, Physics (NY) **1**, 195 (1964); J. S. Bell, in the *Proceedings of the International School of Physics*, "Enrico Fermi" Course 49 (Academic, New York, 1971).

[6] G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley-Interscience, New York, 1972).

[7] Two local algebras, $\mathscr{A}_{\alpha_1}$ and $\mathscr{A}_{\alpha_2}$, are called disjoint if $\alpha_1 \perp \alpha_2$.

[8] See Ref. 1, p. 351.

[9] See Ref. 1, p. 362.

[10] See Ref. 1, p. 363.

[11] Proposition 3.2.4 in Ref. 1.

[12] The steps of proof of Proposition 2.3.24 in Ref. 1 can be repeated with obvious modifications.

[13] See Theorem 2.6.10 in Ref. 1.

[14] Note that the simplicity of $\mathscr{A}$ follows, e.g., if $\mathscr{H}_\alpha$ ($\alpha \in$ I) is a separable, infinite dimensional Hilbert space. See Corollary 2.6.20 in Ref. 1.

[15] See Chap. 4.2.2 in Ref. 1.

[16] See Proposition 4.2.9 and Definition 4.1.20 in Ref. 1.

[17] See Theorem 4.4.9 in Ref. 1.

[18] See Proposition 4.1.22 in Ref. 1.

[19] A. Garuccio and F. Selleri, Lett. Nuovo Cimento **23**, 555 (1978); see also Ref. 20. The term "factorizability" was particularly preferred by A. Fine (see Ref. 20).

[20] A. Fine, in *Philosophy of Science Association 1980*, edited by P. D. Asquith and R. Giere (Philosophy of Science Association, East Lansing, MI, 1981); see also A. Fine, Phys. Rev. Lett. **49**, 1536 (1982).

[21] G. Hellmann, Synthese **53**, 461 (1982).

Addition theorems for spherical wave solutions of the vector Helmholtz equation

B. U. Felderhof
Institut für Theoretische Physik A, RWTH Aachen, Templergraben 55, 5100 Aachen, West Germany

R. B. Jones
Department of Physics, Queen Mary College, University of London, Mile End Road, London E1 4NS, England

(Received 25 September 1986; accepted for publication 26 November 1986)

Addition theorems for spherical wave solutions of the vector Helmholtz equation are discussed. The theorems allow one to expand a vector spherical wave about a given origin into spherical waves about a shifted origin. A simplified derivation of the results obtained earlier by Cruzan [O. R. Cruzan, Q. Appl. Math. **20**, 33 (1962)] is presented.

I. INTRODUCTION

In many physical problems it is necessary to expand a multipole wave centered about a given origin into mulipole waves centered about a shifted origin. In this article we consider the expansion of spherical wave solutions of the vector Helmholtz equation. These are the well-known **L**, **M**, and **N** waves of electromagnetic theory.[1] The expansions are referred to as addition theorems since the expansion coefficients themselves satisfy the scalar wave equation.

Addition theorems for vector spherical waves have been found earlier by Stein[2] and by Cruzan.[3] They started from the addition theorem for scalar spherical waves,[4] which is recalled in Sec. II of this article. The addition theorems for the **L**, **M** and **N** waves are then derived by a tedious calculation in spherical coordinates. We sketch the procedure in Sec. III.

An independent derivation along the same lines for the special case when the origin is shifted in the z direction was given by Langbein,[5] who was led to a different form which is not related in an obvious manner to the results of Stein[2] and Cruzan.[3] Langbein's[5] expansion was generalized to arbitrary directions of the shift vector by Gérardy and Ausloos.[6] In the resulting form of the expansion the coefficients do not obviously satisfy the scalar wave equation. This makes the expansion of Ref. 6 less elegant and less satisfying from a theoretical point of view.

In Sec. IV of this article we show that addition theorems in the desired form may be derived straightforwardly and quickly from an extension of the scalar wave addition theorem to tensor multipole fields. The basic theorem was found by Danos and Maximon,[7] who derived an addition theorem for tensor multipole fields by coupling unit tensors to both sides of the scalar equation and using known quantum mechanical angular momentum algebra. We refer to their article for an interesting review of the history of the problem.

In Sec. V we compare in some detail with Langbein's[5] results. Throughout this article we adhere to the notation used by Edmonds.[8]

II. SCALAR WAVE ADDITION THEOREMS

The scalar wave equation

$$\nabla^2\psi + k^2\psi = 0 \tag{2.1}$$

has spherical wave solutions

$$\Psi_{lm}(\mathbf{r}) = j_l(kr)Y_{lm}(\theta,\varphi)\,, \tag{2.2}$$

where $j_l(kr)$ is a regular spherical Bessel function and $Y_{lm}(\theta,\varphi)$ is a spherical harmonic. We wish to expand the solution $\Psi_{lm}(\mathbf{r})$ into spherical waves centered about a shifted origin. The expansion yields the simplest wave addition theorem. We consider the three vectors $\mathbf{r},\boldsymbol{\rho}$, and $\mathbf{r}'$ related by

$$\mathbf{r} = \boldsymbol{\rho} + \mathbf{r}'\,, \tag{2.3}$$

and expand the corresponding plane-wave identity

$$e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot\boldsymbol{\rho}}e^{i\mathbf{k}\cdot\mathbf{r}'}\,, \tag{2.4}$$

into spherical waves using[9]

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{lm} i^l\Psi_{lm}(\mathbf{r})Y^*_{lm}(\theta_k,\varphi_k)\,. \tag{2.5}$$

Multiplying (2.4) by $Y_{lm}(\theta_k,\varphi_k)$ and integrating over the directions of $\mathbf{k}$ one obtains

$$\Psi_{lm}(\mathbf{r}) = \sum_{l'm'} A^{lm}_{l'm'}(\boldsymbol{\rho})\Psi_{l'm'}(\mathbf{r}')\,, \tag{2.6}$$

where

$$A^{lm}_{l'm'}(\boldsymbol{\rho}) = \sum_{\lambda\mu} c(lm|l'm'|\lambda\mu)\Psi_{\lambda\mu}(\boldsymbol{\rho})\,, \tag{2.7}$$

with coefficients

$$c(lm|l'm'|\lambda\mu) = i^{l'+\lambda-l}(-1)^m[4\pi(2l+1)(2l'+1)(2\lambda+1)]^{1/2} \times \begin{pmatrix} l & l' & \lambda \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l & l' & \lambda \\ -m & m' & \mu \end{pmatrix}. \tag{2.8}$$

The coefficients arise as integrals of products of three spherical harmonics[10]:

$$c(lm|l'm'|\lambda\mu) = 4\pi i^{l'+\lambda-l}\int Y_{lm}Y^*_{l'm'}Y^*_{\lambda\mu}\,d\Omega\,. \tag{2.9}$$

They are related to the coefficients in the expansion of a product of two associated Legendre functions:

$$P^{m'}_{l'}P^{\mu}_{\lambda} = \sum_l a(l'm'|\lambda\mu|l)P^{m'+\mu}_l \tag{2.10}$$

 0022-2488/87/040836-04$02.50

by[3]

$$c(lm|l'm'|\lambda\mu) = i^{l'+\lambda-l}\left[4\pi\frac{(2l'+1)(2\lambda+1)}{2l+1}\right]^{1/2} \times\left[\frac{(l+m)!(l'-m')!(\lambda-\mu)!}{(l-m)!(l'+m')!(\lambda+\mu)!}\right]^{1/2} a(l'm'|\lambda\mu|l)\,, \tag{2.11}$$

where $m = m' + \mu$.

The addition theorem (2.6) may be generalized to singular solutions of the wave equation, which are the product of a singular spherical Bessel function and a spherical harmonic. In the following we generalize (2.2) to

$$\Psi_{lm}(\mathbf{r}) = f_l(kr)\,Y_{lm}(\theta,\varphi)\,, \tag{2.12}$$

where $f_l(\zeta)$ is any of the spherical Bessel functions $j_l(\zeta)$, $y_l(\zeta)$, $h_l^{(1)}(\zeta)$, or $h_l^{(2)}(\zeta)$. The regular solution (2.2) will be distinguished by a superscript: $\Psi_{lm}^{+}(\mathbf{r})$. The generalization of (2.6) then reads[7]

$$\Psi_{lm}(\mathbf{r}) = \sum_{\substack{l'm'\\ \lambda\mu}} c(lm|l'm'|\lambda\mu)\,\Psi_{l'm'}(\mathbf{r}_>)\Psi_{\lambda\mu}^{+}(\mathbf{r}_<)\,, \tag{2.13}$$

where $\mathbf{r}_<$ is the smaller and $\mathbf{r}_>$ is the larger of $\boldsymbol{\rho}$ and $\mathbf{r}'$. It is understood that $\Psi_{lm}(\mathbf{r})$ and $\Psi_{l'm'}(\mathbf{r}_>)$ contain spherical Bessel functions of the same type. Clearly, the expansion (2.13) may be written in two ways. We may either write

$$\Psi_{lm}(\mathbf{r}) = \sum_{l'm'} A_{l'm'}^{lm+}(\mathbf{r}_<)\Psi_{l'm'}(\mathbf{r}_>)\,, \tag{2.14}$$

or

$$\Psi_{lm}(\mathbf{r}) = \sum_{l'm'} A_{l'm'}^{lm}(\mathbf{r}_>)\Psi_{l'm'}^{+}(\mathbf{r}_<)\,, \tag{2.15}$$

where we have used the symmetry of the coefficients $c(lm|l'm'|\lambda\mu)$ in the pairs $(l'm')$ and $(\lambda\mu)$ which follows from (2.9).

III. VECTOR WAVE ADDITION THEOREMS

The vector wave equation

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = 0 \tag{3.1}$$

has spherical wave solutions that may be derived from scalar potentials which are solutions of the scalar wave equation. Thus one finds the three vector spherical waves[1]

$$\begin{aligned}\mathbf{L}_{JM}(\mathbf{r}) &= k^{-1}\nabla\Psi_{JM}(\mathbf{r})\,,\\ \mathbf{M}_{JM}(\mathbf{r}) &= \nabla\times(\mathbf{r}\Psi_{JM}(\mathbf{r}))\,,\\ \mathbf{N}_{JM}(\mathbf{r}) &= k^{-1}\nabla\times[\nabla\times(\mathbf{r}\Psi_{JM}(\mathbf{r}))]\,.\end{aligned} \tag{3.2}$$

The L wave is longitudinal and the M and N waves are transverse. The latter are related by

$$\mathbf{M}_{JM} = k^{-1}\nabla\times\mathbf{N}_{JM},\quad \mathbf{N}_{JM} = k^{-1}\nabla\times\mathbf{M}_{JM}\,. \tag{3.3}$$

More explicitly, the solutions may be written as

$$\begin{aligned}\mathbf{L}_{JM}(\mathbf{r}) &= [1/(2J+1)]f_{J-1}(kr)\mathbf{A}_{JM}(\theta,\varphi) + [1/(2J+1)]f_{J+1}(kr)\mathbf{B}_{JM}(\theta,\varphi)\,,\\ \mathbf{M}_{JM}(\mathbf{r}) &= f_J(kr)\mathbf{C}_{JM}(\theta,\varphi)\,,\\ \mathbf{N}_{JM}(\mathbf{r}) &= [(J+1)/(2J+1)]f_{J-1}(kr)\mathbf{A}_{JM}(\theta,\varphi) - [J/(2J+1)]f_{J+1}(kr)\mathbf{B}_{JM}(\theta,\varphi)\,,\end{aligned} \tag{3.4}$$

where the vector spherical harmonics are given by

$$\begin{aligned}\mathbf{A}_{JM} &= JY_{JM}\mathbf{e}_r + \frac{\partial Y_{JM}}{\partial\theta}\mathbf{e}_\theta + \frac{1}{\sin\theta}\frac{\partial Y_{JM}}{\partial\varphi}\mathbf{e}_\varphi\,,\\ \mathbf{B}_{JM} &= -(J+1)\,Y_{JM}\mathbf{e}_r + \frac{\partial Y_{JM}}{\partial\theta}\mathbf{e}_\theta + \frac{1}{\sin\theta}\frac{\partial Y_{JM}}{\partial\varphi}\mathbf{e}_\varphi\,,\\ \mathbf{C}_{JM} &= \frac{1}{\sin\theta}\frac{\partial Y_{JM}}{\partial\varphi}\mathbf{e}_\theta - \frac{\partial Y_{JM}}{\partial\theta}\mathbf{e}_\varphi\,.\end{aligned} \tag{3.5}$$

The latter are related to the normalized vector spherical harmonics defined by Edmonds[11] by

$$\begin{aligned}\mathbf{A}_{JM} &= \sqrt{J(2J+1)}\;\mathbf{Y}_{J,J-1,M}\,,\\ \mathbf{B}_{JM} &= \sqrt{(J+1)(2J+1)}\;\mathbf{Y}_{J,J+1,M}\,,\\ \mathbf{C}_{JM} &= -i\sqrt{J(J+1)}\;\mathbf{Y}_{JJM}\,.\end{aligned} \tag{3.6}$$

The angular momentum operator for the vector waves is $\mathbf{J} = \mathbf{L} + \mathbf{S}$, where $\mathbf{L} = -i\mathbf{r}\times\nabla$ and $\mathbf{S}$ is the spin operator for spin 1. The vector spherical waves (3.4) are eigenfunctions of J^2 with eigenvalue $J(J+1)$ and eigenfunctions of J_z with eigenvalue M.

By applying the operator $k^{-1}\nabla_>$ to (2.14) and noting that $\mathbf{r} = \mathbf{r}_< + \mathbf{r}_>$, one finds the addition theorem

$$\mathbf{L}_{JM}(\mathbf{r}) = \sum_{J'M'} A_{J'M'}^{JM+}(\mathbf{r}_<)\mathbf{L}_{J'M'}(\mathbf{r}_>)\,. \tag{3.7}$$

In the same manner, by applying the operator $k^{-1}\nabla_<$ to (2.15), one finds

$$\mathbf{L}_{JM}(\mathbf{r}) = \sum_{J'M'} A_{J'M'}^{JM}(\mathbf{r}_>)\mathbf{L}_{J'M'}^{+}(\mathbf{r}_<)\,. \tag{3.8}$$

A similar method may be used for the derivation of addition theorems for the M and N spherical waves. One starts from the definition of $\mathbf{M}_{JM}(\mathbf{r})$ in (3.2) and applies either the operator $\nabla_>\times(\mathbf{r}$ to (2.14) or the operator $\nabla_<\times(\mathbf{r}$ to (2.15). These procedures each yield an addition theorem for $\mathbf{M}_{JM}(\mathbf{r})$. The addition theorems for $\mathbf{N}_{JM}(\mathbf{r})$ then follow from (3.3). The first procedure leads to

$$\begin{aligned}\mathbf{M}_{JM}(\mathbf{r}) &= \sum_{J'M'}\left[F_{J'M'}^{JM+}(\mathbf{r}_<)\mathbf{M}_{J'M'}(\mathbf{r}_>) + G_{J'M'}^{JM+}(\mathbf{r}_<)\mathbf{N}_{J'M'}(\mathbf{r}_>)\right],\\ \mathbf{N}_{JM}(\mathbf{r}) &= \sum_{J'M'}\left[G_{J'M'}^{JM+}(\mathbf{r}_<)\mathbf{M}_{J'M'}(\mathbf{r}_>) + F_{J'M'}^{JM+}(\mathbf{r}_<)\mathbf{N}_{J'M'}(\mathbf{r}_>)\right].\end{aligned} \tag{3.9}$$

The second procedure yields

$$\begin{aligned}\mathbf{M}_{JM}(\mathbf{r}) &= \sum_{J'M'}\left[F_{J'M'}^{JM}(\mathbf{r}_>)\mathbf{M}_{J'M'}^{+}(\mathbf{r}_<) + G_{J'M'}^{JM}(\mathbf{r}_>)\mathbf{N}_{J'M'}^{+}(\mathbf{r}_<)\right],\\ \mathbf{N}_{JM}(\mathbf{r}) &= \sum_{J'M'}\left[G_{J'M'}^{JM}(\mathbf{r}_>)\mathbf{M}_{J'M'}^{+}(\mathbf{r}_<) + F_{J'M'}^{JM}(\mathbf{r}_>)\mathbf{N}_{J'M'}^{+}(\mathbf{r}_<)\right].\end{aligned} \tag{3.10}$$

Explicit expressions for the coefficient functions $F_{J'M'}^{JM}(\boldsymbol{\rho})$

and $G^{JM}_{J'M'}(\rho)$ may be found by a tedious calculation in spherical coordinates. The calculation was first performed by Stein[2] and by Cruzan.[3] An independent derivation for the case where ρ points in the z direction was given by Langbein.[5] Cruzan's[3] result for the coefficient function $F^{JM}_{J'M'}(\rho)$ may be written as

$$F^{JM}_{J'M'}(\rho)=\sum_{\lambda\mu} f(JM|J'M'|\lambda\mu)\Psi_{\lambda\mu}(\rho) \tag{3.11}$$

and his coefficients $f(JM|J'M'|\lambda\mu)$ may be cast in the form

$$f(JM|J'M'|\lambda\mu)=\frac{J(J+1)+J'(J'+1)-\lambda(\lambda+1)}{2J'(J'+1)}\times c(JM|J'M'|\lambda\mu)\,. \tag{3.12}$$

Similarly, the coefficient function $G^{JM}_{J'M'}(\rho)$ may be written in the form

$$G^{JM}_{J'M'}(\rho)=\sum_{\lambda\mu} g(JM|J'M'|\lambda\mu)\Psi_{\lambda\mu}(\rho)\,, \tag{3.13}$$

with coefficients

$$g(JM|J'M'|\lambda\mu)=\frac{[(J+J'-\lambda)(J+\lambda-J')(\lambda+J'+J+1)(\lambda+J'-J+1)]^{1/2}}{2J'(J'+1)}d(JM|J'M'|\lambda\mu)\,, \tag{3.14}$$

where, in analogy to (2.8), the coefficients $d(JM|J'M'|\lambda\mu)$ are given by

$$\begin{aligned}d(JM|J'M'|\lambda\mu)&=i^{J'+\lambda-J}(-1)^{M+1}\\&\times[4\pi(2J+1)(2J'+1)(2\lambda+1)]^{1/2}\\&\times\begin{pmatrix}J-1 & J' & \lambda\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}J & J' & \lambda\\ -M & M' & \mu\end{pmatrix}.\end{aligned} \tag{3.15}$$

Using Edmonds'[8] relation (3.7.16) one sees that (3.14) differs in sign from Cruzan's[3] result. The sign error in Cruzan's result was also noted by other authors.[12,13]

The derivation of the results (3.9) and (3.10) following the method outlined above is lengthy and tedious. In Sec. IV we show how these results may be derived more quickly and more elegantly.

IV. IMPROVED DERIVATION

An improved derivation of the addition theorems for vector spherical waves may be based on the generalization of the scalar addition theorem (2.13) to tensor multipole fields as presented by Danos and Maximon.[7] Tensor wave fields are defined by coupling unit tensors to the scalar fields, for example,

$$\begin{aligned}\Psi^{[J]}_{lSM}(\mathbf{r})&=f_l(kr)\sum_{m's'}(lm'Ss'|lSJM)Y_{lm'}(\theta,\varphi)\hat{\mathbf{e}}^{[S]}_{s'}\\&=f_l(kr)[\mathbf{Y}^{[l]}(\hat{\mathbf{r}})\times\hat{\mathbf{e}}^{[S]}]^{[J]}_M\end{aligned} \tag{4.1}$$

in the notation of Fano and Racah.[14] By coupling unit tensors to both sides of the scalar addition theorem (2.13), one finds

$$\Psi^{[J]}_{lSM}(\mathbf{r})=\sum_{\substack{J'l'M'\\ \lambda\mu}} c^{[S]}(JlM|J'l'M'|\lambda\mu)\times\Psi^{[J']}_{l'SM'}(\mathbf{r}_>)\Psi^{+}_{\lambda\mu}(\mathbf{r}_<)\,, \tag{4.2}$$

with coefficients

$$\begin{aligned}c^{[S]}(JlM|J'l'M'|\lambda\mu)&=i^{l'+\lambda-l}(-1)^{S-M}\\&\times[4\pi(2J+1)(2l+1)(2J'+1)\\&\times(2l'+1)(2\lambda+1)]^{1/2}\\&\times\begin{pmatrix}l & l' & \lambda\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}J & J' & \lambda\\ -M & M' & \mu\end{pmatrix}\begin{Bmatrix}\lambda & J' & J\\ S & l & l'\end{Bmatrix}\end{aligned} \tag{4.3}$$

(we have corrected the prefactor given by Danos and Maximon[7]). An alternative generalization of (2.13) is

$$\Psi^{[J]}_{lSM}(\mathbf{r})=\sum_{\substack{J'l'M'\\ \lambda\mu}} c^{[S]}(JlM|J'l'M'|\lambda\mu)\times\Psi^{[J']}_{l'SM'}(\mathbf{r}_<)\Psi^{+}_{\lambda\mu}(\mathbf{r}_>)\,. \tag{4.4}$$

For $S=1$ the definition (4.1) becomes

$$\Psi^{[J]}_{l1M}(\mathbf{r})\equiv\Psi_{JlM}(\mathbf{r})=f_l(kr)\mathbf{Y}_{JlM}(\theta,\varphi) \tag{4.5}$$

and (4.4) becomes

$$\Psi_{JlM}(\mathbf{r})=\sum_{\substack{J'l'M'\\ \lambda\mu}} c^{[1]}(JlM|J'l'M'|\lambda\mu)\times\Psi^{+}_{J'l'M'}(\mathbf{r}_<)\Psi^{+}_{\lambda\mu}(\mathbf{r}_>)\,. \tag{4.6}$$

Here the orbital quantum numbers l and l' can take the values $J-1, J, J+1$ and $J'-1, J', J'+1$, respectively. It is clear from (3.4), (3.6), and (4.5) that $\Psi_{JlM}(\mathbf{r})$ is just a linear combination of $\mathbf{L}_{JM}(\mathbf{r})$, $\mathbf{M}_{JM}(\mathbf{r})$, and $\mathbf{N}_{JM}(\mathbf{r})$ waves. The addition theorems (3.8) and (3.10) therefore follow from (4.6) by simple operations with 3×3 matrices. We recall that the symmetry properties embodied in the form of the addition theorems (3.8) and (3.10) are due to the vector field relations (3.2) and (3.3). We may use these symmetry properties to simplify the expressions for the coefficients. Thus we find for the coefficient function $F^{JM}_{J'M'}(\rho)$,

$$F^{JM}_{J'M'}(\rho)=\left[\frac{J(J+1)}{J'(J'+1)}\right]^{1/2}\times\sum_{\lambda\mu} c^{[1]}(JJM|J'J'M'|\lambda\mu)\Psi_{\lambda\mu}(\rho) \tag{4.7}$$

and for the coefficient function $G^{JM}_{J'M'}(\rho)$,

$$G^{JM}_{J'M'}(\rho)=i\left[\frac{J(2J+1)}{J'(J'+1)}\right]^{1/2}\times\sum_{\lambda\mu} c^{[1]}(J,J-1,M|J'J'M'|\lambda\mu)\Psi_{\lambda\mu}(\rho)\,. \tag{4.8}$$

Substituting (4.3) for $S=1$ and using expressions for the Wigner $6j$ symbols given by Edmonds,[15] we hence obtain the expressions (3.11) and (3.13).

V. DISCUSSION

An important feature of the addition theorems (3.9) and (3.10) is that the coefficient functions $F^{JM}_{J'M'}(\rho)$ and $G^{JM}_{J'M'}(\rho)$ themselves are solutions of the scalar wave equation. Langbein[5] has derived an addition theorem for vector spherical waves for the special case where ρ is directed along the z axis. His result (5.36) is of the form (3.10), with

$$F^{JM}_{J'M'}(-ae_z) = f^{M}_{JJ'} V^{M*}_{JJ'}(ka)\delta_{MM'}, \qquad G^{JM}_{J'M'}(-ae_z) = f^{M}_{JJ'} W^{M*}_{JJ'}(ka)\delta_{MM'}, \tag{5.1}$$

with coefficients

$$F^{M}_{JJ'} = (-1)^{J+M} \times \left[(2J+1)(2J'+1)\frac{(J-M)!}{(J+M)!}\frac{(J'-M)!}{(J'+M)!}\right]^{1/2} \tag{5.2}$$

and functions

$$V^{M}_{JJ'}(\zeta) = U^{M}_{JJ'}(\zeta) - \{(J'-M+1)[(J'+1)(2J+1)]\}\zeta U^{M}_{JJ'+1}(\zeta) - \{(J'+M)/[J'(2J'+1)]\}\zeta U^{M}_{JJ'-1}(\zeta),$$

$$W^{M}_{JJ'}(\zeta) = iM[J'(J'+1)]^{-1}\zeta U^{M}_{JJ'}(\zeta), \tag{5.3}$$

with $U^{M}_{JJ'} = U^{M}_{J'J}$ and for $J \leqslant J'$,

$$U^{M}_{JJ'}(\zeta) = \left(\frac{2}{\zeta}\right)^{M} \sum_{\nu=0}^{J-M} (-1)^{\nu} \times \frac{\Gamma(J-\nu+\frac{1}{2})\Gamma(J'-\nu+\frac{1}{2})\Gamma(M+\nu+\frac{1}{2})}{\Gamma(J+J'-M-\nu+\frac{3}{2})\Gamma(M+\frac{1}{2})\Gamma(\frac{1}{2})} \times \frac{(J+J'-\nu)!}{(J-M-\nu)!(J'-M-\nu)!\nu!} \times (J+J'-M-2\nu+\tfrac{1}{2})f_{J+J'-M-2\nu}(\zeta). \tag{5.4}$$

(We note that the functions V and W employed by Gérardy and Ausloos[6] are slightly different.)

Langbein's functions $U^{M}_{JJ'}(\zeta)$ are defined from his version of the scalar wave addition theorem. Comparing his Eq. (5.34) with (2.15) for $\rho = -ae_z$ we obtain

$$U^{M}_{JJ'}(\zeta) = [f^{M}_{JJ'}]^{-1}\sum_{\lambda}(-1)^{\lambda}\left(\frac{2\lambda+1}{4\pi}\right)^{1/2} \times c(JM|J'M|\lambda 0)f_{\lambda}(\zeta). \tag{5.5}$$

Cruzan's relations (16) and (19), when specialized to $\rho = -ae_z$, agree with (5.1) and (5.3) when (2.11) and the above expression (5.5) for $U^{M}_{JJ'}(\zeta)$ is used [except for the opposite sign in the function $W^{M}_{JJ'}(\zeta)$]. From (3.11) and (3.12) we find by specializing to $\rho = -ae_z$,

$$V^{M}_{JJ'}(\zeta) = [f^{M}_{JJ'}]^{-1}\sum_{\lambda}(-1)^{\lambda}\left[\frac{2\lambda+1}{4\pi}\right]^{1/2} \times f(JM|J'M|\lambda 0)f_{\lambda}(\zeta), \tag{5.6}$$

with $f^{M}_{JJ'}$ defined by (5.2) and $f(JM|J'M'|\lambda\mu)$ defined by (3.12).

The validity of (5.5) may be proven directly from (5.4). One first proves the identity for $M = J$ and then uses the recursion relation (40) derived by Gérardy and Ausloos[6] to obtain the relation for general values of M. Similarly it should be possible to show (5.6) directly from (5.3). The derivation of (5.5) and (5.6) from the addition theorem is more straightforward.

For general directions of the connecting vector ρ the coefficient functions appearing in the addition theorems (3.9) and (3.10) have the desirable property that they are solutions of the scalar wave equation. Gérardy and Ausloos[6] have generalized Langbein's[5] addition theorem to arbitrary direction of ρ by performing a rotation of axes. This more general form is complicated and it is not evident that the coefficient functions satisfy the scalar wave equation.

In conclusion, we note that the addition theorems (3.7)–(3.10) may be used to derive similar theorems for the solutions of the equations of linearized hydrodynamics[16] and elasticity.[17]

[1] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), p. 414.
[2] S. Stein, Q. Appl. Math. **19**, 15 (1961).
[3] O. R. Cruzan, Q. Appl. Math. **20**, 33 (1962).
[4] B. Friedman and J. Russek, Q. Appl. Math. **12**, 13 (1954).
[5] D. Langbein, *Theory of Van der Waals Attraction* (Springer, Berlin, 1974), p. 79.
[6] J. M. Gérardy and M. Ausloos, Phys. Rev. B **25**, 4204 (1981).
[7] M. Danos and L. C. Maximon, J. Math. Phys. **6**, 766 (1965).
[8] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, 1974).
[9] Reference 8, p. 81.
[10] Reference 8, p. 63.
[11] Reference 8, p. 83.
[12] L. Tsang and J. A. Kong, J. Appl. Phys. **53**, 7162 (1982).
[13] P. C. Waterman and N. E. Pedersen, J. Appl. Phys. **59**, 2609 (1986).
[14] U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1957).
[15] Reference 8, p. 130.
[16] B. U. Felderhof, Physica **84A**, 557 (1976).
[17] R. B. Jones and R. Schmitz, Physica **122A**, 114 (1983).

Exact solution of the Schrödinger equation with noncentral parabolic potentials

Asok Guha[a] and S. Mukherjee
Department of Physics, North Bengal University, Raja-Rammohunpur, Darjeeling, India 734 430

(Received 22 April 1986; accepted for publication 29 October 1986)

The Schrödinger equation with a class of parabolic potentials has been studied. A constant of motion related to angular momentum has been calculated. It has been shown that the scattering states for these potentials can be defined only in some special cases. The Hartmann potential and the Aharonov–Bohm potential are studied as special cases.

I. INTRODUCTION

The study of exact solutions of the Schrödinger equation with a noncentral potential is of considerable interest. The case of a two-center potential is a typical example. The potential has been used to study the scattering of low-energy electrons from a homonuclear diatomic molecule. The problem involves a potential that may be written in prolate spheroidal coordinates as[1]

$$V(\xi,\eta,\varphi) = [U(\xi) + W(\eta)]/(\xi^2 - \eta^2) \tag{1.1}$$

for which the Schrödinger equation is separable. Although angular momentum is not conserved in this potential, one can determine an operator[2]

$$\mathscr{L} = L^2 + \Omega(\xi,\eta),$$

with

$$\Omega = \frac{(\xi^2-1)W(\eta) - (1-\eta^2)U(\xi)}{\xi^2-\eta^2}, \tag{1.2}$$

which is conserved. It may be interesting to look for a similar case in other coordinate systems. We present here a similar case with parabolic coordinates

$$\xi = r - Z, \quad \eta = r + Z, \quad \varphi = \varphi. \tag{1.3}$$

We note that in terms of these coordinates,

$$\nabla^2 = \frac{4}{\xi+\eta}\left[\frac{\partial}{\partial\xi}\left(\xi\frac{\partial}{\partial\xi}\right) + \frac{\partial}{\partial\eta}\left(\eta\frac{\partial}{\partial\eta}\right)\right] + \frac{1}{\xi\eta}\frac{\partial^2}{\partial\varphi^2},$$

$$L^2 = -\hbar^2\left[(\eta-\xi)\left(\frac{\partial}{\partial\xi} - \frac{\partial}{\partial\eta}\right) + \eta\xi\left(\frac{\partial}{\partial\xi} - \frac{\partial}{\partial\eta}\right)^2 + \frac{(\eta+\xi)^2}{4\eta\xi}\frac{\partial^2}{\partial\phi^2}\right]. \tag{1.4}$$

We assume that $\Omega \equiv \Omega(\xi,\eta)$. The condition

$$[H,\mathscr{L}] \equiv [H_0 + V(\xi,\eta),\ L^2 + \Omega(\xi+\eta)] = 0, \tag{1.5}$$

gives

$$[H_0,\Omega] - [L^2,V] = 0, \quad H_0 = -\frac{\hbar^2}{2M}\nabla^2. \tag{1.6}$$

Considering a separable form of the parabolic potential

$$U(\xi,\eta) = [2u(\xi) + 2w(\eta)]/(\xi+\eta),$$

we get from (1.6),

$$\begin{aligned}&\left[\frac{2}{M}\left\{\xi\frac{2\Omega(\eta,\xi)}{\partial\xi}\right\} - 2\eta\xi\{w'(\xi) - u'(\eta)\}\right]\frac{\partial}{\partial\xi} + \left[\frac{2}{M}\left\{\eta\frac{\partial\Omega(\eta,\xi)}{\partial\eta}\right\} + 2\eta\xi\{w'(\xi) - u'(\eta)\}\right]\frac{\partial}{\partial\eta}\\ &+ \left[\frac{1}{M}\left\{\xi\frac{\partial^2\Omega(\eta,\xi)}{\partial\xi^2} + \eta\frac{\partial^2\Omega(\eta,\xi)}{\partial\eta^2}\right\} - \eta\xi\{w''(\xi) + u''(\eta)\}\right]\\ &+ \left[\frac{1}{M}\left\{\frac{\partial\Omega(\eta,\xi)}{\partial\xi} + \frac{\partial\Omega(\eta,\xi)}{\partial\eta}\right\} - (\eta-\xi)\{w'(\xi) - u'(\eta)\}\right] = 0.\end{aligned} \tag{1.7}$$

The condition can be satisfied with the choice

$$\frac{1}{M}\left[\xi\frac{\partial\Omega(\eta,\xi)}{\partial\xi}\right] = -\frac{1}{M}\left[\eta\frac{\partial\Omega(\eta,\xi)}{\partial\eta}\right] = \eta\xi\,[w'(\xi) - u'(\eta)], \tag{1.8}$$

which implies

$$w'(\xi) + \tfrac{1}{2}\xi w''(\xi) = u'(\eta) + \tfrac{1}{2}\eta u''(\eta) = -K. \tag{1.9}$$

The solutions are given by

$$w = -C_1/\xi - K\xi, \quad u = -C_2/\eta - K\eta, \tag{1.10}$$

giving for the potential

$$V = -2(C_2\xi + C_1\eta)/\eta\xi(\xi+\eta), \tag{1.11}$$

where $V = U + 2K$. Again, from (1.8) and (1.10), we have

$$\frac{1}{M}\Omega(\eta,\xi) = -\left(\frac{C_1\eta}{\xi} + \frac{C_2\xi}{\eta}\right) + a(\eta) \tag{1.12}$$

and also

$$\xi\frac{\partial\Omega}{\partial\xi} + \eta\frac{\partial\Omega}{\partial\eta} = 0, \tag{1.13}$$

so that $a(\eta) = 0$. Therefore

$$\Omega(\eta,\xi) = -M(C_1\eta/\xi + C_2\xi/\eta). \tag{1.14}$$

The operator $\mathscr{L} = L^2 + \Omega$ is a constant of motion. In the special case $C_1 = C_2 = -b/2$, we have

[a] On leave from University College, Raiganj-733 134, West Bengal, India.

 0022-2488/87/040840-04$02.50

$$V = b/r^2 \sin^2\theta, \tag{1.15}$$

$$\Omega = Mb(1+\cos^2\theta)/(1-\cos^2\theta). \tag{1.16}$$

The eigenvalue of L_z gives another constant of motion and may be used to specify the energy levels. The potential $V(\xi,\eta)$ is noncentral and also singular along the z axis. The singularity may be restricted to only one-half of the z axis by the choice of either C_1 or C_2 vanishing. The potential (1.15), along with an attractive Coulomb potential, permits perfectly defined bound states with some exceptions, as will be shown in the next section. In Sec. III we shall consider the scattering states. Another application of the results will be made to study the problems with Aharonov–Bohm potentials in the last section.

II. BOUND STATES

It will be useful to consider the Schrödinger equation with the potential

$$V = -2a/(\xi+\eta) + 2[(C_2\xi + C_1\eta)/\xi\eta(\xi+\eta)], \tag{2.1}$$

where a Coulomb potential, also separable in these coordinates, is introduced. For $E<0$, the solutions may be written as

$$\psi(\xi,\eta,\varphi) = Nf_1(\xi)f_2(\eta)e^{\pm im\varphi}, \tag{2.2}$$

where the f_i satisfy

$$\frac{1}{f_i}\frac{\partial}{\partial x_i}\left(x_i\frac{\partial f_i}{\partial x_i}\right) - \frac{1}{4}\left(4M\frac{C_i}{\hbar^2}+m^2\right)\frac{1}{x_i} + \frac{ME}{2\hbar^2}x_i = -\sigma_i, \tag{2.3}$$

with $x_1=\xi, x_2=\eta$. The constants σ_i should satisfy the condition

$$\sigma_1+\sigma_2 = Ma/\hbar^2. \tag{2.4}$$

Equation (2.3) can be written in the form

$$\frac{1}{\zeta_i}\frac{\partial}{\partial\zeta_i}\left(\zeta_i\frac{\partial f_i}{\partial\zeta_i}\right) + \left[\frac{\lambda_i}{\zeta_i} - \frac{1}{4} - \frac{S_i^2}{4\zeta_i^2}\right]f_i = 0, \tag{2.5}$$

where

$$\lambda_i = \sigma_i/\alpha, \quad \alpha = (-2ME/\hbar^2)^{1/2}, \qquad \zeta_i = \alpha x_i, \quad S_i^2 = 4MC_i/\hbar^2 + m^2. \tag{2.6}$$

Considering appropriate boundary conditions, the solutions can be written as

$$f_i = e^{-(1/2)\zeta_i}\zeta_i^{(1/2)|S_i|}L_{\nu_i}^{|S_i|}(\zeta_i) \tag{2.7}$$

if $\nu_i = n_i = \lambda_i - \frac{1}{2}(|S_i|+1)$ are non-negative integers. Note that for arbitrary C_i, the $|S_i|$ are not in general integers and hence the f_i are generalized Laguerre functions. The series solution can still terminate, provided n_i are non-negative integers. The eigenvalue condition, therefore, is given by

$$E = -\frac{Ma^2}{2\hbar^2}\left[n_1+n_2+1+\frac{1}{2}\sqrt{m^2+\frac{4MC_1}{\hbar^2}} + \frac{1}{2}\sqrt{m^2+\frac{4MC_2}{\hbar^2}}\right]^{-2}. \tag{2.8}$$

The energy levels are proportional to a^2. The effect of this noncentral potential can best be illustrated by considering a hydrogen atom with an additional potential of type (1.15). We choose $a=e^2$ and $2C_i = b = (a^2/4)d$, and obtain the energy levels as shown in Table I, where E_{NC^+} and E_{NC^-} are energy levels for potentials with a repulsive or attractive noncentral part of the potential and E_C, that due to the Coulomb potential only. Note that for an attractive noncentral potential (i.e., b negative) there is no level with $m < (-2Mb/\hbar^2)^{1/2}$. The $S-$ state, in particular, does not exist for a negative b.

TABLE I. Energy levels for repulsive and attractive noncentral potential with $d = 1\times10^{-2}$/eV. The dots indicate nonexistent states.

m	n_1+n_2	E_{NC^+} eV	E_{NC^-} eV	E_C eV	n
0	0	7.261 16	···	13.605 82	1
1	0	3.188 03	3.654 62	3.401 45	2
0	1	2.424 63	···		
2	0	1.478 33	1.546 94	1.511 76	3
1	1	1.447 50	1.585 41		
0	2	1.198 83	···		
3	0	0.840 84	0.860 13	0.850 36	4
2	1	0.836 20	0.865 14		
1	2	0.823 04	0.881 16		
0	3	0.712 83	···		
4	0	0.540 56	0.547 96	0.544 23	5
3	1	0.539 35	0.549 22		
2	2	0.536 96	0.551 78		
1	3	0.530 17	0.559 91		
0	4	0.472 02	···		

An interesting application of these results may be found in the studies with the Hartmann potential given by

$$V(r) = \bar{\eta}\sigma^2\left(\frac{2a_0}{r} - \bar{\eta}\frac{a_0^2}{r^2\sin^2\theta}\right)\epsilon_0, \tag{2.9}$$

where a_0 is the Bohr radius, ϵ_0 is the ground state energy of the hydrogen atom, and the parameters $\bar{\eta}$ and σ are both positive real numbers. The potential is useful in the studies of ring-shaped molecules, like the benzene molecule. Hartmann and his collaborators[3–5] have studied a number of problems with this potential. It is obvious that the results obtained above are directly applicable to the Hartmann potential. If one makes the substitution

$$a = -2a_0\bar{\eta}\sigma^2\epsilon_0, \tag{2.10}$$

$$2C_1 = 2C_2 = -\bar{\eta}^2\sigma^2a_0^2\epsilon_0, \tag{2.11}$$

the eigenfunctions and the eigenvalues are given by Eqs. (2.7) and (2.8).

It has been shown by Kibler and Negadi[6] that the problem of motion of a particle in the Hartmann potential can be reduced to that of a coupled pair of two-dimensional anharmonic oscillators with inverse quadratic potentials by considering a nonbijective canonical transformation. This transformation (Kustaanheimo–Stiefel transformation[7]) corresponds to a mapping from R^4 into the physical space R^3 defined by the relations

$$x = 2(u_1u_3 - u_2u_4), \qquad y = 2(u_1u_4 + u_2u_3), \qquad z = u_1^2 + u_2^2 - u_3^2 - u_4^2. \tag{2.12}$$

Kibler and Negadi have shown that the Schrödinger equation in this four-dimensional space for the potential (2.9) is equivalant to two two-dimensional nonharmonic oscillators subject to some constraints. However, it is easy to show that the "radial" parts of the two-dimensional oscillators correspond to the degrees of freedom ξ and η. It follows easily by noting that with the substitution

$$\rho^2 = x_i,$$

Eq. (21) of Ref. (6) transforms into our Eq. (2.3). Thus the Kustaanheimo–Stiefel transformation in the present case leads us to the parabolic coordinates, which seem to be the natural coordinates for studying this problem. Also, the constraint (15) of Ref. 6 obtained for the separation constants is identical with the condition (2.4) of the present work. This shows that the motions of the system in the ξ,η coordinates are constrained just as in two coupled anharmonic oscillators with an inverse quadratic potential.

III. SCATTERING STATES

The wave functions of the nonlocalized states ($E>0$) with the potential (2.1) can also be obtained from Eq. (2.3). Since α is imaginary, the series $L_{v_i}^{|s_i|}$ need not terminate in this case. Putting

$$\alpha = -ik, \quad k = (2ME/\hbar^2)^{1/2}, \qquad \lambda_1 + \lambda_2 = iMa/\hbar^2 k, \tag{3.1}$$

we can write

$$F_1(\xi) = N_1 e^{(1/2)ik\xi} \xi^{(1/2)|S_1|} \times F(\tfrac{1}{2}|S_1| + \tfrac{1}{2} - \lambda_1, |S_1| + 1; -ik\xi), \tag{3.2}$$

$$F_2(\eta) = N_2 e^{(1/2)ik\eta} \eta^{(1/2)|S_2|} \times F(\tfrac{1}{2}|S_2| + \tfrac{1}{2} - \lambda_2, |S_2| + 1; -ik\eta), \tag{3.3}$$

where N_1, N_2 are arbitrary constants.

In the scattering problem with the incident wave along the z axis, one has to put $m = 0$. Considering the asymptotic behavior ($\xi,\eta \to \infty$) of the solutions (3.2) and (3.3), we have

$$\Psi \to e^{i(1/2)k(\xi+\eta)} \Gamma(\gamma_1+1)\Gamma(\gamma_2+1) \times \left[\frac{(ik)^{-(1/2)\gamma_1}(ik\xi)^{\lambda_1 - 1/2}}{\Gamma(\lambda_1 + \frac{1}{2}\gamma_1 + \frac{1}{2})} + \frac{(-ik)^{-(1/2)\gamma_1}(-ik\xi)^{-(1/2)-\lambda_1}}{\Gamma(\frac{1}{2}\gamma_1 + \frac{1}{2} - \lambda_1)} e^{-ik\xi} \right] \times \left[\frac{(ik)^{-(1/2)\gamma_2}(ik\eta)^{\lambda_2 - 1/2}}{\Gamma(\lambda_2 + \frac{1}{2}\gamma_2 + \frac{1}{2})} + \frac{(-ik)^{-(1/2)\gamma_2}(-ik\eta)^{-1/2-\lambda_2}}{\Gamma(\frac{1}{2}\gamma_2 + \frac{1}{2} - \lambda_2)} e^{-ik\eta} \right], \tag{3.4}$$

where

$$\nu = i(\lambda_1 + \lambda_2) = -Ma/\hbar^2 k, \tag{3.5}$$

$$\gamma_i = \sqrt{4MC_i/\hbar^2}, \tag{3.6}$$

with $a<0$ and $C_i>0$. If Ψ is to have the required asymptotic behavior, the coefficient of $e^{-ik\eta}$ in the last term must vanish. This can happen only if $\lambda_2 = n + \frac{1}{2}\gamma_2 + \frac{1}{2}$, where $n = 0$, 1, 2, ..., etc. Using Eq. (3.5) this will give $\lambda_1 = -i\nu - n - \frac{1}{2}\gamma_2 - \frac{1}{2}$ and make the factor $\xi^{-1/2-\lambda_1}$ diverge at infinity unless $\gamma_2 = -2n$. The denominator of the third term then becomes $\Gamma(1-n)$, which gives $n = 0$, and hence $\gamma_2 = 0, \lambda_1 = -\frac{1}{2} - i\nu, \lambda_2 = \frac{1}{2}$. Substituting these, we obtain the asymptotic solution

$$\Psi \to \pi\nu/e^2 k^{-(1/2)\gamma_1} \frac{\Gamma(\gamma_1+1)}{\Gamma(1+i\nu+\frac{1}{2}\gamma_1)} \times \left[e^{i[kZ + \nu \ln k(r-Z) + \pi\gamma_1/4]} + \frac{\Gamma(1+i\nu+\frac{1}{2}\gamma_1)}{i\Gamma(-i\nu+\frac{1}{2}\gamma_1)} \times \frac{e^{-i[\nu \ln \sin^2\theta/2 + \pi\gamma_1/4]}}{2k\sin^2\theta/2} \frac{e^{i[kr - \nu \ln 2kr]}}{r} \right]. \tag{3.7}$$

Hence the differential cross section is found to be

$$\frac{d\sigma}{d\Omega} = \frac{[\nu^2 + (\frac{1}{2}\gamma_1)^2]}{4k^2 \sin^4\theta/2}, \tag{3.8}$$

for the potential

$$V(r) = a/r + C_1/r^2(1 - \cos\theta). \tag{3.9}$$

The noncentral potential enhances the scattering due to the Coulomb potential, without altering the angular distribution. Thus whether the scattering potential is pure Coulombic or not cannot be decided by studying the differential scattering cross section with the incident wave along the z axis only. Note that a scattering state exists even if one puts $a = 0$ in (3.8), unlike the case of a bound state. As expected, the logarithmic phase factors drop out of the amplitude.

It may be mentioned that for $C_1 = C_2$ the only acceptable case corresponds to $C_i = 0$, which is the well-known pure Coulomb case. Thus even for a repulsive noncentral term, one cannot define a scattering state with the potential (1.15). For the case of an attractive noncentral potential ($C_i<0$), the situation, of course, is known to be worse. Since the most singular term in Eq. (2.5) is $(1/\xi_i^2) \times (m^2 + 4MC_i/\hbar^2)$ and we must choose $m = 0$ to satisfy the boundary condition at infinity, we get two oscillatory solutions at the origin, and there is no way of defining a scattering state unambiguously.[8] Of course, one may consider the scattering of a beam incident on the potential along a direction different from the z axis.

IV. AHARONOV–BOHM POTENTIAL

Another application of the results of Sec. II may be found in the problems with the Aharonov–Bohm (AB) potential.[9] We take the AB potential in the spherical polar coordinates as

$$A_r = A_\theta = 0 \quad \text{and} \quad A_\varphi = F/2\pi r \sin\theta,$$

where F is the flux being carried inside a thin infinitely long solenoid along the z axis. In the limit of the radius of the solenoid going to zero, with F fixed, we have an idealized AB flux tube. We also place a point charge e at the origin.

The Hamiltonian for a point charge e' bound to this composite system is given by

$$H = -\frac{\hbar^2}{2M}\nabla^2 - \frac{ee'}{r} + \frac{1}{2M} \times \left[\frac{2ie'\hbar F}{2\pi c r^2 \sin^2\theta} \frac{\partial}{\partial\varphi} + \frac{e'^2F^2}{4\pi^2c^2r^2\sin^2\theta} \right]. \tag{4.1}$$

To be applicable to the Hamiltonian (4.1), the results of

Sec. I need to be generalized, and we now consider the potential operator $V(r)$ given by

$$V(r) = \frac{2u(\xi) + 2w(\eta)}{\xi + \eta} f\left(i\frac{\partial}{\partial\varphi}\right). \tag{4.2}$$

It can be shown that

$$\Omega = -M\left(\frac{C_1\eta}{\xi} + \frac{C_2\xi}{\eta}\right) f\left(i\frac{\partial}{\partial\varphi}\right). \tag{4.3}$$

Although one may define an operator $\mathscr{M}$ by

$$\mathscr{M} = r\times(-i\hbar\nabla - e'(A/c)), \tag{4.4}$$

which gives a conserved operator $\mathscr{M}^2$, it is not possible to identify $\mathscr{M}$ as the generator for the rotation group, since the eigenfunctions of $\mathscr{M}^2$ and $\mathscr{M}_z = L_z - eF/2\pi c$ do not provide a basis for a representation of the rotation group.[10]

Roy *et al.*[11] and Ruijsenaars[12] pointed out that the operator $L_z = -i\hbar(\partial/\partial\varphi)(\varphi\in[0,2\pi[)$ is not essentially self-adjoint. They have defined a one-parameter family of self-adjoint extension by imposing the boundary condition

$$\psi(\varphi = 2\pi) = e^{i\theta_0}\psi(\varphi = 0), \tag{4.5}$$

where θ_0 is a real parameter. The choice $\theta_0 = 2n\pi$, n integer, corresponds to continuous wave functions. The eigenfunctions of H in (4.1) can be written as

$$\psi = F_1(\xi)F_2(\eta)e^{im'\varphi}, \tag{4.6}$$

where F_1 and F_2 are solutions of Eqs. (2.5) with

$$m' = m + \theta_0/2\pi, \quad \nu_i = n_i, \tag{4.7}$$

$$S_1^2 = S_2^2 = S^2 = (m' - e'F/2\pi c\hbar)^2. \tag{4.8}$$

Thus

$$\psi = Ne^{-(1/2)\alpha(\xi+\eta)}(\alpha^2\xi\eta)^{(1/2)|S|}L_{\nu_1}^{|S|}(\alpha\xi)L_{\nu_2}^{|S|}(\alpha\eta)e^{im'\varphi}. \tag{4.9}$$

The eigenvalues are given by (with $\theta_0 = 2n\pi$ or otherwise)

$$E_n = \frac{Me^2e'^2}{2\hbar^2}\frac{1}{(n_1 + n_2 + 1 + |S|)^2}, \tag{4.10}$$

which shows that solutions exist for all m', unlike the case treated in Sec. II.

The problem can be solved in spherical polar coordinates also. However, our results show directly that

$$\mathscr{L} = L^2 + \Omega$$

with

$$\Omega = \frac{1}{2}\left(\frac{2ie'\hbar F}{2\pi c}\frac{\partial}{\partial\varphi} + \frac{e'^2F^2}{4\pi^2c^2}\right)\frac{1+\cos^2\theta}{1-\cos^2\theta} \tag{4.11}$$

is a constant of motion. On the other hand, one needs an explanation why the operator $\mathscr{M}^2$ is a constant of motion, since it does not satisfy the Pauli–Fierz criterion here. The explanation actually comes from the relation

$$\mathscr{M}^2 = \mathscr{L} - \frac{e'F}{2\pi c}L_z + \frac{1}{2}\frac{e'^2F^2}{4\pi^2c^2}, \tag{4.12}$$

which shows that $\mathscr{M}^2$ is also a constant of motion.

To summarize, we have shown that exact solutions of the Schrödinger equation with a class of parabolic noncentral potentials may be useful in the study of many physical problems.

ACKNOWLEDGMENTS

One of the authors (A.G) is thankful to the University Grants Commission, New Delhi for a fellowship under the Faculty Improvement Program. The work has been supported partially by the Indian Space Research Organization.

[1] S. Mukherjee and S. S. Chandel, J. Phys. A: Math. Gen. **11**, 1257 (1978).
[2] H. A. Erikson and E. L. Hill, Phys. Rev. **75**, 29 (1949).
[3] H. Hartmann, Theor. Chim. Acta **24**, 201 (1972).
[4] H. Hartmann, R. Schuck, and J. Radtke, Theor. Chim. Acta **42**, 1 (1976).
[5] H. Hartmann and D. Schuok, Int. J. Quantum Chem. **28**, 125 (1980).
[6] M. Kibler and T. Negadi, Int. J. Quantum Chem. **26**, 405 (1984).
[7] P. Kustaanheimo and E. Stiefel, J. Reine Angew. Math. **218**, 204 (1965).
[8] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford U. P., London, 1965), p. 41.
[9] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[10] S. M. Roy and V. Singh, Phys. Rev. Lett. **51**, 2069 (1983).
[11] S. M. Roy and V. Singh, Nuovo Cimento **79**, 391 (1984).
[12] S. N. M. Ruijsenaars, Ann. Phys. (NY) **146**, 1 (1983).

Commutation relations for creation–annihilation operators associated with the quantum nonlinear Schrödinger equation

Eugene Gutkin
Department of Mathematics, University of Southern California, Los Angeles, California 90089

(Received 18 June 1986; accepted for publication 26 November 1986)

Using the method of intertwining operators, commutation relations are rigorously obtained for the creation–annihilation operators associated with the quantum nonlinear Schrödinger equation.

I. INTRODUCTION

The quantum nonlinear Schrödinger equation (NLS),

$$\sqrt{-1}\psi_t = -\psi_{xx} + 2c\psi^+\psi^2 \tag{1}$$

is the best understood example of an integrable quantum field theory in one plus one dimensions. Solving (1) by the quantum inverse scattering transform (IST), various authors[1,2] have introduced the creation and annihilation operators $b^+(k)$, $b(l)$ for the Bethe ansatz eigenstates (BAE) of the Hamiltonian H of NLS and their companions $a(k)$, $a^+(l)$ which are diagonalized by the BAE. The operators $b(k)$, $b^+(k)$ and $a(k)$, $a^+(k)$ are the quantized scattering data for the classical NLS (Refs. 1–3).

Regarding the transformation $\psi(x,t)\to b(k,t)$ as a linearization of (1), one needs to compute the commutation relations for $b(k)$, $b^+(l)$. Due to the singularity of those, some crucial commutation relations do not exist and one introduces the normalized "quantum reflection coefficients" $R^+(k) = b^+(k)a^{-1}(k)$ and tries to compute the commutation relations for $R(k)$, $R^+(l)$ by quantizing the classical ones. Although everybody believes that the commutation relations between $R^+(l)$ and $R(l)$ given in the literature are correct, in their derivations given so far one finds equations like $\int_0^\infty dx\, e^{ikx} = \delta(k)/2$ (Refs. 4–6) which have no mathematical justification.

In this paper we give a rigorous mathematical derivation of the commutation relations for $R^+(k)$, $R(l)$ both in the repulsive and attractive cases. This derivation has nothing to do with the quantum IST but is based instead on the intertwining operators approach to the quantum NLS as developed in Refs. 7–10. It is important to spell out the distinction between our definition of the operators $a(k)$, $b(l)$ and the definition in the literature on the quantum IST (Refs. 2 and 11). We first write the operators $a_0(k)$, $b_0(k)$ (Ref. 9), which have the desired action on the *free eigenstates*, then conjugate them by a suitable intertwining operator[9] to obtain the operators $a(k)$, $b(l)$ with the same action on the BAE. In the quantum IST approach, the operators $a(k)$ and $b(l)$ are the normal ordered transmission and reflection coefficients, respectively, of the classical NLS (Refs. 2 and 11). Although the explicit formulas for $a(k)$ and $b(l)$ obtained by these two recipes are very different,[10] the operators themselves must be equal (at least in the repulsive case $c>0$) because they have the same action on the BAE which are complete in the Fock space if $c>0$ (Ref. 12).

The paper relies especially on the material of Ref. 9, but can be read independently. The commutation relations between $R^+(k)$ and $R(l)$ are important, in particular, because they alone allow to write simple expressions for the integrals of the quantum NLS (Ref. 13).

II. PRELIMINARIES ON THE NLS

Let $\psi(x,t)$ be the solution of (1) and denote by $\psi^+(x,t)$ the adjoint field. Operators $\psi(x,t)$, $\psi^+(y,t)$ satisfy the standard equal time commutation relations and $\psi^+(x,0) = \psi_0^+(x)$, $\psi(x,0) = \psi_0(x)$ are the standard creation–annihilation operators (in the position representation). Let $\psi_0(x,t)$, $\psi_0^+(x,t)$ be the solution of (1) with $c=0$, i.e., the *free fields* and set

$$\begin{aligned} b_0(k) &= \int_{-\infty}^{\infty} dx\, e^{-\sqrt{-1}kx}\psi_0(x), \\ b_0^+(k) &= \int_{-\infty}^{\infty} dx\, e^{\sqrt{-1}kx}\psi_0^+(x). \end{aligned} \tag{2}$$

Then $b_0^+(k)$, $b_0(l)$ are the standard creation–annihilation operators in the momentum representation and they satisfy

$$\begin{aligned} &[b_0(k), b_0(l)] = [b_0^+(k), b_0^+(l)] = 0, \\ &[b_0(k), b_0^+(l)] = 2\pi\delta(k-l). \end{aligned} \tag{3}$$

Operators ψ, ψ^+, b, b^+ act on the Fock space

$$\mathscr{H} = \bigoplus_{N=0}^{\infty} \mathscr{H}_N$$

where the N-particle sector $\mathscr{H}_N = L_2^{\mathrm{sym}}(R^N)$ consists of functions symmetric in N variables. We denote by w elements of the permutation group S_N (N varies from one to infinity) and by $x\to wx$, $k\to wk$ the natural action of w on N-tuples $x=(x_1,\ldots,x_N)$, $k=(k_1,\ldots,k_N)$. Operators $b_0^+(k)$ create the normalized free eigenstates (plane waves)

$$\begin{aligned} &f_0(x_1,\ldots,x_N|k_1,\ldots,k_N) \\ &\quad = f_0(x|k) = (N!)^{-1/2}\sum_w \exp(\sqrt{-1}\langle wk|x\rangle) \end{aligned} \tag{4}$$

of the free Hamiltonian (i.e., $c=0$)

$$H_0 = -\int_{-\infty}^{\infty} dx\, \psi_0^+\psi_{0xx}. \tag{5}$$

The Hamiltonian

$$H = \int_{-\infty}^{\infty} dx[-\psi^+\psi_{xx} + c\psi^{+2}\psi^2] \tag{6}$$

of Eq. (1) preserves every $\mathscr{H}_N$ and the restriction

$$H|_{\mathscr{H}_N} = -\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} + c\sum_{i\neq j}\delta(x_i - x_j) \tag{7}$$

 0022-2488/87/040844-06$02.50

is the famous Bose delta-gas Hamiltonian.[7,12]

We call an operator A on $\mathscr{H}$ (preserving each $\mathscr{H}_N$) intertwining if A conjugates H_0 with H and sends the free states $f_0(\cdot|k)$ into the BAE of H (in the normalization depending on A). Referring the reader to Ref. 10 for the general theory of intertwining operators we consider the specific intertwining operators P and P^{*-1} introduced in Ref. 9. The BAE $f(\cdot|k)=P^{*-1}f_0(\cdot|k)$ of H are given explicitly by

$$f(\cdot|k)|_{x_1>\cdots>x_N} = (N!)^{-1/2}\sum_w w\left\{\prod_{i<j}\frac{c+\sqrt{-1}(k_i-k_j)}{\sqrt{-1}(k_i-k_j)} \times\exp(\sqrt{-1}\langle k|x\rangle)\right\}, \tag{8}$$

where $k=k_1>k_2>\cdots>k_N$ and w in (8) acts on the k variable. In what follows we denote the region $\{x_1\geqslant x_2\geqslant\cdots\geqslant x_N\}$ by C_+.

The operators

$$b^+(k)=P^{*-1}b_0^+(k)P^* \tag{9}$$

are the creation operators for the BAE (8). Equation (9) implies the obvious commutation relations

$$[b^+(k),b^+(l)]=0 \tag{10}$$

but not the commutation relation between $b^+(k)$ and $b(l)$ because P is not unitary, thus

$$b(k)=Pb_0(k)P^{-1}\neq P^{*-1}b_0(k)P^*. \tag{11}$$

Due to the singularity of $b^+(k)$ [they create BAE in the singular normalization (8)] it is questionable whether $b^+(k)$, $b(l)$ have commutation relations and those written so far are incorrect [including (2.47) in Ref. 9].

Define the operators $a_0(k)$ by Eq. (2.32) in Ref. 9 and let $a_0^+(k)$ be the adjoint operator. Then, by Lemma 2.1 in Ref. 9

$$a_0(k)f_0(\cdot|k_1,\ldots,k_N) = \left[\prod_{i=1}^N\frac{c+\sqrt{-1}(k-k_i)}{\sqrt{-1}(k-k_i)}\right]f_0(\cdot|k_1,\ldots,k_N), \tag{12}$$

$$a_0^+(k)f_0(\cdot|k_1,\ldots,k_n) = \left[\prod_{i=1}^N\frac{c+\sqrt{-1}(k_i-k)}{\sqrt{-1}(k_i-k)}\right]f_0(\cdot|k_1,\ldots,k_N). \tag{13}$$

Thus $a_0^+(k)$, $a_0(l)$ are diagonal on the plane waves, therefore they commute between themselves and with the operator P^*P (Ref. 9). By (12) and (13), $a_0(k)$, $a_0^+(k)$ are singular since they blow up on $f_0(\cdot|k_1,\ldots,k_N)$ if $k=k_i$ for some i. We normalize $a_0(k)$ and $a_0^+(k)$ by introducing

$$A_0(k)=a_0(k)[a_0^+(k)a_0(k)]^{-1/2}, \qquad A_0^+(k)=a_0^+(k)[a_0^+(k)a_0(k)]^{-1/2}. \tag{14}$$

Set for $t\neq 0$

$$\zeta(t)=(c+\sqrt{-1}t)/\sqrt{-1}t \tag{15}$$

and for $k\neq k_1,\ldots,k_n$ denote $\zeta(k-k_1)\cdots\zeta(k-k_N)$ by $\zeta(k;k_1,\ldots,k_N)$. By (12)–(15)

$$A_0(k)f_0(\cdot|k_1,\ldots,k_N) = \frac{\zeta(k;k_1,\ldots,k_N)}{|\zeta(k;k_1,\ldots,k_N)|}f_0(\cdot|k_1,\ldots,k_N), \tag{16}$$

$$A_0^+(k)f_0(\cdot|k_1,\ldots,k_N) = \frac{\bar{\zeta}(k;k_1,\ldots,k_N)}{|\zeta(k;k_1,\ldots,k_N)|}f_0(\cdot|k_1,\ldots,k_N). \tag{17}$$

It follows immediately from (14) or from (16) and (17) that

$$A_0^+(k)=A_0(k)^{-1}. \tag{18}$$

Obviously, $a_0(k)$, $a_0^+(l)$, $A_0(k)$, $A_0^+(l)$ all commute. We set

$$a^+(k)=P^{*-1}a_0^+(k)P^*=Pa_0^+(k)P^{-1}, \tag{19}$$

$$A^+(k)=P^{*-1}A_0^+(k)P^*=PA_0^+(k)P^{-1}. \tag{20}$$

Equations (19) and (20) hold because $a_0^+(k)$ and $A_0^+(k)$ commute with P^*P. Operators $a(k)$, $a^+(l)$, $A(k)$, $A^+(l)$ are diagonalized by the BAE $f(\cdot|k_1,\ldots,k_N)$, therefore they all commute and we have

$$a(k)f(\cdot|k_1,\ldots,k_N)=\zeta(k;k_1,\ldots,k_N)f(\cdot|k_1,\ldots,k_N), \tag{21}$$

$$A(k)f(\cdot|k_1,\ldots,k_n) = \frac{\zeta(k;k_1,\ldots,k_N)}{|\zeta(k;k_1,\ldots,k_N)|}f(\cdot|k_1,\ldots,k_N). \tag{22}$$

By (18) and (20)

$$A^+(k)=A(k)^{-1}. \tag{23}$$

We set

$$R^+(k)=b^+(k)a^{-1}(k). \tag{24}$$

Then

$$R(k)=a^+(k)^{-1}b(k). \tag{25}$$

III. COMMUTATION RELATIONS

First, we establish the commutation relations between $b_0(k)$, $b_0^+(k)$ and $A_0(l)$, $A_0^+(l)$. We fix $c\neq 0$.

Proposition 1: For $k\neq l$,

$$A_0(k)b_0^+(l) = (\zeta(k-l)/|\zeta(k-l)|)b_0^+(l)A_0(k), \tag{26}$$

$$A_0^+(k)b_0^+(l) = (\zeta(l-k)/|\zeta(l-k)|)b_0^+(l)A_0^+(k), \tag{27}$$

$$b_0(l)A_0(k) = (\zeta(k-l)/|\zeta(k-l)|)A_0(k)b_0(l), \tag{28}$$

$$b_0(l)A_0^+(k) = (\zeta(l-k)/|\zeta(l-k)|)A_0^+(k)b_0(l). \tag{29}$$

Proof: We follow the ideas of the proof of Lemma 2.1 in Ref. 9. We set

$$\alpha_0(k)=(2\pi)^{-1}\int_{-\infty}^{\infty}dr[\log\zeta(k-r)]b_0^+(r)b_0(r). \tag{30}$$

Then

$$a_0(k)=\exp[\alpha_0(k)], \quad a_0^+(k)=\exp[\alpha_0^+(k)] \tag{31}$$

and the operators $\alpha_0(k)$, $\alpha_0^+(l)$ commute for all k and l. The commutativity and (31) imply that

$$A_0(k)=\exp\{\tfrac{1}{2}[\alpha_0(k)-\alpha_0^+(k)]\}, \tag{32}$$

$$A_0^+(k) = \exp\{\tfrac{1}{2}[\alpha_0^+(k) - \alpha_0(k)]\} = A^{-1}(k). \quad (33)$$

The commutation relations

$$[\alpha_0(k), b_0(l)] = -[\log \zeta(k-l)]b_0(l), \quad (34)$$

$$[\alpha_0(k), b_0^+(l)] = [\log \zeta(k-l)]b_0^+(l) \quad (35)$$

obtained in Ref. 9, Lemma 2.1, imply that

$$\begin{aligned}&[\tfrac{1}{2}(\alpha_0(k) - \alpha_0^+(k)), b_0(l)] \\ &\quad = -\tfrac{1}{2}\log \zeta(k-l)b_0(l) - \tfrac{1}{2}(-\log \zeta(l-k)b_0(l)) \\ &\quad = \tfrac{1}{2}[\log \zeta(l-k) - \log \zeta(k-l)]b_0(l). \end{aligned} \quad (36)$$

Exponentiating (36) we obtain

$$\begin{aligned}&A_0(k)b_0(l)A_0^{-1}(k) \\ &\quad = [\zeta(l-k)/\zeta(k-l)]^{1/2}b_0(l) \\ &\quad = [\zeta(l-k)^2/\zeta(k-l)\zeta(l-k)]^{1/2}b_0(l) \\ &\quad = [\zeta(l-k)/|\zeta(l-k)|]b_0(l), \end{aligned}$$

where we have used that $\zeta(-t) = \bar{\zeta}(t)$. The last equation means that

$$\begin{aligned}&A_0(k)b_0(l) \\ &\quad = (\zeta(l-k)/|\zeta(l-k)|)b_0(l)A_0(k), \end{aligned} \quad (37)$$

which is equivalent to (28). Equation (26) is proved the same way and (27), (29) immediately follow from (28) and (26), respectively.

By taking the limit $k - l \to 0$ we will extend (26)–(29) to the case $k = l$ but the limit commutation relations depend on whether $k - l \to 0+$ or $k - l \to 0-$. To express these relations in concise form we introduce the following notation. For any function $f(k)$ we set $f(k+) = \lim f(k')$ as $k' \to k^+$ and $f(k-) = \lim f(k')$ as $k' \to k-$. Recall that for $c \neq 0$, $\operatorname{sgn} c = 1$ if $c > 0$ and $\operatorname{sgn} c = -1$ if $c < 0$.

Corollary 1: For a fixed $c \neq 0$ and for any k the following relations hold:

$$A_0(k\pm)b_0^+(k) \pm (\operatorname{sgn} c)\sqrt{-1}\,b_0^+(k)A_0(k\pm) = 0, \quad (38)$$

$$A_0(k\pm)b_0(k) \mp (\operatorname{sgn} c)\sqrt{-1}\,b_0(k)A_0(k\pm) = 0, \quad (39)$$

$$A_0^+(k\pm)b_0^+(k) \mp (\operatorname{sgn} c)\sqrt{-1}\,b_0^+(k)A_0^+(k\pm) = 0, \quad (40)$$

$$A_0(k\pm)b_0(k) \pm (\operatorname{sgn} c)\sqrt{-1}\,b_0(k)A_0(k\pm) = 0. \quad (41)$$

Proof: Since $\zeta(t) = 1 - c\sqrt{-1}t^{-1}$ goes to $-(\operatorname{sgn} c)\sqrt{-1}(+\infty)$ when $t \to 0+$ and to $(\operatorname{sgn} c)\sqrt{-1}(+\infty)$ when $t \to 0-$, we have $\zeta(t)/|\zeta(t)| \to -(\operatorname{sgn} c)\sqrt{-1}$ when $t \to 0+$ and $\zeta(t)/|\zeta(t)| \to (\operatorname{sgn} c)\sqrt{-1}$ when $t \to 0-$. Moreover, $\zeta(t)/|\zeta(t)|$ varies on the unit circle continuously and monotonically from 1 to $(\operatorname{sgn} c)\sqrt{-1}$ when t runs from $-\infty$ to $0-$. Then $\zeta(t)/|\zeta(t)|$ jumps from $(\operatorname{sgn} c)\sqrt{-1}$ to $-(\operatorname{sgn} c)\sqrt{-1}$ and completes the circle in the same fashion as t varies from $0+$ to $+\infty$.

Taking the limit $k - l \to 0\pm$ in (26)–(29) we obtain (38)–(41).

We have remarked earlier that the singularity of the creation operators $b^+(k)$ is due to the singularity of the intertwining operator P^{*-1} which sends the free states $f_0(\cdot|k)$ normalized to the δ-function into the BAE (8) which have a singular normalization. Following an idea of Ref. 9 we take the polar decomposition of P^{*-1},

$$P^{*-1} = [P^{*-1}(P^*P)^{1/2}](P^*P)^{-1/2}, \quad (42)$$

where $Q = P^{*-1}(P^*P)^{1/2}$ is unitary (isometry if $c < 0$) and $(P^*P)^{-1/2}$ is positive definite. The operator Q is a normalized intertwining operator and the operator $(P^*P)^{-1/2}$ is described in the following Lemma.

Lemma 1: (i) The operator $(P^*P)^{-1/2}$ is diagonalized by the free eigenstates $f_0(\cdot|k)$ and

$$\begin{aligned}&(P^*P)^{-1/2}f_0(\cdot|k_1,\ldots,k_N) \\ &\quad \times\left[\prod_{i<j}|\zeta(k_i - k_j)|\right]f_0(\cdot|k_1,\ldots,k_N). \end{aligned} \quad (43)$$

(ii) The operator $(P^*P)^{-1/2}$ commutes with $a_0(k)$, $a_0^+(k)$, $A_0(k)$, $A_0^+(k)$ for all k and

$$b_0(k)(P^*P)^{-1/2} = [a_0^+(k)a_0(k)]^{1/2}(P^*P)^{-1/2}b_0(k), \quad (44)$$

$$\begin{aligned}&(P^*P)^{-1/2}b_0^+(k) \\ &\quad = b_0^+(k)[a_0^+(k)a_0(k)]^{1/2}(P^*P)^{-1/2}. \end{aligned} \quad (45)$$

Proof: Equation (43) follows immediately from Ref. 9, (1.52) and it implies that $(P^*P)^{-1/2}$ commutes with any operators diagonalized by $f_0(\cdot|k)$. Equations (44) and (45) are proved in the same way as (2.53) and (2.54) in Ref. 9.

Recall that the normalized intertwining operator $Q = P^{*-1}(P^*P)^{1/2}$ is unitary if $c \geqslant 0$ (Ref. 9). If $c < 0$, Q is an isometry of $\mathscr{H}$ on the space $\mathscr{H}_{ac}$ of the absolutely continuous spectrum of H and we denote by $Q^{-1} = Q^*$ the left inverse of Q. We denote by 1_{ac} the orthogonal projection of $\mathscr{H}$ on $\mathscr{H}_{ac}$.

Corollary 2: For any k we have

$$R^+(k) = Qb_0^+(k)A_0^+(k)Q^{-1}, \quad (46)$$

$$R(k) = QA_0(k)b_0(k)Q^{-1}. \quad (47)$$

Proof: By (24), (19), and (9)

$$\begin{aligned}R^+(k) &= P^{*-1}b_0^+(k)a_0^{-1}(k)P^* \\ &= Q(P^*P)^{-1/2}b_0^+(k)a_0^{-1}(k)(P^*P)^{1/2}Q^{-1}. \end{aligned}$$

Using (45) and that $(P^*P)^{-1/2}$ commutes with $a_0(k)$, $a_0^+(k)$, we obtain

$$\begin{aligned}R^+(k) &= Qb_0^+(k)a_0^{-1}(k)[a_0^+(k)a_0(k)]^{1/2}Q^{-1} \\ &= Qb_0^+(k)A_0^{-1}(k)Q^{-1} \end{aligned}$$

which proves (46). Taking the adjoint and using (18) we obtain (47). Set

$$R_0(k) = A_0(k)b_0(k), \quad R_0^+(k) = b_0^+(k)A_0^+(k) \quad (48)$$

and rewrite (47) and (46) as

$$R(k) = QR_0(k)Q^{-1}, \quad (49)$$

$$R^+(k) = QR^+(k)Q^{-1}, \quad (50)$$

respectively. Because of the unitarity (isometry if $c < 0$) of Q, it suffices to establish the commutation relations for the operators $R_0(k)$, $R_0^+(l)$.

We fix $c \neq 0$ and set

$$u(t) = [\zeta(t)/|\zeta(t)|]^2. \quad (51)$$

Properties of $\zeta(t)$ which we have discussed in the proof of Corollary 1, imply that $t \to u(t)$ is a continuous mapping of the real line into the unit circle, that when t runs from $-\infty$ to ∞, $u(t)$ runs the circle from 1 to 1 counterclockwise if $c>0$ and clockwise if $c<0$. Finally, $u(-t)=\bar{u}(t)$ and $u(0)=-1$. We will need the following formulas for $u(t)$:

$$\zeta(t)/\zeta(-t)=\zeta(t)^2/\zeta(t)\zeta(-t) = [\zeta(t)/|\zeta(t)|]^2 = u(t). \tag{52}$$

Thus

$$u(t)=\frac{\zeta(t)}{\bar{\zeta}(t)}=\frac{\zeta(t)}{\zeta(-t)} =\frac{c+\sqrt{-1}t}{\sqrt{-1}t}\left(\frac{c-\sqrt{-1}t}{-\sqrt{-1}t}\right)^{-1} = -\frac{c+\sqrt{-1}t}{c-\sqrt{-1}t}=\frac{t-c\sqrt{-1}}{t+c\sqrt{-1}}. \tag{53}$$

Theorem 1: Operators $R_0(k), R_0^+(l)$ satisfy the following commutation relations for all k and l

$$R_0^+(l)R_0^+(k)=u(k-l)R_0^+(k)R_0^+(l), \tag{54}$$

$$R_0(l)R_0(k)=u(k-l)R_0(k)R_0(l), \tag{55}$$

$$R_0(k)R_0^+(l)=u(k-l)R_0^+(l)R_0(k)+2\pi\delta(k-l). \tag{56}$$

Proof: We have, by (48)

$$R_0(k)R_0^+(l)=A_0(k)b_0(k)b_0^+(l)A_0^+(l), \tag{57}$$

$$R_0^+(l)R_0(k)=b_0^+(l)A_0^+(l)A_0(k)b_0(k). \tag{58}$$

Interchanging $A_0^+(l)$ with $A_0(k)$ on the right-hand side of (58) and assuming that $k \neq l$ we use the commutation relations of Proposition 1 to get

$$R_0^+(l)R_0(k) = [\zeta(l-k)/|\zeta(l-k)|]^2 A_0(k)b_0^+(l)b_0(k)A_0^+(l) = u(k-l)^{-1}A_0(k)b_0^+(l)b_0(k)A_0^+(l). \tag{59}$$

Since $u(t)$ is continuous at $t=0$, (59) extends, by continuity, to all k and l. The canonical commutation relations (3) for $b_0(k), b_0^+(l)$ yield

$$R_0^+(l)R_0(k)=u(k-l)^{-1}A_0(k)b_0(k)b_0^+(l)A_0^+(l) - 2\pi\delta(k-l)u(k-l)^{-1}A_0(k)A_0^+(l). \tag{60}$$

Since $u(0)=-1$ and $A_0(k)A_0^+(k)=1$, (60) becomes

$$R_0^+(l)R_0(k)=u(k-l)^{-1}R_0(k)R_0^+(l)+2\pi\delta(k-l), \tag{61}$$

which is equivalent to (56). Equation (54) is proved by an analogous but simpler computation which we leave to the reader and (55) follows from (54).

Theorem 2: The operators $R(k), R^+(l)$ satisfy the commutation relations

$$R^+(l)R^+(k)=u(k-l)R^+(k)R^+(l), \tag{62}$$

$$R(l)R(k)=u(k-l)R(k)R(l), \tag{63}$$

$$R(k)R^+(l)=u(k-l)R^+(l)R(k)+2\pi\delta(k-l)1_{ac}. \tag{64}$$

Proof: By (46) and (47), $R(k)R^+(l) = QR_0(k)R_0^+(l)Q^{-1}$ which, by (56), is equal to

$$Q[u(k-l)R_0^+(l)R_0(k)+2\pi\delta(k-l)]Q^{-1} = u(k-l)QR_0^+(l)Q^{-1}QR_0(k)Q^{-1} + 2\pi\delta(k-l)QQ^{-1} = u(k-l)R^+(l)R(k)+2\pi\delta(k-l)QQ^{-1}.$$

Since $Q^{-1}=Q^*$ is the left inverse of Q, QQ^{-1} is equal to the orthogonal projection on the range of Q which is $\mathscr{H}_{ac}$. Thus $QQ^{-1}=1_{ac}$. This proves (64). Proofs of (62) and (63) are analogous but simpler and we leave them to the reader.

Corollary 3: For $c>0$ (the repulsive case) the operators $R^+(k), R(l)$ satisfy

$$R(k)R^+(l)=u(k-l)R^+(l)R(k)+2\pi\delta(k-l). \tag{65}$$

Proof: For $c>0$ the Hamiltonian H has a purely absolutely continuous spectrum,[12] i.e., $1_{ac}=1$ and (64) becomes (65).

IV. SCATTERING STATES

It is obvious that the operators $R^+(k)$ create BAE in some particular normalization. We show in this section that $R^+(k)$ are the creation operators for the scattering states $f_{\text{in}}(\cdot|k)$ and $f_{\text{out}}(\cdot|k)$ of the Hamiltonian H. Some of the equations below have been obtained earlier using the quantum IST approach (see, e.g., Ref. 11). The purpose of this section is to derive these results using the more rigorous technique of intertwining operators.

Recall (see, e.g., Ref. 9) that for any $k=(k_1>\cdots>k_N)$, the scattering states $f_{\text{in}}(\cdot|k)$ and $f_{\text{out}}(\cdot|k)$ are given by

$$f_{\text{in}}(x|k)|_{C_+} = (N!)^{-1/2}\left[\exp(\sqrt{-1}\langle w_0 k|x\rangle) + \sum_{w\neq w_0} d(w,k)\exp(\sqrt{-1}\langle wk|x\rangle)\right], \tag{66}$$

$$f_{\text{out}}(x|k)|_{C_+} = (N!)^{-1/2}\left[\exp(\sqrt{-1}\langle k|x\rangle) + \sum_{w\neq 1} c(w,k)\exp(\sqrt{-1}\langle wk|x\rangle)\right], \tag{67}$$

where w_0 is the longest permutation of N symbols. For any $k=(k_1>\cdots>k_N)$ the states $f_{\text{in}}(\cdot|k)$ and $f_{\text{out}}(\cdot|k)$ are proportional and

$$f_{\text{out}}(\cdot|k)=S(k_1,\ldots,k_N)f_{\text{in}}(\cdot|k), \tag{68}$$

where $S(k_1,\ldots,k_N)$ is the phase shift. We denote by $|0\rangle$ the vacuum vector.

Theorem 3: (i) Let $k_1>\cdots>k_N$ be arbitrary. Then

$$R^+(k_1)\cdots R^+(k_N)|0\rangle = f_{\text{out}}(\cdot|k_1,\ldots,k_N), \tag{69}$$

$$R^+(k_N)\cdots R^+(k_1)|0\rangle = f_{\text{in}}(\cdot|k_1,\ldots,k_N). \tag{70}$$

(ii) For any $k_1>\cdots>k_N$ and any k,

$$R(k)f_{\text{out}}(\cdot|k_1,\ldots,k_N)$$
$$= 2\pi \sum_{i=1}^{N} u(k-k_1)\cdots u(k-k_{i-1})\delta(k-k_i)$$
$$\times f_{\text{out}}(\cdot|k_1,\ldots,\hat{k}_i,\ldots,k_N). \quad (71)$$

Proof: Equation (69) is trivially true if $N=1$ and we prove it by induction on N. By the inductive assumption

$$R^+(k_2)\cdots R^+(k_N)|0\rangle = f_{\text{out}}(\cdot|k_2,\ldots,k_N). \quad (72)$$

Recall that $f(\cdot|k)$ are the BAE in the singular normalization (8). It is straightforward to check from (8) that for $k=(k_1>\cdots>k_N)$,

$$f(\cdot|k) = \left[\prod_{i<j}\zeta(k_i-k_j)\right] f_{\text{out}}(\cdot|k)$$
$$= \left[\prod_{i<j}\bar{\zeta}(k_i-k_j)\right] f_{\text{in}}(\cdot|k). \quad (73)$$

Thus by (72) and (73)

$$R^+(k_1)\cdots R^+(k_N)|0\rangle$$
$$= R^+(k_1) f_{\text{out}}(\cdot|k_2,\ldots,k_N)$$
$$= R^+(k_1)\left[\prod_{2<i<j<N}\zeta^{-1}(k_i-k_j)\right] f(\cdot|k_2,\ldots,k_N)$$
$$= \left[\prod_{2<i<j<N}\zeta^{-1}(k_i-k_j)\right] b^+(k_1)a^{-1}(k_1)$$
$$\times f(\cdot|k_2,\ldots,k_N)$$
$$= \left[\prod_{2<i<j<N}\zeta^{-1}(k_i-k_j)\right] b^+(k_1)$$
$$\times\left[\prod_{p=2}^{N}\zeta^{-1}(k_1-k_p)\right] f(\cdot|k_2,\ldots,k_N)$$
$$= \left[\prod_{1<i<j<N}\zeta^{-1}(k_i-k_j)\right] b^+(k_1) f(\cdot|k_2,\ldots,k_N)$$
$$= \left[\prod_{i<j}\zeta^{-1}(k_i-k_j)\right] f(\cdot|k_1,\ldots,k_N)$$
$$= f_{\text{out}}(\cdot|k_1,\ldots,k_N),$$

which proves (69). By (62)

$$R^+(k_N)\cdots R^+(k_1)$$
$$= \left[\prod_{i<j}u(k_i-k_j)\right] R^+(k_1)\cdots R^+(k_N). \quad (74)$$

By (69), (73), and (53)

$$R^+(k_N)\cdots R^+(k_1)|0\rangle$$
$$= \left[\prod_{i<j}u(k_i-k_j)\right] R^+(k_1)\cdots R^+(k_N)|0\rangle$$
$$= \left[\prod_{i<j}u(k_i-k_j)\right] f_{\text{out}}(\cdot|k_1,\ldots,k_N)$$
$$= \left[\prod_{i<j}u(k_i-k_j)\zeta^{-1}(k_i-k_j)\bar{\zeta}(k_i-k_j)\right]$$
$$\times f_{\text{in}}(\cdot|k_1,\ldots,k_N)$$
$$= f_{\text{in}}(\cdot|k_1,\ldots,k_N),$$

which proves (70). Equation (71) follows from (64) and (69) by a straightforward induction on N. We spare the details.

The following assertion was shown in the course of proof of Theorem 3.

Corollary 4: The phase shifts $S(k_1,\ldots,k_N)$ (68) for the delta-Bose gas Hamiltonian H are given by

$$S(k_1,\ldots,k_N) = \prod_{i<j} u^{-1}(k_i-k_j). \quad (75)$$

Equation (75), which is well known, means, in particular, that the N-body S matrix of the delta-Bose gas factorizes into the product of two-body S matrices $S(k-l) = u^{-1}(k-l)$. Factorizable S matrices have been much studied lately.[14]

V. DISCUSSION AND CONCLUSION

Using the method of interwining operators we have rigorously obtained the commutation relations for the creation–annihilation operators $R^+(k), R(l)$ much used in the literature on the quantum NLS and the quantum IST (Refs. 5, 6, 11, 13, 15, and 16). These commutation relations depend on the coupling constant c via the coefficient

$$u(k-l) = -\frac{c+\sqrt{-1}(k-l)}{c-\sqrt{-1}(k-l)} = u_c(k-l) \quad (76)$$

[see (53) and (62)–(64)]. The function $t\to u_c(t)$ behaves in a singular way in the $c\to 0$ limit. This fact together with $u_c(0) = -1$ led to speculations about the singular character of the zero coupling limit and about the "fermionic character" of the delta-Bose gas.[13,15] As we have seen in the paper the "fermionic character" of the commutation relations (62)–(64) is due to the very peculiar normalization of the BAE creation operators $R^+(k)$. Moreover, the creation–annihilation operators $R_0^+(k)$, $R_0(l)$ for the free eigenstates satisfy the same "fermionic" commutation relations (54)–(56). Thus the "fermionic character" of the commutation relations has nothing to do with the "fermionic nature" of the delta-Bose gas.

Let us introduce another set of creation–annihilation operators for the BAE of the delta-Bose gas. Recall that $Q = P^{*-1}(P^*P)^{1/2}$ is the normalized intertwining operator and that $b_0^+(k)$, $b_0(l)$ are the creation–annihilation operators for the plane waves (4). Set

$$\psi_n^+(k) = Qb_0^+(k)Q^{-1}, \quad \psi_n(k) = Qb_0(k)Q^{-1} \quad (77)$$

be the normalized BAE creation–annihilation operators.

Theorem 4: (i) The operators $\psi_n(k)$, $\psi_n^+(l)$ satisfy the commutation relations

$$[\psi_n(k),\psi_n(l)] = [\psi_n^+(k),\psi_n^+(l)] = 0, \quad (78)$$
$$[\psi_n(k),\psi_n^+(l)] = 2\pi\delta(k-l)1_{ac}. \quad (79)$$

In particular, for $c\geqslant 0$, $\psi_n(k)$, $\psi_n^+(l)$ satisfy the standard commutation relations (3).

(ii) The operators $\psi_n^+(k)$ create the BAE $g(\cdot|k_1,\ldots,k_N)$ in the following normalization:

$$\psi_n^+(k_1)\cdots\psi_n^+(k_N)|0\rangle$$
$$= g(\cdot|k_1,\ldots,k_N) = (N!)^{-1/2}\sum_w w$$
$$\times\left\{\left[\prod_{i<j}\frac{\zeta(k_i-k_j)}{|\zeta(k_i-k_j)|}\right]\exp(\sqrt{-1}\langle k|x\rangle)\right\}. \quad (80)$$

Proof: Assertion (i) follows immediately from unitarity (isometry for $c<0$) of Q and (3). Assertion (ii) is proved by a straightforward induction on N. It is a special case of a general formula for the intertwining operators and the BAE [see Ref. 10, Eq. (2.31)].

Although $\psi_n^+(k)$, $\psi_n(l)$ are creation–annihilation operators for the quantum NLS, there is nothing fermionic about their commutation relations (79) and it was shown in detail in Ref. 9 that their zero coupling limit is nonsingular.

To conclude, I hope that this paper helps to dispel some of the mysteries of the quantum NLS.

ACKNOWLEDGMENTS

The paper was written while the author was visiting Rockefeller University. It is a pleasure to thank Ken Case and E. G. D. Cohen for their warm hospitality and Diane Lott for the excellent typing. Ken Case also pointed out to me that Eq. (2.47) in (Ref. 9) is incorrect (fortunately it is not used anywhere in Ref. 9) and suggested the subject of the present paper.

This paper was partially supported by NSF Grant No. DMS 84-03238.

[1]E. K. Sklyanin, "Inverse scattering method and the quantum nonlinear Schrödinger equation," Dokl. Akad. Nauk. USSR **244**, 107 (1979).

[2]H. B. Thacker and D. Wilkinson, "Inverse scattering transform as an operator method in quantum field theory," Phys. Rev. D **19**, 3660 (1979).

[3]V. E. Zakharov and S. V. Manakov, "On the complete integrability of a nonlinear Schrödinger equation," Theor. Math. Phys. **19**, 551 (1975).

[4]B. Davies, "Second quantization of the nonlinear Schrödinger equation," J. Phys. A: Math. Gen. **14**, 2631 (1981).

[5]B. Davies, "The quantum inverse method and the nonlinear Schrödinger equation," Inv. Probl. **2**, 141 (1986).

[6]H. B. Thacker, "Exact integrability in quantum field theory and statistical systems," Rev. Mod. Phys. **53**, 253 (1981).

[7]E. Gutkin, "Integrable systems with delta-potential," Duke Math. J. **49**, 1 (1982).

[8]E. Gutkin, "Conservation laws for the nonlinear Schrödinger equation," Ann. Inst. H. Poincaré **2**, 67 (1985).

[9]E. Gutkin, "Quantum nonlinear Schrödinger equation. I. Intertwining operators," Ann. Inst. H. Poincaré **3**, 285 (1986).

[10]E. Gutkin, "Quantum nonlinear Schrödinger equation. II. Explicit solution," to appear in J. Funct. Analysis.

[11]E. K. Sklyanin, "Quantum version of the inverse scattering method," J. Sov. Math. **19**, 1546 (1982).

[12]F. A. Berezin, G. P. Pohil, and V. M. Finkelberg, "Schrödinger equation for a system of one-dimensional particles with point interaction," Vestn. Mosk. Univ. **1**, 21 (1964).

[13]K. M. Case, " 'Polynomial constants' for the quantized NLS equation," J. Math. Phys. **25**, 2306 (1984).

[14]A. B. Zamolodchikov and A. B. Zamolodchikov, "Factorized S-matrices in two dimensions as the exact solutions of certain relativistic quantum field theory models," Ann. Phys. (NY) **120**, 253 (1979).

[15]L. D. Faddeev, "Integrable models in $(1+1)$-dimensional quantum field theory," *Les Houches Summer School, Session 39 (1982)* (Elsevier, New York, 1984).

[16]P. P. Kulish and E. K. Sklyanin, "Quantum spectral transform method. Recent developments," in *Lecture Notes in Physics*, Vol. 151 (Springer, Berlin, 1981).

Unitary implementability of gauge transformations for the Dirac operator and the Schwinger term

P. Falkensteiner and H. Grosse
Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, Austria

(Received 10 June 1986; accepted for publication 10 December 1986)

Recent interest in external field problems led to the determination of the classes of unitarily implementable gauge, axial gauge, and chiral transformations for the Dirac operator on a finite interval. Various charge quantization conditions are obtained. The algebra of charge operators is worked out and the well-known Schwinger term is identified.

I. INTRODUCTION

Since it was realized that fractional charged states may occur within quantum field theoretical models,[1] the quantization of Fermi fields interacting with external solitonlike potentials was studied extensively.[2] It was realized that not only the shape of the potential but also boundary conditions turn out to be essential.

In all the treatments of the external field problem, integral charged excited states occur, but the charge of the ground state may become fractional. It may be fixed using charge conjugation invariance[1]; one may calculate a spectral asymmetry[3] or one may define an index in a suitable manner.[4]

On the other hand it has been realized that already a one-dimensional soliton potential compared to a constant external potential leads to unitarily inequivalent representations of the canonical anticommutation relations (CAR) (Ref. 5), which shows that comparing the ground state charges for such problems has to be done carefully (see Sec. IV).

In this paper we treat the "free" Dirac operator on a finite interval

$$i\frac{\partial}{\partial t}\Phi(t,x)=\sigma_3\frac{1}{i}\frac{\partial}{\partial x}\Phi(t,x), \tag{1.1}$$

where Φ denotes a two-component spinor depending on $x\in[0,1]$ and time; the associated time independent Dirac equation is given by

$$\begin{aligned}&(h\psi)(x)=\sigma_3\frac{1}{i}\frac{d}{dx}\psi(x)=\lambda\psi(x),\\&\Phi(t,x)=e^{-i\lambda t}\psi(x),\end{aligned} \tag{1.2}$$

where we work in the Hilbert space $\mathscr{H}=L^2([0,1],dx)\otimes\mathbb{C}^2$, with scalar product denoted by $\langle\cdot\rangle$. With the help of boundary conditions we simulate interactions. It is easy to see that the deficiency indices of h, on a domain of absolutely continuous (a.c.) functions with $h\psi\in\mathscr{H}$ and vanishing at the ends of the interval are (2,2); therefore there exists a four-parameter family of self-adjoint extensions. Introducing a unitary matrix U allows to parametrize all extensions of h, denoted by h_U, by requiring that

$$\begin{aligned}&\begin{pmatrix}\psi_1(1)\\ \psi_2(0)\end{pmatrix}=U\begin{pmatrix}\psi_2(1)\\ \psi_1(0)\end{pmatrix},\\&U=e^{i\alpha}\begin{pmatrix}\cos\Theta e^{i\beta} & \sin\Theta e^{i\gamma}\\ \sin\Theta e^{-i\gamma} & -\cos\Theta e^{-i\beta}\end{pmatrix},\end{aligned} \tag{1.3}$$

with $0\leqslant\Theta<\pi$, $0\leqslant\alpha,\beta,\gamma<2\pi$. Here h_U is defined by the symbol (1.2) with the domain of definition given by

$$\mathscr{D}_U=\{\psi\in\mathscr{H}\,|\,\psi \text{ a.c., } h\psi\in\mathscr{H},\ \psi \text{ fulfills (1.3)}\}. \tag{1.4}$$

Solving (1.2) with boundary condition (1.3) is trivial and leads to a spectrum $\{\lambda_n{}^j\}$, $n\in\mathbb{Z}$, $j=1,2$, and to associated eigensolutions $(\xi^j_{n,1}(x),\xi^j_{n,2}(x))$:

$$\begin{aligned}&\lambda_n{}^j=\alpha-\sigma_j\epsilon+2\pi n,\quad \sigma_j=+1 \text{ for } j=1,\\&\sigma_j=-1 \text{ for } j=2,\\&\xi^j_{n,a}(x)=f^j{}_a e^{i\sigma_a\lambda_n{}^j x},\quad a=1,2,\end{aligned} \tag{1.5}$$

with $\cos\epsilon=\sin\Theta\cos\gamma$, $0\leqslant\epsilon\leqslant\pi$ and the $f^j{}_a$ fulfill

$$\begin{pmatrix}e^{i\lambda_0{}^j}-\sin\Theta e^{i(\alpha+\gamma)} & -\cos\Theta e^{i(\alpha+\beta-\lambda_0{}^j)}\\ \cos\Theta e^{i(\alpha-\beta)} & 1-\sin\Theta e^{i(\alpha-\gamma-\lambda_0{}^j)}\end{pmatrix}\begin{pmatrix}f^j{}_1\\ f^j{}_2\end{pmatrix}=0. \tag{1.6}$$

Let us remark that a suitable defined ground state charge is given by α/π and varies therefore continuously with α (Ref. 6).

Above we have obtained a four-parameter family of solvable Dirac operators for which we may define projection operators onto the positive and negative energy parts

$$\begin{aligned}&P_+(x,y)=\sum_{\lambda_n{}^j>0}\xi_n{}^j(x)\xi_n{}^{j*}(y),\\&P_-(x,y)=\sum_{\lambda_n{}^j<0}\xi_n{}^j(x)\xi_n{}^{j*}(y).\end{aligned} \tag{1.7}$$

We note that in all of what follows we may shift the zero in energy by a finite amount without changing any conclusion. This will be used later on in order to obtain a suitable expression for the kernels of projection operators.

Starting from the one particle space $\mathscr{H}$ and the C^* algebra $\mathscr{A}$ generated by operators $a(f)$ and $a^\dagger(f)$ with $f\in\mathscr{H}$ and obeying the CAR

$$\begin{aligned}&a(f)a(g)+a(g)a(f)=0,\\&a(f)a^\dagger(g)+a^\dagger(g)a(f)=\langle f,g\rangle,\quad f,g\in\mathscr{H},\end{aligned} \tag{1.8}$$

we may study quasifree states over that algebra, which are determined with the help of the projection P_+:

$$\begin{aligned}&\omega_{P_+}(a(f_n)\cdots a(f_1)a^\dagger(g_1)\cdots a^\dagger(g_m))\\&\quad=\delta_{nm}\det\langle f_i,P_+g_j\rangle.\end{aligned} \tag{1.9}$$

From Ref. 7 we know that the Gel'fand–Naimark–Segal

 0022-2488/87/040850-05$02.50

(GNS) representation π_{P_+} corresponding to the state ω_{P_+} is irreducible, since $0<P_+<1$. We know furthermore that two representations π_{P_+} and $\pi_{P'_+}$ are unitarily equivalent iff $P_+P'_-$ and $P_-P'_+$ belong to $\mathscr{B}_2(\mathscr{H})$, the set of Hilbert–Schmidt operators over $\mathscr{H}$.

Our first task is to check whether different representations defined above are unitarily equivalent. For this problem and the following we shall need the following lemma.

Lemma: Let h_U be a particular self-adjoint extension of h with domain $\mathscr{D}_U$ [Eqs. (1.3) and (1.4)] and eigenvalues $\lambda_p{}^j$ and eigensolutions $\xi_p{}^j(x)$ [Eqs. (1.5) and (1.6)]. Let q denote a quadratic form defined by

$$q(\chi,\psi)=\sum_{j,p}\lambda_p{}^j\langle\chi,\xi_p{}^j\rangle\langle\xi_p{}^j,\psi\rangle,\quad \chi,\psi\in Q(q), \tag{1.10}$$

and form domain $Q(q)$

$$Q(q)=\left\{\psi\in\mathscr{H}\,\middle|\,\sum_{j,p}|\lambda_p{}^j|\,|\langle\xi_p{}^j,\psi\rangle|^2<\infty\right\}. \tag{1.11}$$

Then we deduce that

$$Q(q)\cap\{\psi\in\mathscr{H}|\psi \text{ a.c., } h\psi\in\mathscr{H}\}=\mathscr{D}_U, \tag{1.12}$$

or equivalently if $\psi\in\mathscr{H}$ is a.c. and $h\psi\in\mathscr{H}$ then

$$\sigma_N=\sum_{j,p=0}^{N}\lambda_p{}^j|\langle\xi_p{}^j,\psi\rangle|^2,$$
$$\lim_{N\to\infty}\sigma_N=\sigma<\infty \text{ iff } \psi\in\mathscr{D}_U. \tag{1.13}$$

Proof: (a) Take first $\psi\in\mathscr{D}_U$ and note that finiteness of

$$\|\psi\|^2=\sum_{j,p}|\langle\xi_p{}^j,\psi\rangle|^2<\infty \text{ and } \|h_U\psi\|^2=\|h\psi\|^2$$
$$=\sum_{j,p}|\lambda_p{}^j|^2|\langle\xi_p{}^j,\psi\rangle|^2<\infty$$

together with the Cauchy–Schwarz inequality proves finiteness of σ.

(b) Assume ψ is a.c. and $h\psi\in\mathscr{H}$, but $\psi\notin\mathscr{D}_U$ does not fulfill the boundary condition (1.3). We get

$$\langle\xi_p{}^j,h\psi\rangle=\lambda_p{}^j\langle\xi_p{}^j,\psi\rangle-c_j,\quad c_j=i(A^{j\dagger}A-B^{j\dagger}B),$$
$$A^j=(\xi_{0,1}^{\,j}(1),\xi_{0,2}^{\,j}(0)),\quad B^j=(\xi_{0,2}^{\,j}(1),\xi_{0,1}^{\,j}(0)),$$
$$A=(\psi_1(1),\psi_2(0)),\quad B=(\psi_2(1),\psi_1(0)),$$

where c_j is independent of p and $c_j\neq0$ at least for one j. From the assumed finiteness of $\|h\psi\|$ we conclude that

$$\lambda_p{}^j\langle\xi_p{}^j,\psi\rangle\underset{|p|\to\infty}{\to}c_j\neq0.$$

As a consequence $\lambda_N{}^j|\langle\xi_N{}^j,\psi\rangle|^2$ goes like $1/N$ for N large and σ_N diverges with $N\to\infty$.

Next we are ready to check the conditions implying inequivalence of different representations and state our first result.

Theorem 1: The representations π_{P_+} and $\pi_{P'_+}$ of the CAR, obtained from states ω_{P_+} and $\omega_{P'_+}$ via the GNS construction are unitarily equivalent iff $U=U'$; which means iff the self-adjoint extensions are identical.

Proof: Let P_+,P_- and P'_+,P'_- be the projection operators of Eqs. (1.7) corresponding to unitary matrices U and U'. From the kernel representions we get

$$\mathrm{Tr}\,P'_-P_+=\sum_{\substack{j,k\\m,n}}\Theta(-m-1)\Theta(n)|\langle\xi_n{}^j,\xi'_m{}^k\rangle|^2 \tag{1.14}$$

and finiteness of (1.14) implies that $P'_-P_+\in\mathscr{B}_2(\mathscr{H})$. Note that we have chosen the zero of energy such that $\langle\xi_n{}^j,P_+\xi_m{}^k\rangle=\delta_{jk}\delta_{nm}\Theta(n)$ and similarly for P'_-.

Changing the summation variables leads to

$$\mathrm{Tr}\,P'_-P_+=\sum_{\substack{j,k\\p>0}}p|\langle\xi_p{}^j,\xi'_0{}^k\rangle|^2. \tag{1.15}$$

From the Lemma [Eq. (1.13)] we deduce finiteness iff $\xi'_0{}^k\in\mathscr{D}_U$ for $k=1,2$, which means that $A'^k=UB'^k$ for $k=1,2$ with $A'^k=(\xi'^k_{0,1}(1),\xi'^k_{0,2}(0))$ and $B'^k=(\xi'^k_{0,2}(1),\xi'^k_{0,1}(0))$. On the other hand, $A'^k=U'B'^k$ for $k=1,2$ is assumed. Since $\xi'_0{}^k(x)$ for $k=1,2$ are orthonormal in $\mathscr{H}$ it follows that A'^1 is orthonormal in $\mathbb{C}^2$ to A'^2 and B'^1 to B'^2, too, which implies that $U=U'$.

II. IMPLEMENTABILITY OF GAUGE TRANSFORMATIONS

Let G be a group acting on $\mathscr{H}$ by unitary operators $\{V_\alpha,\alpha\in G\}$, and let τ_α be the unique * automorphism of the algebra $\mathscr{A}$, which reduces to V_α on the one-particle operators $a(f)$,

$$\tau_\alpha\,a(f)=a(V_\alpha f),\quad \forall f\in\mathscr{H}. \tag{2.1}$$

In order to have the automorphism τ_α to be implemented in the GNS representation corresponding to the state (1.9), there must exist a unitary operator $\Gamma(V_\alpha)$ acting on the Hilbert space $\mathscr{F}$ of the GNS construction (being the antisymmetric Fock space $\mathscr{F}$ over $\mathscr{H}$), such that

$$\pi_{P_+}(\tau_\alpha(a))=\Gamma(V_\alpha)\pi_{P_+}(a)\Gamma(V_\alpha)^{-1}. \tag{2.2}$$

Therefore it is necessary that the two representations determined by P_+ and by $V_\alpha P_+V_\alpha^{-1}$ are unitarily equivalent, which is true iff [8]

$$X_\alpha=P_+-V_\alpha P_+V_\alpha^{-1}\in\mathscr{B}_2(\mathscr{H}). \tag{2.3}$$

The map $\alpha\to X_\alpha$ is a $\mathscr{B}_2(\mathscr{H})$ valued one-cocycle for G, which obeys the cocycle condition

$$V_\alpha X_\beta V_\alpha^{-1}=X_{\alpha\beta}-X_\alpha. \tag{2.4}$$

Following Ref. 8 one may define one coboundaries as maps $G\to\mathscr{B}_2(\mathscr{H})$ of the form $X_\alpha=C-V_\alpha CV_\alpha^{-1}$ with $C\in\mathscr{B}_2(\mathscr{H})$, and the first Hilbert–Schmidt cohomology group of G relative to the representation V_α as the vector space of cocycles obeying (2.4) modulo coboundaries.

We remark that extensive work has been done on the question of implementing gauge transformations for the free massive or massless Dirac operator on the full line.[9–11] Recently we have been dealing with similar questions for certain external field problems on the line.[12]

Typically we shall consider gauge, axial gauge, or chiral transformations of the form

$$V_\phi=\begin{pmatrix}e^{i\phi_1(x)} & 0\\ 0 & e^{i\phi_2(x)}\end{pmatrix}, \tag{2.5}$$

where we always assume $\phi_i(x)$ to be C^1. Note that condition (2.3) and the analog for P_- are equivalent to the conditions

$$P_+V_{\underline{\phi}}P_-\in\mathcal{B}_2(\mathcal{H}) \quad \text{and} \quad P_-V_{\underline{\phi}}P_+\in\mathcal{B}_2(\mathcal{H}) \tag{2.6}$$

since

$$\begin{aligned}\operatorname{Tr} PVP^{\perp}V^{-1}P &= \operatorname{Tr}(P-PVPV^{-1}) \\ &= \tfrac{1}{2}\operatorname{Tr}(P-VPV^{-1})(P-VPV^{-1})\end{aligned}$$

holds for any projection operator P. Therefore we shall study the group

$$\begin{aligned}\mathcal{G} = \{V_{\underline{\phi}} \mid & V_{\underline{\phi}} \text{ unitary, } P_+V_{\underline{\phi}}P_-\in\mathcal{B}_2(\mathcal{H}), \\ & P_-V_{\underline{\phi}}P_+\in\mathcal{B}_2(\mathcal{H})\};\end{aligned} \tag{2.7}$$

similarly to Ref. 9 one can equip $\mathcal{G}$ with a topology and results on the connected components of $\mathcal{G}$ and their relation to the Fredholm index of $P_+V_{\underline{\phi}}P_+$ apply to our case as well. In the following we shall determine necessary and sufficient conditions on $\phi_1(x)$ and $\phi_2(x)$ such that the local transformations are implementable.

We will find it convenient to distinguish three kinds of boundary conditions:

$$\text{(A) } \Theta=\pi/2, \quad \text{(B) } \Theta=0, \quad \text{(C) } \Theta\neq 0,\pi/2.$$

Case (A) corresponds to two uncoupled modes, case (B) means coupling the modes independently at the ends of the interval; and case (C) means mixed boundary conditions.

Case (A) opposed to (B) and (C) differs already if one takes rigid transformations

$$\tilde{V}_{\underline{\phi}} = \begin{pmatrix} e^{i\phi_1} & 0 \\ 0 & e^{i\phi_2} \end{pmatrix}, \quad \phi_i\in\mathbb{R}. \tag{2.8}$$

Invariance is implied in case (A) for any value of ϕ_1 and ϕ_2, while for cases (B) and (C) ϕ_1 has to be equal to ϕ_2 mod 2π.

Theorem 2: $V_{\underline{\phi}}$ of Eq. (2.5) is unitary implementable in the representation π_{p_+} iff $\phi_1(x)$ and $\phi_2(x)$ fulfill the following boundary conditions:

$$\begin{aligned}&\text{(A) } \phi_1(1)=\phi_1(0)+2m\pi, \quad \phi_2(1)=\phi_2(0)+2n\pi, \\ &\qquad m,n\in\mathbb{Z}, \\ &\text{(B) } \phi_1(1)=\phi_2(1)+2m\pi, \quad \phi_1(0)=\phi_2(0)+2n\pi, \\ &\qquad m,n\in\mathbb{Z}, \\ &\text{(C) } \phi_1(1)=\phi_1(0)+2m\pi, \quad \phi_2(1)=\phi_2(0)+2n\pi, \\ &\qquad m,n\in\mathbb{Z}, \\ &\qquad \phi_1(1)=\phi_2(1)+2p\pi, \\ &\qquad \phi_1(0)=\phi_2(0)+2(p+n-m)\pi, \quad p\in\mathbb{Z}.\end{aligned}$$

Proof: In order to fulfill (2.6) we require finiteness of

$$\kappa = \operatorname{Tr}(P_+V_{\underline{\phi}}P_-V_{\underline{\phi}}^{-1}P_+) < \infty. \tag{2.9}$$

In the basis of orthogonal eigenfunctions (1.5) we get for the kernel K of $P_+V_{\underline{\phi}}P_-$:

$$\begin{aligned}K(x,y) = \sum_{j,m;n,k} & \xi_m{}^j(x)\Theta(m)\Theta(-n-1)\langle \xi_m{}^j, V_{\underline{\phi}}\xi_n{}^k\rangle \\ & \times \xi_n{}^{k*}(y).\end{aligned} \tag{2.10}$$

From (2.10) we get for κ

$$\begin{aligned}\kappa &= \sum_{j,m;n,k}\Theta(m)\Theta(-n-1)|\langle \xi_m{}^j, V_{\underline{\phi}}\xi_n{}^k\rangle|^2 \\ &= \sum_{j,k;p>0} p|\langle \xi_p{}^j, V_{\underline{\phi}}\xi_0{}^k\rangle|^2,\end{aligned} \tag{2.11}$$

where again a change of the summation variable has been made. As in Theorem 1 we use next the lemma to conclude that $\kappa<\infty$ iff $V_\phi\xi_0{}^k\in\mathcal{D}_U$ for $k=1,2$.

Define

$$\begin{aligned}&A^k = (\xi^k_{0,1}(1), \xi^k_{0,2}(0)), \quad B^k = (\xi^k_{0,2}(1), \xi^k_{0,1}(0)), \\ &V_A = \begin{pmatrix} e^{i\phi_1(1)} & 0 \\ 0 & e^{i\phi_2(0)} \end{pmatrix}, \quad V_B = \begin{pmatrix} e^{i\phi_2(1)} & 0 \\ 0 & e^{i\phi_1(0)} \end{pmatrix}.\end{aligned} \tag{2.12}$$

From orthonormality of $\xi_0{}^k(x)$ in $\mathcal{H}$ we deduce orthonormality of A^1 to A^2 and B^1 to B^2 in $\mathbb{C}^2$. Rewriting the requirements $\xi_0{}^k\in\mathcal{D}_U$ and $V_{\underline{\phi}}\xi_0{}^k\in\mathcal{D}_U$, means that $A^k=UB^k$ and $V_AA^k=UV_BB^k$ for $k=1,2$ and therefore $V_AU=UV_B$. Inserting (1.3) yields the stated boundary conditions.

Remark: From Theorem 2 it is trivial to deduce boundary conditions for gauge and axial gauge transformations and chiral transformations

$$\begin{aligned}&V_\phi = \begin{pmatrix} e^{i\phi(x)} & 0 \\ 0 & e^{i\phi(x)} \end{pmatrix}, \quad V_{\phi_5} = \begin{pmatrix} e^{i\phi(x)} & 0 \\ 0 & e^{-i\phi(x)} \end{pmatrix}, \\ &V_{\phi_L} = \begin{pmatrix} e^{i\phi(x)} & 0 \\ 0 & 1 \end{pmatrix}, \quad V_{\phi_R} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi(x)} \end{pmatrix},\end{aligned} \tag{2.13}$$

which we summarize in Table I (where $m,n\in\mathbb{Z}$). Gauge transformations in case (B) are special. The fact that we get no restriction on ϕ means that we also have invariance for rigid transformations.

III. SCHWINGER TERM

In this section we shall treat the Schwinger term for our problem and follow the work of Lundberg.[13] We start as in Ref. 9 with the uniformly continuous one-parameter group of unitary transformations

$$V_{s\underline{u}} = \begin{pmatrix} e^{isu_1(x)} & 0 \\ 0 & e^{isu_2(x)} \end{pmatrix}, \quad s\in\mathbb{R}, \tag{3.1}$$

and define self-adjoint generators j_u,

TABLE I. Boundary values for gauge, axial gauge, and chiral transformations implying that the transformations are unitarily implementable.

Case	V_ϕ	V_{ϕ_5}	V_{ϕ_L}	V_{ϕ_R}
(A)	$\phi(1)=\phi(0)+2\pi m$	$\phi(1)=\phi(0)+2\pi m$	$\phi(1)=\phi(0)+2\pi m$	$\phi(1)=\phi(0)+2\pi m$
(B)	no restriction on ϕ	$\phi(0)=m\pi$, $\phi(1)=n\pi$,	$\phi(0)=2\pi m$, $\phi(1)=2\pi n$,	$\phi(0)=2\pi m$, $\phi(1)=2\pi n$,
(C)	$\phi(1)=\phi(0)+2\pi m$	$\phi(0)=m\pi$, $\phi(1)=(m+2n)\pi$	$\phi(0)=2\pi m$ $\phi(1)=2\pi n$	$\phi(0)=2\pi m$ $\phi(1)=2\pi n$

$$j_{\underline{u}} = \begin{pmatrix} u_1(x) & 0 \\ 0 & u_2(x) \end{pmatrix}, \tag{3.2}$$

such that $V_{s\underline{u}} = \exp(isj_{\underline{u}})$, $V_{s\underline{u}}$ is implementable for all $s \in \mathbb{R}$ iff $P_+ j_{\underline{u}} P_- \in \mathscr{B}_2(\mathscr{H})$. The operators $\Gamma(V_{s\underline{u}})$ form then a strongly continuous one parameter group of unitaries on $\mathscr{F}$, and there exists by Stones theorem a generator $J_{\underline{u}}$ such that

$$\Gamma(V_{s\underline{u}}) = e^{isJ_{\underline{u}}}. \tag{3.3}$$

In the next theorem we shall quote first conditions on $u_i(x)$ such that the Hilbert–Schmidt criterion is fulfilled, and second, evaLuate the commutator algebra, which following Ref. 13 is given by

$$i[J_{\underline{u}}, J_{\underline{w}}] = -2 \operatorname{Im} \operatorname{Tr}(P_- j_{\underline{u}} P_+ j_{\underline{w}}) = S_{\underline{uw}}. \tag{3.4}$$

Theorem 3: (a) $P_+ j_{\underline{u}} P_- \in \mathscr{B}_2(\mathscr{H})$ iff $u_i(x)$ fulfill the following conditions:

(A) $u_1(1) = u_1(0)$ and $u_2(1) = u_2(0)$,
(B) $u_2(0) = u_1(0)$ and $u_2(1) = u_1(1)$,
(C) $u_1(1) = u_1(0)$, $u_2(1) = u_2(0)$
and $u_2(0) = u_1(0)$, $u_2(1) = u_1(1)$.

(b) If $P_+ j_{\underline{u}} P_-$ and $P_+ j_{\underline{w}} P_- \in \mathscr{B}_2(\mathscr{H})$, we obtain for the Schwinger term [Eq. (3.4)] explicitly

$$S_{\underline{uw}} = \frac{1}{2\pi} \int_0^1 dx (u_1(x) w_1'(x) - u_2(x) w_2'(x)). \tag{3.5}$$

Proof: (a) We follow the proof of Theorem 2 and get

$$\operatorname{Tr}(P_+ j_{\underline{u}} P_- j_{\underline{u}}) = \sum_{\substack{j,k \\ p>0}} p |\langle \xi_p{}^j, j_{\underline{u}} \xi_0{}^k \rangle|^2. \tag{3.6}$$

According to the lemma we get finiteness in (3.6) iff $j_{\underline{u}} \xi_0{}^k \in \mathscr{D}_U$ for $k = 1,2$. Similar reasoning as in Theorem 2 applies now replacing $\exp(i\phi_i(x))$ by $u_i(x)$.

(b) From the kernel representation,

$$(P_- j_{\underline{u}} P_+)(x,y) = \sum_{m,j;n,k} \xi_m{}^j(x) \Theta(n) \Theta(-m-1) \times \langle \xi_m{}^j, j_{\underline{u}} \xi_n{}^k \rangle \xi_n{}^{k*}(y), \tag{3.7}$$

we first obtain

$$\operatorname{Tr}(P_- j_{\underline{u}} P_+ j_{\underline{w}}) = \sum_{\substack{j,k \\ p>0}} p \langle \xi_{-p}{}^j, j_{\underline{u}} \xi_0{}^k \rangle \langle j_{\underline{w}} \xi_0{}^k, \xi_{-p}{}^j \rangle, \tag{3.8}$$

and second, by a change of summation index,

$$S_{\underline{uw}} = i \sum_{k,j,p} p \langle j_{\underline{u}} \xi_0{}^k, \xi_p{}^j \rangle \langle \xi_p{}^j, j_{\underline{w}} \xi_0{}^k \rangle \tag{3.9}$$

is obtained using (3.4). Since $\lambda_p{}^j = 2\pi p + \alpha - \sigma_j \epsilon$ we may rewrite (3.9) as

$$S_{\underline{uw}} = \frac{i}{2\pi} \sum_k \{ \langle j_{\underline{u}} \xi_0{}^k, h j_{\underline{w}} \xi_0{}^k \rangle - (\alpha - \sigma_k \epsilon) \langle j_{\underline{u}} \xi_0{}^k, j_{\underline{w}} \xi_0{}^k \rangle \} \tag{3.10}$$

and finally get (3.5) after straightforward calculations.

Remark: It is trivial to discuss special cases like

$$j_u = \begin{pmatrix} u(x) & 0 \\ 0 & u(x) \end{pmatrix}, \quad j_{u_5} = \begin{pmatrix} u(x) & 0 \\ 0 & -u(x) \end{pmatrix}, \tag{3.11}$$

$$j_{u_L} = \begin{pmatrix} u(x) & 0 \\ 0 & 0 \end{pmatrix}, \quad j_{u_R} = \begin{pmatrix} 0 & 0 \\ 0 & u(x) \end{pmatrix};$$

results are listed in Table II.

Remark: Similar as in Ref. 14 we may identify the formal Wick-ordered fermion currents with the generators $J_{\underline{u}}$ and introduce the time zero fermion field as

$$\psi(x) = \sum_{j,m} \xi_m{}^j(x) A_m{}^j \quad \text{with } A_m{}^j = \pi_{P_+}(a(\xi_m{}^j)) \tag{3.12}$$

and the current as

$$\tilde{J}_{\underline{u}} = \sum_{a=1,2} \int_0^1 dx :\psi_a^\dagger(x) \psi_a(x): u_a(x). \tag{3.13}$$

Inserting (3.12) into (3.13) and rearranging terms yields

$$\tilde{J}_{\underline{u}} = \sum_{j,k,p} \langle \xi_0{}^k, j_{\underline{u}} \xi_p{}^j \rangle \tilde{J}_p{}^{kj}. \tag{3.14}$$

The operators

$$\tilde{J}_p{}^{kj} = \sum_n :A_n{}^{k\dagger} A^j{}_{n+p}: \tag{3.15}$$

leave invariant a dense domain of vectors in $\mathscr{F}$ with only a finite number of particles present which are fast decreasing in momentum space. They satisfy a Kac–Moody algebra with central extension

$$[\tilde{J}_p{}^{jk}, \tilde{J}_q{}^{rs}] = \tilde{J}^{js}_{p+q} \delta_{rk} - \tilde{J}^{rk}_{p+q} \delta_{js} + p \delta_{p,-q} \delta_{js} \delta_{rk}. \tag{3.16}$$

Using (3.14) and (3.16) for the calculation of the commutator $i[\tilde{J}_{\underline{u}}, \tilde{J}_{\underline{w}}]$ yields immediately (3.9).

IV. CONCLUSIONS

In this paper we studied the Dirac operator on a finite interval; an "interaction" is simulated with the help of boundary conditions. This leads to solvable but nontrivial models. It turns out that two representations of the canonical anticommutation relations determined by two Hamiltonians with different boundary conditions are always inequivalent.

For each representation we studied the charged sectors

TABLE II. Boundary values for functions with which a charge, axial charge, and chiral charges have to be smeared such that the one-parameter groups generated by the charges are unitarily implementable.

Case	j_u	j_{u_5}	j_{u_L}	j_{u_R}
(A)	$u(1) = u(0)$	$u(1) = u(0)$	$u(1) = u(0)$	$u(1) = u(0)$
(B)	no restriction on u	$u(0) = u(1) = 0$	$u(0) = u(1) = 0$	$u(0) = u(1) = 0$
(C)	$u(1) = u(0)$	$u(0) = u(1) = 0$	$u(0) = u(1) = 0$	$u(0) = u(1) = 0$

which can be reached by applying local gauge transformations. Implementable transformations are obtained iff the gauge functions fulfill certain boundary conditions. These are therefore models which determine certain quantum numbers "by themselves." Concerning boundary conditions on the gauge function, Case (A), for example, is similar to the massless Dirac operator on an infinite interval. Case (B), on the other hand, is similar to the massive Dirac operator on an infinite interval. Gauge transformations need no restrictions; axial gauge transformations are characterized by integers m and n. For all three cases $n - m$ corresponds to the Fredholm index of $P_+ V_\alpha P_+$ and is equal to the winding number of the axial gauge function.[9,11]

In addition we worked out the current algebra. As one expects the Fourier components of currents obey an infinite-dimensional algebra of the Kac–Moody type with central extension; the algebra of currents leads to a Schwinger term corresponding to an anomalous vector, axial vector commutator. The Schwinger term turns out to be independent of the boundary conditions.

Clearly, by studying generators of gauge transformations it is impossible to fix the charge of the ground state of our systems. One possible way is given by the spectral asymmetry[3,6]

$$\Delta Q = -\tfrac{1}{2} \lim_{t \to \infty} \mathrm{Tr}(P_+ - P_-) e^{-t|H|^2}. \tag{4.1}$$

The fractional part of ΔQ turns out to be equal to α/π for our models.[6] We are therefore dealing with systems with continuous varying ground state charge. Here ΔQ corresponds to an effective charge, and to the charge difference of the system relative to the free one; it is well defined after using the regularization in (4.1), for instance.

ACKNOWLEDGMENT

One of us (P.F.) was supported in part by "Fonds zur Förderung der wissenschaftlichen Forschung in Österreich," Project No. P5588.

[1] R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).
[2] R. Rajaraman and J. Bell, Phys. Lett. B **116**, 151 (1982); R. Mackenzie and F. Wilcek, Phys. Rev. D **30**, 2194 (1984).
[3] A. J. Niemi and G. W. Semenoff, Phys. Rep. **135**, 99 (1986).
[4] D. Boyanovsky and R. Blankenbecler, Phys. Rev. D **30**, 1821 (1984); D. Bollé, F. Gesztesy, H. Grosse, and B. Simon, "Kreins spectral shift function and Fredholm determinants as efficient methods to study supersymmetric quantum mechanics," Lett. Math. Phys. (to be published).
[5] H. Grosse and G. Karner, Phys. Lett. B **172**, 231 (1986).
[6] S. M. Roy and V. Singh, "Soliton and boundary condition induced fractional fermion number," TIFR preprint 1984-10.
[7] R. T. Powers and E. Størmer, Commun. Math. Phys. **16**, 1 (1970).
[8] K. Kraus and R. F. Streater, J. Phys. A: Math. Gen. **14**, 2467 (1981).
[9] A. L. Carey, C. A. Hurst, and D. M. O'Brien, J. Funct. Anal. **48**, 360 (1982).
[10] F. Gallone, A. Sparzani, G. Ubertone, and R. F. Streater, J. Phys. A: Math. Gen. **19**, 241 (1986).
[11] A. L. Carey and S. N. M. Ruijsenaars, "On fermion gauge groups, current algebra and Kac–Moody," Canberra preprint, 1985.
[12] H. Grosse and G. Karner, J. Math. Phys. **28**, 371 (1987).
[13] L. E. Lundberg, Commun. Math. Phys. **50**, 103 (1976).
[14] A. L. Carey, C. A. Hurst, and D. M. O'Brien, J. Math. Phys. **24**, 2212 (1983).

Weak rigidity in the PPN formalism

V. del Olmo
Departamento de Matemática Aplicada, E. U. de Informática, Universidad Politécnica de València, València, Spain

J. Olivert
Department de Mecànica i Astronomia, Facultat de Matemàtiques, Burjassot, València, Spain

(Received 25 March 1986; accepted for publication 3 December 1986)

The influence of the concept of weakly rigid almost-thermodynamic material schemes on the classical deformations is analyzed. The methods of the PPN approximation are considered. In this formalism, the equations that characterize the weak rigidity are expressed. As a consequence of that, an increase of two orders of magnitude in the strain rate tensor is obtained.

I. INTRODUCTION

In a preceding paper,[1] a definition of a weakly rigid almost-thermodynamic material scheme was proposed. To verify the Ferrando and Olivert relativistic incompressibility condition,[2] the concept given in Ref. 1 is affected by dynamical and kinematical considerations. Some geometrical consequences on the space-time are also deduced from it.

The methods on the Hadamard discontinuities, used in Ref. 3, allow us to obtain absence of perturbations in a weakly rigid almost-thermodynamic material scheme.

To make this study more realistic, several elasticity hypotheses often used in the literature were analyzed. Their compatibility with the incompressibility condition was checked in Ref. 3. The study of the conditions of weak rigidity in such schemes led to the consistency of both formulations for the principal shock waves, assuming the nonvanishing of one of the Lamé coefficients. To improve the obtained results, a relativistic generalization of the Hooke's law, in which the internal rotations were removed, was considered in Ref. 4. The weak rigidity in an elastic scheme of this kind leads to the Born condition and, so, to the absence of shock waves of any type.

We think that our study about the weak rigidity must be completed. It seems interesting to us to analyze in which way a rigid motion (a weakly rigid almost-thermodynamic material scheme) particularizes in classical mechanics; i.e., to check if it has an appreciable effect on the classical deformations.

With respect to this point, let us say that some of the relativistic formulations that improve the Born condition have considered this matter. For instance, the work of Bona (Ref. 5) in special relativity recovers the classical rigidity in the nonrelativistic limit: inertial synchronization of $(\mathbb{R}^4,\eta)$ obtained by parallel transport of a spatial hyperplane.[6,7] On the other hand, Ehlers and Rudolph,[8] when analyzing the concept of Dixon dynamical rigidity,[9] propose the concept of pseudorigid motions and prove that they have the generality of the classical ones.

In the present work our aim is to study the conditions of weak rigidity by applying to them some techniques of the PPN formalism in general relativity. We will check the way in which they modify the strain rate tensor.

Section II gathers together the basic formulas in the PPN approximation useful for the subsequent study. In Sec. III we give the PPN expressions for the weak rigidity equations and the ones for the strain rate tensor. Some algebraic manipulations lead us to increase the order of the strain initial calculation.

The notation used in this paper is, basically, the one considered in Ref. 4. Greek indices range from 1 to 3 and Latin ones have values from 1 to 4. As usually in the almost-thermodynamic material schemes, we admit the Taub decomposition

$$\rho = r(1+\epsilon), \tag{1.1}$$

ρ being the proper mass-energy density of the scheme, with four-velocity u; r and ϵ symbolize the matter density and the specific internal energy, respectively.

For the energy-momentum tensor of the scheme we write, in the local form,

$$T_{ij} = \rho u_i u_j + t_{ij}, \tag{1.2}$$

the t_{ij} being the covariant components of the relativistic stress tensor.

The metric tensor field g of the space-time manifold is of signature (3,1). The second-order Christoffel symbols associated with the linear connection ∇, compatible with g and without torsion, will be noted as Γ^i_{jk}.

Finally, as is known for the covariant components of the strain rate tensor of the scheme, we write

$$d_{ij} = \tfrac{1}{2}(\nabla_i u_j + \nabla_j u_i + u_i \nabla_u u_j + u_j \nabla_u u_i). \tag{1.3}$$

II. SOME EXPRESSIONS IN THE PPN APPROXIMATION

Henceforth we will use the notation and results presented in Ref. 10, the baryon "mass" density ρ_0 in the notation of Misner–Thorne–Wheeler being replaced here by the matter density r and the specific internal energy being symbolized as ϵ.

We will consider[10] that the following quantities are affected by the approximation order, $O(2)$, of the magnitude v^2 of the coordinate velocity field:

U (Newtonian potential),

$t_{\alpha\beta}$ (components of the stress tensor),

$p = \frac{1}{3}(t_{11} + t_{22} + t_{33})$ (pressure),

ϵ (specific internal energy).

The tensorial expressions used here are given in a co-

 0022-2488/87/040855-04$02.50

moving orthonormal frame with components (x_1,x_2,x_3,t).

We will assume that the matter density r is of order $O(0)$. The spatial partial derivatives will preserve the original approximation orders while the time derivatives will increase these orders in one unit. The partial derivatives will be represented by means of commas; d/dt will symbolize the operator time derivative following the matter, given by $\partial/\partial t + v_\alpha \partial/\partial x_\alpha$.

We will now collect the post-Newtonian expressions useful to develop the weak rigidity equations.

For the covariant components of the metric tensor field, we have[10]

$$g_{44} = -1 + 2U - 2U^2 + 4\Psi + O(6), \tag{2.1}$$

$$g_{4\alpha} = -\tfrac{7}{2} V_\alpha - \tfrac{1}{2} W_\alpha + O(5), \tag{2.2}$$

$$g_{\alpha\beta} = \delta_{\alpha\beta}(1 + 2U) + O(4), \tag{2.3}$$

where V_α, W_α are the vector functionals of order $O(3)$ and Ψ the scalar of order $O(4)$ given, respectively, by Eqs. (39.23c), (39.23d), and (39.23g) of Ref. 10.

For the connection parameters, we can write[10]

$$\Gamma^4_{44} = -U_{,t} + O(5), \tag{2.4}$$

$$\Gamma^4_{4\alpha} = -U_{,\alpha} - 2\Psi_{,\alpha} + O(6), \tag{2.5}$$

$$\Gamma^4_{\alpha\beta} = U_{,t}\delta_{\alpha\beta} + \tfrac{7}{4}(V_{\alpha,\beta} + V_{\beta,\alpha}) + \tfrac{1}{4}(W_{\alpha,\beta} + W_{\beta,\alpha}) + O(5), \tag{2.6}$$

$$\Gamma^\alpha_{44} = -U_{,\alpha} + (2U^2 - 2\Psi)_{,\alpha} - \tfrac{7}{2} V_{\alpha,t} - \tfrac{1}{2} W_{\alpha,t} + O(6), \tag{2.7}$$

$$\Gamma^\alpha_{4\beta} = U_{,t}\delta_{\alpha\beta} + 2(V_{\beta,\alpha} - V_{\alpha,\beta}) + O(5), \tag{2.8}$$

$$\Gamma^\alpha_{\beta\gamma} = -(U_{,\alpha}\delta_{\beta\gamma} - U_{,\beta}\delta_{\alpha\gamma} - U_{,\gamma}\delta_{\alpha\beta}) + O(4). \tag{2.9}$$

Here, in Eq. (2.5), we have inserted the terms of order $O(4)$, not considered in the approximation of Ref. 10.

The contravariant components of the four-velocity u are given[10] by

$$u^4 = 1 + \tfrac{1}{2} v^2 + U + O(4), \tag{2.10}$$

$$u^\alpha = v_\alpha u^4, \tag{2.11}$$

in terms of the coordinate velocity.

Next, we write the post-Newtonian expressions for the components of the energy-momentum tensor. For the contravariant ones, the formulas given in Misner–Thorne–Wheeler are written, in our notation, as

$$T^{44} = r(1 + \epsilon + v^2 + 2U) + O(4), \tag{2.12}$$

$$T^{4\alpha} = r(1 + \epsilon + v^2 + 2U)v_\alpha + t_{\alpha\beta}v_\beta + O(5), \tag{2.13}$$

$$T^{\alpha\beta} = t_{\alpha\beta}(1 - 2U) + r(1 + \epsilon + v^2 + 2U)v_\alpha v_\beta + \tfrac{1}{2}(v_\alpha t_{\beta\gamma}v_\gamma + v_\beta t_{\alpha\gamma}v_\gamma) + O(6). \tag{2.14}$$

From Eqs. (2.12)–(2.14) and using Eqs. (2.1)–(2.3), we obtain the PPN expression for the trace of the energy-momentum tensor:

$$T \equiv T^{ij}g_{ij} = -r(1 + \epsilon) + 3p + O(4). \tag{2.15}$$

Now we are ready to obtain the PPN equations for the Ricci tensor. Let us consider the field equations

$$ET_{ij} = R_{ij} - \tfrac{1}{2} R g_{ij}, \tag{2.16}$$

E being the Einstein constant, R_{ij} the covariant components of the Ricci tensor, and R the scalar curvature. After contracting with the metric tensor field, we get

$$-R = ET; \tag{2.17}$$

thus it is possible to express Eq. (2.16) in the equivalent form

$$R_{ij} = E(T_{ij} - \tfrac{1}{2} T g_{ij}). \tag{2.18}$$

From Eqs. (2.18), (2.15), (2.1)–(2.3), and the covariant expressions of Eqs. (2.12)–(2.14), we deduce

$$R_{44}/E = (r/2)(1 + \epsilon + 2v^2 - 2U) + \tfrac{3}{2} p + O(4), \tag{2.19}$$

$$R_{4\alpha}/E = -r(1 + \epsilon + v^2 + 2U)v_\alpha - t_{\alpha\beta}v_\beta + (r/4)(7V_\alpha + W_\alpha) + O(5), \tag{2.20}$$

$$R_{\alpha\beta}/E = t_{\alpha\beta} + (r/2)[(1 + \epsilon + 2U)\delta_{\alpha\beta} + 2v_\alpha v_\beta] - \tfrac{3}{2} p\delta_{\alpha\beta} + O(4). \tag{2.21}$$

Finally, we write the PPN approximation of the baryon number conservation law (continuity equation), given by $\nabla_i(ru^i) = 0$. We get[10]

$$\frac{dr}{dt}\left(1 + \frac{1}{2}v^2 + U\right) + r\left(\frac{1}{2}\frac{dv^2}{dt} + 3\frac{dU}{dt}\right) + rv_{\alpha,\alpha}\left(1 + \frac{1}{2}v^2 + U\right) = O(5). \tag{2.22}$$

Moreover, we will use the expression

$$r\frac{d\epsilon}{dt} + t_{\alpha\beta}v_{\alpha,\beta} = O(5), \tag{2.23}$$

obtained in Ref. 10 after using Eq. (2.22), the conservation equation $\nabla_i(T^{4i}) = 0$, and the Newtonian equations of motion.

III. PPN EQUATIONS OF WEAK RIGIDITY. CONSEQUENCES

Let us remember that the equations which define the weakly rigid almost-thermodynamic material schemes, proposed in Ref. 1, are given by

$$F_u t_{ij} = 0,$$

$$\nabla_u r = \nabla_u \epsilon = 0,$$

in which F_u symbolizes the Fermi derivative[11] with respect to the four-velocity u of the scheme. In Ref. 1 an equivalent expression for the weak rigidity was deduced, given by

$$F_u R_{ij} = 0, \tag{3.1}$$

$$\nabla_i u^i = 0. \tag{3.2}$$

Our present purpose is to express, in the PPN formalism, Eqs. (3.1)–(3.2). We begin with the vanishing of the expansion velocity scalar.

By virtue of Eqs. (2.4)–(2.11), Eq. (3.2) can be written as

$$\frac{1}{2}\frac{dv^2}{dt} + 3\frac{dU}{dt} + v_{\alpha,\alpha}\left(1 + \frac{1}{2}v^2 + U\right) = O(5). \tag{3.3}$$

With respect to the PPN expressions of Eq. (3.1), let us take into account the formulas

$$F_u R_{ij} = \nabla_u R_{ij} - R_{kj}(u_i \nabla_u u^k - u^k \nabla_u u_i) - R_{ik}(u_j \nabla_u u^k - u^k \nabla_u u_j).$$

By considering in the former equations the timelike, mixed, and purely spatial components and using the PPN approximation of the Christoffel symbols, the corresponding ones for the four-velocity, and Eqs. (2.19)–(2.21), we obtain

$$F_u R_{44} = 0$$

$$\Rightarrow \frac{dr}{dt}\left(1 + \epsilon + \frac{5}{2} v^2 - U\right) + r\frac{d\epsilon}{dt} + 3\frac{dp}{dt} = O(5), \tag{3.4}$$

$$F_u R_{4\alpha} = 0$$

$$\Rightarrow 2v_\alpha \frac{dr}{dt}\left(1 + \epsilon + \frac{3}{2} v^2 + 3U\right) - \frac{dr}{dt}\left(\frac{7}{2} V_\alpha + \frac{1}{2} W_\alpha\right) + 2rv_\alpha \frac{d\epsilon}{dt} + 2v_\beta \frac{dt_{\alpha\beta}}{dt} = O(6), \tag{3.5}$$

$$F_u R_{\alpha\beta} = 0$$

$$\Rightarrow \frac{dr}{dt}\left[\left(1 + \epsilon + \frac{1}{2} v^2 + 3U\right)\delta_{\alpha\beta} + 2v_\alpha v_\beta\right] + r\frac{d\epsilon}{dt}\delta_{\alpha\beta} - 3\frac{dp}{dt}\delta_{\alpha\beta} + 2\frac{dt_{\alpha\beta}}{dt} = O(5). \tag{3.6}$$

We will now verify the way in which the weak rigidity conditions modify the strain rate tensor.

Previously we expressed the components of such a tensor field in the PPN formalism. By virtue of Eqs. (2.4)–(2.11) we get, from Eq. (1.3),

$$d_{44} = v_\alpha v_\beta v_{\beta,\alpha} + O(5), \tag{3.7}$$

$$d_{4\alpha} = -\tfrac{1}{2} v_\beta (v_{\alpha,\beta} + v_{\beta,\alpha}) + O(4), \tag{3.8}$$

$$d_{\alpha\beta} = \frac{1}{2}\left(v_\alpha \frac{dv_\beta}{dt} + v_\beta \frac{dv_\alpha}{dt}\right) + \frac{1}{2}(v_{\alpha,\beta} + v_{\beta,\alpha})\left(1 + \frac{1}{2} v^2 + 3U\right) + \frac{1}{2}(v_\beta v_\gamma v_{\gamma,\alpha} + v_\alpha v_\gamma v_{\gamma,\beta}) + \frac{dU}{dt}\delta_{\alpha\beta} + O(5). \tag{3.9}$$

By subtracting Eqs. (2.22) and (3.3), the latter multiplied by the matter density r, we obtain

$$\frac{dr}{dt}\left(1 + \frac{1}{2} v^2 + U\right) = O(5). \tag{3.10}$$

If we consider Eq. (3.4) and its corresponding contravariant form, subtract both, and take into account Eq. (3.10), we derive

$$\frac{dr}{dt} = O(5). \tag{3.11}$$

From Eqs. (3.4) and (3.15), we get

$$r\frac{d\epsilon}{dt} + 3\frac{dp}{dt} = O(5). \tag{3.12}$$

If we impose $\beta = \alpha$ in Eq. (3.6) and sum for α ranging from 1 to 3 we obtain, after considering Eq. (3.12),

$$\frac{dp}{dt} = O(5). \tag{3.13}$$

By means of the latter, use of Eq. (3.11) allows us to express Eq. (3.4) in the reduced form

$$\frac{d\epsilon}{dt} = O(5), \tag{3.14}$$

from which Eq. (2.23) yields

$$t_{\alpha\beta} v_{\alpha,\beta} = \tfrac{1}{2} t_{\alpha\beta}(v_{\alpha,\beta} + v_{\beta,\alpha}) = O(5), \tag{3.15}$$

after considering the symmetry of the relativistic stress tensor. Given that, in the approximation use here, the velocities are independent of the material structure, we deduce from Eq. (3.15),

$$v_{\alpha,\beta} + v_{\beta,\alpha} = O(3). \tag{3.16}$$

As a consequence of this result, the PPN expressions for the strain rate tensor, Eqs. (3.7)–(3.9), reduce to

$$d_{44} = \tfrac{1}{2} v_\alpha v_\beta (v_{\alpha,\beta} + v_{\beta,\alpha}) = O(5), \tag{3.17}$$

$$d_{4\alpha} = O(4), \tag{3.18}$$

$$d_{\alpha\beta} = O(3); \tag{3.19}$$

so, all the components of this tensor increase by two orders the initial approximation when applying the weak rigidity conditions. In particular, Eq. (3.9) contains the relativistic corrections to the Newtonian deformations; these terms drop out in the considered formalism.

IV. DISCUSSION

As was pointed out in Ref. 1, the weak rigidity conditions lead us, after considering substantial restrictions in the almost-thermodynamic material scheme given, to the Born rigidity condition. We have just obtained, however, the vanishing of the Newtonian deformations from the weak rigidity without additional considerations. This fact would manifest that our concept of rigidity is not a pure replacement of the Born rigidity since, being in a certain way a dynamical definition, our concept is more general than the Born condition. Let us remember the comments that Bona[5] makes about this point.

With respect to the development of Sec. III, let us note that Eq. (3.5) has not been used. Nevertheless, from Eqs. (3.11) and (3.14),

$$v_\beta \frac{dt_{\alpha\beta}}{dt} = O(6), \tag{4.1}$$

a result compatible with

$$\frac{dt_{\alpha\beta}}{dt} = O(5), \tag{4.2}$$

which follows from Eqs. (3.6), (3.11), (3.13), and (3.14).

On the other hand, let us indicate that the analysis of the weak rigidity conditions in the Newtonian approximation would not be enough for our study. In the matter, if we take into account the results presented in Ref. 10 for the Newtonian expressions of the metric tensor field, the energy-momentum tensor and the four-velocity, we get

$$R_{44}/E = \tfrac{1}{2} r + O(2), \tag{4.3}$$

$$R_{4\alpha}/E = -rv_\alpha + O(3), \tag{4.4}$$

$$R_{\alpha\beta}/E = \tfrac{1}{2} r\delta_{\alpha\beta} + O(2), \tag{4.5}$$

for the Ricci tensor covariant components.

Besides, the weak rigidity equations (3.1) and (3.2) can be written, at this level of approximation, as

$$F_u R_{44} = 0 \Rightarrow \frac{dr}{dt} = O(3), \tag{4.6}$$

$$\nabla_i u^i = 0 \Rightarrow v_{\alpha,\alpha} = O(3). \tag{4.7}$$

Note that, by virtue of the time component of the Newtonian local energy-momentum conservation law[10]

$$\frac{dr}{dt} + rv_{\alpha,\alpha} = O(3), \tag{4.8}$$

Eqs. (4.6) and (4.7) are equivalent.

With respect to the strain rate tensor, the Newtonian approximation gives us

$$d_{44} = O(3), \tag{4.9}$$

$$d_{4\alpha} = O(2), \tag{4.10}$$

$$d_{\alpha\beta} = \tfrac{1}{2}(v_{\alpha,\beta} + v_{\beta,\alpha}) + O(3). \tag{4.11}$$

Thus Eq. (4.7) is not applicable here: the absence of the higher-order corrections does not allow us to specify the real order of magnitude of the terms $v_{\alpha,\beta} + v_{\beta,\alpha}$ when the weak rigidity equations are applied.

[1] V. del Olmo and J. Olivert, "Weak rigidity in almost-thermodynamic material schemes," J. Math. Phys. **26**, 1311 (1985).

[2] J. J. Ferrando and J. Olivert, "Incompressibility in relativistic continuous media," J. Math. Phys. **22**, 2223 (1981).

[3] V. del Olmo and J. Olivert, "Compatibility of rigid motions with the relativistic incompressibility condition," J. Math. Phys. **24**, 1172 (1983).

[4] V. del Olmo and J. Olivert, "Compatibility of weak rigidity with some types of elastic schemes," J. Math. Phys. **27**, 1048 (1986).

[5] C. Bona, "Rigid-motion conditions in special relativity," Phys. Rev. D **27**, 1243 (1983).

[6] B. Coll and C. Moreno, "On a class of synchronized observers attached to the Lorentzian structures," J. Math. Phys. **24**, 1245 (1983).

[7] B. Coll, "Sur la notion de rigidité relative à une synchronisation de l'espace-temps," C. R. Acad. Sci. Paris, II. **295**, 103 (1982).

[8] J. Ehlers and E. Rudolph, "Dynamics of extended bodies in general relativity. Center-of-mass description and quasirigidity," Gen. Relativ. Gravit. **8**, 197 (1977).

[9] W. G. Dixon, "Extended bodies in General Relativity: Their description and motion. Isolating gravitating systems in General Relativity," in *Proceedings of the International School of Physics "Enrico Fermi,"* Course LXVII, 1976 (North-Holland, Amsterdam, 1979), p. 156.

[10] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[11] S. W. Hawking and G. F. R. Ellis, *The Large-scale Structure of Space-Time* (Cambridge, U.P., Cambridge, 1973).

Bundle structures in the Kerr space-time

Brian Punsly
Department of Physics, University of California, Los Angeles, California 90024

(Received 26 August 1986; accepted for publication 3 December 1986)

A new extension of the Kerr space-time that was found by this author in a previous paper [J. Math. Phys. **26**, 1728 (1985)] is analyzed in this article. The case of $a^2 > m^2$ is of particular interest. The hypersurface at $r = 0$ acts as if it were a "source" for the geometry. There exists a natural mapping of the hypersurface to a double covering of the null sphere at asymptotic infinity via the principal null congruences. After a study of the obstructions to fibering the constant time cross sections of the hypersurface at $r = 0$, one finds that this two-surface transforms like a spinorialized version of Penrose's null flags. This connection is made by studying the effects of Lorentz transformations at asymptotic infinity on the doubly covered null sphere. Since there exists an inequivalent bundle structure on the two-surface at $r = 0$ that corresponds to an inequivalent spinorialized null flag, one concludes that to observers at stationary infinity the two-surface transforms as if it were a direct sum of two inequivalent Weyl spinors, a four-spinor.

I. INTRODUCTION

This article is an investigation of the global implications on the Kerr space-time that are induced by the geometry of the hypersurface at $r = 0$ (in Boyer–Lindquist coordinates which are used throughout this paper). As was discussed in a previous paper,[1] the geometry of this hypersurface is highly dependent on the manner in which the space-time is extended through the ring singularity. This paper is essentially a development of the final section of Ref. 1 where it was shown that the hypersurface is not homotopically trivial as is commonly assumed in earlier attempts at extending the metric through the singularity.[2] The emphasis of this paper will be on the case $a^2 > m^2$, where a is the angular momentum per unit mass associated with the source of the gravitational field and m is the mass of the source. It will be shown in the course of this paper that a product topology can be assumed for this hypersurface. Consequently, the geometry of the hypersurface is essentially given by the two-dimensional spacelike surface resulting from constant time cross sections.

The new extensions found in Ref. 1 revealed a two-dimensional spacelike surface at $r = 0$ that is topologically a Klein bottle. The Klein bottle appeared as if it were a "source" for the external gravitational field. The global implications of this extension will be addressed in the following sense: What are the induced transformation properties of the hypersurface at $r = 0$ under the action of the Lorentz group as viewed in the stationary frames at asymptotic infinity? This is an instructive way of stating the problem if one is to interpret the Klein bottle as a source for the external geometry.

There is a natural connection between the hypersurface and the null sphere at asymptotic infinity, which allows the transformations induced from the action of the Lorentz group to be readily interpretable. It was found in Ref. 1 that to a set of preferred freely falling observers at $r = 0$, the principal null congruences formed a set of corotating normal vectors to the Klein bottle. These vector fields can be integrated out to the null sphere at infinity. It is through this mapping that a relationship between the transformational properties of the hypersurface and Penrose's null flags (defined in Ref. 3) can be established. For instance, the flagpole corresponds to the principal null geodesic with a spacelike component that is parallel to the total angular momentum vector of the gravitational field. The flag is associated with the azimuthal orientation of the Klein bottle.

In order to understand the details of this mapping to the null sphere, one must investigate the cohomological obstructions to trivializing various fiber bundles over the two-dimensional spacelike surface at $r = 0$. In the process, the obstruction to finding a continuous nonzero cross section for the principal null congruences appears. It is these obstructions that lead to a double covering of the space of Penrose's null flags by the set of orientations of the Klein bottle at $r = 0$. This set of orientations gives the spinorialized versions of the null flags that Penrose had to introduce in an *ad hoc* manner.[3]

This paper begins with a brief recap of what was originally deduced in Ref. 1. The particular extension through the ring singularity that was discussed in some detail in Ref. 1 is rigorously analyzed in Sec. II. First, the question of causality is studied. In spite of what is commonly stated in the literature for the Kerr solution with $a^2 > m^2$, the global space-time is essentially causal. There is a class of acausal timelike trajectories. However, they are not physically reasonable as they require diverging values of acceleration near the ring singularity. Earlier discussions of causality were totally dependent on the previously suggested extension of the space-time through the ring singularity proposed in Ref. 2. As a consequence, the global space-time had the property of being highly acausal.[4]

The emphasis of the main body of Sec. II is to extract all of the information that is contained in the geometry of the two-surface at $r = 0$. Saying that there is a Klein bottle is not specific enough for the purposes of this paper. By virtue of the method of extension through the ring singularity the two-surface is endowed with additional structure. Specifically, the geometry of the equatorial region determines how the surface rotates and how frames can be transported along curves. There is a detailed derivation of these properties.

 0022-2488/87/040859-29$02.50

Since the Klein bottle cannot be embedded in E^3, one cannot simply draw this two-surface on a piece of paper. In keeping with the geometrical spirit of this article, numerous illustrations are used to elucidate the complicated twisted features of this two-surface.

Various fiber bundles are constructed over the Klein bottle that are consistent with the fact that U_2 (the topological symbol for the Klein bottle) is immersed in a four-dimensional space-time. In particular, the tangent, normal, and four-frame or $\mathbb{R}^4$ bundles over U_2 are studied by means of characteristic classes and obstruction cocycles. This leads to the construction of spinorialized null flags. There exists a set of inequivalent bundle structures on U_2 that induce an inequivalent set of spinorialized null flags at asymptotic infinity. After making Penrose's equivalence between spinorialized null flags and Weyl spinors, one can conclude that the two-surface at $r=0$ as viewed from asymptotic infinity transforms as if it were a four-spinor that was obtained by combining two inequivalent Weyl two-spinors in a manner that depends on the representation of the spinor algebra.

II. THE HYPERSURFACE AT $r=0$

In Ref. 1 the Kerr space-time was interpreted as a set of local Lorentz coordinate patches that were integrated in a global hypersurface orthogonal frame so as to foliate space-time. The insight that was gained proved to be particularly useful for analyzing the geometry of the hypersurface at $r=0$. At $r=0$, the hypersurface orthogonal frame is particularly natural as it is stationary with respect to asymptotic infinity. In the hypersurface orthogonal frame, the constant time cross sections appear to be spheres that have been "twisted" (see Fig. 1). The surface area would be precisely that of a sphere except that null geodesics in the hypersurface (not geodesics in the four-manifold) that are projected onto the two-dimensional spacelike surface are not meridians but twisted meridians as pictured in Fig. 1. The null geodesics in the hypersurface are given by (2.1),

$$\frac{\partial}{\partial\theta}+\frac{\partial}{\partial\phi}+\frac{\partial}{\partial t}, \tag{2.1a}$$

$$\frac{\partial}{\partial\theta}-\frac{\partial}{\partial\phi}+\frac{\partial}{\partial t}, \tag{2.1b}$$

$$\frac{\partial}{\partial\theta}+\frac{\partial}{\partial\phi}-\frac{\partial}{\partial t}, \tag{2.1c}$$

$$\frac{\partial}{\partial\theta}-\frac{\partial}{\partial\phi}-\frac{\partial}{\partial t}. \tag{2.1d}$$

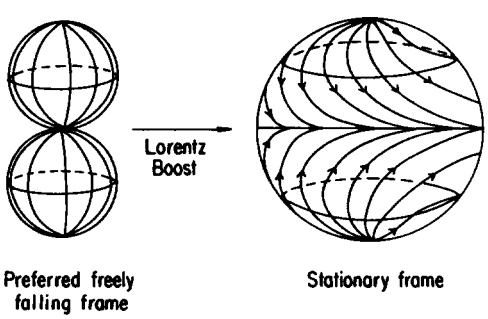


FIG. 1. The global mapping of the two-surface at $r=0$ as viewed by locally inertial observers to its image in the stationary frames.

The meridians are twisted since the sphere as viewed in the stationary frames appears to be rotating so that the equator (which must be the region of maximum angular velocity for the rigid rotation of a sphere) is moving azimuthally at the speed of light. Consequently, the null meridians of the sphere receive an appreciable headlight effect as viewed in the stationary frames.

It turns out that the twisted geodesics identify antipodal points on the equator of the twisted sphere. One concludes that the surface is topologically the real projective space $\mathbb{R}P^2$. It was further observed from the structure equations of Cartan's method of moving frames for submanifolds,

$$dw^{\phi}{}_{\theta} = -K\,dA = w^{\phi}{}_{\alpha}\wedge w^{\alpha}{}_{\theta} + R^{\phi}{}_{\theta}\,, \tag{2.2}$$

that K, the Gaussian curvature of the two-dimensional spacelike surface, was zero. This is the correct version of (B3) of Ref. 1 which was written down incorrectly in Ref. 1. Fortunately, all of the results in Ref. 1 due to (B3) were written down correctly in that paper. For a description of the quantities in (2.2) see Ref. 1 or (3.1) of this paper.

The conclusion based on (2.2) is that one appears to have a contradiction. $\mathbb{R}P^2$ is compact and the only compact two-surfaces with $K=0$ are the torus and the Klein bottle. The problem is remedied by attaching another $\mathbb{R}P^2$ associated with the other sheet of space-time that is obtained by extension through the ring singularity. The connected sum $\mathbb{R}P^2\#\mathbb{R}P^2$ is topologically the Klein bottle, resolving the paradox. It should be noted that this discussion is just a sketch of what was done in the last section of Ref. 1 with many crucial details omitted.

The extension of interest in this paper is the same one that is illustrated in Ref. 1. The second sheet of space-time is obtained from the first by the discrete isometry of the Kerr metric $\phi\to-\phi$ and $t\to-t$. The two nonorientable $\mathbb{R}P^2$'s are equivalent except that the geometric boundaries ($\mathbb{R}P^1$) rotate in opposite senses and one of them goes backward in time. The interpretation that the first space-time is actually the same as the second will be made clear in this section. As indicated in Sec. III, this is also implied by the Stiefel–Whitney classes of the two-surface and its cobordism class. The Klein bottle at $r=0$ turns space-time inside out upon itself.

In order to understand how this happens, one can look at the extension of trajectories that strike the Klein bottle through this two-surface. Recalling that U_2 is $\mathbb{R}P^2\#\mathbb{R}P^2$ and that $\mathbb{R}P^2$ is the sphere, S^2, with antipodal points identified, this is very straightforward. The first step is to decompose any tangent vector to a trajectory that strikes the two-surface into a component that lies in the tangent space of U_2, $T(U_2)$, and a component in the normal space, $N(U_2)$. A parity transformation acts on the Boyer–Lindquist coordinates as follows:

$$\theta\to\theta+\pi,\quad \phi\to\phi,\quad r\to-r,\quad t\to t\,. \tag{2.3}$$

The quotient map that takes S^2 to $\mathbb{R}P^2$ identifies the points on the sphere $(\theta+\pi,\phi)\sim(\theta,\phi)$. Thus, one can conclude that the tangent vector to the trajectory emerges from the two-surface at the antipodal point to which it entered (which is actually the same point at which it entered by the quotient map in the previous sentence) with the same component in $T(U_2)$ that it had upon striking the surface and the

spacelike component in $N(U_2)$ reversed. For instance, the vector labeled k_{in} as it reached the two-surface corresponding to an ingoing principal null geodesic is just relabeled k_{out}, an outgoing principal null geodesic, as it leaves the surface. In general, normal vectors that pass through the two-surface are just relabeled as they propagate back out towards asymptotic infinity. Similarly, for the past directed trajectories that strike the other $\mathbb{R}P^2$ as exemplified by the identification $-k_{\text{out}} \leftrightarrow -k_{\text{in}}$, since under the discrete isometry

$$k_{\text{out}} \to -k_{\text{in}} \quad \text{and} \quad k_{\text{in}} \to -k_{\text{out}} .$$

This is equivalent to the fact that the Klein bottle is a one-sided surface in E^3; its "inside" and "outside" are the same space.[5]

The existence of past directed geodesics that are on an equal footing with future directed geodesics leads to the question of causality of the space-time associated with this extension. Notice that the discussion in the previous paragraph implies that past directed trajectories remain past directed as they pass through the two-surface and similarly for future directed curves. Thus, an observer cannot see himself at earlier times after passing through the two-surface. But, what about observers that remain tangent to the two-surface and pass through the ring singularity? In principle, they could detect themselves at an earlier time on returning to the original sheet of space-time. In order to understand the precise physical implications of this, one must say something about the manner in which the connected sum $\mathbb{R}P^2 \# \mathbb{R}P^2$ is constructed. The connected sum is accomplished by cutting out a region that is homeomorphic to a disk from each $\mathbb{R}P^2$. This leaves two Möbius bands. Then the two Möbius bands are connected along the edges of these holes. Assume that the region that is cut out is an infinitesimal neighborhood of a portion of the ring singularity. It cannot contain the whole ring singularity for then it would not be homeomorphic to a disk. Then by (C2) and (C8) of Ref. 1, it would take a rocket capable of supplying a diverging value of acceleration to overcome the centrifugal force experienced by a timelike observer that attempts to remain tangent to the two-surface as the observer approaches the junction between the two sheets of space-time. Consequently, even though causality violation is mathematically possible, it may be considered physically to be highly unlikely. It should be noted that if the connected sum were in a region away from the equator, physically reasonable observers could easily pass from sheet to sheet, making for a very acausal space-time.

This extension seems to result in a physically reasonable space-time. The conclusion here is in contrast to that of Carter in studying the Boyer–Lindquist extension in Ref. 4. In that case the coordinate r is allowed to take on negative values and therefore so does the metric coefficient $g_{\phi\phi}$. This enables observers to go backward in time by rotating about the symmetry axis for certain negative values of the coordinate r and then meet oneself at an earlier time by going back out towards asymptotic infinity at $r \to +\infty$. There is no such problem with the extension that is discussed here. Similarly, the timelike and null lines that are studied in Ref. 4 that lie in the equatorial plane and strike the singularity resulting in acausality in the Boyer–Lindquist extension just pass through and are relabeled in this extension due to the quotient space structure of U_2 and its constituent $\mathbb{R}P^2$'s as mentioned previously.

There is one more piece of information on the global geometry of the two-surface that can be obtained without an explicit construction of the connected sum. One can describe global geodesics as was done for the $\mathbb{R}P^2$'s in Ref. 1. These same geodesics can be pieced together to form a global geodesic on U_2. As a guide as to how this should be done, one uses the fact that since $K = 0$ on the two-surface, there can be no conjugate points. However, there is a cut locus since U_2 can be defined as a quotient space of the torus.[5]

The geodesics on the hypersurface containing $\mathbb{R}P^2$ are given by Eq. (2.1). Note that the discrete isometry acting on one of these geodesics creates another geodesic. Begin the global geodesic on the future directed sheet near the equator in the form (2.1a) as drawn in Fig. 2. The spacelike projection of the geodesic integrates toward the north pole of the twisted sphere (see Figs. 1 and 2). Continuing through the pole, (2.1a) integrates down the "front" side of the "sphere" towards the equator. In the case of $\mathbb{R}P^2$ as discussed in Ref. 1, it would rejoin back to its starting point on the equator, a conjugate point. However, the conjugate point was removed when the connected sum was performed. The spacelike projection of the geodesic passes through the equatorial region and emerges on the past directed sheet of space-time in the form (2.1d). Recall that these are geodesics in the hypersurface and not in the four-dimensional manifold. Thus, it would take a diverging value of acceleration for a physical observer to travel this same path. The spacelike component of the geodesic can be continued to the pole of the past directed hemisphere in Fig. 2 and back to the equator as it did on the other sheet of space-time. The entire global geodesic is indicated in Fig. 2. However, Fig. 2 is intended to be merely a schematic as the Klein bottle cannot be embedded in E^3. The past directed sheet of the Klein bottle does not actually lie "below" the future directed sheet as is indicated in Fig. 2.

In order to get a clear understanding of the geometry of the two-surface, one must describe in complete detail the structure of the connected sum of the two projective spaces in the equatorial region. As is the case of the model of $\mathbb{R}P^2$ as a sphere with antipodal points on the equator identified, the geometric structure of the region where the two $\mathbb{R}P^2$'s join determines many of the global properties of the two-surface at $r = 0$.

One item that must be considered when dealing with nonorientable surfaces (such as $\mathbb{R}P^2$, the Möbius band, and

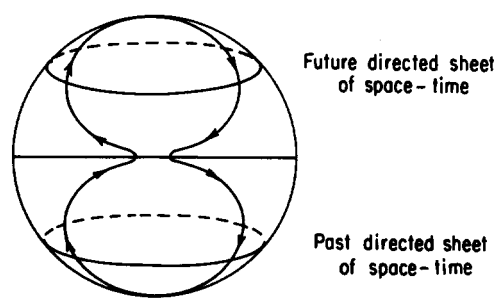


FIG. 2. A global closed geodesic on the two-surface at $r = 0$.

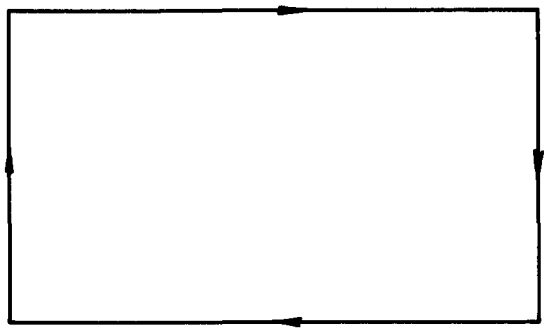


FIG. 3. A standard model of $\mathbb{R}P^2$ as a rectangle with identifications of the edges.

the Klein bottle) is that if one translates a two-frame (i.e., parallel translates a basis of tangent vectors) along a closed curve the orientations of the two-frame can change upon returning to its starting point. Consequently, a preliminary discussion of the orientation bundle over the two-surface becomes relevant. Although many properties are independent of the open covering of the base manifold, in this particular case there is a class of open coverings that bring out more details of the geometry. In order to construct an explicit representative of this class of open coverings, one must first go back to a detailed study of the various models of $\mathbb{R}P^2$.

$\mathbb{R}P^2$ cannot be embedded in Euclidean three-space, E^3. Thus, one must resort to models in three-space that are topologically equivalent to $\mathbb{R}P^2$. Since this paper is not concerned so much with the space $\mathbb{R}P^2$ in general but for a particular space that is topologically equivalent to $\mathbb{R}P^2$, it is best to think of these models as parameter spaces that can be mapped 1 to 1 to the actual surface under consideration.

The first model to be introduced is the square with opposite sides identified as pictured in Fig. 3. For further details of this model and the ones to follow see Refs. 5 and 6. As is standard in topological models, the arrows define which edges should be identified. Parallel arrows imply that the two sides can be simply glued together. Antiparallel arrows imply that one side is to be flipped over before being glued to the opposite side (see Figs. 3–6). One can also define $\mathbb{R}P^2$ as a disk with antipodal points on the boundary identified. A similar model that will be used is the hemisphere with antipodal points on the equator identified (a cross cap). There are other models for $\mathbb{R}P^2$ but these are adequate for the purposes of this paper. The topologist's technique of going from

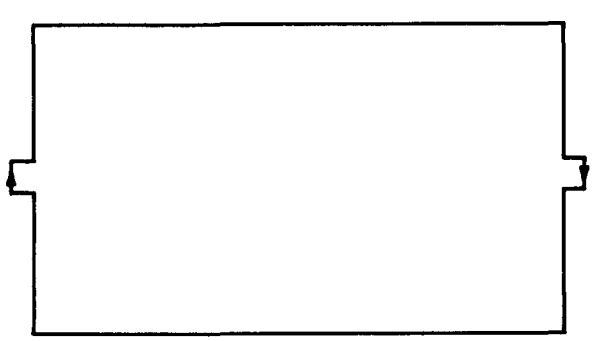


FIG. 4. A model of the Möbius band that is based on Fig. 3.

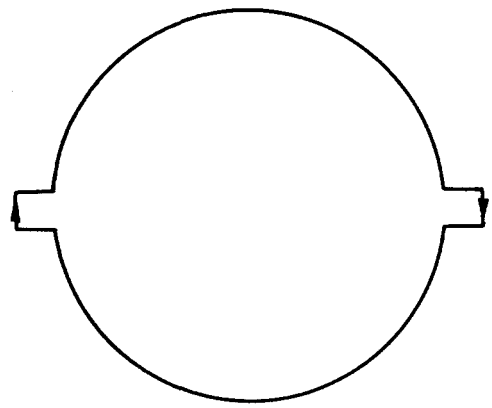


FIG. 5. A model of the Möbius band.

one model to another depending on the property of the surface under investigation will be used in the following.

In constructing the connected sum, a region homeomorphic to a disk is removed from each $\mathbb{R}P^2$ in a neighborhood of the equator. It turns out that an interesting model for U_2 is one in which most of the equator has been removed from each of the constituent $\mathbb{R}P^2$'s. This is the Klein bottle of interest in this paper and the result of this excision on one of the $\mathbb{R}P^2$'s is illustrated for the various models of $\mathbb{R}P^2$ in Figs. 4–6.

All of the identifications that are indicated by the arrows in Figs. 4–6 are manifested in an infinitesimal neighborhood of the equator that was not removed when the hole was cut in $\mathbb{R}P^2$. This infinitesimal neighborhood is like a "bridge" that connects the left edge (with a half twist) to the right edge in Figs. 4–6. For this reason, it will sometimes be referred to in the following as the bridge. This bridge is greatly exaggerated in these illustrations for the sake of added definition. It is actually of infinitesimal extent. In some of the following illustrations it will be exaggerated even more so that additional details can be drawn.

It is easy to see from Figs. 4–6 and the discussion of the bridge in the previous paragraph that the space resulting from excising the hole from $\mathbb{R}P^2$ is topologically a Möbius band. Since it is only the resulting Möbius bands that are used in the construction of U_2, the orientation bundle will be described for these with a particular emphasis on an illumi-

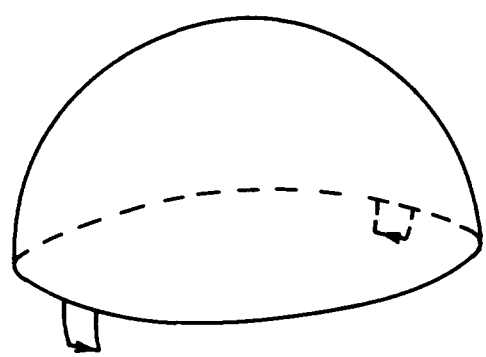


FIG. 6. A model of the Möbius band.

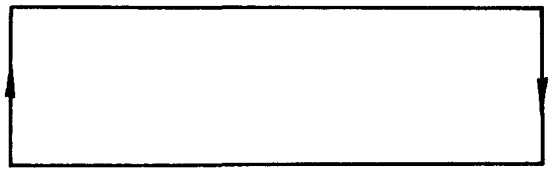


FIG. 7. A standard model of the Möbius band as a rectangle with two of the sides identified.

nating open cover. Preliminary to this, observe how the Möbius bands are joined to form U_2. A model for the Möbius band is given by Fig. 7. Combining two of these and attaching along the edges one has a standard model for the Klein bottle (Fig. 8).

Define two open sets on the future directed Möbius band that is obtained from the future directed $\mathbb{R}P^2$ by removing a set that is homeomorphic to a disk. The first set, V_1, can be defined in terms of the models in Figs. 5 and 6 as the unshaded area in Fig. 9. Notice that there is a segment that is removed (half of a meridian). Define another open set, V_3, as the unshaded region of Fig. 10. Note that there is no segment removed from V_3 and $V_1 \cup V_3$ is the future directed Möbius band.

The intersection of V_1 and V_3 contains two disjoint open sets labeled $(V_1 \cap V_3)_1$ and $(V_1 \cap V_3)_2$ that are denoted by the unshaded regions in Figs. 11 and 12, respectively. The missing meridian is an unusual construction in the open covering of the Möbius band. The motivation for removing the meridian will be elucidated when the structure near the equator has been studied in detail.

The orientation bundle over the Möbius band can now be defined as a line bundle in the standard construction as in Ref. 7. One can orient each one of the open sets V_1 and V_3 separately. For instance, the coordinate transformation from the tangent space at a point of V_1 to the tangent space at another point of V_1 can be chosen to have positive determinant for each pair of points in V_1. The same statement can be made for pairs of points in V_3. The line that is a fiber over the Möbius band can be described by a $+1$ or -1 (line points "up" or "down") depending upon whether the determinant of the transformation from some base point is positive or negative. One can trivialize this line bundle on V_1 and on V_3

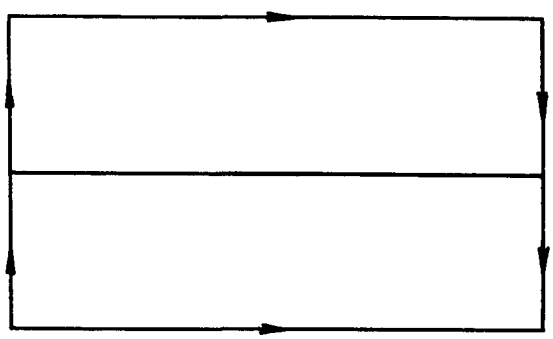


FIG. 8. A standard model of the Klein bottle as a rectangle with the sides identified. This model is obtained by gluing two of the Möbius bands in Fig. 7 along the edges.

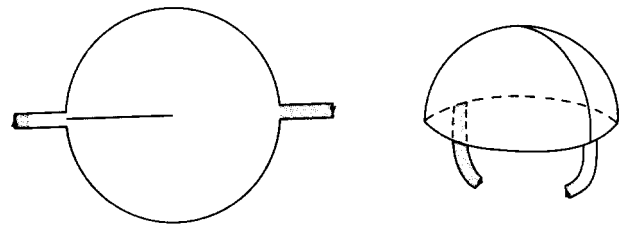


FIG. 9. The open set V_1, and element of the open covering of the future directed Möbius band, is represented by the unshaded region of these two different models of the Möbius band.

separately. For a good discussion of the orientation line bundle see Bott and Tu.[7] However, one finds that on the overlap of the two open sets, the trivialization over V_1 cannot agree with that over V_3 on both components of the overlap. This is best described by the transition functions of the line bundle. For any covering of the Möbius band by two open sets, Steenrod shows that the transition functions are $g_{11} = g_{22} = g_{21} = g_{12} = 1$ on one component of the intersection and $g_{11} = g_{22} = -g_{21} = -g_{12} = 1$ on the other component.[8] See Ref. 8 for definitions and discussion of transition functions. In this paper one chooses

$$g_{11} = g_{33} = g_{31} = g_{13} = 1 \quad \text{for } (V_1 \cap V_3)_1, \tag{2.4a}$$

$$g_{11} = g_{33} = -g_{31} = -g_{13} = 1 \quad \text{for } (V_1 \cap V_3)_2. \tag{2.4b}$$

The trivialization over V_1 can be arbitrarily associated with a "$+1$" and the trivialization over V_3 is therefore associated with a "-1" on the region of intersection, $(V_1 \cap V_3)_2$. This is a useful way of conceptualizing the orientation bundle on this set. It allows for the interpretation of V_1 as the positive orientation sheet of $(V_1 \cap V_3)_2$ and V_3 represents the negative orientation sheet. Since these sheets are only subsets of V_1 and V_3, one introduces a new notation to describe these subsets. The positive and negative orientation sheets of $(V_1 \cap V_3)_2$ will be denoted as regions I and III, respectively.

One can similarly cover the past directed Möbius band by the two open sets V_2 and V_4. The transition functions for the orientation bundle are the same as (2.4) with the appropriate relabeling of indices. V_2 is chosen to represent the positive orientation sheet on the hemisphere minus a meridian and the equator, $(V_2 \cap V_4)_2$ (see Fig. 12), which is the past directed analog of $(V_1 \cap V_3)_2$. Here V_4 is defined analogous to the definition for V_3. One can continue the analogy to the future directed sheet by making the corresponding defi-

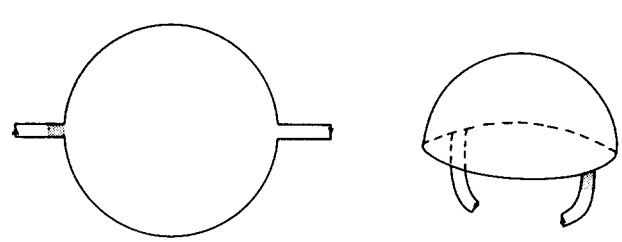


FIG. 10. The open set V_3, an element of the open coverings of the future directed Möbius band, is represented by the unshaded region of these two different models of the Möbius band.

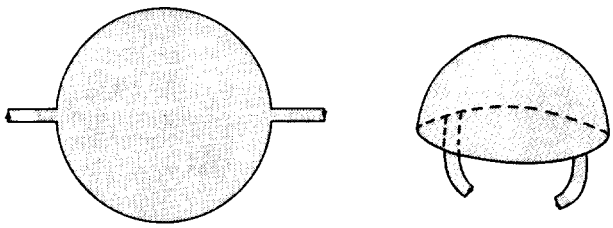


FIG. 11. The open set $(V_1 \cap V_3)_1$, one of the disjoint components of the intersection of V_1 and V_3, is indicated by the unshaded region of these two different models for the future directed Möbius band.

nitions for regions II and IV as positive and negative orientation sheets, respectively.

After making the connected sum, one infinitesimally enlarges the open sets V_i to be larger sets W_i, such that $V_i \subset W_i$ and $\cup_i W_i = U_2$. Then one can find the transition functions of the orientation bundle for this covering of U_2 by four open sets. As an explicit example, $W_1 \cap W_2$ is an open set with one connected component as drawn in Fig. 13.

Figure 13 is just a schematic in a sense. It is homeomorphic to the actual intersection and it is based on the schematic diagram of $\mathbb{R}P^2$ in Fig. 6. The equator represents where the two sheets are joined in the connected sum. "Above" the equator is the future directed sheet and "below" the equator is the past directed sheet. It is at this point that one realizes that the diagrams can be more confusing than helpful when the second sheet of space-time is attached. The figures no longer speak for themselves and must be supplemented with words to clarify their meaning. One sheet is not really above the other.

The transition functions on the overlap of the trivializations of the orientation bundles on W_1 and W_2 are chosen to be

$$g_{11} = g_{22} = g_{12} = g_{21} = 1. \tag{2.5a}$$

Note that $g_{12} = g_{21}$ could have been chosen to equal -1, but the choice of $+1$ simplifies the notation. The other choice will be revealed in Sec. III to be representative of an inequivalent bundle structure. Once this choice is made, the transition functions on the other intersections are fixed. On $W_3 \cap W_4$ the intersection has one component with transition functions

$$g_{33} = g_{44} = g_{34} = g_{43} = 1. \tag{2.5b}$$

As before, $W_1 \cap W_3$ is a set with two components. On

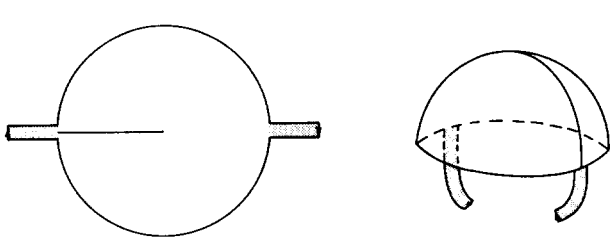


FIG. 12. The open set $(V_1 \cap V_3)_2$, one of the disjoint components of the intersection of V_1 and V_3, is indicated by the unshaded region of these two different models for the future directed Möbius band.

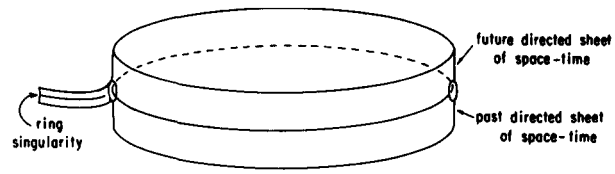


FIG. 13. The intersection of U_1 and U_2. Note the holes in the main cylindrical region on the far left and the far right. There is an adjoining flap on the left side that connects along part of the edge of one of these holes. A portion of this appendage intersects the ring singularity.

$(W_1 \cap W_3)_1$ (see Fig. 11) one has the transition functions [see (2.4a)]

$$g_{11} = g_{33} = g_{13} = g_{31} = 1. \tag{2.5c}$$

Similarly, on $(W_1 \cap W_3)_2$ (see Fig. 12) one has [see (2.4b)]

$$g_{11} = g_{33} = -g_{13} = -g_{31} = 1. \tag{2.5c'}$$

The remaining transition functions are compiled below. On

$$\begin{aligned} &(W_1 \cap W_4)_1, \quad g_{11} = g_{44} = g_{41} = g_{14} = 1, \\ &(W_1 \cap W_4)_2, \quad g_{11} = g_{44} = -g_{41} = -g_{14} = 1, \end{aligned} \tag{2.5d}$$

$$\begin{aligned} &(W_2 \cap W_4)_1, \quad g_{22} = g_{44} = g_{42} = g_{24} = 1, \\ &(W_2 \cap W_4)_2, \quad g_{22} = g_{44} = -g_{42} = -g_{24} = 1, \end{aligned} \tag{2.5e}$$

$$\begin{aligned} &(W_2 \cap W_3)_1, \quad g_{22} = g_{33} = g_{23} = g_{32} = 1, \\ &(W_2 \cap W_3)_2, \quad g_{22} = g_{33} = -g_{23} = -g_{32} = 1. \end{aligned} \tag{2.5f}$$

One can construct a model for U_2 along the lines of Fig. 6 as a surface consisting of two concentric hemispherical shells that are joined along the equator with a handle connected to the equatorial region. This is depicted in Fig. 14. As with the bridge of the Möbius band in Figs. 4–6, the handle is very much exaggerated. It actually has an infinitesimal diameter and is infinitesimally long.

One can associate the inner hemisphere of Fig. 14 with the past directed sheet and the outer hemisphere with the future directed sheet of space-time. As drawn, the surface is topologically a torus (a sphere with one handle). However,

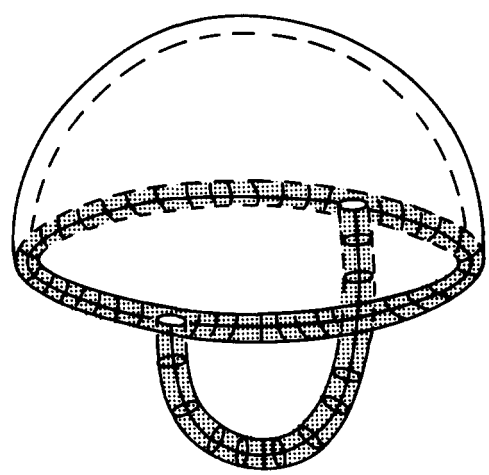


FIG. 14. A model of the Klein bottle. Note that the handle has been greatly exaggerated.

if in the middle of the circular tube that forms the handle, where there is a heavy line signifying a junction in Fig. 14, the circular cross section is reflected through a diameter as it passes through the junction, then the surface is topologically a Klein bottle.[8] Figure 14 is a little deceiving in that the neighborhood of the equatorial junction of the two hemispheres is not topologically a cylinder. A major portion of the remainder of this section is an investigation of this region.

This construction is useful as one can associate the set $W_1 \cup W_2$ with the positive orientation sheet of the two hemispheres that comprise part of the Klein bottle as pictured in Fig. 14 (remember that a meridian has been removed from each hemisphere) as was done for the orientation sheets on the corresponding subsets of the two constituent $\mathbb{R}P^2$'s. Similarly, $W_3 \cup W_4$ correspond to the negative orientation sheet of the two hemispheres. Normally, only two open sets would be necessary to cover the Klein bottle in order to make such a construction for the orientation bundle.[8] Four sets have been chosen here so that it is possible to keep track of whether a trajectory is on the past or future directed sheet of space-time as well as the positive or negative orientation sheet.

The motivation for drawing the Klein bottle as in Fig. 14 is twofold. Ultimately, the Klein bottle will be associated with the null sphere at infinity by means of a set of normal vectors to U_2, the principal null congruences. Second, it was found in Ref. 1 that the spacelike two-surface at $r = 0$ is just like a sphere as viewed by observers that are stationary with respect to asymptotic infinity, except that the rotation of the equator at the speed of light gives the equatorial region a complicated twisted and knotted structure. It is what goes on at the equator that changes the topology of the entire two-surface. This was also the motivation for choosing the open sets as was done in Figs. 9 and 10. The goal was to make the component of the intersection that is depicted in Fig. 12 look as much like a hemisphere as is possible. To understand why a meridian was removed from W_1 and W_2, one must study the equatorial region of Fig. 14 where the connected sum is formed in detail.

Finally, one has enough constructions and tools to actually study how the connected sum is formed. Again, one must go back to the original models of $\mathbb{R}P^2$. The boundary of $\mathbb{R}P^2$ (the equator of the twisted sphere) is topologically $\mathbb{R}P^1$. This can most easily be seen from the model of $\mathbb{R}P^2$ that was used to draw Fig. 5. An open neighborhood of the equatorial singularity is therefore homeomorphic to the canonical open line bundle over $\mathbb{R}P^1$ (i.e., line bundle minus the boundary) as defined in Ref. 9. This neighborhood is topologically a Möbius band. To make the connected sum of interest a region that is homeomorphic to a disk is removed from this neighborhood that contains all of the equator except for an infinitesimal region. What remains of the Möbius band is pictured in Fig. 15.

The connected sum is formed by attaching two of these regions that are depicted in Fig. 15 along the boundary of the excised region. It should be noted that since the central circle of the Möbius band is $\mathbb{R}P^1$, it only takes a parameter value of $\phi = \pi$ radians to complete the circular path. Due to the difficulty of drawing both of the Möbius bands with holes in them complete with their connection along the boundary of the excised holes, a schematic that represents this join is drawn in Fig. 16.

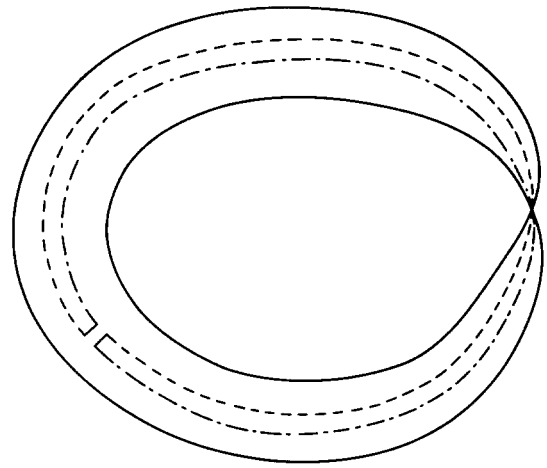


FIG. 15. A Möbius band with a long slit excised. The excision removes most of the center circle.

It should be noted that in Fig. 16 when the left end is rejoined to the right end (the Möbius band had to be severed so that it was possible to draw the junction between the two sheets of space-time in a coherent manner) of each Möbius band that there must be a half twist to get the actual Möbius band structure that arises from the canonical line bundle over $\mathbb{R}P^1$. There is an added complication in joining the left-hand side to the right-hand side of Fig. 16 that cannot be deduced at this point. It will be shown in the discussion of Fig. 17 that the junction between the left- and right-hand sides of Fig. 16 actually involves the right-hand side of the neighborhood of the future directed $\mathbb{R}P^1$ be attached to the left-hand side of the neighborhood of the past directed $\mathbb{R}P^1$ and vice versa. This is related to the quadratic branch point nature of the ring singularity.[11] There will be more said about this in the discussion that follows Fig. 17 from which this is much easier to see.

The advantage of drawing Fig. 16 is that the critical point structure of the manifold is readily revealed. These critical points are circled in Fig. 16. It appears that if one traces a path along where the connected sum is made, one need not specify which sheet of space-time it is on. However,

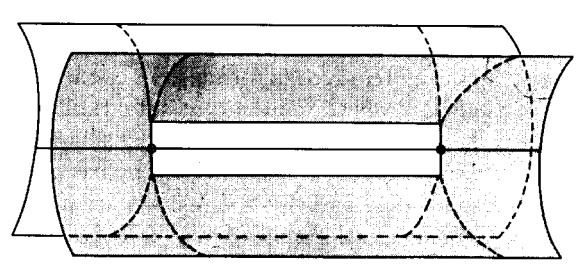


FIG. 16. The connected sum of the two sheets of space-time is made along the edge of the rectangular hole. There is an identification of points on the left edge with points on the right edge. The small circles indicate the location of the two critical points.

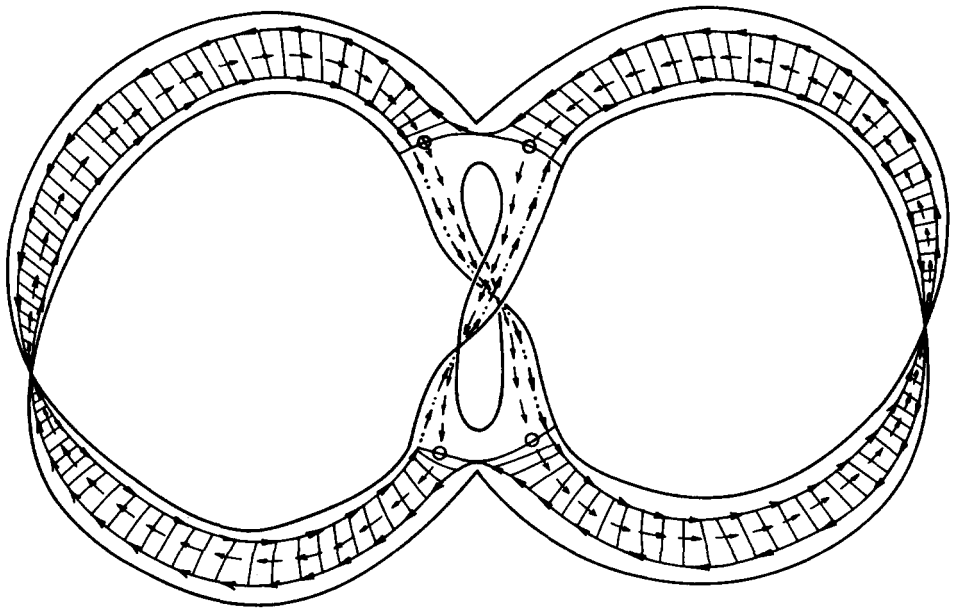


FIG. 17. The equatorial region of the two-surface at $r = 0$. Note that the region has been sliced open along the join of the connected sum. The location of the critical points are indicated by the small circles.

if one is moving from left to right there is a problem at the right end. If one continues to move azimuthally, there is a splitting of the manifold into the two sheets of space-time in the regions of the original Möbius bands that were not removed when the connected sum was formed. These regions are part of the handle of the Klein bottle in Fig. 14. It should be remembered that the whole equatorial region cannot be removed when the connected sum is formed as this region is not homeomorphic to a disk (the topology of such a region was described in the discussion leading to Fig. 15). At the right end, the path has a choice as to which sheet of space-time it will emerge on. In order to understand this situation, one can refer to the classic work of Whitney.[10] A critical point at which two Möbius bands intersect is classified as type $1 \rightarrow 1$. There is another critical point where the path exits the bridge region on the left-hand side of Fig. 16. It is of type $1 \rightarrow 1$ as well. There are two critical points that are very close together in this construction. Whitney shows in Ref. 10 that two type $1 \rightarrow 1$ critical points on a manifold is consistent with that manifold having a vanishing Euler characteristic, χ. The Klein bottle has $\chi = 0$.

Before trying to draw the structure of the manifold in the vicinity of the critical points, there is one more detail that should be noted. Kerr and Schild in Ref. 11 found that the ring singularity was essentially a quadratic branch point between two sheets of space-time. Even though most of the ring singularity has been removed in forming the connected sum, not all of it can be removed. Thus, the manifold still has structure that is dictated by the nature of the ring singularity. If you circle around the singularity once you pass onto another sheet of space-time and if you circle around again you come back to the original sheet of space-time that you started on.[12] Adding this into the structure of the connected sum, one can draw the critical point structure as in Fig. 17.

A few things should be noted about the schematic nature of the diagram. Most obviously, the Möbius bands have been disconnected along the region where the connected sum is made to make a drawing possible. The big loop of the Möbius band on the left can be associated with the past directed Möbius band and the big loop on the right-hand side of Fig. 17 with the future directed Möbius band. In actuality, the two Möbius bands are wrapped around each other. However, if this were drawn one would not see the critical point structure. There are four critical points indicated in the drawing. However, due to the identification of points along the boundary of the hole where the connected sum is formed, there are really two critical points that are counted twice. The holes in the Möbius bands are designated by a series of parallel lines that mark off the evacuated areas of the holes. Where the holes meet the edge of the bridge there is drawn a very heavy line. The ring singularity is drawn as a dashed line even though the region that was evacuated by the holes. It appears as a center circle for the Möbius bands in the holes. The quadratic branch point structure is implicit in the fact that as one traces out the ring singularity, it is on the past directed sheet for one revolution and it is on the future directed sheet on the next revolution and so on. Also indicated on the drawing is a directional curve that traces out azimuthal curves along which the connected sum is made. Since these azimuthal curves trace out four disjoint arcs in Fig. 17 (in actuality there are only two arcs when the two Möbius bands are glued together) one can join them through the bridge by a unique continuous path that is indicated by a directional dotted line in the bridge region. This is the unique curve that is consistent with both the quadratic branch point structure of the singularity as well as the structure of the connected sum. This is the solution to the ambiguity of the critical point structure of Fig. 16, where it was not clear which sheet of space-time an azimuthal curve was on when it reached the bridge. It should also be noted that it takes a rotation of $\phi = \pi$ radians (due to the structure of RP^1) to circle one of the big loops and come back to the two critical points.

In order to analyze Fig. 17, one must refer back to the

structure of the canonical open line bundle over $\mathbb{R}P^1$ in one of the constituent $\mathbb{R}P^2$'s, for the sake of argument say the future directed $\mathbb{R}P^2$. This is topologically a Möbius band that has a boundary that can be thought of as a double covering of the central circle. In this case, the central circle is $\mathbb{R}P^1$. Therefore, the boundary is S^1 corresponding to a parameter value of $\phi = 2\pi$ to complete a tracing of this circle. The Möbius band is not orientable. If one were to transport a tangent frame along the bounding S^1, one could say that on the first half of the circle one is on the positive orientation sheet of the hemisphere minus a meridian, $(V_1 \cap V_3)_2$, in the model of $\mathbb{R}P^2$ corresponding to region I. On the second half of the circle it is in region III. It returns to region I after completing a rotation of $\phi = 2\pi$ radians. See the discussion on the line bundle over the Möbius band in the next section for the bundle structure of this construction. These regions were defined in the discussion of the orientation bundle earlier in this section. Similarly, for the past directed sheet, the bounding circle to the open line bundle over $\mathbb{R}P^1$ passes through regions II and IV.

From this discussion, one can analyze Fig. 17. As indicated previously, the curve that contains the four azimuthal arcs along which the connected sum is made, is clearly indicated in Fig. 17. As noted earlier there are actually only two arcs counted twice. However, if the curve is associated with the sheet of space-time it was on as it left the bridge region, the two arcs are double covered by the two sheets of space-time and stating that there are actually four arcs becomes well-defined. The curve that contains these four arcs is topologically a circle that is a double covering of S^1. It takes a parameter value of $\phi = 4\pi$ radians to complete the circle. Based on the previous paragraph, it passes through the various regions of space-time at $r = 0$ that have been defined in this section. The trajectory of this curve through these regions is tabulated in Table I as a function of the parameter ϕ.

The structure of the equatorial region fixes the global geometry of the spacelike two-surface at $r = 0$ just as the structure of the equator in the model of $\mathbb{R}P^2$ as a sphere with antipodal points on the equator identified fixes the geometry of the entire surface. This is due to the fact that the two circles that form the two equators of the hemispheres in the model of U_2 in Fig. 14 are subsets of the circle of 4π radians that contains the four disjoint azimuthal arcs of the connected sum that is drawn in Fig. 17.

Similarly, for this same curve near the equator at $\theta = 3\pi/2$ (which is distinct from the curve near the equator at $\theta = \frac{1}{2}\pi$ as implied by the remarks at the end of the previous paragraph), one has the same circle as in Table I rotated by π radians. This must be since at $\theta = 3\pi/2$ the circle must pass through the various regions of the two-surface in the same order as it did for $\theta = \frac{1}{2}\pi$. Furthermore, due to the branch point nature of the singularity, if at $\theta = \frac{1}{2}\pi$ a portion of the equator is on the future directed sheet of space-time, say region I, then after a rotation of π radians to $\theta = 3\pi/2$ this equatorial arc must pass onto the other sheet of space-time, say region II. This is consistent with the fact that each time a trajectory passes through the singularity it must change sheets of space-time. The arbitrary choice of region II as opposed to region IV is discussed in Sec. III. The alternate choice corresponds to an inequivalent tangent bundle for the two-surface at $r = 0$. The curve at $\theta = 3\pi/2$ is described in Table II.

One can construct similar tables for $\theta = 5\pi/2$ in Table III and for $\theta = 7\pi/2$ in Table IV. The actual motivation for constructing the tables as they are can only be made clear after Fig. 19 is analyzed. The results of Tables I–IV can be combined in the lattice structure that is drawn in Fig. 18.

Figure 18 contains a complete coordinate description of the curve that contains the four azimuthal arcs of the connected sum. Thus, Fig. 18 determines the global geometry of the spacelike two-surface at $r = 0$. Each time a trajectory crosses the equator it changes sheets of space-time by the quadratic branch point property of the ring singularity. Thus, a portion of the curve that contains the four azimuthal arcs of the connected sum that lies in region I can only bound region I itself or the past directed regions II and IV. Similar-

TABLE I. The structure of the curve that contains the four azimuthal arcs along which the connected sum is formed as a function of the parameter ϕ for $\theta = \frac{1}{2}\pi$.

		$\theta = \frac{1}{2}\pi$	
Region of surface at $r = 0$	Sheet of space-time	Spatial orientation	Azimuthal angular displacement
I	$t > 0$	positive	$0 < \phi < \pi$
II	$t < 0$	positive	$\pi < \phi < 2\pi$
III	$t > 0$	negative	$2\pi < \phi < 3\pi$
IV	$t < 0$	negative	$3\pi < \phi < 4\pi$

TABLE II. The structure of the curve that contains the four azimuthal arcs along which the connected sum is formed as a function of the parameter ϕ for $\theta = 3\pi/2$.

		$\theta = 3\pi/2$	
Region of surface at $r = 0$	Sheet of space-time	Spatial orientation	Azimuthal angular displacement
II	$t < 0$	positive	$0 < \phi < \pi$
III	$t > 0$	negative	$\pi < \phi < 2\pi$
IV	$t < 0$	negative	$2\pi < \phi < 3\pi$
I	$t > 0$	positive	$3\pi < \phi < 4\pi$

TABLE III. The structure of the curve that contains the four azimuthal arcs along which the connected sum is formed as a function of the parameter ϕ for $\theta = 5\pi/2$.

		$\theta = 5\pi/2$	
Region of surface at $r = 0$	Sheet of space-time	Spatial orientation	Azimuthal angular displacement
III	$t > 0$	negative	$0 < \phi < \pi$
IV	$t < 0$	negative	$\pi < \phi < 2\pi$
I	$t > 0$	positive	$2\pi < \phi < 3\pi$
II	$t < 0$	positive	$3\pi < \phi < 4\pi$

TABLE IV. The structure of the curve that contains the four azimuthal arcs along which the connected sum is formed as a function of the parameter ϕ for $7\pi/2$.

		$\theta = 7\pi/2$	
Region of surface at $r=0$	Sheet of space-time	Spatial orientation	Azimuthal angular displacement
IV	$t<0$	negative	$0\leqslant\phi\leqslant\pi$
I	$t>0$	positive	$\pi\leqslant\phi\leqslant2\pi$
II	$t<0$	positive	$2\pi\leqslant\phi\leqslant3\pi$
III	$t>0$	negative	$3\pi\leqslant\phi\leqslant4\pi$

ly, the portion of this equatorial curve that lies in region II can only border regions I, II, or III and so on. Thus, one can fill in the squares in Fig. 18 to give the unique region that is bounded by this equatorial curve for different ranges of values of the parameters θ and ϕ as is done in Fig. 19.

The motivation for choosing Table I distinct from Table III and Table II distinct from Table IV can be made clear. For instance if this equatorial curve at $\theta = 3\pi/2$ was the same as for $\theta = -\frac{1}{2}\pi = 7\pi/2$, then one could not place a region II between the equatorial curve at $\theta = \frac{1}{2}\pi$ and the equatorial curve at $\theta = 3\pi/2$. It would have to be region I again. This would contradict the fact that the curve changes sheets upon crossing the equator.

Before interpreting Fig. 19, it is probably best to digress briefly in order to expand out what is really meant by the chart. Each one of the $\pi\times\pi$ squares represents part of the Möbius band that is drawn in Fig. 4. The bridge region has been severed as indicated in Fig. 20.

Figure 19 is a convenient abbreviation for the more detailed expanded chart that is given in Fig. 21. The difference between Figs. 19 and 21 is that the infinitesimal regions that form the handle of the Klein bottle have been cut off in Fig. 19. The motivation for this is merely the infinitesimal extent of the handle. Before dismissing the handle, it is appropriate

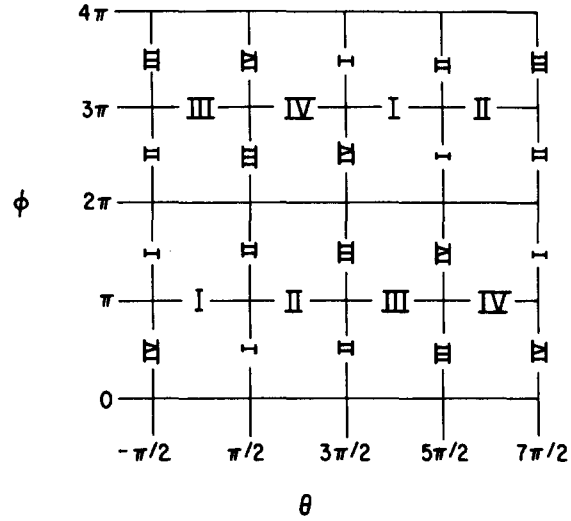


FIG. 19. The global geometry of the two-surface at $r = 0$ that is dictated by Fig. 18.

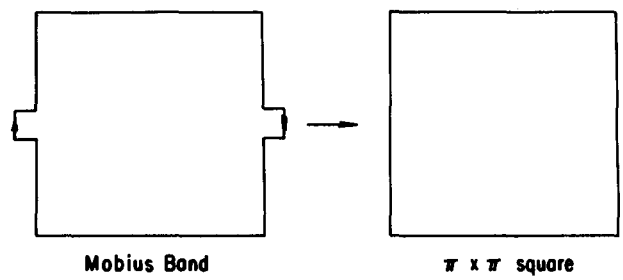


FIG. 20. A mapping of a Möbius band to one of the squares in Fig. 19.

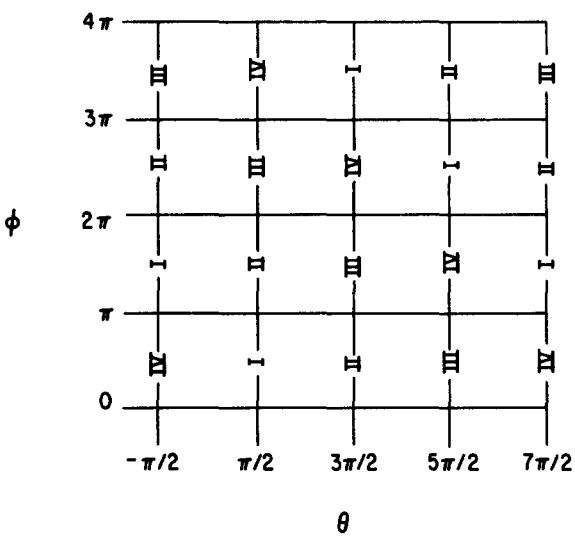


FIG. 18. This lattice structure gives a coordinate description of the equatorial curve that contains the four azimuthal arcs along which the connected sum is formed.

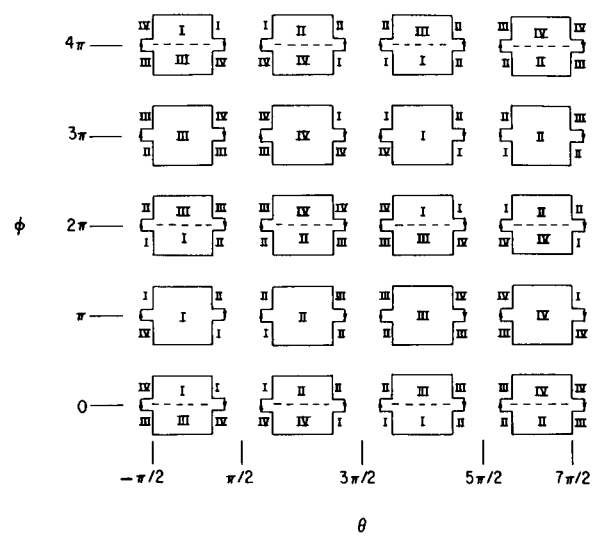


FIG. 21. An expanded version of Fig. 19.

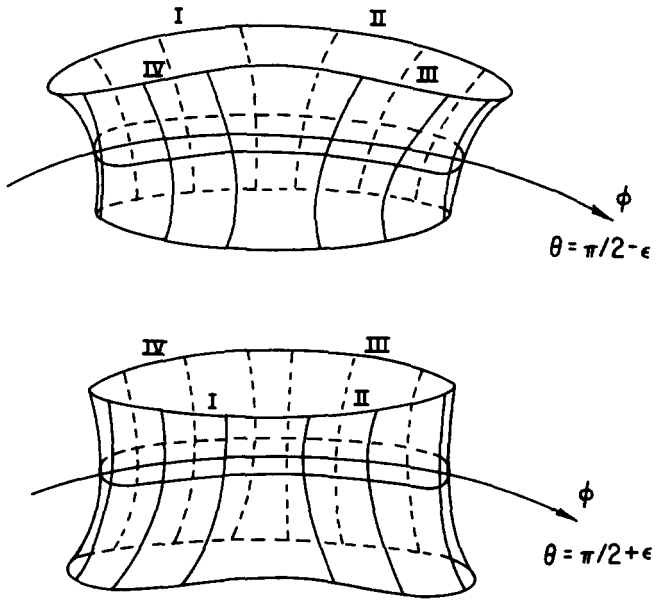


FIG. 22. The mouths of the handle of the Klein bottle.

to give a complete description of this in terms of the constituent bridges. By observing Figs. 17 and 21, one can deduce that at the mouth of the handle (i.e., where the handle connects to the hemispheres in Fig. 14) all four regions of the two-surface are used to cover the cross-sectional curve. This is illustrated in Fig. 22 as the cross section at $\theta = \frac{1}{2}\pi - \epsilon$, where ϵ is considered to be a small positive number. The topology of the Klein bottle results from reflecting this cross section through a diameter before the handle reaches $\theta = \frac{1}{2}\pi + \epsilon$ in Fig. 22. The only diameter that can be chosen for this reflection that is consistent with the nonorientability of the Klein bottle is the diameter that is parallel to the azimuthal direction that is generated by the vector field $\partial/\partial\phi$. This construction is consistent with the mandatory changing of sheets of space-time when crossing the ring singularity as dictated by the quadratic branch point structure. In Fig. 17 the handle would be manifested as a tubular loop which includes the aforementioned reflection that is connected to the rest of the manifold in the region between the critical points.

In Fig. 21 it is implicit that the adjacent parallel boundaries without arrows should all be connected. They were not so that the contributions to the handle could be drawn.

Figures 19 and 21 have some interesting global implications for the 2-D spacelike surface at $r = 0$. It is instructive to assume that one of the hemispherical shells in Fig. 14 lies on a single orientation sheet (i.e., the region admits a trivial line bundle) so that it is all of the orientation type of region I. The Kerr metric and therefore its restriction to the two-surface is independent of ϕ. One concludes that under an azimuthal rotation the surface is mapped into itself. The left-hand column of Fig. 19 implies that a rotation about the symmetry axis maps region I into region III continuously in a rotation of $\phi = 2\pi$ radians. The hemispheres in Fig. 14 are rotated from the positive orientation sheet to the negative orientation sheet. This is a totally unexpected behavior of the two-surface that results from the structure of the equatorial region. The other columns of Fig. 19 give a similar result for the other regions that cover portions of U_2. It is this change in orientation that results from azimuthal rotation that motivated the construction of the open sets W_1 and W_2 with a meridian removed. This is the behavior of each of the four regions that is dictated by the columns of Fig. 19. A rotation azimuthally of 2π radians changes the orientation of the region and a rotation of 4π radians returns the region to its original orientation. Note that azimuthal rotation does not translate a particular region into the other sheet of space-time (i.e., it preserves the local direction of time). This is consistent with the discussion at the beginning of this section on causality.

The rows of Fig. 19 have a different interpretation. The surface is not mapped into itself under a change in the coordinate θ. Otherwise, combined with the azimuthal symmetry, one would have spherical symmetry (which is clearly not manifest). The rows can be interpreted as describing what happens when a tangent two-frame is translated along a curve with $\partial/\partial\theta$ as a tangent vector. Every π radians the tangent two-frame passes from one sheet of space-time to the other as required by the quadratic branch point structure of the ring singularity. Unexpectedly, after a translation along the curve for a parameter value corresponding to 2π radians when the two-frame reemerges onto its original sheet of space-time, it has its spatial orientation reversed. Again, this results from the structure of the equatorial region through which the translation described above keeps passing through.

It is very important to note that the columns do not imply that if a tangent two-frame is translated along an azimuthal curve that it will change orientation sheets. The columns arise as a result of a rigid rotation of both a particular region and the equatorial curve that is attached to it. If the equator does not rotate, then azimuthal translation alone does not allow the two-frame to leave its original orientation sheet.

Since the Klein bottle cannot be embedded in E^3, Figs. 17, 19, 21, and 22 when combined with Fig. 14 are probably the most concise drawings that can be made that contain all of the geometric information of this particular Klein bottle. In them, it is claimed that all of the geometry of the two-surface can be ascertained.

III. THE NORMAL BUNDLE

To understand the details of the immersion of the Klein bottle in the Kerr manifold, one can study the normal bundle over U_2. This is a two-plane bundle with fibers that can be coordinated by the Boyer–Linquist coordinates r and t. Besides elucidating whether one has an immersion or an embedding, one is also interested in how many vector or line sections can be found. The two problems are somewhat related. This study will be conducted through the use of characteristic classes and obstruction cocycles. It is through these concepts that one can analyze what happens as a principal null geodesic is translated around the two-surface.

The first thing that one should look at is the top obstruction cocycle, the Euler number of the normal bundle [de-

noted by $\chi(N)$]. In Ref. 1, $\chi(N)$ was evaluated in (B5). However, this may or may not be correct as there is an ambiguity in the defining relation. In the following, a more detailed account of how this result was calculated is given along with a discussion of where the ambiguity arises. After this elaboration, one does the next best thing. A calculation is performed that is more indirect in order to obtain a similar result. This author was only able to verify a weaker though unambiguous statement about the normal bundle by means of an alternative construction to that of (B5). In Ref. 1 it was "found" that $\chi(N) = 0$. In the alternative treatment to be discussed in the following, it will be shown that, to say the least, $\chi(N) \bmod 2 = 0$.

The result in Ref. 1 was based on the equation of structure for submanifolds

$$dw^r{}_t = N\,dA = w^r{}_\alpha \wedge w^\alpha{}_t + R^r{}_t, \tag{3.1}$$

where the $w^i{}_j$ are the connection one-forms and $R^r{}_t$ is the curvature two-form of the Kerr manifold. The quantity $dA = |w^\theta \wedge w^\phi|$ is the area element of the submanifold and N is the normal curvature. In Ref. 1 these quantities are calculated in the zero angular momentum frames. After solving the structure equations in this frame and restricting the results to the hypersurface at $r = 0$, one has on the future directed sheet of space-time:

$$dw^r{}_t = -(m/a^3)\sec^5\theta(2 + \sin^2\theta)dA = R^r{}_t. \tag{3.2a}$$

By the first equation of structure $w^r{}_t \to -w^{\hat{r}}{}_{\hat{t}}$ under the discrete isometry $t \to -t$ and $\phi \to -\phi$ since the basis covectors defined in Ref. 1 transform as $w^t \to -w^{\hat{t}}$ and $w^r \to w^{\hat{r}}$ under the discrete isometry. Thus, on the past directed sheet of space-time one has

$$dw^{\hat{r}}{}_{\hat{t}} = (m/a^3)\sec^5\theta(2 + \sin^2\theta)dA. \tag{3.2b}$$

In (B5), the Euler characteristic is computed by a Gauss–Bonnet-type integral,

$$\chi(N) = \int_\Omega dw^r{}_t = \int_\Omega N\,dA = 0. \tag{3.3}$$

The first problem with this formalism is that if the Kerr manifold is orientable, the normal fiber cannot be oriented over U_2 since the base manifold is nonorientable. There is no Gauss–Bonnet theorem for nonorientable surfaces or fibers. Even though there is no theorem that is proved or stated in the mathematical literature, an extension of the Gauss–Bonnet theorem to the nonorientable case with two-dimensional fibers is known among a community of mathematicians.[13] The reason for the extension of the theorem in the two-dimensional case is straightforward. One can define an area on a nonorientable surface by dividing the surface into simplices and adding up the areas of the constituent simplices. In the two-dimensional case, the Euler class is $dw^r{}_t$, which is just a function (the curvature) times the area on each simplex (see Milnor and Stasheff[9] for a discussion of the Euler class). Thus, one does not have to worry about densities or orientations. One just adds up the contributions from each simplex. Clearly, if the simplicial structure is made fine enough this can be incorporated into the Riemannian definition of the integral of this curvature function. The integral over the whole surface can be thought of as a sum of numbers times infinitesimal area elements. This simplicial method is essentially the way that Whitney originally defined the normal Euler number in Ref. 10 (in his notation the normal Euler number is $W^2 \cdot M$) for any two-manifold immersed in a four-dimensional space. Since the Euler class is more complicated in higher dimensions, it is not at all clear if this argument can be extended to the higher-dimensional case.

Applying this result to (3.3), the result follows directly from (3.2). Since each of the Möbius bands in the connected sum are identical with the same areas and the normal curvatures are identical except for the sign, the integral in (3.3) must vanish identically. However, there is an ambiguity. The normal curvatures diverge at the ring singularity. On one side, the normal curvature approaches to $+\infty$ and on the other side $-\infty$. When the connected sum is formed, the region where the normal curvature is infinite is removed except for an infinitesimal set. Where the singularity has not been removed, one is integrating infinite quantities and there can be some ambiguity as to what the integral means, even though there is an identical cancellation of the divergences since there is an equally negative divergent contribution on the future directed sheet of space-time to cancel the positive divergence on the past directed sheet by (3.2a) and (3.2b).

The fact that this infinitesimal set cannot vanish and maintain the global topology of the two-surface, one has to be wary of any limiting process that tries to sweep this problem under the rug. As this is a crucial result that is just a starting point for the following analysis, one needs a statement with a stronger foundation even though (3.3) may actually be correct.

Fortunately, one can get information on the Euler number without resorting to (3.3). For the purposes of this paper, this alternate treatment will reveal all of the information that is needed in the following. It is shown in the Appendix that the existence of a normal line section is equivalent to the fact that $\chi(N) \bmod 2 = 0$. There are more direct ways of obtaining this result that rely on the Whitney product theorem (3.11).[14] However, the treatment in the Appendix involves a Gauss–Bonnet-type theorem for nonorientable surfaces. A surface integral of a density (an element of twisted cohomology) mod 2 gives the Euler number mod 2. As there is no discussion of such a theorem in the mathematical literature that this author could find, it is presented in the Appendix.[15] Furthermore, this kind of theorem seems easily extendible to higher dimensions, while the previous discussion of the Gauss–Bonnet theorem for two-dimensional nonorientable fibers that involved simplices may not be.

In the spirit of the Appendix, one simply proceeds to prove the existence of a trivial normal line bundle. This bundle is physically significant as it allows one to find a constant time cross section that contains both sheets of space-time. One can find a line bundle on each sheet of space-time in a straightforward manner everywhere except at the ring singularity. Since there is a pseudo-Riemannian metric on the manifolds (the two sheets of space-time at $r = 0$), $g_{\mu\nu}(x)$, there exists a splitting.[16] One can put a Riemannian metric on the same manifold, $\tilde{g}_{\mu\nu}(x)$. The splitting is defined by the eigenvalue equations

$$\tau^{\mu} g_{\mu\nu}(x)\tau^{\nu} = (-1)\tau^{\mu}\tilde{g}_{\mu\nu}(x)\tau^{\nu}, \tag{3.4a}$$

$$\lambda_i^{\mu} g_{\mu\nu}(x)\lambda_i^{\nu} = (+1)\lambda_i^{\mu}\tilde{g}_{\mu\nu}(x)\lambda_i^{\nu},$$

$$i = 1,2,3 \text{ with no sum over } i, \tag{3.4b}$$

and $\tau_i \cdot \lambda_i = 0$, $\forall i$.

The solutions to (3.4a) form a line bundle. Here τ^{ν} is the timelike component of the orthogonal tetrad completed by the λ_i's. Both $\pm\tau^{\nu}$ are solutions to (3.4a). Thus, one has a vector section defined by (3.4a) up to a sign. This is a line section. For example, τ^{ν} might be the vector field $\partial/\partial t$. The problem with performing such a splitting at the ring singularity at $\theta = \pi/2$ and $r = 0$ is that the metric coefficient g_{tt} becomes singular and positive,

$$g_{tt} = -(1 - 2mr/(r^2 + a^2\cos^2\theta))\big|_{\substack{r=0 \\ \theta=(1/2)\pi}} = +\infty. \tag{3.5}$$

A limit is not used in (3.5) as it reaches this value discontinuously for fixed r. Some of the other metric coefficients become either zero or infinite at the ring singularity. The metric does not even retain its signature of (3,1) at the singularity. There is no splitting here.

In order to deal with this pathology, notice that there is a splitting above and below a neighborhood of the equatorial region. One can therefore trivialize the normal line bundle in the region above the equatorial neighborhood and in the region below independently. One wants to know if these two trivializations can be joined continuously through the equatorial region. It was found in the discussion dealing with Fig. 17 that the boundary of this region passes through both sheets of space-time, i.e., the regions above and below. Thus, if the normal line bundle can be trivialized on this boundary, the line section can be joined continuously from one sheet of space-time to the other. There is one more region where the line bundle must be extended, namely, the interior of the equatorial neighborhood. The splitting exists all the way down to but not including those parts of the equator on each Möbius band that still remains after the holes were cut to make the connected sum. Thus, if the two trivializations agree on the boundary of the equatorial region, one can trivialize everywhere in the equatorial neighborhood except possibly on what remains of the two center circles of the two Möbius bands. In fact this trivialization can be extended to the center circles before the holes were cut. There is no difficulty in constructing a line bundle over the center circle of a Möbius band as is evidenced by the line section formed by the tangent vector to the center circle (see Fig. 7). Any line bundle that does not involve the rotation between the tangent plane and the spacelike normal vector that is described at the beginning of Sec. IV can clearly be trivialized over this center circle. If the trivialization can be extended to the center circle before the holes were cut, it certainly can be after the holes are cut. The difference between this line bundle and the orientation bundle is that the orientation bundle cannot be trivialized all of the way down to the center circles as one can with the normal bundle due to the splitting and the abrupt discontinuity in (3.5). The problem of trivializing a normal line bundle over U_2 has been reduced to whether the line bundle can be trivialized on the boundary of the two

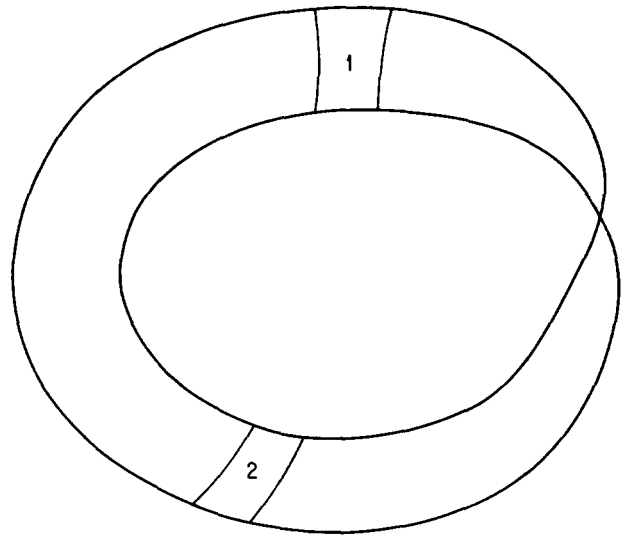


FIG. 23. The two components of the intersection of a two set open covering of a Möbius band are labeled by "1" and "2."

intersecting Möbius bands in Fig. 17.[13] One does not even have to worry about the disklike regions that were removed. This does not affect the boundary of the neighborhood.

First of all one describes a relevant example that has all of the concepts and constructions in the argument to follow, but has the advantage that it is easy to visualize and can be described explicitly in terms of easy to follow figures. As a simple example, one looks at the problem of trivializing a line bundle on the boundary of a Möbius band. This was encountered in the discussion of the orientation bundle over S^1, the boundary of the canonical open line bundle over $\mathbb{R}P^1$. If one were to cover a Möbius band with two open sets, as in Fig. 23, with the intersection denoted by the two open sets that are labeled 1 and 2, then there are two transition functions for any line bundle that can take on the values 1 or -1. Call these functions $a_1 = +1$ or -1 and $a_2 = +1$ or -1. If one restricts these open sets to the boundary of the Möbius band then there are four open one-dimensional sets in the intersection as indicated in Fig. 24. There are now four transition functions that need to be specified to define a line bundle over the boundary of the Möbius band. These transition functions are labeled in Fig. 24. One has $a_1 = a_1' = a_1''$ and $a_2 = a_2' = a_2''$. Clearly, the number of minus signs is an even number for any line bundle over the Möbius band that is restricted to the boundary. The various possibilities are tabulated in Table V.

One can conclude that if there is a line bundle over a Möbius band, then there exists a continuous nonzero line section on the boundary.[13] This is a justification of the discussion of the orientation bundle in the last section. It was stated that on the boundary of the canonical line bundle over $\mathbb{R}P^1$ that if a tangent two-frame was transported around this circle then it must return to its starting point with the same orientation.

This same argument can be applied to the neighborhood of the equator. However, there are many more intersections. Drawing these on top of Fig. 17 would be very confusing. The equatorial region can be covered by four open sets, two

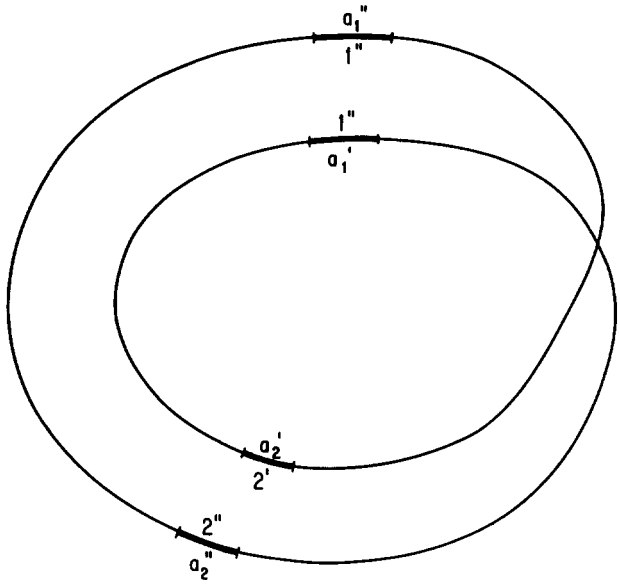


FIG. 24. The two components of the intersection of the two open sets that are labeled in Fig. 23 have been restricted to the boundary of the Möbius band yielding four one-dimensional open sets.

for each Möbius band. Call these sets $V_{\rm I}$ and $V_{\rm III}$ for one of the Möbius bands and $V_{\rm II}$ and $V_{\rm IV}$ for the other. The Roman numeral subscripts are meant to be suggestive of the orientation bundle that was described earlier. $V_{\rm I}$ and $V_{\rm III}$ intersect in only one open set since the two Möbius are cut and spliced between the critical points (see Fig. 17). The intersection must be outside of the region of the handle of the Klein bottle that contains the two critical points (i.e., in the big loop of the Möbius band as drawn in Fig. 17). Similarly, $V_{\rm II}$ and $V_{\rm IV}$ intersect in one open set. All other pairs of sets intersect in two disjoint open sets. One of these disjoint sets lies between the two critical points. The other envelopes a region of the curve along which the connected sum is made. This cannot be pictured in Fig. 17 as the connected sum was unraveled to illustrate the critical point structure. There are a total of ten transition functions, $\{a_i | i = 1,...,10\}$, that must be specified for a normal line bundle over this region. As with the Möbius band, when these functions are restricted to the bounding circle there are 20 transition functions that need to be specified, $\{a_i^1, a_i^2 | i = 1,...,10\}$, where $a_i^1 = a_i^2 = a_i$. Consequently, for these 20 transition functions there are an even number of minus signs, as in Table V. Hence, one can trivialize a normal line bundle over the boundary of the equatorial region. Consequently, one can choose a time direction continuously on the Klein bottle that is never zero. This is a scalar time that can be chosen positive everywhere on the two-surface at $r = 0$. For instance, if this time direction appears future directed at asymptotic infinity then it will appear to point backwards in time on the past directed sheet of the Klein bottle. One can also conclude from the existence of this line section and the results of the Appendix that $\chi(N) \mod 2 = 0$.

TABLE V. The transition functions for various line bundles over the boundary of a Möbius band.

Values of the transition functions of the line bundle over the Möbius band		Number of minus signs of the transition functions of the line bundle on the boundary
$a_1 > 0$	$a_2 > 0$	0
$a_1 > 0$	$a_2 < 0$	2
$a_1 < 0$	$a_2 > 0$	2
$a_2 < 0$	$a_2 < 0$	4

Now that it has been verified that the line bundle of time is trivializable, it is of interest to look at the hypersurface that is formed by this line bundle over U_2 (the hypersurface given by $r = 0$). The nature of the immersion of this surface in the Kerr space-time will yield information on the orthogonal complement of the coordinate t in the normal fiber on U_2. The conclusion will be that $\chi(N) \neq 0$. Consequently, the concern over the divergence in the integrand of the Gauss–Bonnet integral for normal fibers that was expressed at the beginning of this section was justified.

In order to study the immersion of the hypersurface at $r = 0$ in the Kerr space-time, one introduces the extrinsic curvature tensor of the hypersurface, $K_{\alpha\beta}$. In terms of the basis covectors, w^α, and the connection one-forms, $w^\alpha{}_\beta$, in the orthonormal frames of the ZAMO's (the zero angular momentum observers, see Ref. 1), the extrinsic curvature tensor is defined by

$$w^r{}_\alpha = K_{\alpha\beta} w^\beta.$$

It is not difficult to convince oneself that this is the same extrinsic curvature tensor that is defined in Ref. 17 by the relation

$$dn^\alpha = K^\alpha{}_\beta \, dx^\beta,$$

where n^α is the unit normal to the hypersurface and dx^β is a one-form on the hypersurface. Therefore, dn^α is a vector-valued one-form. The two expressions for the extrinsic curvature are related by writing

$$dn^\alpha = \nabla_\beta n^\alpha \, dx^\beta,$$

and using the definition of the covariant derivative in the ZAMO frame.

The equations of structure were solved in Ref. 1 and restricted to the hypersurface $r = 0$. Noting the corrections of signs from (B1) of Ref. 1,

$$K_{\phi\phi} = M \sin^2\theta / (a^2 |\cos^3\theta|),$$
$$K_{\phi t} = -M \sin^2\theta / (a^2 |\cos^3\theta|),$$
$$K_{t\phi} = M \sin\theta / (a^2 |\cos^3\theta|),$$
$$K_{tt} = -M / (a^2 |\cos^3\theta|),$$

all the other components of the tensor vanish.

The tensor has three eigenvalues, the principal curvatures of the immersion,

$$k_1 = k_2 = 0 \text{ and } k_3 = -M / (a^2 |\cos\theta|).$$

There are three corresponding eigenvectors, the principal directions,

$$e_1 = \hat{e}_\theta, \quad e_2 = |\cos\theta|^{-1} (\hat{e}_\phi + \sin\theta \hat{e}_t),$$
$$e_3 = |\cos\theta|^{-1} (\sin\theta \hat{e}_\phi + \hat{e}_t),$$

where $\hat{e}_\alpha$ are the ZAMO basis vectors. It is interesting that e_3 is the four-velocity of the Klein bottle in space-time (see Ref.

1). The only curvature in the immersion is along the four-velocity of the Klein bottle. The world lines of the points of the Klein bottle are lines of curvature (the only lines of curvature) of the immersion. This suggests that the slicing of space-time at $r=0$ and $t=$ const is not just a random slicing of space-time. This fact lends more credence to the conjecture that the two-surface behaves as if it were a source for the geometry.

It should be noted that k_3 diverges when $\theta=\frac{1}{2}\pi$. Since the entire equatorial region is not removed in the formation of the connected sum, this divergence must be investigated. This divergence is what gives rise to the normal curvature divergence that was discussed at the beginning of this section, as well as the divergence of $R^{\mu\lambda\sigma\delta}R_{\mu\lambda\sigma\delta}$ for the Kerr space-time.

On the equator, e_3 is a null vector. From the definition of the extrinsic curvature in Ref. 17, it is obvious that one cannot trivialize the spacelike normal $\hat{e}_r$ over an open set that contains the equator. This is clear since a displacement of $\hat{e}_r$ along the world line of a point of the two-surface at $\theta=\frac{1}{2}\pi$ induces an infinite change in $\hat{e}_r$ (k_3 diverges). There is no continuous nonzero spacelike normal section on any open set that includes the equator.

Two things can be happening that explain this phenomenon. First, the hypersurface may not be well behaved at $\theta=\frac{1}{2}\pi$. Second, the Kerr manifold is not well behaved ($R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma}$ diverges). This may be inherent to the space external to the hypersurface and therefore the external space might be entirely responsible for the pathology that is observed in the extrinsic curvature. The fact that there are only two distinct principal directions of the extrinsic curvature tensor implies that a careful analysis is necessary to make sense out of the divergence of k_3. Since it has already been proved that the timelike line bundle is trivializable, one would suspect that the hypersurface is well behaved. It will be shown in the following how this can happen. The fact that k_3 diverges in a smooth manner as the equator is approached from another point of the hypersurface makes one believe that the divergence is real and significant. The breakdown of the eigenspace of the curvature tensor should be understandable in a way that is consistent with the nature of this divergence.

In order to ascertain the truthfulness of these conjectures, one must sort out two problems. First, the Boyer–Lindquist coordinates are not well behaved at $r=0$. Second, there is an actual pathology in the manifold structure ($R^{\mu\lambda\delta\sigma}R_{\mu\lambda\delta\sigma}$ diverges). Thus, there is no change of coordinates that allows the metric to be extended to the set $r=0$, $\theta=\frac{1}{2}\pi$.

The Kerr metric can be expressed as the following line element in Boyer–Lindquist coordinates:

$$\begin{aligned} ds^2 = &-(1-2Mr/\rho^2)dt^2+\rho^2\,d\theta^2+(\Delta^2/\rho^2)dr^2 \\ &+[(r^2+a^2)+(2Mra^2/\rho^2)\sin^2\theta\,]\sin^2\theta\,d\phi^2 \\ &-(4Mra/\rho^2)\sin^2\theta\,d\phi\,dt, \end{aligned}$$

where

$$\rho^2=r^2+a^2\cos^2\theta,\quad \Delta=r^2+a^2-2Mr.$$

At $r=0$, $\theta=\frac{1}{2}\pi$: $\rho^2\to 0$ and the metric coefficients go haywire, some of them discontinuously. Also, the volume element $dV\sim\rho^2\sin^2\theta\to 0$.

Denote the orthogonal stationary frame at infinity by the usual spherical coordinate basis, $\tilde{e}_i$. Let $\hat{e}_j$ denote a leg of the tetrad that is carried by a ZAMO. The coordinate transformation between the two frames is

$$\begin{bmatrix}\hat{e}_t\\ \hat{e}_\phi\end{bmatrix}=\begin{bmatrix}|g_{tt}-\Omega^2 g_{\phi\phi}|^{-1/2} & \Omega|g_{tt}-\Omega^2 g_{\phi\phi}|^{-1/2}\\ 0 & g_{\phi\phi}^{-1/2}\end{bmatrix}\begin{bmatrix}\tilde{e}_t\\ \tilde{e}_\phi\end{bmatrix},$$

$\hat{e}_r=(\Delta^{1/2}/\rho)\tilde{e}_r$ and $\hat{e}_\theta=(1/\rho)\tilde{e}_\theta$,

where

$$\Omega\equiv -g_{\phi t}/g_{\phi\phi}.$$

The basis convectors transform as

$$\begin{bmatrix}w^t\\ w^\phi\end{bmatrix}=\begin{bmatrix}|g_{tt}-\Omega^2 g_{\phi\phi}|^{1/2} & 0\\ -\Omega g_{\phi\phi}^{1/2} & g_{\phi\phi}^{1/2}\end{bmatrix}\begin{bmatrix}dt\\ d\phi\end{bmatrix}.$$

This definition of the ZAMO basis in terms of Boyer–Lindquist coordinates is not well defined when $\rho\to 0$ as well. This does not imply that there is no extension of the ZAMO frames (or at least some of the legs of the ZAMO tetrad may be extendible in a well-defined manner) to $\rho=0$. One might be able to define the legs of the tetard in another coordinate system.

In order to see just how much of the pathology of the space-time can be removed by a coordinate transformation, the well-behaved nature of the following area element should be noted in the ϕ-t plane:

$$dA_{\phi t}=\left\{\det\begin{vmatrix}g_{tt} & g_{\phi t}\\ g_{\phi t} & g_{\phi\phi}\end{vmatrix}\right\}^{1/2}d\phi\,dt=\Delta^{1/2}\sin\theta\,d\phi\,dt,$$

$$\lim_{\rho\to 0}dA_{\phi t}=a\sin\theta\,d\phi\,dt.$$

The implication is that the pathology of the metric in this two-dimensional slicing of space-time is due entirely to a poor choice of coordinates. A change of coordinates can be used to get a well-behaved two-dimensional space of signature (1,1).

Since $\hat{e}_\phi$ and $\hat{e}_t$ only depend on $\partial/\partial t$ and $\partial/\partial\phi$ and the metric coefficients g_{tt}, $g_{\phi t}$, and $g_{\phi\phi}$, one concludes that a change of basis can be chosen so that these two legs of the ZAMO tetrad can be extended to $\rho=0$ in a well-defined manner. Thus, the four-velocity of the two-surface is still defined (as well as e_3) when $\rho=0$.

One also needs to check that the concept of the two-surface is still defined as $\rho\to 0$. Going back to the definition of the Gaussian curvature of the two-surface:

$$dw^\theta{}_\phi=K\,dA,$$

$K\equiv 0$, except that it changes discontinuously to ∞ at $\rho=0$, since

$$\begin{aligned} w^\theta{}_\phi = &-\frac{\sin\theta\cos\theta}{g_{\phi\phi}^{1/2}}\left[(r^2+a^2)+\frac{2Mra^2\sin^2\theta}{\rho^2}\right. \\ &\left.+\frac{2Mra^2(r^2+a^2)\sin^2\theta}{\rho^4}\right]d\phi \\ &+(\text{a function of }\theta\text{ and }\phi)dt \end{aligned}$$

is not well-defined at $\rho=0$. However, neither are the basis vectors $\hat{e}_\theta$ and $\hat{e}_\phi$. It was shown that by a change of coordi-

nates $\hat{e}_\phi$ can be extended to $\rho = 0$. But, what about $\hat{e}_\theta$ and $w^\theta{}_\phi$?

The fact that $w^\theta{}_\phi$ is discontinuous is actually not a cause for alarm. The problem is totally analogous to a misconception that Chern had as a young man when he thought that he had "proved" that all two-surfaces are flat.[18] His reasoning went as follows. On a two-surface, S, $w^\theta{}_\phi$ is a one-form. Therefore $dw^\theta{}_\phi$ is a two-form. Hence, $dw^\theta{}_\phi$ is proportional to the area element ($dw^\theta{}_\phi = K\,dA$). The form $dw^\theta{}_\phi$ represents the Euler class in cohomology. But, if $w^\theta{}_\phi$ is defined in a global basis then $dw^\theta{}_\phi$ represents 0 in the cohomology module $H^2(S)$ ($dw^\theta{}_\phi$ is exact). This implies that the Euler class is zero or equivalently that the surface is flat.

The conclusion is wrong because by assuming that $w^\theta{}_\phi$ is defined in a global basis assumes the conclusion (the basis vectors $\hat{e}_\theta$ and $\hat{e}_\phi$ are two continuous vector sections). The correct way of defining the connection one-form is in terms of an open cover $\{U_\alpha\}$ of S,

$$w^\theta{}_\phi = \sum_\alpha (w^\theta{}_\phi)_\alpha,$$

where $(w^\theta{}_\phi)_\alpha$ is a one-form on U_α that can be pieced together to form a global one-form on S. Obviously, one needs more than one open set to cover U_2. The behavior of $w^\theta{}_\phi$ at $\rho = 0$ is due to the fact that the basis vector $\hat{e}_\theta$ cannot be a global continuous nonzero vector field on U_2 if $\hat{e}_\phi$ has already been chosen to be one (U_2 admits only one independent continuous nonzero vector section) as well as the pathological nature of the coordinates at $\rho = 0$.

The conclusion is that the two-surface is indeed flat and coordinates can be defined on the surface locally but not globally. There is no nonzero extension of $\hat{e}_\theta$ to $\rho = 0$, yet the two-surface is perfectly well behaved. This is manifested in the vanishing of $g_{\theta\theta}$ at $\rho = 0$. The only way of extending $g_{\theta\theta}$ continuously through $\rho = 0$ on the two-surface is if it vanishes. Remember that there is only one global nonzero continuous vector field on U_2. But, if the orthogonal vector field is allowed to vanish at certain points it can be defined continuously as well.

The inescapable conclusion of these arguments is that the hypersurface is well behaved and it is the spacelike normal $\hat{e}_r$ that is pathological. The divergences in $R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma}$ and k_3 imply that the space-time is not a manifold at $\rho = 0$. As a matter of fact the entire concept of an $\hat{e}_r$ direction breaks down on a two-dimensional set that is spanned by the null vectors, $\hat{e}_\phi + \hat{e}_t$ (the four-velocity of the two-surface at $\theta = \frac{1}{2}\pi$) and $-\hat{e}_\phi + \hat{e}_t$. Since the hypersurface is well defined, the divergence in k_3 implies that $\hat{e}_r$ cannot be well defined at $\rho = 0$. Thus, there is no nonzero continuous section of $\hat{e}_r$ on any open set of the hypersurface, or therefore the two-surface, that contains the set $\rho = 0$. Finally, one arrives at the conclusion that $\chi(N) \neq 0$ even though it was previously found that $\chi_2(N) = 0$.

The pathology of the radial coordinate at $\rho = 0$ can be caused by one of two things. The space-time could be "cusped" or "bumpy" in the radial direction. Alternatively, the radial direction might be "curled-up" in a continuous manner along the two-surface near $\rho = 0$ as indicated by the continuity of k_3. Finally, at $\rho = 0$ the radius of curvature of this curling-up process actually becomes zero and the dimension in the spacelike normal direction vanishes. The resulting space-time would be four-dimensional everywhere except along a two-dimensional subset where the dimension of space-time is reduced to 3 (only the hypersurface dimensions remain). Thus, the space-time is not a manifold. However, one can still define subspaces that are manifolds. Since the larger space is a CW complex (see Ref. 19 for a definition) as stated in Ref. 9, one can still define characteristic classes on subspaces that are manifolds as is done in this paper. Finally, it should be noted that all of these results involving the immersion of the hypersurface are unchanged when the open set is removed in the formation of the connected sum.

Now that the normal Euler number has been analyzed, it can be used with confidence in the following. A key assumption that was mentioned earlier in this section is that the global four-manifold should be orientable. This is a physically reasonable assumption that is compatible with conservation of angular momentum. If the space were not orientable a particle with positive helicity can travel around a closed path and return to its starting point with negative helicity. This does not seem physically reasonable. If the four-manifold is orientable then there exists a local $\mathbf{Z}$ (the set of integers) orientation of M (the Kerr manifold) along U_2 (Ref. 19). With these two facts one can construct most of the relevant characteristic classes and obstruction cocycles.

These two facts can be implemented in a study of the Stiefel–Whitney classes of the normal and the tangent bundles. These are the relevant characteristic classes since the bundles to be considered are nonorientable. The Stiefel–Whitney classes of a fiber bundle, B, are equivalence classes in a sequence of cohomology modules with coefficients in $\mathbf{Z}_2$ (integers mod 2),[9]

$$w^i \in H^i(B,\mathbf{Z}_2), \quad i = 0,1,2,\ldots . \tag{3.6}$$

The Stiefel–Whitney classes can be defined by a set of axioms as in Milnor and Stasheff.[9] They also have a geometrical significance in terms of obstruction theory. This is how Whitney first discovered them in Ref. 10.

In order to elaborate on the interpretation as an obstruction to forming a cross section, assume that the bundle, B, is a finite cell complex (Milnor and Stasheff say in Chap. 12 of Ref. 9 that this argument is also true for CW complexes). If L is a subcomplex with a cross section defined on it then it is trivial to extend the cross section to the remaining 0-skeleton, K^0, of B that is not contained in L (Steenrod, p. 148).[8] One can try to continue the process as follows. A cross section is now defined on $L \cup K^0$, but can it be extended to the one-cells, $L \cup K^1$? This can happen if and only if a certain obstruction cocycle, $c^1(B)$, vanishes. Similarly, if the cross section can be defined on the $q-1$ skeleton, $L \cup K^{q-1}$, it can be extended to $L \cup K^q$ if and only if the obstruction cocycle, $c^q(B)$, represents the 0 element in the q-dimensional cohomology module of the bundle.[8] As a consequence of this, if B is a trivial bundle then $c^i(B) = 0\ \forall i$.

When i is even $c^i(B)$ is an element of the cohomology module $H^i(B,\mathbf{Z}_2)$. When i is odd or $i = n$, where n is the dimension of the fiber, the coefficients of the cohomology

module are infinite cyclic.[8] However, as Milnor and Stasheff point out they are not canonically isomorphic to the integers.[9] They are a twisted bundle of coefficients as described in Ref. 8. These coefficients will be labeled ∞. The discussion in this paragraph is tabulated in (3.7):

$$\begin{aligned} &c^i(B)\in H^i(B,\infty), \quad i=\text{odd or } i=n, \\ &c^i(B)\in H^i(B,\mathbf{Z}_2), \quad i=\text{even}. \end{aligned} \tag{3.7}$$

The Stiefel–Whitney classes alleviate the complications of the twisted bundle of coefficients by taking the unique nontrivial homomorphism of $\infty \to \mathbf{Z}_2$. The Stiefel–Whitney classes are the obstruction cocycles mod 2,

$$\begin{aligned} &w^i(B)=c^i(B) \bmod 2, \quad i=\text{odd or } i=n, \\ &w^i(B)=c^i(B), \quad i=\text{even}. \end{aligned} \tag{3.8}$$

The cost of taking the obstruction cocycles mod 2 is a loss of information. The gain is that there are formulas that are derivable from the ring structure of $H^i(B,\mathbf{Z}_2)$, namely, the Whitney product formula, (3.11). This digression is necessary because the w^i's do not contain enough information in themselves and a knowledge of the c^i's is needed in some of the following.

Every manifold has a $\mathbf{Z}_2$ orientation. Define the fundamental homology class of a manifold M of dimension n corresponding to the orientation as $\mu\in H_n(M,\mathbf{Z}_2)$.[19] If one takes the Kronecker product $[w^2(N),\mu]$ then the resulting number $\chi_2(N)$ is the normal Euler number modulo 2, $w^2(N)$ is the second Stiefel–Whitney class of the normal bundle over the Klein bottle, and μ is the fundamental class on U_2,[10]

$$[w^2(N(U_2)),\mu]=\chi_2(N) \tag{3.9a}$$

and similarly

$$[w^2(T(U_2)),\mu]=\chi_2(U_2), \tag{3.9b}$$

where $T(U_2)$ is the tangent bundle over U_2.

For a Klein bottle $\chi=0$. Thus, $\chi_2(U_2)\equiv\chi_2(T(U_2))=0$ and by the nondengeneracy of the Kronecker product,[19]

$$w^2(T(U_2))=0. \tag{3.10a}$$

It was also deduced in this section that $\chi_2(N(U_2))=0$, hence

$$w^2(N(U_2))=0. \tag{3.10b}$$

In Corollary 11.4 of Ref. 9 Milnor and Stasheff show that the vanishing of $w^2(N)$ is consistent with but does not prove that the immersion of the base submanifold (in this case U_2) is an embedding. There are a couple of reasons to conjecture that this might be an embedding. First, the Gaussian curvature vanishes everywhere making this a flat Klein bottle. It is hard to understand how a surface can have self-intersections if the curvature vanishes. Second, by the method of moving frames for a submanifold (2.2) and (3.1), $K=R^{\theta}{}_{\phi}$ and the normal curvature $N=R^{r}{}_{t}$. In this sense, the two-dimensional surface at $r=0$ is molded to the geometry of the Kerr manifold at $r=0$. This lends more credence to the conjecture that the hypersurface at $r=0$ is a physically significant slicing of the global four-manifold (i.e., a source).

It is not hard to convince oneself that a bundle is orientable if and only if $c^1(B)=0$ (Steenrod, Theorem 38.12). But U_2 is not orientable so $c^1(T(U_2))\neq 0$. Since this is an element of odd-dimensional cohomology, then by (3.8), $c^1(T(U_2)) \bmod 2 = w^1(T(U_2))$. Thus, $w^1(T(U_2))$ may or may not be zero. Write $w^1(T(U_2))=a$. By Theorem 4.10 of Ref. 9, if all of the Stiefel–Whitney classes of a manifold, M, are zero then M can be realized as a boundary of a smooth compact manifold. Take the standard immersion of U_2 in E^3. This is a one-sided surface.[5] It cannot bound anything. Thus, $a\neq 0$ for all Klein bottles.

One can express $w^1(N(U_2))$ in terms of $w^1(T(U_2))$ by the Whitney product formula. Defune two bundles ξ and η over the same base space then the Whitney product formula can be written as[9]

$$w^k(\xi\oplus\eta)=\sum_{i=0}^{k} w^i(\xi)\smile w^{k-i}(\eta), \tag{3.11}$$

where $\smile$ is the cup product and $w^0=1\in H^0(B,\mathbf{Z}_2)$ and the symbol $\oplus$ denotes the Whitney sum of ξ and η (see Milnor and Stasheff,[9] p. 27). This is a useful relation because the tangent bundle of the Kerr manifold, M, along U_2 is isomorphic to the Whitney sum of $T(U_2)$ and $N(U_2)$ (Corollary 3.4 in Milnor and Stasheff[9]),

$$T(U_2)\oplus N(U_2)\cong T(M)_{|U_2} \tag{3.12}$$

where $T(M)_{|U_2}$ is the tangent bundle to M along U_2 (i.e., an $\mathbf{R}^4$ bundle over U_2).

It was previously assumed that the Kerr manifold is orientable. Thus, there exists a local $\mathbf{Z}$ orientation of M along U_2. Therefore, $w^1(T(M)_{|U_2})=0$ by the previous remarks about $c^1(B)$ and orientability. According to the Whitney product formula, this condition can only be satisfied if $w^1(N(U_2))=-a=a$ (since there are $\mathbf{Z}_2$ coefficients $a=-a$), which is consistent with the fact that the normal fiber over U_2 must be nonorientable if U_2 is embedded in an orientable space. The Stiefel–Whitney classes are compiled in Table VI. Note that $a\smile a$ is not necessarily zero. The cup product with $\mathbf{Z}_2$ coefficients implies that $2(a\smile a)=0$.[19]

Another use of (3.11) is that it tells one that w^1 of the hypersurface at $r=0$ is nonzero. Hence by Theorem 4.9 of Ref. 9 it cannot bound a smooth compact four-manifold. Thus, if one were to take a large compact four-dimensional subset of the Kerr manifold, M (i.e., one that includes times well into the past and the future as well as large spatial dimension), that intersects the hypersurface at its end points in time, then there is no compact surface in M that can be bounded by the hypersurface at $r=0$ that is restricted to this range of times. This is equivalent to the statement that was

TABLE VI. The Stiefel–Whitney classes for various fiber bundles over the two-surface at $r=0$.

Tangent bundle	Normal bundle	$\mathbf{R}^4$ bundle
$w^0(T(U_2))=1$	$w^0(N(U_2))=1$	$w^0(T(M)_{\|U_2})=1$
$w^1(T(U_2))=a$	$w^1(N(U_2))=a$	$w^1(T(M)_{\|U_2})=0$
$w^2(T(U_2))=0$	$w^2(N(U_2))=0$	$w^2(T(M)_{\|U_2})=a\smile a$
All other w^i's that are not tabulated are identically zero by definition.[a]		$w^3(T(M)_{\|U_2})=0$ $w^4(T(M)_{\|U_2})=0$

[a] See Ref. 9.

made in the last section that the hypersurface turns space-time inside out onto itself.

Stiefel–Whitney classes can be interpreted as obstructions to forming independent nonzero vector cross sections. A bundle admits a continuous vector cross section only if the top obstruction, the Euler characteristic, vanishes. $T(U_2)$ admits at least one continuous section, as is well known. But, neither the normal or the tangent bundle can be trivialized (i.e., admit two continuous cross sections) because $w^1 \neq 0$ (see Milnor and Stasheff,[9] p. 39). Thus, $T(U_2)$ admits one cross section and $N(U_2)$ admits at most one cross section. Similarly, if $a \cup a \neq 0$ then $T(M)_{|U_2}$ cannot admit three cross sections. If it could, one could break up $T(M)_{|U_2}$ into its trivializable part, ϵ, and its orthogonal complement $\epsilon^{\perp}$. Then the Whitney product theorem (3.11) implies that $w^2 = w^3 = w^4 = 0$, which is a contradiction. This same idea can be used to prove the proposition on p. 39 of Milnor and Stasheff[9] that was cited a few sentences back. This is a preview of the discussion of the $\mathbb{R}^4$ bundle over U_2 to be presented in the next section.

With the obstruction information as a guide, one can construct representatives of equivalence classes of normal bundles over the Klein bottle. As was noted in Sec. II, one needs to cover U_2 with four open sets in order to get all of the geometric information on the two-surface at $r = 0$. All of the equivalence classes of normal bundles should result from looking at the possible sets of transition functions that can result on the intersections of the four sets W_i, $i = 1,\ldots,4$, that were defined in Sec. II. It will turn out that the bundle structure for the normal bundle is mirrored in the tangent bundle as a result of the orientability of the Kerr manifold in which U_2 is immersed.

There is a reduction of the normal bundle to the associated principal bundle over U_2 that has a fiber that is the group $O(1,1)$.[16] This group has four connected components that are all disjoint from each other. The four components can be associated with the identity, time reversal, a parity (actually the reflection through an axis in the plane; the term parity will be used as it is a word that is familiar in physics), and a combination of a parity and a time reversal. In matrix form an arbitrary element of $O(1,1)$ can be written as a product of the following kinds of elements. An element of the group that is in the component that is connected to the identity, I, is of the form

$$g_I(\alpha) = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}, \quad 0 \leqslant \alpha < \infty. \tag{3.13a}$$

A time reversal, T, and a parity, P, are given by the matrices

$$T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.13b}$$

Any element of $O(1,1)$ can be obtained by multiplying matrices of the type given by (3.13). Referring back to the statement in the previous paragraph about determining all of the equivalence classes of normal bundles from the covering by four open sets, if one takes a refinement of the open cover $\{W_i\}$, then one gets a bundle that is equivalent in the group $O(1,1)$ to the one generated by the cover $\{W_i\}$ (Steenrod, Lemma 2.8). The reason that this is true is that it will be shown that the transition functions for the cover $\{W_i\}$ are just constants, either I, P, T, or PT. A refinement of any one of the open sets will introduce transition functions that all lie on the same connected component of $O(1,1)$. All of the transition functions among the sets that make up the refinement of one of the W_i will be of the form (3.13a). The problem of finding equivalence classes for the normal bundle is reduced to studying the equivalence classes that are induced by the cover $\{W_i\}$.

Noting that there exists a $\mathbb{Z}$ orientation of M along U_2 (for the sake of argument say it is positive) and that the transition functions of the orientation bundle over U_2 are given by (2.5a)–(2.5f) as well as the discrete isometry that relates the two sheets of space-time, one can write the following transition functions for a normal bundle: On

$$W_1 \cap W_2\text{: } g_{11} = g_{22} = I, \quad g_{12} = g_{21} = TP, \tag{3.14a}$$

$$W_3 \cap W_4\text{: } g_{33} = g_{44} = I, \quad g_{34} = g_{43} = TP, \tag{3.14b}$$

$$\begin{aligned} &(W_1 \cap W_3)_1\text{: } g_{11} = g_{33} = g_{13} = g_{31} = I, \\ &(W_1 \cap W_3)_2\text{: } g_{11} = g_{33} = I, \quad g_{13} = g_{31} = P, \end{aligned} \tag{3.14c}$$

$$\begin{aligned} &(W_1 \cap W_4)_1\text{: } g_{11} = g_{44} = I, \quad g_{14} = g_{41} = TP, \\ &(W_1 \cap W_4)_2\text{: } g_{11} = g_{44} = I, \quad g_{14} = g_{41} = T, \end{aligned} \tag{3.14d}$$

$$\begin{aligned} &(W_2 \cap W_4)_1\text{: } g_{22} = g_{44} = g_{24} = g_{42} = I, \\ &(W_2 \cap W_4)_2\text{: } g_{22} = g_{44} = I, \quad g_{24} = g_{42} = P, \end{aligned} \tag{3.14e}$$

$$\begin{aligned} &(W_2 \cap W_3)_1\text{: } g_{22} = g_{33} = I, \quad g_{23} = g_{32} = TP, \\ &(W_2 \cap W_3)_2\text{: } g_{22} = g_{33} = I, \quad g_{23} = g_{32} = T. \end{aligned} \tag{3.14f}$$

Since the transition functions lie on four disconnected components of $O(1,1)$, it is clear that there can be no continuous cross section of the associated principal bundle to the normal bundle. Hence, the normal bundle is not equivalent to the trivial bundle in the group $O(1,1)$ by Theorem 8.3 and Corollary 8.4 of Steenrod.[8] This is equivalent to saying that the normal bundle is not parallelizable, which is consistent with the discussion of characteristic classes.

Similarly, one can consider the tangent bundle by the reduction to the group $O(2)$. The group $O(2)$ has two connected components that are disjoint from each other. The group is generated by a parity and rotations in the plane. In a manner that is consistent with the treatment of the orientation bundle in (2.5), the transition functions are chosen to be the same as (3.14) with the substitutions, $TP \rightarrow$ rotation by π radians, and $T \rightarrow$ parity composed with a rotation by π radians. The transition functions for the tangent bundle are given by (3.15),

$$\text{(3.14) with } TP \rightarrow \text{rotation by } \pi, \quad T \rightarrow P \circ (\text{rotation by } \pi). \tag{3.15}$$

In direct analogy to the normal bundle, since the transition functions lie on two different components of $O(2)$, the Klein bottle is not parallelizable as is well known.

It can be seen by the constructions in (3.14) and (3.15) that by orientability of M, the transition functions of the tangent bundle determine those in the normal bundle and vice versa. Thus, it is clear that there is a one to one correspondence between equivalence classes of normal bundles and of tangent bundles.

Let B be a bundle with group G over a space X. Define G_e as the component of G connected to the identity and

$\pi_1(X)$ is the fundamental group of X. One defines the characteristic class, $\chi(B)$, of the bundle by an equivalence class of homomorphisms (Steenrod, 13.10)

$$\chi\colon \pi_1(X)\to G/G_e. \tag{3.16}$$

For O(2), $G/G_e\cong \mathbf{Z}_2$ and for O(1,1), G/G_e is a group with four elements that is not isomorphic to $\mathbf{Z}_4$. Its elements are given by the commuting matrices

$$a=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad b=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad c=\begin{pmatrix}-1&0\\0&1\end{pmatrix},\quad d=\begin{pmatrix}-1&0\\0&-1\end{pmatrix}, \tag{3.17a}$$

or equivalently,

$$a=I,\quad b=P,\quad c=T,\quad d=PT. \tag{3.17b}$$

There are numerous characteristic homomorphisms (12) from the fundamental group of U_2 to G/G_e of O(1,1) that are inequivalent. However, the only ones that are consistent with the global orientation of the Kerr manifold and the discrete isometry at $r=0$ can be determined by their one to one correspondence with the equivalence classes of tangent bundles that was mentioned earlier.

$\pi_1(U_2)$ has two noncommuting generators, $\tilde{a}$ and $\tilde{b}$, that obey the single relation $\tilde{a}\tilde{b}=\tilde{b}^{-1}\tilde{a}$.[8] Consequently, there are two nontrivial equivalence classes of tangent bundles. This follows by an application of 13.9 of Steenrod to the reduction of the tangent bundle to the principal bundle with fiber O(2). These classes correspond to the inequivalent homomorphisms

$$\chi\colon \tilde{a}\to\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \tilde{b}\to\begin{pmatrix}+1&0\\0&-1\end{pmatrix}, \tag{3.18a}$$

$$\chi\colon \tilde{b}\to\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \tilde{a}\to\begin{pmatrix}+1&0\\0&-1\end{pmatrix}, \tag{3.18b}$$

where the matrices in (3.18) are a standard representation of $\mathbf{Z}_2$. The inequivalence of the homomorphism is in the following sense: there is no inner automorphism $g\chi(\tilde{a})g^{-1}$, $g\in\mathbf{Z}_2$, that can map $\chi(\tilde{a})$ in (3.18a) to $\chi(\tilde{a})$ in (3.18b) (see Steenrod, 13.5).

If (3.18a) corresponds to the bundle that was constructed in (3.15), then it is easy to see how to construct the tangent bundle that corresponds to (3.18b) from the discussion of the orientation bundle in Sec. II. There was one arbitrary construction in the formation of the orientation bundle. This was the choice of W_1 and W_2 to have the same orientation. This was the same arbitrary choice that was made in Table II when region II was chosen instead of region IV. This choice of orientation fixed all of the other transition functions. The inequivalent tangent bundle is obtained by choosing W_1 and W_2 to have opposite orientations. Then the transition functions of the new bundle are the same as (3.15) with the substitutions: rotation by $\pi\leftrightarrow$parity composed with rotation by π, or this can be expressed by (3.19),

$$\begin{aligned}&\text{(3.14) with the substitutions } TP\to P\circ(\text{rotation by }\pi)\\ &\qquad\qquad T\to(\text{rotation by }\pi).\end{aligned} \tag{3.19}$$

Similarly, there are two inequivalent homomorphisms of $\pi_1(U_2)$ to G/G_e of O(1,1) that are exactly the same as those in (3.18). One can identify the matrices in (3.18) with elements in (3.17). These homomorphisms represent the unique equivalence classes of normal bundles that are consistent with the global orientation of the Kerr manifold and respects the existence of two sheets of space-time at $r=0$ that have opposite local time directions. The new equivalence class of normal bundle has transition functions that are given by

$$\text{(3.14) with the substitution } T\leftrightarrow TP. \tag{3.20}$$

IV. THE $\mathbf{R}^4$ BUNDLE OVER U^2

One can use the results of the previous section to give a quick study of the $\mathbf{R}^4$ bundle over U_2. The four-frame bundle is particularly interesting since the four legs of a pseudo-orthogonal tetrad can be associated with the measuring rods and clocks of physical observers at $r=0$. Fortunately, almost all of the obstruction information was derived in Sec. III and is contained in Table VI.

First of all, the $\mathbf{R}^4$ bundle is orientable since there exists a local orientation of M along U_2. This implies that the $\mathbf{R}^4$ bundle can be reduced to a principal bundle with fiber SO(3,1). Here SO(3,1) is isomorphic to $\mathbf{R}P^3\times\mathbf{R}^3$ which is homotopically equivalent to $\mathbf{R}P^3$. Therefore, $\pi_1(\mathrm{SO}(3,1))\cong\pi_1(RP^3)\cong\mathbf{Z}_2$. Thus, by (3.18) there exist two nontrivial inequivalent classes of $\mathbf{R}^4$ bundles over U_2, corresponding to the constructions that were made for the tangent and normal bundles at the end of the last section.

The Stiefel–Whitney classes for the $\mathbf{R}^4$ bundle were tabulated in Table VI of the last section. There was one ambiguity, was $w^2(T(M)_{|U_2})$ zero or not? In this section it will be shown that $c^2(T(M)_{|U_2})\neq 0$. If $c^2(T(M)_{|U_2})\neq 0$ then there can exist at most two linear independent four-vector cross sections over U_2. As an example of why this needs to be investigated is the Möbius band that is embedded in E^3. The tangent bundle admits one vector section and the normal bundle does not admit any, but the R^3 bundle is trivializable. In order to see how this happens, define a Möbius band with a constant twist. Describe the geometry in spherical coordinates. The center circle of the Möbius band is a circle about the z axis in the equatorial plane given by

$$S(\phi)=\cos(\phi)\hat{x}+\sin(\phi)\hat{y}. \tag{4.1}$$

The tangent vector to the Möbius band along the center circle that is orthogonal to the center circle is given by

$$t(\phi)=\cos(\tfrac{1}{2}\phi)\hat{z}+\sin(\tfrac{1}{2}\phi)(\cos(\phi)\hat{x}+\sin(\phi)\hat{y}) \tag{4.2}$$

and the normal vector field along the center circle is given by

$$n(\phi)=-\sin(\tfrac{1}{2}\phi)\hat{z}+\cos(\tfrac{1}{2}\phi)(\cos(\phi)\hat{x}+\sin(\phi)\hat{y}). \tag{4.3}$$

$n(\phi)$ and $t(\phi)$ do not form a global system of coordinates over the Möbius band as they are double valued. Define global orthonormal coordinates in $\mathbf{R}^3$ along the Möbius band as in (4.4),

$$\hat{e}_1=\cos(\tfrac{1}{2}\phi)t(\phi)+\sin(\tfrac{1}{2}\phi)n(\phi), \tag{4.4a}$$

$$\hat{e}_2=-\sin(\tfrac{1}{2}\phi)t(\phi)+\cos(\tfrac{1}{2}\phi)n(\phi), \tag{4.4b}$$

$$\hat{e}_3=\cos(\phi)\hat{x}+\sin(\phi)\hat{y}. \tag{4.4c}$$

Then the three vector fields,

$$\hat{e}_1 = \cos(\phi)\hat{z} + \sin(\phi)(\cos(\phi)\hat{x} + \sin(\phi)\hat{y}), \quad (4.5a)$$

$$\hat{e}_2 = -\sin(\phi)\hat{z} + \cos(\phi)(\cos(\phi)\hat{x} + \sin(\phi)\hat{y}), \quad (4.5b)$$

$$\hat{e}_3 = \cos(\phi)\hat{x} + \sin(\phi)\hat{y}, \quad (4.5c)$$

are three independent cross sections since they define single-valued coordinates. A rotation in the plane that is spanned by $t(\phi)$ and $n(\phi)$ was used to trivialize the bundle. This is the same as saying that the bundle is equivalent to the trivial bundle in SO(2). Here $SO(2) \subset SO(3)$ implies that the $\mathbb{R}^3$ bundle is trivializable.

One might think that it could be possible to do a similar thing with the $\mathbb{R}^4$ bundle on U_2 by rotating in the (ϕ,t) plane and the (r,θ) plane independently. This is the same as trying to find a section for the associated principal bundle with fiber SO(3,1). However, a continuous rotation in the (ϕ,t) plane rotates spacelike regions through the light cone into timelike regions. This is inconsistent with the pseudo-Riemannian nature of the manifold. Thus, one cannot do the same thing that was done for the Möbius band in E^3. This was hinted at by the fact that there are only two equivalence classes for the $\mathbb{R}^4$ bundle. This means that all of the information in the $\mathbb{R}^4$ bundle can be obtained by its decomposition into its normal and tangent components unlike the $\mathbb{R}^3$ bundle over the Möbius band. Thus there is an obstruction to continuing a cross section of the $\mathbb{R}^4$ bundle to the two-skeleton of M. One concludes that $c^2(T(M)_{|U_2}) \neq 0$.

The other obstruction cocycles can be evaluated more directly. The existence of the vector cross section in the tangent bundle implies that there is at least one vector cross section of the $\mathbb{R}^4$ bundle. Thus, $c^4(T(M)_{|U_2}) = 0$.[8] The orientability of the bundle as stated earlier implies that $c^1(T(M)_{|U_2}) = 0$.

In order to proceed further, construct a four-dimensional manifold by taking a tubular of U_2 (as Whitney does in Ref. 10). This is a four-manifold, M', that must be orientable since it is a four-dimensional submanifold of M. Hence, $c^3(T(M')_{|U_2}) \equiv c^3(T(M')_{|U_2}) = c^3(T(M)_{|U_2})$ must vanish by the orientability of M' as discussed in Whitney,[10] p. 137. It should be noted that the breakdown of the four-manifold structure at $\rho = 0$ is a technicality that is not always mentioned. It is irrelevant to these arguments since the spaces are CW complexes for which the same conclusions will hold.

The results of the investigation of the obstruction cocyles are tabulated for the $\mathbb{R}^4$ bundle over U_2 in (4.6),

$$c^1(T(M)_{|U_2}) = w^1(T(M)_{|U_2}) = 0, \quad (4.6a)$$

$$c^2(T(M)_{|U_2}) = w^2(T(M)_{|U_2}) = a \smile a, \quad (4.6b)$$

by (3.8) and $a \smile a$ cannot be zero by the discussion in this section

$$c^3(T(M)_{|U_2}) = w^3(T(M)_{|U_2}) = 0, \quad (4.6c)$$

$$c^4(T(M)_{|U_2}) = w^4(T(M)_{|U_2}) = 0. \quad (4.6d)$$

V. SPINORIALIZED NULL FLAGS

Penrose has attempted to define the Weyl spinors of the Lorentz group SO(3,1) [corresponding to fundamental representations of SL(2,C)] by a purely geometric realization. His construction, a null flag, involves a null vector and a spacelike half-plane. The reader is referred to Chap. 1 of Ref. 3 as the details of the invariance of this construction are too lengthy to be duplicated here. Although this construction is successful in many ways, spinors of opposite sign have been identified with each other. He is forced to introduce "spinorial" objects in an *ad hoc* and not well-defined fashion in order to resolve this ambiguity. In essence, the fundamental geometrical difficulty in interpreting the 2 to 1 mapping from $SL(2,\mathbb{C}) \to SO(3,1)$ is glossed over. The aim of this section is to show how a spinorialized version of null flags arise naturally in the Kerr manifold.

The reason why Penrose's construction fails can be seen by the following interpretation in terms of fiber bundles over the two-sphere. A null vector, n^μ, is given by $n^\mu = \psi\sigma^\mu\bar{\psi}$, where σ^μ is a 2×2 Pauli spin matrix and ψ is a Weyl two-spinor. This is the mapping that associates a spinor with a null direction. A particular null vector, n^μ, determines ψ up to a phase, $e^{i\alpha}$. The set of spatial orientations of spinors is isomorphic to the rotation group, SU(2), by the transitive action of SU(2) on the space of two-spinors. SU(2) is topologically S^3. Here SL(2,C) is topologically $S^3 \times R^3$, where R^3 represents the three possible boost directions. The boosts will be ignored for now so that one can study the effects of spatial rotations. The set of unboosted future directed null vectors span the null sphere at infinity S^2_+ [by unboosted one means in Penrose's construction that the extent of the null vectors (see Ref. 3 for a definition) is constant or equivalently one has taken a constant time section of S^2_+ with $t = 1$ in his coordinates]. The phase ambiguity in the mapping from null vectors to spinors, $e^{i\alpha}$, corresponds to a circle, S^1. If one fibers the sphere S^2 with circles, S^1, by the Hopf fibration then one has the set of spatial orientations of spinors, namely, S^3 (Ref. 8). Penrose almost did this. The set of spatial orientations of null flags of fixed extent is an S^1 bundle over S^2. The null vectors or flagpoles span S^2_+, the base space. Penrose fibered this with his half-plane flags which are spanned by the flagpole and a vector in the tangent space at a point $p \in S^2_+$, where p is the point at which the flagpole pierces the null sphere. The set of null flags of fixed extent is therefore isomorphic to the tangent bundle over S^2_+, $T(S^2_+)$. If the extent of the flagpoles were not held fixed then the space of possible configurations of null flags is isomorphic to $T(S^2_+) \times R^3$. As is well known the tangent bundle to the two-sphere is the real projective three-space.[8] The set of spatial orientations of null flags of fixed extent is therefore $\mathbb{R}P^3 \cong T(S^2)$. $\mathbb{R}P^3$ can be realized as a circle bundle over S^2 (Ref. 8). This is clear since the tangent bundle can be reduced to a principal SO(n) bundle. SO(2) is topologically S^1. Thus, the set of spatial orientations of null flags of fixed extent is the S^1 bundle over S^2 that is topologically $\mathbb{R}P^3$. The circle bundle that corresponds to the spatial orientations of the spin vectors is the Hopf fibration of S^3. Here S^3 maps 2 to 1 to $\mathbb{R}P^3$; hence the sign ambiguity in Penrose's construction.

If one constructs a line bundle over $\mathbb{R}P^3$, one can clearly obtain S^3 ($\mathbb{R}P^3 \cong S^3/\mathbb{Z}_2$). Such a line bundle arises naturally in the Kerr geometry. One has a mapping from the two-surface at $r = 0$ to the null sphere at infinity by means of the principal null congruences. The line bundle can be thought of as arising from the orientation bundle along the set of

"base" points of the principal null congruences as they leave the two-surface at $r = 0$ (i.e., the orientation bundle that was constructed in Sec. II). The flag is essentially (for flagpoles of fixed extent) a point on a circle in the tangent space at some point of S^2_+. This circle is replaced in this construction by a Möbius band (i.e., a nontrivial line bundle over S^1). Similarly, one can use the boundary of a Möbius band to give a double covering of S^1 in order to describe the possible orientations of the flag half-plane about the flagpole. More relevantly, this could also correspond to the circle of 4π radians that bounds the equatorial region of the two-surface at $r = 0$ that was described in Fig 17. As discussed in Sec. II, this circle passes through the positive orientation sheet of the two-surface for half of the circle and the negative orientation sheet on the other half. Thus, for each flagpole, the equatorial circle covers the possible orientations of the null flags in a two to one manner.

The two to one covering of $T(S^2_+)$ can be traced to the decomposition of U_2 into its two constituent $\mathbb{R}P^2$'s. By transporting a normal vector around closed paths in $\mathbb{R}P^2$ one can obtain a normal vector of opposite sign after completing the circuit (nonorientability of $\mathbb{R}P^2$). Thus, the inward and outward pointing normals of $\mathbb{R}P^2$ can be related by continuous translations around the surface unlike the case for S^2 (since S^2 is orientable, translating an outward pointing normal around a close curve yields an outward pointing normal). Thus, in the immersion of $\mathbb{R}P^2$ in a pseudo-Riemannian four-space the set of spacelike normals has one connected component that can be mapped one to one to the points of the two-sphere (not just a hemisphere!). Similarly, the spacelike component of the normal space of two $\mathbb{R}P^2$'s can double cover S^2. This mapping is realized by the component of the principal null congruences in the normal space to U_2 (Ref. 1). The double covering of the null sphere at asymptotic infinity is accomplished by means of covering S^2 with both ingoing and outgoing principal null congruences.

There is a global consistency requirement of this construction. If a spinorialized null flag is rotated by $\theta = 2\pi$ radians, it is the same as rotating by $\phi = 2\pi$ radians as in (5.3)–(5.5). One must be on the same sheet of the line bundle over $\mathbb{R}P^3$. This is guaranteed by the global geometry of the two-surface at $r = 0$ that is given in Figs. 19 and 21. The remainder of this section is an attempt to make the global consistency requirement and the actual mapping to spinorialized null flags more explicit. To do this one needs to invoke many of Penrose's constructions for null flags.

This construction can be considered as a double covering of the Riemannian sphere, where the tangent spaces are double covered as well. The interesting thing is that the Klein bottle has a hole in it and the sphere does not. If one pictures the Klein bottle as a sphere with a handle as in Fig. 14, then the hole is the region between the handle and the sphere. The origin of the hole in U_2 is the two holes (regions homeomorphic to a disk) that were cut in the two $\mathbb{R}P^2$'s in making the connected sum. These holes were like two infinitesimally thin slits at the equator at $r = 0$ before the connected sum is made. However, after the connected sum is formed, the resulting topology implies that the hole must be of the same size as the handle (see Fig. 14). One concludes

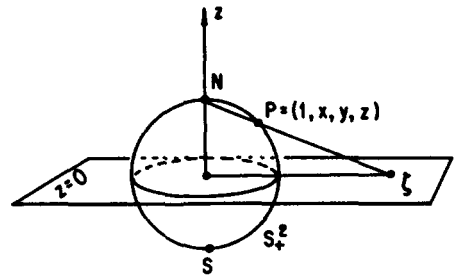


FIG. 25. The stereographic projection from the north pole of S^2_+ into the Argand plane at $z = 0$ with complex coordinate ζ.

that the hole is of infinitesimal extent in every direction. The hole itself does not turn up on the doubly covered sphere at infinity but the mouths of the handle do (i.e., where the handle joins the sphere in Fig. 14). The mouths of the handle are illustrated in Fig. 22. These mouths turn up as infinitesimal holes in the doubly covered sphere (see Figs. 26 and 27). On the doubly covered sphere at infinity, they simply appear as junctions between sheets of the double covering.

The Riemannian sphere results from the stereographic projection of the null sphere that is spanned by the flagpoles into a plane. The coordinates on the plane are given by (5.6) in terms of complex numbers making this an Argand plane (See Fig. 25). The north pole cannot be projected into the plane and must be attached to the entire boundary of the plane of infinity resulting in the Riemannian sphere.

Figure 14 leads straightforwardly to a mapping from the two-surface at $r = 0$ to a doubly covered sphere at null infinity as pictured in Fig. 26 by just integrating out along the principal null geodesics. The dashed lines represent a negative spatial orientation at the base points of the principal null geodesics that span this covering of the null sphere at infinity. The solid lines represent the covering of the null sphere at infinity that has a positive orientation at the base points. Note that in this construction the future directed sheet of the two-surface at $r = 0$ maps to the northern hemisphere (2 to 1) and the past directed sheet to the southern hemisphere.

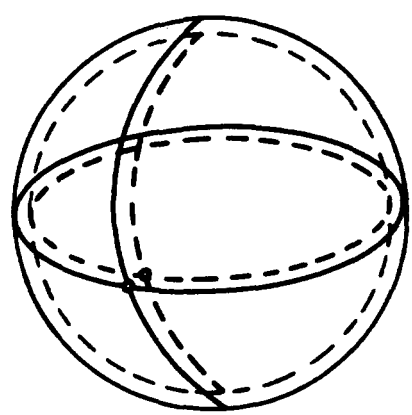


FIG. 26. The doubly covered sphere. The dashed inner sphere is the negative orientation sheet and the solid line outer sphere is the positive orientation sheet. The junctions between the two sheets are given by Fig. 19. The small circles represent the image of the mouths of the handle of the Klein bottle at $r = 0$.

There is an equivalent (on the level of fiber bundles) construction where the mappings are reversed.

The infinitesimally small mouths of the handle of the Klein bottle at $r = 0$ integrates so as to represent the intersections of the meridianal junction and the equatorial junction on the doubly covered sphere in Fig. 26. The equatorial junction is between hemispherical regions that are distinguishable in that the null generators that span these regions have origins on different sheets of space-time at $r = 0$. It was established in Sec. III that there exists a trivial normal line bundle over the Klein bottle. This corresponds to choosing a future time direction over the whole surface as viewed in the stationary frame at asymptotic infinity. Thus, all of the null geodesics that strike the doubly covered null sphere can be chosen to be future directed. This implies that at the points of origin on the past directed sheet these null geodesics must appear past directed to local observers at $r = 0$. One can represent the two sheets of the doubly covered sphere as corresponding to different orientations of S^2_+. The positive orientation is represented by an outward pointing normal vector field corresponding to k_{out} (using the notation for the principal null geodesics that was introduced in Sec. II). The negatively oriented sheet of the doubly covered sphere has k_{in} as the normal vector field. The meridian that is signified in Fig. 26 (a junction between the two sheets of the doubly covered sphere) represents the fact that one sheet rotates into the other after an azimuthal rotation of 2π radians.

The statements in the previous paragraph are analogous to the bundle structures on U_2. If one of the hemispherical shells in the model of the Klein bottle in Fig. 14 is rotated by 2π radians it will change orientation sheets as is stated in Fig. 19. In complete analogy, one can cover the doubly covered sphere by four open sets $W_i'(i = 1,...,4)$, that can be considered as the same open sets that were used to cover U_2 that have been translated along the principal null goedesics to the doubly covered sphere at infinity. One can go as far as to define transition functions for the orientation bundle as having the same values as those that were found for U_2 in (2.5). However, one has the added simplification that one no longer needs to consider the disjoint nature of some of the intersections of the W_i's. The small sets that are labeled with a subscript 1 as in $(W_i \cap W_j)_1$ are not manifested on the doubly covered sphere at infinity. The changing of the time and spatial orientation sheets at the equatorial junctions is just the analog of what was found in Fig. 19 when a tangent two-frame was transported along a curve through the equatorial region of the two-surface at $r = 0$. In essence, the geometry of the doubly covered sphere is contained in Fig. 19.

Analogous to the double covering of the Riemannian sphere, there is a double covering of the Argand plane as indicated in Fig. 27. The shaded regions correspond to null geodesics that originate on the future directed sheet at $r = 0$ and the unshaded regions to those that originated on the past directed sheet. The meridian that is drawn in Fig. 27 and the equatorial junction have the same interpretations as those in Figs. 19 and 26.

The tangent space at a point on the doubly covered sphere is drawn in Fig. 28. It is a double covering of the tangent plane. A rotation designated by the polar angle α

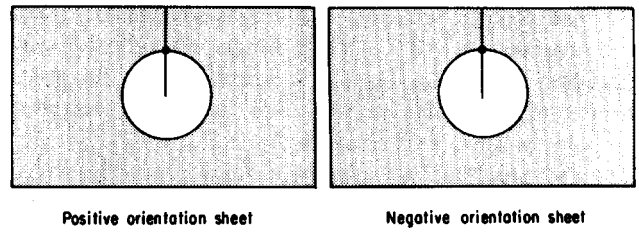


FIG. 27. The doubly covered Argand plane that is obtained by a stereographic projection of the doubly covered sphere in Fig. 26. The shaded region is associated with principal null geodesics that originate on the future directed sheet of space-time at $r = 0$. Similarly, the unshaded region is associated with the past directed sheet of space-time at $r = 0$. The small circles represent the image of the mouths of the handle of the two-surface at $r = 0$.

about the base point in the tangent space of 2π radians rotates one sheet into the other as indicated by the radial line junction between the sheets that are labeled $\alpha = 0$ in Fig. 28. The two sheets of the tangent space correspond to the two tangent planes at each point of S^2_+ that are induced by the double covering. This rotational structure follows from the fact that rotation in the tangent space corresponds to a rotation of the two-surface at $r = 0$ about its symmetry axis, where the base point of the tangent space is defined to lie on this symmetry axis. The double sheeted nature of the tangent space then follows from Fig. 19. Thus one has a double covering of the set of flags for each flagpole. The set of vectors in this doubly covered plane when combined with the flagpole that passes through the base point of the tangent space defines the set of spinorial flags that can be associated with this particular flagpole.

Base states for the spinorialized null flags can be defined in a straightforward manner. A spin $+\frac{1}{2}$ state, $\binom{1}{0}$, can be given by the north pole of the k_{out} sheet of the doubly covered sphere at infinity with the spinorial flag aligned with the demarcated meridian in Fig. 26 (this corresponds to $\alpha = 0$ on the positive orientation sheet of the tangent space). The state $-\binom{1}{0}$ would correspond to the north pole of the k_{in} sheet of the doubly covered sphere with spinorial flag pointing along the demarcated meridian on the negative orientation sheet of Fig. 26 (this corresponds to $\alpha' = 0$ on the negative orientation sheet of the tangent space, where $\alpha' = \alpha + 2\pi$). It is implicit in this construction that if spinors represent spin angular momentum that the angular momentum vector points along the flagpole on the doubly covered sphere [i.e., the states $\binom{1}{0}$ and $-\binom{1}{0}$ have the same angular momentum vector]. The other base state, spin $-\frac{1}{2}$ or $\binom{0}{1}$, is given by the south pole of the positive orientation

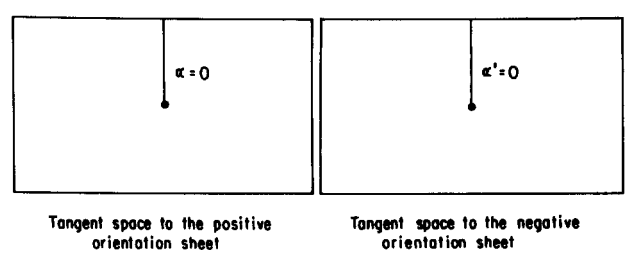


FIG. 28. The tangent space of the doubly covered sphere is a doubly covered tangent plane.

sheet of the doubly covered sphere with the spinorial flag pointing in the opposite direction of the demarcated meridian at the south pole of Fig. 26 (this corresponds to $\alpha = 0$ on the positive orientation sheet of the tangent space at the south pole). Its angular momentum vector is antiparallel to that of the spin $+\frac{1}{2}$ state. One defines a $-\binom{0}{1}$ state analogously to the definition of $-\binom{1}{0}$.

Penrose has shown that if one takes the complex number, ζ, which is a coordinate on the Argand plane and signify it by a ratio of complex numbers, $\zeta = \xi/\eta$, then ξ and η can be chosen to transform like the components of the Weyl spinor $\binom{\xi}{\eta}$ and are defined by a null flag up to a sign.[3] The reader is referred to Ref. 3 for the details of Penrose's treatment as it is too lengthy to be reproduced here.

Briefly, the analog of what Penrose did with the Riemannian sphere can be sketched here for the doubly covered Riemannian sphere. One can replace the complex coordinate, ζ, on the Argand plane by a density $\hat{\zeta} \equiv \zeta e_\alpha \equiv \xi e_\alpha / \eta e_\alpha$, where e_α represents the line fiber of the trivialization of the orientation bundle on each of the four open sets W_i' defined in the discussion following Fig. 26 and in (2.5). Here $\hat{\zeta}$ can be construed as a coordinate on the doubly covered Argand plane. It should be noted that this is the minimal number of open sets that can be used to cover the doubly covered sphere (as is well known one needs two for each sphere). One interprets $\hat{\zeta}$ on the positive orientation sheet to be

$$\hat{\zeta} = (\xi e^{i(1/2)\alpha})/(\eta e^{i(1/2)\alpha}) \sim \binom{\xi}{\eta} e^{i(1/2)\alpha}, \quad \text{where } 0 \leqslant \alpha \leqslant 2\pi, \tag{5.1a}$$

where α gives the rotation angle in the spinorialized flag plane on the positive sheet of the doubly covered tangent space. Note that if $\alpha > 2\pi$ the flagpole must be on the other orientation sheet of the doubly covered sphere. The base point of the tangent space can be determined from ξ and η by (5.6) and (5.7). The density on the negative orientation sheet is given by

$$\hat{\zeta} = (\xi e^{i(1/2)\alpha})/(\eta e^{i(1/2)\alpha}) \sim \binom{\xi}{\eta} e^{i(1/2)\alpha}, \quad \text{where } 2\pi \leqslant \alpha \leqslant 4\pi, \tag{5.1b}$$

or if α' is the angle that is defined on the negative sheet of the tangent space then (5.1b) can be written as

$$\hat{\zeta} \equiv (\xi e^{i(1/2)\alpha'})/(\eta e^{i(1/2)\alpha'}) \sim \binom{\xi}{\eta} e^{i(1/2)\alpha'}, \quad \text{where } 0 \leqslant \alpha \leqslant 2\pi. \tag{5.1b'}$$

Consequently, to define the density one needs a global definition of $\alpha = 0$ and $\alpha' = 0$. By the transitive action of SU(2) [the rotation subgroup of SL(2,C)] on the set of Pauli spinors, every spatial orientation of a spinorialized null flag of fixed extent can be obtained by a rotation applied to the base state corresponding to the Pauli spin $\binom{1}{0}$. These rotations can be described by the three Euler angles ϕ, θ, and γ.[20] For the spinorialized null flags the rotation group is reparametrized by ϕ, θ, and α, the rotation angle in the tangent space. Choose $\alpha = 0$ for the tangent space at the north pole to be along the tangent vector that is mapped by the exponential map to the demarcated meridian on the positive orientation sheet of the doubly covered sphere in Figs. 26 and 27. Similarly, the radial line in Fig. 28 that represents $\alpha' = 0$ exponentiates onto the corresponding demarcated meridian on the negative sheet. By transitivity, one can define $\alpha = 0$ for other flagpoles and the associated spinorialized flag planes by parallel transporting this tangent vector at the north pole to the tangent space at other points of the same sheet of the doubly covered sphere. Since $\partial/\partial\theta$ and $\partial/\partial\phi$ are an orthogonal coordinate frame (not orthonormal) on the sphere (i.e., $[\partial/\partial\theta, \partial/\partial\phi] = 0$), the flagpole that is obtained by translating first in the θ direction then in the ϕ direction is the same as for the reverse order. Similarly, the vector in the tangent space that labels $\alpha = 0$ is invariant under the order of parallel translation on the positive orientation sheet minus the south pole. The validity of this claim can be seen by the fact that the sphere minus the south pole can be thought of as the one to one image of the conformal mapping of the sphere into the Argand plane by projection from the south pole. The plane is flat, so the translation of the vector signifying $\alpha = 0$ is independent of the path. The ambiguity in defining $\alpha = 0$ at the south pole by the translation of a vector from the north pole arises because the Euler number of the sphere is 2. A discussion of the implications of $\chi = 2$ is given in the analysis of (5.5). There is obviously a relationship between this phenomena and the need to independently define a new base state, $\binom{0}{1}$, at the south pole. The definition of this state that was given earlier in this section depends on differentiating between the various possible paths that are available for parallel translation. The unique path that was chosen corresponds to a rotation about the y axis [coordinates of Fig. 25 and (5.6) and (5.7)] of the $\alpha = 0$ tangent vector of the state $\binom{1}{0}$ by π radians to get the $\alpha = 0$ tangent vector for $\binom{1}{0}$ at the south pole [see (5.4)].

If one considers the tangent bundle of the doubly covered sphere as was done for the orientation bundle then one can just take the tangent bundle structure that was defined in Sec. III to describe this tangent bundle at asymptotic infinity. As with the orientation bundle on the doubly covered sphere, one can ignore one of the components of the disjoint intersections, namely, those of the form $(W_i \cap W_j)_1$. The remaining transition functions are the same as those in Sec. III. The group of the tangent bundle with this covering by four sets can be considered to be a four element discrete group that is spanned by the transition functions in (3.15). The four group elements are I, P, rotation by π and P° (rotation by π). Steenrod shows in Theorem 13.2 of Ref. 8 that translation of a fiber along a curve is a unique operation for a bundle with a discrete group. Thus, one can relate the $\alpha' = 0$ tangent vectors for flagpoles on the negative orientation sheet of the doubly covered sphere that are obtained by parallel translating from the north pole of this sheet to the $\alpha = 0$ spinorial flags that are similarly defined on the positive orientation sheet in a unique way. If one makes the unique matching of the parallel translates from both south poles and both north poles then one has the global definition of α and α' that is desired.

Penrose shows that from the topology of SO(3,1) $\cong \mathbb{R}P^3 \times \mathbb{R}^3$, where $\mathbb{R}^3$ represents the boosts and $\mathbb{R}P^3$ the rotations, that one need only to establish the transformation laws for spatial rotations and z boosts (toward the north

pole) in order to describe all of the Lorentz transformational properties as any Lorentz transformation can be decomposed into a product of these. The rotation law for the spinorialized null flags can be obtained in the same way that Penrose derived them for null flags.

For a rotation about the z axis of ψ radians, one finds from the stereographic projection from S^2_+ that the coordinate in the doubly covered Argand plane transforms as[3]

$$\hat{\zeta}' = e^{i\psi}\hat{\zeta}, \tag{5.2}$$

from which one concludes that

$$\begin{pmatrix}\xi'\\ \eta'\end{pmatrix} = \begin{pmatrix}e^{i(1/2)\psi} & 0\\ 0 & e^{-i(1/2)\psi}\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix} e^{i(1/2)\tilde{\alpha}(\psi)}. \tag{5.3a}$$

But $\tilde{\alpha}(\psi)$ is actually zero since the matrix by itself is consistent with the changing of sheets that is defined by Fig. 19. For example, if one has a pure $\binom{1}{0}$ or $\binom{0}{1}$ state then a rotation about the z axis is just a rotation of the spinorial flagpole. The change in α, the polar angle in the tangent plane, is just given by the matrix in (5.3a) with the substitution $\psi \to \alpha$. Clearly, this description of α as a phase (actually the orbit of the one-parameter group that is generated by α, a doubly covered circle in the tangent space, is merely being mapped to a singly covered circle by a half-angle mapping in this identification of α with a phase) describes the changing of orientation sheets that is associated with the spinorialized flags in a manner that is consistent with the definition of the density, $\hat{\zeta}$, in (5.1). Since one can decompose any Weyl spinor into a linear combination of base states, one can conclude that the transformation on the doubly covered Argand plane induced by rotation about the z axis induces a transformation in $\mathbb{C}^2$, the two-dimensional complex space that is spanned by ξ and η, that can be described in matrix notation as

$$\begin{pmatrix}\xi'\\ \eta'\end{pmatrix} = \begin{pmatrix}e^{i(1/2)\psi} & 0\\ 0 & e^{-i(1/2)\psi}\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \tag{5.3b}$$

As stated above, (5.3b) is consistent with the definition of the coordinate $\hat{\zeta}$ given in (5.1). Namely, a rotation of 2π about the z axis causes $\hat{\zeta}$ to change orientation sheets. Similarly, no matter which direction the flagpole that is associated with a spinorialized null flag points in space, a rotation about the direction along the flagpole just rotates the spinorialized half-plane. After 2π radians, the spinorialized null flag lies on a different sheet of the doubly covered sphere. This represents the coordinate independence of this construction. Note that (5.3) is inconsistent with the rotational properties of Penrose's flags (half-planes) for the states $\binom{1}{0}$ and $\binom{0}{1}$.

For a rotation about the y axis of the Riemannian sphere by θ radians, the stereographic projection implies that[3]

$$\begin{pmatrix}\xi'\\ \eta'\end{pmatrix} = \begin{pmatrix}\cos(\frac{1}{2}\theta) & -\sin(\frac{1}{2}\theta)\\ \sin(\frac{1}{2}\theta) & \cos(\frac{1}{2}\theta)\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \tag{5.4}$$

This is consistent with the structure of the junction conditions between the two sheets of the doubly covered sphere as given by (2.5) with the minor modifications that are indicated earlier in this section. By the positions of the junctions and Fig. 19 it is implicit that all of the spinorialized null flags of the form $R_z(\phi)R_y(\theta)\binom{1}{0} = \psi$ for the parameter ranges $(0 \leqslant \phi \leqslant 2\pi,\ -\frac{1}{2}\pi \leqslant \theta \leqslant 3\pi/2)$ or $(2\pi \leqslant \phi \leqslant 4\pi,\ 3\pi/2 \leqslant \theta \leqslant 7\pi/2)$ all have images on the same sheet of the doubly covered sphere. The parameter ranges $(0 \leqslant \phi \leqslant 2\pi,\ 3\pi/2 \leqslant \theta \leqslant 7\pi/2)$ and $(2\pi \leqslant \phi \leqslant 4\pi,\ -\frac{1}{2}\pi \leqslant \theta \leqslant 3\pi/2)$ have images on the other orientation sheet.

When there is a rotation about the x axis by $\bar{\chi}$ radians Penrose finds that

$$\begin{pmatrix}\xi'\\ \eta'\end{pmatrix} = \begin{pmatrix}\cos(\frac{1}{2}\bar{\chi}) & i\sin(\frac{1}{2}\bar{\chi})\\ i\sin(\frac{1}{2}\bar{\chi}) & \cos(\frac{1}{2}\bar{\chi})\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \tag{5.5}$$

Spinorialized null flags that are defined by $\psi = R_z(\phi) \times R_x(\bar{\chi})\binom{1}{0}$ can lie on different sheets of the doubly covered sphere. The parameter ranges for each sheet are not as easy to write as they were for the y rotation. The reason being that a rotation about the x axis of the spinorialized null flag that corresponds to $\binom{1}{0}$ does not leave the angle of the spinorialized half-plane, α, equal to zero. This results from the topology of S^2. The ambiguity in defining a global $\alpha = 0$ at the south pole of the null sphere by parallel transport that was discussed earlier in this section is of the same origin as this problem.

If one rotates the tangent plane with $\alpha = 0$ at the north pole by π radians about the y axis, then $\alpha = 0$ in the tangent plane at the south pole in accordance with the global definition of α that was made earlier in this section in conjunction with the definition of the spin-$\frac{1}{2}$ base state. If one now rotates the same tangent space at the north pole about the x axis by π radians, one finds that the tangent space at the south pole has been rotated by π radians with respect to the plane that was obtained by rotation about the y axis. Hence, $\alpha = \pi$. This is due to the singular point structure of the tangent bundle of the sphere (the sphere has Euler number, $\chi = 2$).[7] The fact that the singular point at the south pole has degree equal to 2 is manifested in the fact that the x axis is obtained from the y axis by an azimuthal rotation of $\frac{1}{2}\pi$ radians, yet $\alpha = 2(\frac{1}{2}\pi) = \pi$. As a further consequence of the fact that the singular point is of degree 2, one can see that as one rotates about different axes in the equatorial plane at asymptotic infinity by π radians then the tangent plane with $\alpha = 0$ at the north pole rotates into a distinct spatial configuration of the tangent plane (different α) at the south pole for each one of these axes. As a matter of fact, as the axis of rotation in the equatorial plane is changed continuously the resultant tangent plane at the south pole continuously changes spatial configuration. The angle α varies continuously over its entire range from 0 to 4π as the rotation axis is varied from $0 \leqslant \phi \leqslant 2\pi$ (degree 2 nature of the singular point). The conclusion of this discussion of the degree of the singular point is that the parameter range of $\bar{\chi}$ that corresponds to different orientation sheets for spinorialized null flags of the form $\psi = R_z(\phi)R_x(\theta)\binom{1}{0}$, though implied by (5.3), (5.4), and Fig. 19, will be more complicated than it was for θ.

The discussion in the previous paragraph accounts for the different possible identifications of spinorialized null flags that are obtained by rotating $\binom{1}{0}$ about the x axis by $\bar{\chi}$ radians and the z axis by ϕ_x radians and the same state that can be described by an equivalent representation as a rotation by θ radians about the y axis and ϕ_y radians about the z axis. These equivalences are obtained by equating the ac-

TABLE VII. Equivalent representations of spinorialized null flags that are obtained by the action of elements of SU(2), with the indicated set of group parameters, on the spin $+\frac{1}{2}$ base state.

(1) $\theta=0$	$\bar{\chi}=0$	(2) $\theta=\pi$	$\bar{\chi}=\pi$
$\sim$		$\sim$	
ϕ_y	$\phi_x=\phi_y$	ϕ_y	$\phi_x=\phi_y+\pi$
(3) $\theta=\pi$	$\bar{\chi}=3\pi$	(4) $\theta=3\pi$	$\bar{\chi}=\pi$
$\sim$		$\sim$	
ϕ_y	$\phi_x=\phi_y-\pi$	ϕ_y	$\phi_x=\phi_y-\pi$
(5) $\theta=3\pi$	$\bar{\chi}=3\pi$	(6) $\theta=2\pi$	$\bar{\chi}=2\pi$
$\sim$		$\sim$	
ϕ_y	$\phi_x=\phi_y$	ϕ_y	$\phi_x=\phi_y$
(7) $\theta=0$	$\bar{\chi}=2\pi$	(8) $\theta=2\pi$	$\bar{\chi}=0$
$\sim$		$\sim$	
ϕ_y	$\phi_x=\phi_y+2\pi$	ϕ_y	$\phi_x=\phi_y+2\pi$

tions of (5.4) and (5.5) applied to the state $\binom{1}{0}$. The sets of parameters that yield the same spinor are tabulated in Table VII.

As stated earlier, all of the boost transformation laws can be deduced by studying the transformation rules for boosts along the z axis. To do this, one must look at the topology of the product space of spatial orientations of the spinorialized null flags. In particular, the component $\mathbb{R}^3$ corresponding to the three boost directions must be incorporated into the construction of the spinorialized null flags. To this point, the radius of the doubly covered sphere has been fixed (this corresponds to fixing the extent of Penrose's flagpole in Ref. 3). This was the condition that was needed in the study of the effects of rotating the doubly covered sphere. It was implicit that one was looking at the null sphere in a constant time cross section chosen by Penrose to be $t=1$. Now, it is of interest to study the effects of time dilation and Lorentz contraction. Thus, t cannot be chosen to be the constant 1 any longer.

The projection from S^2_+ to the Argand plane at $t=1$ can be described in terms of the embedding of S^2_+ in E^3 (Fig. 25) by the mapping of the rectangular coordinates, $(x,y,z)\to\zeta$. Penrose gives this mapping

$$x=\frac{\xi\bar{\eta}+\eta\bar{\xi}}{\xi\bar{\xi}+\eta\bar{\eta}},\quad y=\frac{\xi\bar{\eta}-\eta\bar{\xi}}{i(\xi\bar{\xi}+\eta\bar{\eta})},\quad z=\frac{\xi\bar{\xi}-\eta\bar{\eta}}{\xi\bar{\xi}+\eta\bar{\eta}}. \tag{5.6}$$

These are the relations that Penrose used to verify (5.3)–(5.5) with $t=\xi\bar{\xi}+\eta\bar{\eta}=1$. To look at the effects of boosts one has to make t variable, so multiplying (5.6) by $(2)^{-1/2}(\xi\bar{\xi}+\eta\bar{\eta})$, one has

$$\begin{aligned} &T=(2)^{-1/2}(\xi\bar{\xi}+\eta\bar{\eta}),\quad X=(2)^{-1/2}(\xi\bar{\eta}+\eta\bar{\xi}),\\ &Y=-i(2)^{-1/2}(\xi\bar{\eta}-\eta\bar{\xi}),\quad Z=(2)^{-1/2}(\xi\bar{\xi}-\eta\bar{\eta}). \end{aligned} \tag{5.7}$$

A z boost on the spin $+\frac{1}{2}$ state does not cause a rotation or translation of the flagpole on the doubly covered Riemannian sphere. Hence, it does not change orientation sheets. Similarly, there is no rotation induced in the tangent space. One concludes that this spinorialized null flag does not change sheets under the action of a z boost. The same is true for the spin-$(-\frac{1}{2})$ base state. Consequently, the effect of a z boost on the density $\hat{\zeta}$ is the same as for the coordinate ζ on the Argand plane that was deduced by Penrose. Define the exponential of the rapidity by $w=(1+v)^{1/2}/(1-v)^{1/2}$, where v is the velocity of the boost along the z direction. Then according to Ref. 3

$$\hat{\zeta}'=w\hat{\zeta}, \tag{5.8a}$$

$$\begin{pmatrix}\xi'\\ \eta'\end{pmatrix}=\begin{pmatrix}w^{1/2} & 0\\ 0 & w^{-1/2}\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \tag{5.8b}$$

Note that the $+$ sign has been chosen in front of the transformation matrices (5.3)–(5.5) and (5.8) that were derived by Penrose up to a sign. It is desired that if the parameter of a transformation is zero then the spinorialized null flag should remain unchanged.

One concludes that to observers at asymptotic infinity that the two-surface at $r=0$ transforms under Lorentz transformation as if it were a spinorialized version of Penrose's null flag. Penrose shows, as was basically reproduced here, that these transform as a Weyl spinor under the action of the group SO(3,1). It should be noted that there was an inequivalent normal bundle (and therefore tangent bundle) that exists over U_2 that was defined at the end of Sec. III. This results in an inequivalent construction of the doubly covered sphere at asymptotic infinity. The construction is totally analogous to what is done here and is left to the reader. One can conclude that the two-surface at $r=0$ actually transforms as a direct sum of two inequivalent Weyl spinors, a four-spinor, as viewed from asymptotic infinity. Since there are only two inequivalent spinor representations of the Lorentz group [i.e., fundamental representations of SL(2,C)], the direct sum must be as for the Dirac spinor, a spinor representation of $O^{\uparrow}(3,1)$. The vertical arrow is used to represent the future directed component of O(3,1) (the doubly covered sphere at asymptotic infinity is future directed). The manner in which the direct sum is made and its physical realization depends on the representation of the spinor algebra that is implemented.

The spinorial nature can be thought of as arising from three things. First, the natural mapping of the fibering of the two-surface at $r=0$ by principal null congruences and their integration to the null sphere at asymptotic infinity. Second, the nonorientability of the two-surface at $r=0$. Also, the quadratic branch point nature of the ring singularity, which along with the nonorientability of U_2 is manifested in the tangent space at every point of the doubly covered null sphere at asymptotic infinity.

The sign ambiguity in Penrose's null flags is resolved here by the obstruction to trivializing the bundle of principal null congruences at their points of "origin" on the two-surface at $r=0$. It is the component of these null geodesics in the normal space which experiences the obstruction. The fact that the two-surface at $r=0$ turns space-time inside out on itself gives rise to the obstruction and therefore to the spinor structure associated with the normal bundle of this surface as viewed from asymptotic infinity.

The rotational properties of the two-surface have an interesting implication for the hypersurface at $r=0$. The two-surface is rotating as a function of time. This induces a phase transformation on the spinorialized null flag by (5.3). It was

shown in Ref. 1 that the two-surface at $r=0$ appears to stationary observers to be a sphere of radius a that is rotating with a frequency w such that $v_\phi = wa\sin(\theta) = c\sin(\theta)$. On the equator this velocity equals c. It is this equivalence with c that forces space-time to knot and twist along the equatorial region so as to give the two-surface a topology other than S^2 (a Klein bottle). The rotation matrix for azimuthal rotations is given by (5.3). Choose the spatial configuration of the two-surface at $r=0$ to coincide with the spinorialized null flag that is represented by the spin-$\frac{1}{2}$ base state. The azimuthal angle of rotation as a function of time is $\phi(t) = wt$, so

$$\psi_{+1/2}(t) = e^{i(1/2)wt}\psi_{+1/2}(t=0) \tag{5.9a}$$

$$= e^{i(1/2)ct/a}\begin{pmatrix}1\\0\end{pmatrix}. \tag{5.9b}$$

For the sake of argument, choose the component of the angular momentum of the source about its symmetry axis to be $\hbar/2$, then it is also $\hbar/2$ about the z axis. Therefore, $a = \hbar/(2mc)$. This allows (5.9) to be rewritten in the provocative form

$$\psi_{+1/2}(t) = e^{i(mc^2/\hbar)t}\begin{pmatrix}1\\0\end{pmatrix}. \tag{5.10}$$

Similarly, for the spin down state (5.3) implies the following effect of rotation:

$$\psi_{-1/2}(t) = e^{-i\phi(t)/2}\psi_{-1/2}(t=0)\,. \tag{5.11a}$$

Since the spin down state has been rotated about the y axis by π radians with respect to the spin up state, $\phi(t) = -wt$ and

$$\psi_{-1/2}(t) = e^{i(mc^2/h)t}\begin{pmatrix}0\\1\end{pmatrix}. \tag{5.11b}$$

By the structure of the spinorialized flag plane any spatial configuration of the two-surface at $r=0$ will have this same time dependence induced by its rotation as viewed in the stationary frames at asymptotic infinity.

The four-spinor nature of this source for the Kerr geometry and the quantum wave nature that is implied by (5.10) and (5.11) makes one interested in the physical interpretation of this source. If it were to represent a lepton on some level then there would have to be an electric charge as viewed from asymptotic infinity. One is led to Wheeler's description of electric charge as resulting from worm holes or handles on space-time.[17] Recall that there is a handle in this geometry on the two-surface at $r=0$ (see Fig. 22). The fact that the handle needed to be of infinitesimal extent in order to get the desired geometry might be consistent with the fact that the handle's physical dimension might be on the order of the Planck length, 10^{-33} cm. It would be interesting to speculate as to whether there is any connection between the geometry that is described in this paper and Wheeler's notion of electric charge.

ACKNOWLEDGMENTS

I would like to thank Edwin Spanier for many enlightening conversations about fiber bundles. I would also like to thank Robert Greene and Ronald Stern for their insight into the nonorientable Gauss–Bonnet theorem on two-dimensional manifolds and their ideas concerning line bundles.

APPENDIX: THE NONORIENTABLE GAUSS–BONNET THEOREM

This Appendix develops a construction of a Gauss–Bonnet type of integral for a nonorientable two-dimensional fiber bundle. The Euler characteristic modulo 2 is equated to the integral of a density (for a discussion of densities see Ref. 7 or 21) modulo 2.

The description of densities that one needs in this discussion depends on their interpretation as elements of a twisted cohomology module. A twisted cohomology is a generalization of De Rahm cohomology that is specifically designed to deal with nonorientable fiber bundles and manifolds (see Bott and Tu, Chap. 7).[7] In the orientable case, the twisted De Rahm complex has elements that are differential forms and the twisted cohomology is the ordinary De Rahm cohomology. For nonorientable manifolds, the elements of the top twisted De Rahm complex are densities.

A twisted De Rahm complex, $\Omega^*(M,E)$ is given by differential forms on a manifold, M, with values in a vector bundle, E. On an open set U_α on M an E-valued q-form can be written locally as $\Sigma_i\, w^i \otimes e^i_\alpha$, where w^i is a q-form on U_α and e^i_α is a vector in E. For a flat bundle (i.e., one in which the transition functions are locally constant, or equivalently, $de^i_\alpha = 0$) Bott and Tu show that the notion of a twisted cohomology module makes sense.[7] The line bundle is obviously a flat bundle. Define L to be the $\mathbb{R}^1$ or line bundle on the manifold M of dimension n that is associated with the orientation bundle (as defined in Sec. II). Here $H^n(M,L)$ is the cohomology of densities. This module is of interest because one can prove a version of Stoke's theorem for nonorientable manifolds in terms of densities.[7] This section utilizes elements of the twisted cohomology of the form $w\otimes e_\alpha \in H^*(M,L)$, where $w\in H^*(M)$ and $e_\alpha\in\mathbb{R}^1$.

In order to obtain the nonorientable version of the Gauss–Bonnet formula, the construction of a twisted global angular form is desirable. This will be done by essentially rewriting the construction of the global angular form that is given in Chap. 6 of Bott and Tu.[7] Assume that the rank of the bundle E is 2 and the base manifold, M, is covered by open sets $\{U_\alpha\}$. The case of interest in this paper has a pseudo-Riemannian connection on the normal bundle. However, it was discussed in Sec. III how a Riemannian structure can be put on the fiber. Define E^0 as the complement of the zero section of E. With the Riemannian connection, one can define polar coordinates r_α and θ_α in the fiber and $x_1,\ldots,x_n$ (for more generality assume that the dimension of the base manifold is an arbitrary number, n, not necessarily 2 as is the case in this paper) are coordinates on U_α. Then r_α, θ_α, $\pi^*x_1,\ldots,\pi^*x_n$ are coordinates on $E^0_{|U_\alpha}$, where π is the projection to the base space in the bundle and the pullback is denoted by π^*.

On the overlap of two open sets of the open covering of the base manifold the angular coordinates θ_α and θ_β in general differ by a rotation angle. If $e_\alpha = e_\beta$, then it appears in the α-coordinate system that the β-coordinate system is rotated by

$$\pi^*\hat{\phi}_{\alpha\beta} = \theta_\beta - \theta_\alpha,\quad 0\leqslant\hat{\phi}_{\alpha\beta}\leqslant 2\pi\,, \tag{A1}$$

on $U_\alpha\cap U_\beta$. If $e_\alpha \neq e_\beta$, then in the α-coordinate system θ_β

appears as $-\theta_\beta+\pi$ (different orientations imply that there must be a reflection of a circle through a diameter that relates the two coordinate systems as well as a possible composition with an orientation preserving rotation). The change of coordinates on the intersection can be given by the following "rotation":

$$\pi^*\hat{\phi}_{\alpha\beta}=-(\theta_\beta+\theta_\alpha)+\pi,\quad 0\leqslant\hat{\phi}_{\alpha\beta}\leqslant 2\pi. \tag{A2}$$

Equations (A1) and (A2) can be combined in one expression by rewriting and implementing twisted cohomology. Let $e_\alpha=1$ and e_β takes on the numerical values $+1$ or -1. Define $J_{\alpha\beta}$ as the sign of the Jacobian determinant of the coordinate transformation that relates U_α and U_β. Then even if e_α is not 1 it is always true that $J_{\alpha\beta}=e_\alpha\cdot e_\beta$. Define $\hat{\phi}_{\alpha\beta}=J_{\alpha\beta}\,\phi_{\alpha\beta}$, where $\phi_{\alpha\beta}$ is a function on $U_\alpha\cap U_\beta$. Then (A1) and (A2) imply the combined relation

$$\pi^*\hat{\phi}_{\alpha\beta}=\theta_\beta e_\beta-\theta_\alpha e_\alpha+\Theta(-J_{\alpha\beta})\pi,\quad 0\leqslant\hat{\phi}_{\alpha\beta}\leqslant 2\pi, \tag{A3}$$

where Θ is the step function that has the value of 1 for positive argument and is zero otherwise.

It is straightforward to verify the condition on the triple intersection $U_\alpha\cap U_\beta\cap U_\gamma$

$$\hat{\phi}_{\alpha\beta}+\hat{\phi}_{\beta\gamma}-\hat{\phi}_{\alpha\gamma}=2\pi z,\quad z\in\mathbb{Z}, \tag{A4}$$

for the four possible situations

(1) $e_\alpha=1,\quad e_\beta=1,\quad e_\gamma=-1$;

(2) $e_\alpha=1,\quad e_\beta=-1,\quad e_\gamma=1$;

(3) $e_\alpha=1,\quad e_\beta=-1,\quad e_\gamma=-1$;

(4) $e_\alpha=1,\quad e_\beta=1,\quad e_\gamma=1$.

Thus, in general the $\{\hat{\phi}_{\alpha\beta}\}$ do not satisfy the cocycle condition [the right-hand side of (A4) is not necessarily zero].

Let ρ_γ be a partition of unity subordinate to the cover $\{U_\gamma\}$. Define ξ_α to be a one-form on U_α and $\xi_\alpha e_\alpha=\hat{\xi}$ is an element of the twisted complex over U_α, ξ_β, and $\hat{\xi}_\beta=\xi_\beta e_\beta$ are defined similarly on U_β by

$$2\pi\hat{\xi}_\beta=\sum_\gamma\rho_\gamma\,d\hat{\phi}_{\beta\gamma}=\sum_\gamma\rho_\gamma J_{\beta\gamma}\,d\phi_{\beta\gamma}.$$

Then by differentiating (A4) one has

$$(1/2\pi)d\hat{\phi}_{\alpha\beta}=\hat{\xi}_\beta-\hat{\xi}_\alpha. \tag{A5}$$

Consequently, $d\hat{\xi}_\alpha=d\hat{\xi}_\beta$. Therefore, the $d\hat{\xi}^\alpha$ can be pieced together to give a global twisted form $\hat{e}\in H^2(M,L)$. The twisted form, $\hat{e}$, will be called the twisted Euler class of the bundle.

Also by (A3) and (A5)

$$(1/2\pi)d\theta_\alpha\,e_\alpha-\pi^*\hat{\xi}_\alpha=(1/2\pi)d\theta_\beta\,e_\beta-\pi^*\hat{\xi}_\beta \tag{A6}$$

on $E^0_{|U_\alpha\cap U_\beta}$. Thus, the twisted forms $(1/2\pi)d\theta_\beta\,e_\beta-\pi^*\hat{\xi}_\beta$ can be pieced together to form a global twisted form $\hat{\psi}$ on E^0. The twisted one-form $\hat{\psi}\in H^1(E^0,L)$ will be called the twisted global angular form. Note that the restriction to each fiber is $\pm(1/2\pi)d\theta$ and

$$d\hat{\psi}=-\pi^*\hat{e}. \tag{A7}$$

By proposition 11.14 of Bott and Tu, one can form a "partial" section of a bundle which has the orthogonal group as its structure group. This means that one can form a global section everywhere on M except in the neighborhood of (a finite number of) the singular points $\{y_1,\ldots,y_m\}$. Look at the section, s, near one of the singular points y_i. One can choose a punctured neighborhood, D^i_r, sufficiently small so that it is diffeomorphic to a punctured open ball of radius r in $\mathbf{R}^n$ (the origin, y_i, is removed) and E is trivial over this set. Choose a covering $\{U_\gamma\}$ of M so that one of the sets covers D^i_r, providing a local orientation on D^i_r and its boundary $\partial\overline{D}^i_r\cong S^{n-1}_i$. The orientation can be chosen on S^{n-1} so that $E_{|D^i_r}\cong D^i_r\times S^{n-1}_i$ is an orientation preserving isomorphism.[7] The local degree of the section, s, at y_i is the degree of the composite map[7]

$$\partial\overline{D}^i_r\xrightarrow{s}E_{|D^i_r}=\overline{D}^i_r\times S^{n-1}\xrightarrow{\rho}S^{n-1}, \tag{A8}$$

where ρ is the projection.

Define M_r to be M minus m (one for each singular point) such balls of radius r then

$$\int_M\hat{e}=\lim_{r\to 0}\int_{M_r}\hat{e}=\lim_{r\to 0}\int_{M_r}s^*\pi^*\hat{e}, \tag{A9}$$

since s is a global section over M_r. By (A7)

$$\int_M\hat{e}=-\lim_{r\to 0}\int_{M_r}s^*\,d\hat{\psi}$$

then an application of Stoke's theorem for densities yields

$$\int_M\hat{e}=-\lim_{r\to 0}\int_{\partial M_r}s^*\hat{\psi}. \tag{A10}$$

Let σ be a generator of S^{n-1}. If one defines $\hat{\sigma}=\sigma e_i$ and $n=2$, then by the common restriction to the same generator on each fiber, $\rho^*\hat{\sigma}$ is cohomologically equivalent in $H^*(E_{|D^i_r},L)$ to $\hat{\psi}$. Equivalently,

$$\hat{\psi}-\rho^*\hat{\sigma}=d\hat{\tau} \tag{A11}$$

for some $(n-2)$ twisted form, $\hat{\tau}$, in $\Omega^*(E_{|D^i_r},L)$. For small r one has

$$s^*\hat{\psi}-s^*\rho^*\hat{\sigma}=s^*\,d\hat{\tau}. \tag{A12}$$

Another application of Stoke's theorem for densities yields

$$\int_{\partial\overline{D}^i_r}s^*\hat{\psi}=\int_{\partial\overline{D}^i_r}s^*\rho^*\hat{\sigma}. \tag{A13}$$

By comparing (A13) and (A8), the integral in (A13) is equal to the local degree of the section, s, at y_i modulo a sign since $\hat{\sigma}=\pm\sigma$. On one of the open sets of the covering of M about one of the singular points (it is assumed that the covering is chosen so that no open set contains more than one singular point) the orientation bundle can be trivialized so that $\hat{\sigma}=\sigma$. Note that if $\hat{\sigma}=-\sigma$ it corresponds to a reflection of a circle about a diameter as mentioned in the discussion of (A2) for the two-plane bundle. For a nonorientable surface, one cannot trivialize the orientation bundle over M_r unless the singular points of the orientation bundle, $\{z_1,\ldots,z_k\}$, are a subset of the singular points of the two-plane bundle, $\{y_1,\ldots,y_m\}$. If this condition is not true then $\hat{\sigma}=-\sigma$ at some (at least one) of the other singular points.

Since $\partial\overline{D}^i_r$ has an opposite local orientation to ∂M_r, one has

$$-\lim_{r\to 0}\int_{M_r} s^*\hat{\psi} = \lim_{r\to 0}\sum_{y_i}\int_{\partial\bar{D}_r^i} s^*\hat{\psi} = \begin{cases} \sum_i (\text{local degree of } s \text{ at } y_i) + \text{ an even number}, & \text{if } \{z_1,\ldots,z_k\} \not\subset \{y_1,\ldots,y_m\}, \\ \sum_i (\text{local degree of } s \text{ at } y_i), & \text{if } \{z_1,\ldots,z_k\} \subset \{y_1,\ldots,y_m\}, \end{cases}$$

or by (A9) and (A10)

$$\int_M \hat{e} = \begin{cases} \chi(E) + \text{ an even number}, & \text{if } \{z_1,\ldots,z_k\} \not\subset \{y_1,\ldots,y_m\}, \\ \chi(E), & \text{if } \{z_1,\ldots,z_k\} \subset \{y_1,\ldots,y_m\}. \end{cases} \tag{A14}$$

It is always true $\int_M \hat{e} \bmod 2 = \chi(E) \bmod 2$.

An obvious example of (A14) is the orientable Gauss–Bonnet theorem. Since the dimension of the fiber was chosen to be two, if one can form a vector cross section up to a reflection in the fiber at some of the singular points ($\hat{\sigma} = -\sigma$ at some of the singular points), then one has a vector section modulo a reflection, a line section. For the nonorientable case, by the construction leading to (A14), the existence of a line section implies that $\int_M \hat{e} = 0 \bmod 2$. In order to see why this must be true one can make the following observations. The integral of the twisted Euler class over M is independent of the covering of M by open sets. Choose the cover of M, $\{U_\gamma\}$, to include small sets $\{U_i\} \subset \{U_\gamma\}$ that are disjoint from each other and each one contains exactly one of the singular points, y_i.

In this analysis, the vector bundle is studied in terms of degrees of singular points by equating the bundle to a circle bundle.[7] One can consider the line bundle as a bundle of diameters of these circles that are associated with the vector bundle. If one can trivialize the vector bundle over M_r then it is possible to trivialize the line bundle over M_r.

If the absolute value of the degree of the singular point y_i is k_i then cover U_i by $2k_i$ sets U_i^t, $t = 1,\ldots,2k_i$ such that each U_i^t contains y_i and $\cup_t U_i^t = U_i$. Choose the U_i^t's small enough so that one can trivialize the line bundle on each of the sets U_i^t separately for all i and t. Choose r small enough so that $M_r \cap U_i$ is topologically an annulus. The intersections $U_i^{t'} \cap U_i^{t''}$ have only one connected component. The geometry of this situation is indicated for a singular point of degree of absolute value equal to 1 in Fig. 29. The sets have been chosen so that the transition functions for the line bundle on the intersections, $\{U_i - \cup_{t'} U_i^{t'}/t' \neq t\} \cap M_r$, are $+1$. This can be done for the following reason. If one transports the circle fiber about $\partial\bar{D}_r^i$ then the circle will execute k_i rotations (if the absolute value of the local degree of s at y_i is k_i) upon completion of the circuit. By choosing $2k_i$ sets, one can isolate a half of a rotation in each $U_i^t \cap M_r$. This allows one to match the trivialization on each U_i^t to that on M_r in such a way that the information on the degree of singularity is reflected in the intersections of the U_i^t's among themselves. There are k_i $(2k_i - 1)$ such intersections. Thus, one must specify k_i $(2k_i - 1)$ transition functions. It is easy to convince oneself that k_i^2 of these functions have minus signs and k_i $(k_i - 1)$ will have plus signs. Thus, a singular point of odd degree will have an odd number of minus signs in the set of transition functions for this covering. If the singular point is of even degree there will be an even number of minus signs.

The integral of $\hat{e}$ over M is also independent of the vector field that is used to compute this number (i.e., it is independent of the trivial vector bundle that is chosen on M_r). For example, the total degree of the singular points of a vector field over S^2 can be thought of as arising from one singular point of degree 2 at the south pole or two singular points of degree 1 placed at both poles. Choose the vector field on M that is used to compute $\int_M \hat{e}$ so that the degrees of the singular points have absolute value equal to unity. The manifold minus these points will still be called M_r.

Define a closed curve that passes through each singular point exactly once and through each of the two sets that cover these singular points of degree of magnitude 1 (see Fig. 29) exactly once. By the previous statements in this discussion, the transition function on the intersection of each of these two set coverings of a neighborhood of a singular point must be negative since the local degree is odd. Thus, if one can trivialize the line bundle over this closed curve then $\Sigma_i \pm$ (local degree of s at y_i) is an even number. One con-

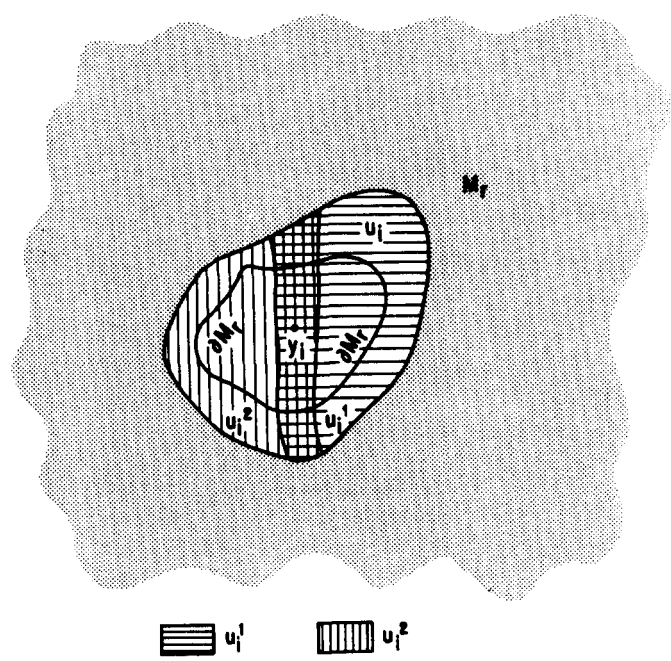


FIG. 29. The point y_i is a singular point with a local degree of magnitude 1. Here M_r, indicated by the shading, is a large manifold that extends far beyond the boundaries of the illustration. The open set U_i^1 is indicated by an array of horizontal lines and U_i^2 by the vertical lines.

cludes that if there exists a global line section over M that this sum must be even. Then by (A14), the existence of a line section over M implies that $\int_M \hat{e} = 0 \bmod 2$. One concludes that $\chi(E) \bmod 2 = 0$ if there exists a line section of the two-plane bundle. This is a strange theorem to prove for a two-dimensional fiber as it was shown that one can define a Gauss–Bonnet theorem without densities to find the Euler number for a nonorientable bundle in this case (see the beginning of Sec. III). For the purposes of this paper the integral in that treatment was undefined. One could construct a normal line section and by the results of this Appendix, $\chi(N) \bmod 2 = 0$.

Another advantage of this construction is that a Gauss–Bonnet theorem like (A14) for twisted cohomology elements may be extendible to higher-dimensional nonorientable fibers. From this discussion, if $\hat{e}$ can be found no matter what the dimensions then $\int_M \hat{e} \bmod 2$ gives the Euler characteristic mod 2. For the orientable case the Euler class, e, can be determined by the Chern–Weil homomorpism. One defines differential forms with values in a group. This group is $SO(n)$ for the Euler class. The homomorphism does not work for $O(n)$.[22] Hence, there is no nonorientable Gauss–Bonnet theorem. This author believes that it can be proved that if one defines elements of the twisted cohomology that are associated with the line fiber of orientation as was done here and gives them values in $SO(n)$, then one can proceed in complete analogy to the Chern–Weil homomorphism to obtain $\hat{e}$. This results in a Gauss–Bonnet type of theorem for nonorientable bundles in the form

$$\int_M \hat{e} \bmod 2 = \chi(B) \bmod 2 .$$

[1]B. Punsly, J. Math. Phys. **26**, 1728 (1985).
[2]R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).
[3]R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U. P., New York, 1984), Vol. I.
[4]B. Carter, Phys. Rev. **174**, 1559 (1968).
[5]D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination* (Chelsea, New York, 1952).
[6]W. S. Massey, *Algebraic Topology: An Introduction* (Harcourt, New York, 1967).
[7]R. Bott and L. Tu, *Differential Forms in Algebraic Topology* (Springer, New York, 1982).
[8]N. Steenrod, *The Topology of Fiber Bundles* (Princeton U. P., Princeton, NJ, 1974).
[9]J. Milnor and J. Stasheff, *Characteristic Classes* (Princeton U. P., Princeton, NJ, 1974).
[10]H. Whitney, "On the topology of differentiable manifolds," *Lectures in Topology* (University of Michigan, Ann Arbor, 1941).
[11]R. P. Kerr and A. Schild, "Some algebraically degenerate solutions of Einstein's gravitational field equations," presented at the American Mathematical Society Symposium, New York, 1964.
[12]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., New York, 1977).
[13]R. Greene (private communication).
[14]R. Stern (private communication).
[15]Loring Tu has indicated to this author in a letter that the nonorientable Gauss–Bonnet theorem formulated in terms of densities is not in any mathematical literature that he has seen and may be of interest to the mathematical community.
[16]S. Chern, An. Acad. Brasil. Ciênc. **35**, 16 (1963).
[17]C. Misner, K. Thorne, and J. Wheeler *Gravitation* (Freeman, San Francisco, 1973).
[18]M. Green (private communication).
[19]M. J. Greenberg and J. R. Harper, *Algebraic Topology: A First Course* (Benjamin/Cummings, Don Mills, Ontario, 1981).
[20]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Menlo Park, CA, 1981).
[21]F W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Scott and Foresman, Glenview, IL, 1971).
[22]J. L. Dupont, *Curvature and Characteristic Classes* (Springer, New York, 1978).

Computer-aided study of a class of Riemannian space-times

M. J. Rebouças
Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud, 150-Urca, 22290 Rio de Janeiro-RJ, Brazil
J. E. Åman
Institute of Theoretical Physics, University of Stockholm, Vanadisvägen 9, S-11346 Stockholm, Sweden

(Received 23 September 1986; accepted for publication 26 November 1986)

The Riemannian Gödel-type manifolds are examined in the light of the equivalence problem techniques by using algebraic computing. The conditions for space-time homogeneity of a Riemannian manifold with a Gödel-type metric are derived, generalizing previous work on this subject matter. A classification of the Gödel-type Riemannian manifolds based on the two relevant parameters m and Ω is formulated. It is shown that apart from the $m^2 = 4\Omega^2$ case they are all Petrov type D with a five-parameter group of motions. The special $m^2 = 4\Omega^2$ manifold is shown to be conformally flat and to admit a seven-parameter group of isometries. An algebraic classification of the Ricci spinor of all Gödel-type Riemannian manifolds is discussed. Possible sources for these space-times are examined. A generalization of the Rebouças–Tiomno theorem on Gödel-type manifolds is given.

I. INTRODUCTION

The arbitrariness in the choice of coordinates in general relativity gives rise to the problem of deciding whether or not two Riemannian space-times, given explicitly in terms of coordinates, viz.,

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta , \tag{1.1}$$

$$ds^2 = \tilde{g}_{\alpha\beta}\, d\tilde{x}^\alpha\, d\tilde{x}^\beta , \tag{1.2}$$

are locally the same. This is the so-called equivalence problem (see, for example, Refs. 1–5).

Most relativists' first approach to cope with that problem would probably be through scalar invariants such as $R^\alpha{}_\alpha$, $R_{\alpha\beta}R^{\alpha\beta}$, $R_{\alpha\beta\gamma\rho;\sigma}R^{\alpha\beta\gamma\rho;\sigma}$, etc. However, this attempt cannot work so long as there exist *curved* space-times in which all scalar invariants vanish—indistinguishable therefore from the *flat* space-time of special relativity.

On the other hand, from the fact that any metric can be written locally in the form of a Taylor series whose coefficients are functions of $R_{\alpha\beta\gamma\rho}$ and its covariant derivatives, one may infer that what does matter in fact are the Riemann tensor and its covariant derivatives. Actually, the equivalence problem can be reduced to that of finding out whether the finite set of equations

$$\begin{aligned} &R_{ABCD} = \tilde{R}_{ABCD} , \\ &R_{ABCD;E_1} = \tilde{R}_{ABCD;E_1} , \\ &R_{ABCD;E_1\cdots E_p} = \tilde{R}_{ABCD;E_1\cdots E_p} , \end{aligned} \tag{1.3}$$

is consistent or not. Here the curvature tensor as well as its covariant derivatives are referred to a canonically chosen frame.

It is worth noting that in calculating the derivatives of the Riemann tensor one can stop as soon as one reaches a step at which the pth derivatives (say) are algebraically expressible in terms of the previous ones, since further differentiation will not yield any new piece of information.[6] Actually, if the pth derivative is expressible in terms of its predecessors, for all $k > p$ the k th derivatives can all be expressed in terms of 0th, 1st,..., $(p-1)$th derivatives.

Handling distinctively space-time coordinates x^α, and tetrad rotation parameters ξ^A, Karlhede[2] refined upon the equivalence problem: rephrased it in the language of spinors, pointed out that the maximum number of derivatives needed for a complete characterization of a given space-time depends on the corresponding Petrov type, and discussed how to fix the frame for each Petrov type in a canonical way.

Karlhede's procedure was implemented using the computer algebra system SHEEP,[7,8] at first for vacuum, by Åman.[9,10] More recently, programs to extend the treatment to the nonvacuum case have been written by MacCallum, Joly, and Åman and are described in Refs. 3 and 11–13.

In 1949 Gödel[14] presented a cosmological solution of Einstein's equations of the form

$$ds^2 = [dt + H(x)dy]^2 - D^2(x)dy^2 - dx^2 - dz^2 , \tag{1.4}$$

which is known as a Gödel-type metric. The Gödel solution and its energy-momentum tensor can be given by

$$H = e^{mx}, \quad D = e^{mx}/\sqrt{2} , \tag{1.5}$$

and

$$T_{\mu\nu} = \rho V_\mu V_\nu, \quad V^\mu = \delta^\mu{}_0 , \tag{1.6}$$

$$\kappa\rho = -2\Lambda = m^2 = 2\Omega^2 , \tag{1.7}$$

where κ is the Einstein constant, V^μ is the fluid four-velocity, Ω is the rotation of matter, and m is a constant.

Despite its various striking properties, the cosmological solution discussed by Gödel has a well recognized historical (and even philosophical) importance[15] and has to a large extent motivated the investigation on rotating cosmological space-times. Particularly, the search for rotating Gödel-type models (1.4) has received more attention in recent years, and the literature on those kinds of space-times is fairly large today.[16]

However, apart from Gödel's paper, the problem of space-time homogeneity of Gödel-type Riemannian manifolds was considered only in 1980 by Raychaudhuri and Thakurta.[17] They found the necessary conditions for a Riemannian manifold to be homogeneous in space and time

 0022-2488/87/040888-05$02.50

(hereafter called ST homogeneous).

Two years later Rebouças and Tiomno[16] proved that the Raychaudhuri–Thakurta conditions are also sufficient for ST homogeneity of Gödel-type Riemannian manifolds. However, not only Raychaudhuri and Thakurta but also Rebouças and Tiomno have explicitly or implicitly restricted their study to the class of time-independent Killing vector fields. Perceiving this restriction in both papers (Refs. 16 and 17) Teixeira *et al.*[18] extended these investigations so as to include the time-dependent isometries. Nevertheless, their study is yet limited in that they have used equations [Eqs. (19)–(21)] borrowed from Ref. 16, obtained under a restrictive assumption on the form of the Killing vector fields.

In the light of the equivalence problem techniques, as formulated by Karlhede[2] and embodied in a suite of computer programs (CLASSI[19]) written in SHEEP by Åman, MacCallum, and Joly,[3,11,13] we examine all Riemannian Gödel-type manifolds. The necessary and sufficient conditions for a Gödel-type Riemannian manifold to be ST homogeneous are derived without assuming any limiting hypothesis on the isometries. The Rebouças–Tiomno major results are recast in a more general context. The Gödel-type ST-homogeneous Riemannian manifolds are shown to be characterized by two parameters m and Ω: identical pairs (m^2,Ω) correspond to equivalent manifolds. A classification of all Gödel-type Riemannian space-times based on these parameters is presented. We also show that apart from the $m^2=4\Omega^2$ case they are all Petrov type D and have a five-parameter maximal group of motion (G_5). The special case $m^2=4\Omega^2$ is shown to be Petrov type 0 and to admit a G_7 of isometries, in agreement with the results of Teixeira *et al.*[18] obtained through a procedure completely different from ours. The particular case $m=\Omega=0$ is shown to be nothing but the flat space-time. As for the Segré type of the Ricci tensor we prove that, in general, it is [1,1(11)] unless $m^2=2\Omega^2$ or $m^2=4\Omega^2$. The $m^2=2\Omega^2$ case corresponds to the Gödel solution and is of the algebraic Segré type [1,(111)] whereas the $m^2=4\Omega^2$ case is shown to be of Segré type [(1,11)1]. Possible sources of Gödel-type metrics are discussed. The degenerated Gödel-type metrics ($\Omega=0$, $m^2\neq 0$) are also studied: they admit a G_6 of motions and are of Segré type [(1,1)11]. There emerges from our study a generalization of a theorem proved in Ref. 16.

Our major aim in the next section is to present a brief summary of important practical results and prerequisites on the equivalence problem required in Sec. III, to set our framework, define the notation, and make our text to a certain extent self-contained. Detailed and quite good reviews on that subject can be found in Karlhede,[2] MacCallum,[3,5] and Ehlers.[4]

In Sec. III we state, prove, and discuss our main results.

II. PRACTICAL PROCEDURE AND PREREQUISITES

An important practical point to be considered, once one wishes to test the equivalence of two metrics, is that before attempting to solve Eqs. (1.3) one can extract and compare partial pieces of information: the subgroup of H_q of G_r under which the set $R_q=\{R_{ABCD}, R_{ABCD;E_1}, \ldots, R_{ABCD;E_1\ldots E_q}\}$ is invariant, and the number t_q of functionally independent functions of space-time coordinates contained in R_q. They must be the same (at each q) if the matrics are equivalent.

The procedure for comparing two metrics which results from the above considerations is as follows.

(1) Let $q=0$.

(2) Compute R_q.

(3) Find the set of frame transformations under which R_q is invariant (isotropy group H_q).

(4) Fix the frame up to H_q by choosing a canonical form for R_q and find t_q.

(5) If $t_q=t_{q-1}$ and $H_q=H_{q-1}$, stop (let us denote such a q by p). Otherwise, increment q by 1 and return to step (2).

To compare two space-times we first test if they have the same t_q and H_q for each q up to p. If they differ, so do the metrics. If not, it is necessary to check the consistency of Eqs. (1.3).

The isotropy group of each space-time will be H_p and the group of motions will have dimension r given by

$$r=4-t_p+\dim(H_p)\,, \tag{2.1}$$

acting on an orbit with dimension

$$d=r-s=4-t_p\,, \tag{2.2}$$

where

$$s=\dim(H_p)\,. \tag{2.3}$$

It is worth emphasizing that although the equivalence problem is couched in terms of equivalence of two metrics, by the above procedure we actually compute one invariant characterization of each metric and at the end of this procedure in addition to t_q's and H_q's we shall have a number of consequent data as, for example, the dimension of the isometry group [given by (2.1)], the Petrov type of the Weyl tensor, and the Segré type of the Ricci tensor.[20]

Particularly stimulated by the simplicity of the Petrov classification in spinorial language, Karlhede, rather than using frame as such, rephrased the equivalence problem in terms of spinor equivalents: ψ_{ABCD} (Weyl spinor), $\phi_{AB\dot{C}\dot{D}}$ (Ricci spinor), $\Lambda=R/24$ (Ricci scalar), and their covariant derivatives. He has also prescribed a way of fixing (as far as possible) the frame for each Petrov type. For Petrov type I, for example, a canonical frame is fixed by demanding that the nonvanishing components of the Weyl spinor ψ_A are such that $\psi_1=\psi_3\neq 0$, $\psi_2\neq 0$, with $\mathrm{Re}[\psi_1]>0$.

An important point to be taken into account when one needs to compute derivatives of the curvature tensor is that they are interrelated by the Ricci and Bianchi identities. It is therefore of importance to be able to single out a set of quantities which are algebraically independent and from which the Riemann tensor and its covariant derivatives are obtainable by algebraic operations. Penrose[21] showed that for vacuum and electrovacuum space-times only the symmetrized spinor derivatives of the Weyl and Maxwell spinors are required. However this result does not hold for the general case. Recently MacCallum and Åman[22] have generalized Penrose's result stating that one must instead take a large set of components, which for the qth derivative of the curvature can be taken to be (i) the totally symmetrized

spinor qth derivatives of the Weyl and Ricci spinors, and the Ricci scalar, viz.,

$$\nabla_{(A}^{\dot{X}}\nabla_{B}^{\dot{Y}}\cdots\nabla_{C}^{\dot{Z})}\psi_{MNPQ)}\,, \tag{2.4}$$

$$\nabla_{(A}^{\dot{X}}\nabla_{B}^{\dot{Y}}\cdots\nabla_{C}^{\dot{Z}}\Phi_{MN)}{}^{\dot{P}\dot{Q})}\,, \tag{2.5}$$

$$\nabla_{(A}^{\dot{X}}\nabla_{B}^{\dot{Y}}\cdots\nabla_{C)}^{\dot{Z}}\Lambda\,, \tag{2.6}$$

together with (ii) for $q\geqslant 1$, the totally symmetrized $(q-1)$th derivatives of

$$\chi_{DEFW}=\nabla_{W}^{C}\psi_{CDEF}\,,$$

i.e.,

$$\nabla_{(A}^{\dot{X}}\nabla_{B}^{\dot{Y}}\cdots\nabla_{C}^{\dot{Z}}\chi_{DEF)}{}^{\dot{W}}\,, \tag{2.7}$$

and (iii) for $q\geqslant 2$, the d'Alembertian of all quantities used at the $(q-2)$th step, viz.,

$$\nabla^{A}{}_{\dot{X}}\nabla^{\dot{X}}{}_{A}Q\,, \tag{2.8}$$

where Q stands for a generic quantity at the $(n-2)$th step.

To close this section we remark that in the CLASSI implementation of the above results, used in the next section, a notation similar to that adopted for ψ_A and $\Phi_{A\dot{B}}$ by Newman–Penrose is employed. In this notation the indices are all subscripts, components are labeled by one dotted and one undotted index whose numerical values are the total number of corresponding spinor indices. Thus, for example, one has

$$\nabla\psi_{2\dot{0}}\equiv D\psi_{2\dot{0}}=\psi_{(1000;1)\dot{0}}=\nabla_{(A}^{\dot{X}}\psi_{BCDE)}\iota^{A}\iota^{B}o^{C}o^{D}o^{E}\bar{o}_{\dot{X}}\,, \tag{2.9}$$

where the pair ι^{A},o^{B} constitutes an orthonormal spinor basis.

III. HOMOGENEOUS GÖDEL-TYPE METRICS: RESULTS AND DISCUSSIONS

Let us state from the beginning our general setting. Throughout this section we shall consider a four-dimensional Riemannian manifold M, endowed with a Gödel-type metric (1.4).

For arbitrary functions $H(x)$ and $D(x)$ the Gödel-type metrics are Petrov type I. Accordingly the null tetrads ω^{A} which turn out to be appropriate (canonical) for our discussions are

$$\begin{aligned}
\omega^{0}&=(1/\sqrt{2})(dt+H\,dy+dz)\,,\\
\omega^{1}&=(1/\sqrt{2})(dt+H\,dy-dz)\,,\\
\omega^{2}&=(1/\sqrt{2})(D\,dy-i\,dx)\,,\\
\omega^{3}&=(1/\sqrt{2})(D\,dy+i\,dx)\,.
\end{aligned} \tag{3.1}$$

By using this basis the Gödel-type line element (1.4) may be written as

$$ds^{2}=2(\omega^{0}\omega^{1}-\omega^{2}\omega^{3})\,. \tag{3.2}$$

It is worth mentioning that, not only the Petrov type but also the canonical frame were obtained by interaction with CLASSI, starting from one arbitrary null frame and bearing in mind Karlhede's results on how to fix the frame for a Petrov type I metric.

Now using the suite of SHEEP classification programs (CLASSI) referred to in the Introduction one finds

$$\psi_{1}=\psi_{3}=-\tfrac{1}{8}(H'/D)'\,, \tag{3.3}$$

$$\psi_{2}=\tfrac{1}{6}[D''/D-(H'/D)^{2}]\,, \tag{3.4}$$

$$\Phi_{0\dot{0}}=\Phi_{2\dot{2}}=\tfrac{1}{8}(H'/D)^{2}\,, \tag{3.5}$$

$$\Phi_{0\dot{1}}=\Phi_{1\dot{2}}=\tfrac{1}{8}(H'/D)'\,, \tag{3.6}$$

$$\Phi_{1\dot{1}}=\tfrac{1}{4}[\tfrac{3}{4}(H'/D)^{2}-D''/D\,]\,, \tag{3.7}$$

$$\Lambda=-\tfrac{1}{12}[D''/D-\tfrac{1}{4}(H'/D)^{2}]\,, \tag{3.8}$$

where the prime denotes derivative with respect to x.

For ST homogeneity from Eq. (2.2) one must have

$$t_{p}=0\,, \tag{3.9}$$

that is, the number of functionally independent functions of space-time coordinates contained in the set R_{p} must be zero. Accordingly, the spinor quantities ψ_{A}, $\Phi_{A\dot{B}}$, and Λ must be constant. Therefore from Eqs. (3.3)–(3.8) we conclude that for ST homogeneity it is necessary that

$$H'/D=\text{const}\equiv 2\Omega\,, \tag{3.10}$$

$$D''/D=\text{const}\equiv m^{2}\,, \tag{3.11}$$

where a notation similar to that of the Rebouças and Tiomno paper[16] is used.

The above necessary conditions are also sufficient for ST homogeneity. Under those conditions the Weyl and Ricci spinors and the Ricci scalar reduce to

$$\psi_{2}=m^{2}/6-\tfrac{2}{3}\Omega^{2}\,, \tag{3.12}$$

$$\Phi_{0\dot{0}}=\Phi_{2\dot{2}}=\tfrac{1}{2}\Omega^{2}\,, \tag{3.13}$$

$$\Phi_{1\dot{1}}=\tfrac{3}{4}\Omega^{2}-(m^{2}/4)\,, \tag{3.14}$$

$$\Lambda=\tfrac{1}{12}(\Omega^{2}-m^{2})\,. \tag{3.15}$$

We group the Gödel-type Riemannian manifolds, according to the relevant parameters m and Ω, into three classes:

(i) $m^{2}\neq 4\Omega^{2}$, $m,\Omega\neq 0$,

(ii) $m^{2}=4\Omega^{2}$, $m,\Omega\neq 0$,

(iii) $m^{2}\neq 0$, $\Omega=0$.

We remark that the particular case $m=\Omega=0$ is not included in our study inasmuch as, from Eqs. (3.12)–(3.15), it is obviously the flat space-time.

We proceed by carrying out the procedure of Sec. II for each class of Gödel-type Riemannian metrics.

For the first class we have $\psi_{2}\neq 0$ and therefore all metrics are Petrov type D. The Weyl and Ricci spinors are both invariant under the spatial rotation

$$\begin{pmatrix} e^{i\theta} & 0\\ 0 & e^{-i\theta}\end{pmatrix}, \tag{3.16}$$

where θ is a real parameter. Thus

$$\dim(H_{0})=1\,. \tag{3.17}$$

Following the scheme outlined in Sec. II, we next calculate the algebraically independent parts of the first convariant derivative of the Riemann spinor. Now CLASSI gives

$$D\psi_{2\dot{0}}=-D\psi_{3\dot{1}}=(2i\sqrt{2}/5)\Omega(\Omega^{2}-\tfrac{1}{4}m^{2})\,, \tag{3.18}$$

$$\chi_{1\dot{0}}=\chi_{2\dot{1}}=i\sqrt{2}\Omega(\Omega^{2}-\tfrac{1}{4}m^{2})\,, \tag{3.19}$$

$$D\Phi_{A\dot{B}}=0,\quad \text{for all } A \text{ and } \dot{B}\,, \tag{3.20}$$

$$D\Lambda = 0\,. \tag{3.21}$$

Since $t_0 = t_1 = 0$ and the invariance group is yet the same, the process terminates. The group of motions is therefore five dimensional with a one-dimensional isotropy group, as one may easily find by using Eq. (2.1). Thus the necessary conditions (3.10) and (3.11) are also sufficient for ST homogeneity for the present class.

For the next class ($m^2 = 4\Omega^2; m,\Omega \neq 0$) the Weyl spinor ψ_A vanishes identically, which means conformal flatness. We also have

$$\Phi_{0\dot{0}} = \Phi_{2\dot{2}} = -2\Phi_{1\dot{1}} = m^2/8\,, \tag{3.22}$$

$$\Lambda = -m^2/16\,. \tag{3.23}$$

Here the group of invariance of $\Phi_{A\dot{B}}$ is the three-dimensional Lorentz group SO(2,1). Thus we so far have $t_0 = 0$ and $\dim(H_0) = 3$. Since all derivatives turn out to vanish identically, the process terminates and we have a G_7 of motions for this class of Gödel-type Riemannian manifolds.

As for the last class ($\Omega = 0$, $m^2 \neq 0$), using CLASSI, we similarly find that

$$\psi_2 = m^2/6, \quad \Phi_{1\dot{1}} = -m^2/4\,, \tag{3.24}$$

$$\Lambda = -m^2/12, \quad t_0 = t_1 = 0\,, \tag{3.25}$$

and that all derivatives as well as $\chi_{DEF\dot{W}}$ vanish identically. Moreover, from Eqs. (3.24) and (3.25), we learn that the Weyl and Ricci spinors are both invariant under spatial rotations and boosts, namely

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix}, \tag{3.26}$$

where θ and r are real parameters. Therefore from Eqs. (2.1) and (2.2) we readily find that there exists a G_6 of motions acting on the manifold. The conditions (3.10) and (3.11) are once more sufficient for ST homogeneity.

We collect the results so far obtained in the following theorems.

Theorem 1: The necessary and sufficient conditions for a Riemannian Gödel-type manifold to be ST homogeneous are those given by Eqs. (3.10) and (3.11).

Theorem 2: All ST-homogeneous Gödel-type Riemannian manifolds are characterized by two parameters m and Ω: identical pairs (m^2,Ω) specify equivalent manifolds.

Theorem 3: Except for the $m^2 = 4\Omega^2$ manifold, all Riemannian Gödel-type manifolds are Petrov type D with either a G_5 (when $\Omega \neq 0$) or a G_6 (when $\Omega = 0$) of motions. The special $m^2 = 4\Omega^2$ manifold is Petrov type 0 and admits a G_7 of isometries.

It is worth emphasizing that in Theorem 1 above we recover the Raychaudhuri–Thakurta–Rebouças–Tiomno necessary and sufficient conditions for ST homogeneity of a Gödel-type Riemannian manifold. However, contrary to their proof, we have made no hypothesis on the nature of the isometries.

It should also be noticed that Theorem 2 is a generalization of a theorem given in Ref. 16, in that it is now proved in a much more general setting.

Before proceeding to the discussion of the types of Ricci spinor, let us state the problem and fix our notation. The algebraic classification of the symmetric second-order Ricci tensor $R_{\mu\nu}$, in general relativity, is a linear algebra problem with an underlying four-dimensional space-time having an indefinite norm, which can be given in terms of the Segré characteristics or Segré types. It turns out that only the types [1,111] and [2,11] and their specializations are consistent with both the local Lorentzian character of general relativity and the dominant energy conditions. In referring to the Segré types we use a notation where the individual digits inside square brackets are related to the multiplicity of the corresponding real eigenvalue, equal eigenvalues are enclosed in round brackets, the first digit corresponds with a timelike or null eigenvector and is separated from the spacelike ones by a comma.

Now making use of a CLASSI program written in SHEEP by Joly and MacCallum[13] one finds that the ST-homogeneous Gödel-type manifolds have a Ricci spinor or Segré type [1,1(11)] unless

$$m^4 - 6\Omega^2 m^2 + 8\Omega^4 = 0\,.$$

The roots of this equation correspond to the Gödel manifold $m^2 = 2\Omega^2$, which is Segré type [1,(111)], and the Rebouças–Tiomno space-time $m^2 = 4\Omega^2$, which is of type [(1,11)1] of the Segré classification.[23]

As for the possible sources for the [1,1(11)] algebraic type one can show that [24] it may be generated by either a nonviscous fluid ($\eta = \xi = 0$) with heat flow or combination of fields as, e.g., two radiation fields, two perfect fluids, or even a radiation field plus perfect fluid.

It seems worth stressing that energy-momentum tensors of quite different matter distributions may, in fact, have precisely the same Segré type. Tupper,[25,26] for example, showed that the Gödel model $m^2 = 2\Omega^2$, originally found as a perfect fluid solution, may also be produced by a magnetohydrodynamic fluid.

To conclude, we mention that although originally found as generated by a massless scalar field, the $m^2 = 4\Omega^2$ Gödel-type space-time is also a solution for a rotating matter with spin.[27,28] The fact that it is of Segré type [(1,11)1] implies that the related energy momentum tensor may yet be understood as that of a tachyon fluid.[29]

ACKNOWLEDGMENTS

We are grateful to Professor M. A. H. MacCallum for the suggestion of the present way of looking into this problem, for fruitful correspondence, and for his very kind hospitality at the School of Mathematical Sciences, Queen Mary College. One of us (M. J. R.) is also grateful to Professor I. W. Roxburgh for his hospitality at that school, and to Professor B. Laurent for providing the opportunity of working at the Institute of Theoretical Physics, University of Stockholm to proceed with earlier collaboration.

Special thanks are due to Coordenação de Aperfeiçoamento de Nível Superior (CAPES) for financial support making our collaboration possible. This work was also partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq.

[1] E. Cartan, *Leçons sur la géométrie des espaces de Riemann* (Gauthier-Villars, Paris, 1946) [English translation by J. Glazebrook, *Geometry of*

Riemannian Spaces, with notes by R. Hermann (Math. Sci. Press, Brookline, MA, 1983)].
[2]A. Karlhede, Gen. Relativ. Gravit. **12**, 683 (1980).
[3]M. A. H. MacCallum, "Classifying metrics in theory and practice," in *Unified Field Theory in More Than 4 Dimensions, Including Exact Solutions*, edited by V. de Sabbata and E. Schmutzer (World Scientific, Singapore, 1983).
[4]J. Ehlers, "Christoffel's work on the equivalence problem for Riemannian spaces and its importance for modern field theories of physics," in *E. B. Christoffel*, edited by P. L. Butzer and F. Feher (Birkhauser, Basel, 1981).
[5]M. A. H. MacCallum, "Computer-aided classification of exact solutions in general relativity," lecture delivered at the XIV Yamada Conference, Kyoto, Japan, 1986; to appear in the *Proceedings of the Conference*, edited by T. Nakamura and H. Sato (World Scientific, Singapore).
[6]T. Y. Thomas, *Differential Invariants of Generalized Spaces* (Cambridge U. P., Cambridge, 1934).
[7]I. Frick, "SHEEP users guide," Institute of Theoretical Physics, University of Stockholm Report 77-14, 1977.
[8]I. Frick, "The computer algebra system SHEEP, what it can and cannot do in general relativity," Institute of Theoretical Physics, University of Stockholm Report 77-15, 1977.
[9]J. E. Åman and A. Karlhede, Phys. Lett. A **80**, 229 (1981).
[10]J. E. Åman and A. Karlhede, "An algorithmic classification of geometries in general relativity," in *SYMSAC'81, Proceedings of the 1981* ACM *Symposium on Symbolic and Algebraic Computation*, edited by P. S. Wang (Association of Computing Machinery, New York, 1981).
[11]M. A. H. MacCallum, "Algebraic computer in general relativity," in *Classical General Relativity*, edited by W. B. Bonnor, J. N. Islam, and M. A. H. MacCallum (Cambridge U. P., Cambridge, 1984).
[12]I. Cohen, I. Frick, and J. E. Åman, "Algebraic computing in general relativity," in *General Relativity and Gravitation*, edited by B. Bertotti, A. Pascolini, and F. de Felice (Reidel, Dordrecht, 1984).
[13]G. C. Joly and M. A. H. MacCallum, "Computer-aided classification of Ricci tensor in general relativity," to appear in Class. Quant. Gravit.
[14]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[15]See, for example, J. Pfarr, Gen. Relativ. Gravit. **9**, 393 (1981), and references therein quoted on philosophical discussions which came about after Gödel's paper.
[16]M. J. Rebouças and J. Tiomno, Phys. Rev. D **28**, 1251 (1983); and references therein mentioned on Gödel-type solutions.
[17]A. K. Raychaudhuri and S. N. Guha Thakurta, Phys. Rev. D **22**, 802 (1980).
[18]A. F. F. Teixeira, M. J. Rebouças, and J. E. Åman, Phys. Rev. D **32**, 3309 (1985).
[19]J. E. Åman, "Manual for CLASSI—Classification programs for geometries in general relativity," University of Stockholm Report, 1982.
[20]See, for example, G. S. Hall, Diff. Geom. **12**, 53 (1983); Arab. J. Sci. Eng. **9**, 87 (1984) for useful reviews on the classification of a second-order symmetric tensor in general relativity.
[21]R. Penrose, Ann. Phys. (NY) **10**, 171 (1960).
[22]M. A. H. MacCallum and J. E. Åman, Class. Quant. Gravit. **3**, 1133 (1986).
[23]As a matter of fact, the Rebouças–Tiomno space-time is the unique Gödel-type metric with an algebraic structure of this type, as was proved in Ref. 29.
[24]G. S. Hall and D. A. Negum, J. Theor. Phys. **25**, 405 (1986).
[25]B. O. J. Tupper, Gen. Relativ. Gravit. **15**, 47 (1983).
[26]B. O. J. Tupper, Gen. Relativ. Gravit. **15**, 849 (1983).
[27]J. Duarte de Oliveira, A. F. F. Teixeira, and J. Tiomno, contributed paper to the *10th International Conference on General Relativity and Gravitation*, edited by B. Bertotti, F. de Felice, and A. Pascolini (Padova, Italy, 1983).
[28]J. Duarte de Oliveira, A. F. F. Teixeira, and J. Tiomno, Phys. Rev. D **34**, 3661 (1986).
[29]M. J. Rebouças, J. E. Åman, and A. F. F. Teixeira, J. Math. Phys. **27**, 1370 (1986).

Soliton transformations for axially symmetric higher-dimensional gravity. I. Generalized Neugebauer–Kramer transformations

S.-C. Lee
Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China

(Received 17 July 1986; accepted for publication 10 December 1986)

The field equations for axially symmetric gravitational fields in $4+K$ dimensions with K-dimensional toroidal compatification can be reduced to those of a generalized nonlinear sigma model. The dual symmetry of this model is considered first. Then the soliton transformations for these equations are derived using the method of Belinskii and Zakharov consequently generalizing their results to the higher-dimensional case. From the results of the one-soliton transformation, a series of discrete symmetries generalizing the Neugebauer–Kramer transformation in the four-dimensional case are obtained. The basic properties of these new transformations are studied.

I. INTRODUCTION

Recently, some progress has been made toward the solutions of vacuum Einstein equations in higher dimensions under the Kaluza–Klein ansatz.[1–7] First Sorkin[1] and Gross and Perry[2] discovered a monopole solution in five dimensions. Then we found that the field equations for spherically symmetric monopoles[3] and dyons[4] in this case can be completely integrated and explicit integrations were carried out. The complete integrability was generalized to $4+K$ dimensions by discovering a Lax form for the field equations.[5] The explicit integration in this case was accomplished later.[6] The spherically symmetric case represents solutions of vacuum Einstein equations in $4+K$ dimensions with $2+K$ *commuting* Killing vectors. It is natural to expect that the complete integrability might be extended to such general cases. In particular, we found[7] that the field equations for axially symmetric Abelian Kaluza–Klein monopoles and dyons can be separated into two sets. The first set can be integrated by quadrature once the second set is solved while the second set reduces to a generalized nonlinear sigma model,[8] which possesses dual symmetry. When K equals zero, i.e., in four dimensions, this dual symmetry is known[9] and is related to a Lax pair proposed by Belinskii and Zakharov (BZ)[10] as was shown by Cosgrove.[11] A Lax pair of the form of BZ is also valid in the general case when K is nonzero. It is then possible to work out soliton transformations following the method of BZ.[10,12] In the five-dimensional case, Belinsky and Ruffini[13] used the BZ two-soliton transformation to generate asymptotic flat solutions from the trivial background. They found that the solutions represent electric monopoles with magnetic dipole moments. Burzlaff and Maison[14] derived the linear eigenvalue problem by first constructing infinitely many conservation laws. In the series of works, we shall consider the generalization to arbitrary K. This may not be of immediate interest as a theory of higher-dimensional gravity but it is quite interesting when considered as a theory generalizing the nonlinear sigma models.[8] In particular, one may ask, for example, if the rather successful formulation of Hauser and Ernst[15] for the $K=0$ case can be generalized to higher dimensions and used for sigma models. One key ingredient in the Hauser–Ernst formulation is the Neugebauer–Kramer (NK) mapping.[16] Our main result in the present work is to show that this mapping admits generalizations to higher dimensions.[17] The generalized NK transformations will play the central role in the theory of soliton transformations for axially symmetric higher dimensional gravity and for sigma models. In Sec. II we shall write down the field equations and establish their relations to the generalized nonlinear sigma model. The notations follow closely those of Ref. 7. The dual symmetry transformation is then treated in detail. In Sec. III, we derive the BZ one-soliton transformation for all the relevant quantities so that iterations can be carried out readily. These two sections represent straightforward generalizations of known results.[10–14] We include them for completeness and for setting up our notations for subsequent works. We then point out the fact that the BZ one-soliton transformation can be written as successive transformations of a dual transformation followed by a certain discrete transformation and then a dual transformation again. This generalizes the known result of Cosgrove[11] for the vacuum Einstein equations in four dimensions. The discrete symmetries can then be identified as the generalized Neugebauer–Kramer (NK)[16] transformations. Explicit formulas for them are given in Sec. IV. In Sec. V, we derive the basic properties of the generalized NK transformations. A brief discussion is given in Sec. VI.

II. FIELD EQUATIONS AND DUAL SYMMETRY

We shall consider the following metric in $4+K$ dimensions:

$$\bar{g}=g_{\mu\nu}\,dx^{\mu}\otimes dx^{\nu}+\phi_{ab}(x)\theta^{a}\otimes\theta^{b}, \tag{1}$$

where

$$\theta^{a}=dy^{a}+A^{a}{}_{\mu}(x)dx^{\mu}. \tag{2}$$

The y^{a} are coordinates of the internal space that is isomorphic to an Abelian Lie group. For the four-dimensional part of the metric, we assume that

$$g_{\mu\nu}\,dx^{\mu}\,dx^{\nu}=f(d\rho^{2}+dz^{2})+h_{\bar{\mu}\bar{\nu}}\,dx^{\bar{\mu}}\,dx^{\bar{\nu}},\quad \bar{\mu}=t,\varphi, \tag{3}$$

where f and $h_{\bar{\mu}\bar{\nu}}$ are functions of ρ,z only. We shall represent ρ,z by $x^{\mu'}$.

The curvature tensor of the metric can be found in Ref. 18. The field equations are of the form of conserved currents. Explicitly, let us introduce the current[7]

 0022-2488/87/040893-08$02.50

$$J_{\mu'} = -\frac{\tau}{2}\begin{bmatrix} (\partial_{\mu'} h\, h^{-1})_{\bar\mu}{}^{\bar\nu} + \phi_{ab} h^{\bar\nu\bar\lambda} A^{a}{}_{\bar\mu} \partial_{\mu'} A^{b}{}_{\bar\lambda}\,, & X_{\mu'\bar\mu}{}^{b} \\ \phi_{ab} h^{\bar\mu\bar\nu} \partial_{\mu'} A^{b}{}_{\bar\mu}\,, & (\partial_{\mu'}\phi\phi^{-1})_{a}{}^{b} - \phi_{ac} h^{\bar\mu\bar\nu} A^{b}{}_{\bar\nu} \partial_{\mu'} A^{c}{}_{\bar\mu} \end{bmatrix}, \tag{4}$$

where

$$X_{\mu'\bar\mu}{}^{b} = -(\partial_{\mu'} h\, h^{-1})_{\bar\mu}{}^{\bar\nu} A^{b}{}_{\bar\nu} + (\partial_{\mu'}\phi\,\phi^{-1})_{a}{}^{b} A^{a}{}_{\bar\mu} + \partial_{\mu'} A^{b}{}_{\bar\mu} - \phi_{ac} h^{\bar\nu\bar\lambda} A^{a}{}_{\bar\mu} A^{b}{}_{\bar\nu}\, \partial_{\mu'} A^{c}{}_{\bar\lambda}\,. \tag{5}$$

Then the field equation $\overline{R}_{a\mu} = 0$ corresponds to the conservation of the lower left block of $J_{\mu'}$, while the equations $\overline{R}_{\bar\mu\bar\nu} = 0$ and $\overline{R}_{ab} = 0$ correspond to the conservation of the upper and lower diagonal blocks, respectively. The conservation of $X_{\mu'\bar\mu}{}^{b}$ follows from these equations as an algebraic identity. The remaining field equations $\overline{R}_{\mu'\nu'} = 0$ can be written as

$$\begin{aligned} &\partial_\xi^2 \tau - \partial_\xi \ln(-\bar g)^{1/2}\, \partial_\xi \tau + \tau^{-1} \operatorname{Tr} J_\xi^2 = 0\,, \\ &\partial_\eta^2 \tau - \partial_\eta \ln(-\bar g)^{1/2}\, \partial_\eta \tau + \tau^{-1} \operatorname{Tr} J_\eta^2 = 0\,, \end{aligned} \tag{6}$$

$$\partial_\xi\, \partial_\eta \ln(-\bar g)^{1/2} = -\tau^{-2} \operatorname{Tr} J_\xi\, J_\eta\,. \tag{7}$$

Here we have introduced the complex coordinates

$$\xi = z - i\rho\,, \quad \eta = z + i\rho$$

and the notations

$$(-\bar g)^{1/2} = (-\det \bar g)^{1/2}\,, \quad \tau = (-\det h \cdot \det \phi)^{1/2}\,. \tag{8}$$

The conservation of the trace of $J_{\mu'}$ implies

$$(\partial_\rho^2 + \partial_z^2)\tau = 0\,. \tag{9}$$

We shall introduce the conjugate harmonic function σ

$$\partial_z \tau = \partial_\rho \sigma\,, \quad \partial_\rho \tau = -\partial_z \sigma\,, \tag{10}$$

and w, $\overline{w}$ such that

$$w = \tau + i\sigma\,, \quad \overline{w} = \tau - i\sigma\,. \tag{11}$$

Note that w is an analytic function of η while $\overline{w}$ is an analytic function of ξ.

Once $J_{\mu'}$ is known, Eq. (6) can be used to get $(-\bar g)^{1/2}$ and hence f by quadrature. Equation (7) follows from (6) and the conservation of $J_{\mu'}$. Indeed, it is the consistency condition of Eq. (6).

To see the relation with sigma models, let us introduce the vielbein

$$\begin{aligned} &h_{\bar\mu\bar\nu} = e_{\bar\mu}{}^{\bar m} e_{\bar\nu}{}^{\bar m}\,, \quad (e^{-1})_{\bar m}{}^{\bar\mu} e_{\bar\mu}{}^{\bar n} = \delta_m{}^n\,, \\ &\phi_{ab} = e_a{}^m e_b{}^m\,, \quad (e^{-1})_m{}^a e_a{}^n = \delta_m{}^n\,, \end{aligned} \tag{12}$$

and the matrix G,

$$G = \begin{bmatrix} e_{\bar\mu}{}^{\bar m}\,, & A^{a}{}_{\bar\mu} e_a{}^m \\ 0\,, & e_a{}^m \end{bmatrix}. \tag{13}$$

Let σ denote the involutive automorphism on invertible matrices

$$\sigma(M) = (M^T)^{-1}\,, \quad \sigma^2 = 1\,, \tag{14}$$

then

$$Q = \sigma(G) G^{-1} = (GG^T)^{-1} \tag{15}$$

depends only on the metric, indeed,

$$Q = \begin{bmatrix} h^{\bar\mu\bar\nu}\,, & -h^{\bar\mu\bar\nu} A^{a}{}_{\bar\nu} \\ -h^{\bar\mu\bar\nu} A^{a}{}_{\bar\mu}\,, & \phi^{ab} + h^{\bar\mu\bar\nu} A^{a}{}_{\bar\mu} A^{b}{}_{\bar\nu} \end{bmatrix}. \tag{16}$$

The current $J_{\mu'}$ can be written as

$$J_{\mu'} = (\tau/2) Q^{-1} \partial_{\mu'} Q\,. \tag{17}$$

It follows that $J_{\mu'}$ satisfies the constraint

$$\partial_\xi J_\eta - \partial_\eta J_\xi + 2\tau^{-1}[J_\xi, J_\eta] - \tau^{-1}\partial_\xi \tau J_\eta + \tau^{-1}\partial_\eta \tau J_\xi = 0 \tag{18}$$

and the conservation equation

$$\partial_\xi J_\eta + \partial_\eta J_\xi = 0\,. \tag{19}$$

Equations (18) and (19) can be linearized. Following BZ,[10] we introduce the vector fields L_ξ, L_η:

$$\begin{aligned} L_\xi &= \partial_\xi + 2\lambda(\tau + \lambda)^{-1} \partial_\xi \tau\, \partial_\lambda\,, \\ L_\eta &= \partial_\eta + 2\lambda(\tau - \lambda)^{-1} \partial_\eta \tau\, \partial_\lambda\,. \end{aligned} \tag{20}$$

It is easy to see that

$$[L_\xi, L_\eta] = 0\,. \tag{21}$$

The Lax pair is

$$\begin{aligned} L_\xi \psi(\cdot;\lambda) &= 2(\tau + \lambda)^{-1} \psi(\cdot;\lambda)\, J_\xi\,, \\ L_\eta \psi(\cdot;\lambda) &= 2(\tau - \lambda)^{-1} \psi(\cdot;\lambda)\, J_\eta\,. \end{aligned} \tag{22}$$

From Eq. (17), it follows that we can choose the normalization

$$\psi(\cdot;0) = Q\,. \tag{23}$$

We now change variables from λ to s such that $s(\cdot;\lambda)$ satisfies

$$ds(L_\xi) = 0\,, \quad ds(L_\eta) = 0\,. \tag{24}$$

These equations can be integrated to give

$$s = f(\lambda + (w - \overline{w}) + \tau^2/\lambda)\,, \tag{25}$$

where $f(x)$ is an arbitrary smooth function. We choose $f(x) = x^{-1}$ so that

$$s = \lambda/\{\lambda^2 + (w - \overline{w})\lambda + \tau^2\}\,. \tag{26}$$

The Lax pair, Eq. (20), becomes in terms of ξ, η, s:

$$\begin{aligned} \partial_\xi U(\cdot;s) &= \tau^{-1}(1 + \gamma^{-1}) U(\cdot;s)\, J_\xi\,, \\ \partial_\eta U(\cdot;s) &= \tau^{-1}(1 + \gamma) U(\cdot;s)\, J_\eta\,, \end{aligned} \tag{27}$$

where

$$\psi(\cdot;\lambda) = U(\cdot;s)|_s \text{ as in (26)}\,, \tag{28}$$

$$\gamma^{-1} = (1 - 2sw)^{1/2}(1 + 2s\overline{w})^{-1/2}\,. \tag{29}$$

Note that λ and γ are double-valued functions of s so that $\lambda, \tau^2/\lambda$ correspond to the same s and so do γ, $-\gamma$. The value of $\psi(\cdot;\lambda)$, $U(\cdot;s)$ on the second sheet of the s surface will be denoted by $\tilde\psi(\cdot;\lambda)$, $\tilde U(\cdot;s)$ and they satisfy, respectively, Eqs. (22) and (27) with $\lambda \to \tau^2/\lambda$, $\gamma \to -\gamma$.

Equation (27) is the linearized equation expressing dual symmetry for the generalized nonlinear sigma models on symmetric spaces.[8,19] Indeed, once Eq. (27) is solved, a new solution is given by

$$G^{(s)} = U(\cdot;s)G\,, \quad \tau^{(s)} = \tau/S^2\,, \qquad S^2 = (1-2sw)(1+2s\bar{w}) \tag{30}$$

and

$$\begin{aligned} J_\xi{}^{(s)} &= -\tau^{-1}\tau^{(s)}\gamma^{-1}U(\cdot;s)\,J_\xi U(\cdot;s)^{-1} \\ &= \tfrac{1}{2}\partial_\xi(\partial_s U\,U^{-1})\,, \\ J_\eta{}^{(s)} &= -\tau^{-1}\tau^{(s)}\gamma U(\cdot;s)\,J_\eta U(\cdot;s)^{-1} \\ &= -\tfrac{1}{2}\partial_\eta(\partial_s U\,U^{-1})\,, \end{aligned} \tag{31}$$

with

$$w^{(s)} = w/(1-2sw)\,, \quad \bar{w}^{(s)} = \bar{w}/(1+2s\bar{w})\,. \tag{32}$$

The transformed U is given by

$$U^{(s)}(\cdot;s') \doteq \tilde{U}(\cdot;s'+s)U(\cdot;s)^{-1}\,, \tag{33}$$

where a dot means equal up to multiplication on the left by a constant matrix. The new Q follows from Eqs. (15) and (30) and also from Eqs. (23) and (33):

$$Q^{(s)} = \sigma(U(\cdot;s))QU(\cdot;s)^{-1} \doteq \tilde{U}(\cdot;s)U(\cdot;s)^{-1}\,, \tag{34}$$

which is manifestly symmetric.

In going from Eq. (30) to (32), we have fixed an arbitrary additive constant. By fixing this constant, we have made the Abelian character of the dual symmetry transformation manifest as is clear from Eq. (33), which can be used to generate successive dual transformations. In the four-dimensional case (i.e., $K=0$), this dual symmetry transformation corresponds to Neugebauer's I_2 transformation, or, more precisely, Cosgrove's $\tilde{Q}_{4s}$ transformation.[11,20] Let us also note that for successive dual transformations, we can choose either sheet of the s surface at each stage so that there are two variants of Eq. (34) and four variants of Eq. (33). We shall consistently use a tilde to denote quantities on the second sheet.

Note that the last equality in Eq. (31) gives the potential for the conserved currents $J_{\mu'}{}^{(s)}$. If we expand $J_{\mu'}{}^{(s)}$ around $s=0$, we get infinitely many nonlocal conserved currents with the corresponding potentials given by

$$\tfrac{1}{2}\partial_s U\,U^{-1} = R_0 + 2sR_1 + 2s^2R_2 + \cdots\,, \tag{35}$$

$$U = (1 + 2sR_0 + 2s^2(R_0^2 + R_1) + \cdots)Q\,. \tag{36}$$

The R_i's will be called the "higher potentials." They are introduced mainly for later use.

From Eqs. (8), (13), and (15), we find that

$$\det Q = -1/\tau^2\,. \tag{37}$$

However, $\det Q^{(s)}$ does not satisfy the corresponding constraint. Indeed, it follows from the trace of Eq. (27) that

$$\det U(\cdot;s) \doteq s/\lambda \tag{38}$$

so that

$$\det Q^{(s)} \doteq -\lambda^2/s^2\tau^2\,. \tag{39}$$

A solution that satisfies the proper constraint can be obtained by renormalization:

$$\hat{Q}^{(s)} = (sS^2/\lambda)^{2/n}Q^{(s)}\,, \quad n = K+2\,, \tag{40}$$

with the corresponding change

$$\begin{aligned} \hat{J}_\xi{}^{(s)} &= J_\xi{}^{(s)} - (1/n)\tau^{(s)}\tau^{-1}\gamma_s^{-1}(1+\gamma_s^{-1})\,, \\ \hat{J}_\eta{}^{(s)} &= J_\eta{}^{(s)} - (1/n)\tau^{(s)}\tau^{-1}\gamma_s(1+\gamma_s)\,, \end{aligned} \tag{41}$$

where the subscript s for γ means evaluating at s. Denoting the transformed γ by $\gamma^{(s)}$, we find from Eqs. (29) and (32)

$$\gamma_{s'}^{(s)} = \gamma_s^{-1}\gamma_{s+s'}\,. \tag{42}$$

Similarly, we find the transformation law for S to be

$$S_s^{(s)2} = S_s^{-2}S_{s+s'}^2\,. \tag{43}$$

The renormalization necessary for $U^{(s)}$ is found to be

$$\hat{U}^{(s)}(\cdot;s) \doteq \{2sS(\gamma_s+\gamma_{s+s'})/(1-\gamma_s) \times (1-\gamma_{s+s'})\}^{2/n}U^{(s)}(\cdot;s')\,. \tag{44}$$

Substituting Eq. (41) into Eq. (6), we find that

$$(-\bar{g}^{(s)})^{1/2} \doteq (-\bar{g})^{1/2}(\lambda^2/s^2S)^{2/n}S^{-4}\,. \tag{45}$$

This completes our derivation of the dual symmetry transformation. The formulas are clearly ready for iteration. Apart from the renormalization factor, the symmetry is essentially Abelian. As we shall see in the following section, dual symmetry is closely related to the BZ soliton transformation.

III. SOLITON TRANSFORMATIONS

Given Q, $J_{\mu'}$, and $\psi(\cdot;\lambda)$ satisfying Eq. (22), we seek a new solution $\psi'(\cdot;\lambda)$ by dressing ψ[10,12]:

$$\psi'(\cdot;\lambda) = \psi(\cdot;\lambda)\chi(\cdot;\lambda)\,, \quad Q' = \psi'(\cdot;0)\,. \tag{46}$$

It is clear that Q' in general need not be symmetric. How can we get a symmetric Q'? Note that the symmetry requirement of Q can be expressed as

$$\sigma(Q)Q = 1\,. \tag{47}$$

If we normalize ψ as in Eq. (23) and normalize $\tilde{\psi}(\cdot;\lambda)$ by

$$\tilde{\psi}(\cdot;0) = 1\,, \tag{48}$$

then it is clear that to ensure condition (45), we have only to require

$$\sigma(\psi(\cdot;\lambda))Q = \tilde{\psi}(\cdot;\lambda)\,. \tag{49}$$

That this is possible can be seen from the second equality of Eq. (34). Requiring ψ', $\tilde{\psi}'$ to satisfy Eqs. (48) and (49) is equivalent to requiring

$$\tilde{\chi}(\cdot;0) = 1 \tag{50}$$

and

$$\sigma(\chi(\cdot;\lambda))Q' = Q\tilde{\chi}(\cdot;\lambda)\,. \tag{51}$$

The dressing matrix $\chi(\cdot;\lambda)$ in general can be represented as a sum over discrete poles on the λ plane together with a line integral over a closed curve.[12] When only the pole terms exist, the transformation is called pure solitonic. For pure soliton transformations, we write

$$\begin{aligned} \chi(\cdot;\lambda) &= 1 + \sum_{i=1}^{N}\frac{Q_i(\cdot)}{(\lambda-\lambda_i)}\,, \\ \chi(\cdot;\lambda)^{-1} &= 1 + \sum_{i=1}^{M}\frac{R_i(\cdot)}{(\lambda-\mu_i)}\,. \end{aligned} \tag{52}$$

In general, N need not equal M, but the constraint (51) requires that they be equal and form pairs:

$$\mu_i = \tau^2/\lambda_i\,, \quad i = 1,\ldots,N\,. \tag{53}$$

Substituting Eqs. (46) and (52) into Eq. (22), and requiring the residues of the double poles at $\lambda = \lambda_i, \lambda = \mu_i$ to vanish, we find that λ_i, μ_i have to satisfy the partial differential equation

$$\partial_\xi y = 2y(\tau + y)^{-1}\partial_\xi \tau\,, \quad \partial_\eta y = 2y(\tau - y)^{-1}\partial_\eta \tau\,, \tag{54}$$

which can be integrated so that

$$\lambda_i + (w - \bar{w}) + \tau^2/\lambda_i = 1/s_i \tag{55}$$

and μ_i satisfies the same equation with the same constant s_i, following from Eq. (53). It is now clear why we made the particular choice when going from Eq. (25) to Eq. (26).

In the following, we shall consider the case $N = 1$. The general N-soliton transformations can be obtained by iteration. Their explicit formulas will be presented in a subsequent work. The one-soliton transformation that we shall consider preserves the symmetric tensor character of Q but need not preserve reality. The reality condition is easy to satisfy and we shall comment on this later.

Putting $N = M = 1$ in Eq. (52) and multiplying the two expressions, we find the algebraic constraints

$$\begin{aligned} &R_1 = -Q_1\,, \quad Q_1 = Q_1^2/(\lambda_1 - \mu_1)\,, \\ &R_1 = R_1^2/(\mu_1 - \lambda_1)\,, \end{aligned} \tag{56}$$

which imply

$$P_1 = Q_1/(\lambda_1 - \mu_1) = R_1/(\mu_1 - \lambda_1)\,, \quad P_1^2 = P_1\,, \tag{57}$$

is a projection operator. Using P_1, we can write

$$\chi(\cdot;\lambda) = 1 - P_1 + (\lambda - \mu_1)(\lambda - \lambda_1)^{-1}P_1\,. \tag{58}$$

Let the rank of P_1 be q_1, then we have

$$\operatorname{Tr} P_1 = q_1\,, \quad \det\chi(\cdot;\lambda) = (\lambda - \mu_1)^{q_1}(\lambda - \lambda_1)^{-q_1}\,. \tag{59}$$

The matrix P_1 can be written as

$$\begin{aligned} &P_1 = X_1(F_1^+ X_1)^{-1}F_1^+\,, \\ &X_1, F_1 = \text{complex } n \times q_1 \text{ matrices}\,. \end{aligned} \tag{60}$$

From the partial differential equations satisfied by $\chi(\cdot;\lambda)$, we find that P_1 satisfies

$$\begin{aligned} &(1 - P_1)(\partial_\xi + 2(\tau + \lambda_1)^{-1}J_\xi)P_1 = 0\,, \\ &(1 - P_1)(\partial_\eta + 2(\tau - \lambda_1)^{-1}J_\eta)P_1 = 0\,, \end{aligned} \tag{61}$$

and

$$\begin{aligned} &P_1(\partial_\xi - 2(\tau + \mu_1)^{-1}J_\xi)(1 - P_1) = 0\,, \\ &P_1(\partial_\eta - 2(\tau - \mu_1)^{-1}J_\eta)(1 - P_1) = 0\,. \end{aligned} \tag{62}$$

It follows that we may require

$$\begin{aligned} &(\partial_\xi + 2(\tau + \lambda_1)^{-1}J_\xi)X_1 = 0\,, \\ &(\partial_\eta + 2(\tau - \lambda_1)^{-1}J_\eta)X_1 = 0\,, \end{aligned} \tag{63}$$

so that

$$X_1 = \psi^{-1}(\cdot;\lambda_1)X_1^{(0)} \tag{64}$$

for some constant matrix $X_1^{(0)}$.

To find the remaining part of P_1, we use Eq. (51). Requiring Q' to have no poles on the λ plane, we find

$$P_1^T Q(1 - P_1) = 0\,, \quad P_1^T Q = QP_1\,, \tag{65}$$

which can be solved for $(F_1^+ X_1)^{-1}F_1^+$ to give

$$P_1 = X_1(X_1^T Q X_1)^{-1}X_1^T Q\,. \tag{66}$$

Using Eqs. (34) and (64), we find

$$P_1 = \psi^{-1}(\cdot;\lambda_1)P_1^{(s_1)}\psi(\cdot;\lambda_1)\,, \tag{67}$$

where

$$P_1^{(s_1)} = X_1^{(0)}(X_1^{(0)T}Q^{(s_1)}X_1^{(0)})^{-1}X_1^{(0)T}Q^{(s_1)}\,. \tag{68}$$

Note that from Eqs. (67) and (34) one can see that Eq. (62) is satisfied.

It follows from Eqs. (46) and (58) that

$$\begin{aligned} \psi'(\cdot;\lambda) = \psi(\cdot;\lambda)\psi(\cdot;\lambda_1)^{-1} \\ \times\left(1 - P_1^{(s_1)} + \frac{\lambda - \mu_1}{\lambda - \lambda_1}P_1^{(s_1)}\right)\psi(\cdot;\lambda_1) \end{aligned} \tag{69}$$

and

$$\begin{aligned} \hat{Q}' = \left(\frac{\lambda_1}{\mu_1}\right)^{q_1/n} Q\psi(\cdot;\lambda_1)^{-1} \\ \times\left(1 - P_1^{(s_1)} + \frac{\mu_1}{\lambda_1}P_1^{(s_1)}\right)\psi(\cdot;\lambda_1)\,, \end{aligned} \tag{70}$$

where proper renormalization has been imposed on Q'. The renormalized ψ' is then

$$\hat{\psi}'(\cdot;\lambda) = ((\lambda - \lambda_1)/(\lambda - \mu_1))^{q_1/n}\psi'(\cdot;\lambda)\,. \tag{71}$$

The transformation law for the current J_μ, follows from the differential equations for $\chi(\cdot;\lambda)$. By examining the behavior of these equations near $\lambda = \pm\tau$, we find

$$\begin{aligned} \hat{J}'_\xi &= \chi(\cdot;-\tau)^{-1}\left[J_\xi + \tau\partial_\xi\tau\frac{\mu_1 - \lambda_1}{(\tau + \lambda_1)(\tau + \mu_1)}\left(\frac{q_1}{n} - P_1\right)\right]\chi(\cdot;-\tau)\,, \\ \hat{J}'_\eta &= \chi(\cdot;\tau)^{-1}\left[J_\eta - \tau\partial_\eta\tau\frac{\mu_1 - \lambda_1}{(\tau - \lambda_1)(\tau - \mu_1)}\left(\frac{q_1}{n} - P_1\right)\right]\chi(\cdot;\tau)\,. \end{aligned} \tag{72}$$

Substituting Eq. (72) into Eq. (6) and using the relations

$$\begin{aligned} \operatorname{Tr}(X_1^T Q X_1)^{-1}\partial_\xi(X_1^T Q X_1) &= -\frac{2}{\tau}\frac{\tau - \lambda_1}{\tau + \lambda_1}\operatorname{Tr} J_\xi P_1\,, \\ \operatorname{Tr}(X_1^T Q X_1)^{-1}\partial_\eta(X_1^T Q X_1) &= -\frac{2}{\tau}\frac{\tau + \lambda_1}{\tau - \lambda_1}\operatorname{Tr} J_\eta P_1\,, \end{aligned} \tag{73}$$

which follow from Eq. (63), we obtain

$$\begin{aligned} (-\hat{\bar{g}}')^{1/2} = (-\bar{g})^{1/2}\det(X_1^{(0)T}\hat{Q}^{(s_1)}X_1^{(0)}) \\ \times(\tau/S_1^2)^{q_1(n - q_1 + 2)/n}\,, \end{aligned} \tag{74}$$

where

$$S_1^2 = (1 - 2s_1 w)(1 + 2s_1\bar{w})\,. \tag{75}$$

Equations (70), (71), and (74) represent the full results for

one-soliton transformation that can be iterated any number of times.

Now we shall say a few words about the condition for the metric to be real. From Eqs. (22) and (27), one can see that to ensure reality, we can require

$$\psi^*(\cdot;\lambda)=\psi(\cdot;-\lambda)\,,\quad U^*(\cdot;s)=U(\cdot;-s)\,. \tag{76}$$

It follows that for $Q^{(s)}$ in Eq. (34) to be real, we have to choose s to be pure imaginary. For soliton transformations, insisting condition (76) for $\psi'(\cdot;\lambda)$ and hence $\chi(\cdot;\lambda)$, we see from Eq. (52) that if λ_i is a pole, $-\lambda_i^*$ must also be a pole. This also implies, from Eq. (55), that the constants s_i chosen must form $(s_i,-s_i^*)$ pairs unless s_i is pure imaginary. Moreover, for $(s_i,-s_i^*)$ pairs, the corresponding Q_i, R_i have to form $(Q_i,\ -Q_i^*)$ and $(R_i,\ -R_i^*)$ pairs. For one-soliton transformation, the only choice is to have s_1,λ_1, Q_1 pure imaginary so that P is real. This can be achieved by requiring

$$X^{(0)*}=cX^{(0)}\,, \tag{77}$$

where c is an arbitrary constant.

Finally, let us make the following observation. In Eq. (69) for BZ one-soliton transformation, we replace λ by $\mu=\tau^2/\lambda$ and notice that

$$\frac{\mu-\mu_1}{\mu-\lambda_1}=-(\gamma_{s_1}+1)(\gamma_{s_1}-1)^{-1}\times(\gamma_{s-s_1}^{(s_1)}-1)(\gamma_{s-s_1}^{(s_1)}+1)^{-1}\,. \tag{78}$$

Equation (69) can then be written as

$$\tilde{\psi}'(\cdot;\lambda)=\Phi(\cdot;s-s_1)\Phi(\cdot;-s_1)^{-1}\,, \tag{79}$$

where

$$\Phi(\cdot;s-s_1)=\tilde{U}(\cdot;s)U(\cdot;s_1)^{-1}\times\left(1-P_1^{(s_1)}+\frac{\gamma_{s-s_1}^{(s_1)}-1}{\gamma_{s-s_1}^{(s_1)}+1}P_1^{(s_1)}\right) \tag{80}$$

and we have used the relation

$$\gamma_s^{-1}=\gamma_{-s}^{(s)}\,, \tag{81}$$

which follows from Eq. (42).

Recalling the definition of $P_1^{(s_1)}$ in Eq. (68) and comparing Eqs. (79) and (80) with the dual symmetry transformation Eq. (33) and its variants, we see that the BZ one-soliton transformation consists of successive transformations of the form $I_2 I^{(q_1)} I_2$, where I_2 is a dual symmetry transformation and $I^{(q_1)}$ is a certain discrete transformation that can be read off from Eq. (80). Let us recall that for the four-dimensional ($K=0$) case, Cosgrove[11] showed that the BZ one-soliton transformation can be written as $I_2 I I_2$, where I is the Neugebauer–Kramer[16] transformation. Hence we see that the Neugebauer–Kramer symmetry generalizes to a series of discrete symmetries.[17] In the following section, we derive the formulas for these symmetry transformations.

IV. GENERALIZED NEUGEBAUER–KRAMER MAPPING

The discussions at the end of the previous section suggest that we seek a new solution by writing the new generating function $U_1(\cdot;s)$ in the form

$$U_I(\cdot;s)=A(s)\tilde{U}(\cdot;s)[1-P+(\gamma+1)(\gamma-1)^{-1}P]Y^{-1}(\cdot)\,, \tag{82}$$

where $Y(\cdot)$ is a matrix function of ξ, η, undetermined from the soliton transformation. Here $A(s)$ is an invertible matrix function of s only and P is a projection operator defined by

$$P=X_0(X_0^TQX_0)^{-1}X_0^TQ\,, \tag{83}$$

where X_0 is a constant complex $n\times q$ matrix of rank q with $1\leqslant q<n$.

As before, Q_I will be given by $U_I(\cdot,0)$. From Eq. (82), we find that in order for Q_I to exist, we must require

$$A(0)P=0\,. \tag{84}$$

This is achieved by the minimum choice

$$A(s)=1-P_0+2sP_0\,, \tag{85}$$

where

$$P_0=X_0(X_0^Tk_0X_0)^{-1}X_0^Tk_0 \tag{86}$$

with k_0 an arbitrary invertible matrix. This choice is minimum in the sense that any other choice can be obtained by multiplying $A(s)$ on the left by another matrix $B(s)$ which is analytic and invertible at $s=0$. Such modifications, as we shall see, affect the formulas for Q_I only trivially.

As in the case of soliton transformation, to ensure that Q_I be symmetric, we shall require the validity of Eqs. (48) and (49) for $\tilde{U}_I, U_I$. This together with the fact that $\tilde{U}_I$ is the analytic continuation of U_I onto the second sheet so that it is essentially determined by U_I with the change $\gamma\to-\gamma$, $\tilde{U}\to U$, we obtain

$$\tilde{U}_I(\cdot;s)=\sigma(A(s))U(\cdot;s)\left(1-P+\frac{\gamma-1}{\gamma+1}P\right)Y^{-1}(\cdot) \tag{87}$$

and

$$Q_I=\sigma(Y)QY^{-1} \tag{88}$$

with $Y(\cdot)$ given by

$$Y(\cdot)=\lim_{s\to0}\left\{\sigma(A(s))U(\cdot;s)\left(1-P+\frac{\gamma-1}{\gamma+1}P\right)\right\}=[1-P^T+\tfrac{1}{2}\tau P_0^T+P_0^TR_0(1-P^T)]Q\,, \tag{89}$$

where R_0 is the first potential defined in Eq. (35). The inverse of $Y(\cdot)$ is easily found to be

$$Y^{-1}(\cdot)=Q^{-1}[1-P_0^T+2\tau^{-1}P^T-2\tau^{-1}P^TR_0(1-P_0^T)]\,. \tag{90}$$

One can check by straightforward computation that

$$\partial_\xi U_I(\cdot;s)=\tau^{-1}(1+\gamma^{-1})U_I(\cdot;s)\,J_{I\xi}\,,\qquad \partial_\eta U_I(\cdot;s)=\tau^{-1}(1+\gamma)U_I(\cdot;s)\,J_{I\eta}\,, \tag{91}$$

with

$$J_{I\xi}=-Y[(1-P)J_\xi(1-P)+PJ_\xi P-PJ_\xi(1-P)-(1-P)J_\xi P+\partial_\xi\tau P]Y^{-1}\,,\qquad J_{I\eta}=-Y(J_\eta+\partial_\eta\tau P)Y^{-1}\,. \tag{92}$$

From Eq. (6), we find the transformation formula for $\bar{g}$:

$$(-\bar{g}_I)^{1/2}=(-\bar{g})^{1/2}\tau^4\det(X_0^TQX_0)\,. \tag{93}$$

The lack of symmetry between ξ and η in Eq. (92) shows the complex character of generalized NK transformations. Indeed, if we call the above transformation $I^{(q)}$, then there exists the conjugate transformation $I_*^{(q)}$ given by

$$U_{I_*}(\cdot;s) = A_*(s)\tilde{U}(\cdot;s) \times [1-P-(\gamma+1)(\gamma-1)^{-1}P]Y_*^{-1}(\cdot) \tag{94}$$

choosing

$$A_*(s) = 1 - P_0 - 2sP_0, \tag{95}$$

we have

$$Y_*(\cdot) = \lim_{s\to 0}\sigma(A_*(s))U(\cdot;s)\left(1-P-\frac{\gamma-1}{\gamma+1}P\right) = [1-P^T+\tfrac{1}{2}\tau P_0^T - P_0^T R_0(1-P^T)]Q. \tag{96}$$

The currents now transform as

$$\begin{aligned} J_{I_*\xi} &= -Y_*(J_\xi + \partial_\xi \tau P)Y_*^{-1}, \\ J_{I_*\eta} &= -Y_*[(1-P)J_\eta(1-P) + PJ_\eta P \\ &\quad - PJ_\eta(1-P) - (1-P)J_\eta P + \partial_\eta \tau P]Y_*^{-1}. \end{aligned} \tag{97}$$

If we change $A(s)$ by multiplying the minimum form on the left by $B(s)$, $Y(\cdot)$ will be modified to $B(0)Y(\cdot)$. This amounts to applying an additional symmetry transformation, namely, the global rotation of the $K+2$ commuting Killing vectors by $B(0)$.

Notice that the potential R_0 appears in the generalized NK transformations. For completeness and for later use, we give the transformation law for R_0, which follows from Eq. (82) and the definition of the potential,

$$\begin{aligned} R_{I0} &= P_0(P\tilde{R}_0 - i\sigma)P_0 + (1-P_0)\tilde{R}_0(1-P)(1-P_0) \\ &\quad - (1-P_0)[i\sigma\tilde{R}_0 + \tfrac{1}{2}(\tilde{R}_0^2 - \tilde{R}_1) - \tilde{R}_0 P\tilde{R}_0] \\ &\quad \times P_0 - P_0 P(1-P_0), \end{aligned} \tag{98}$$

where σ is defined in Eq. (10) and $\tilde{R}_0$, $\tilde{R}_1$ are potentials on the second sheet. In fact, due to Eq. (47), we have the relations

$$\tilde{R}_i = -R_i^T \quad \forall i. \tag{99}$$

The generalized NK transformations $I^{(q)}(k_0,X_0)$ depends on the matrix parameters k_0 and X_0. One can show that the inverse transformation is

$$I^{(q)}(k_0,X_0)^{-1} = I^{(q)}(k_0^{-1},k_0^T X_0) \tag{100}$$

so that $I^{(q)}$ is involutive when $k_0 = 1$. Although we have "derived" the NK transformations from the BZ one-soliton transformations, the exact relation between them requires more careful treatment and will be given in the next section.

V. BASIC PROPERTIES OF NK TRANSFORMATIONS

Let us begin by introducing some obvious symmetry transformations. Let Q be a solution and $U(\cdot;s)$ the corresponding generating function. Then

$$Q\to\sigma(Q) = Q^{-1}, \quad U\to\sigma(U), \quad \tilde{U}\to\sigma(\tilde{U}) \tag{101}$$

is also a solution. We shall also use σ to denote this symmetry transformation despite possible confusions. From Eqs. (33) and (34), we can see that

$$I_2^{(0)} = \sigma. \tag{102}$$

It is convenient to define $\tilde{I}_2^{(s)}$ by

$$\tilde{I}_2^{(s)} = \sigma I_2^{(s)}. \tag{103}$$

Similarly, we shall define

$$\hat{I}^{(q)}(k_0,X_0) = \sigma I^{(q)}(k_0,X_0). \tag{104}$$

Let $C\in \mathrm{SL}(n,R)$, $n = K+2$. Then we have the symmetry transformation $R(C)$:

$$\begin{aligned} &R(C)Q = \sigma(C)QC^{-1}, \quad R(C)U = \sigma(C)UC^{-1}, \\ &R(C)\tilde{U} = C\tilde{U}C^{-1}. \end{aligned} \tag{105}$$

Finally, let $G(s)\in \mathrm{GL}(n,C)$ be a function of s only and $G(0) = 1$. Then we have the trivial gauge transformation $L(G)$,

$$L(G)U = G(s)U(\cdot;s), \quad L(G)\tilde{U} = \sigma(G)\tilde{U}, \tag{106}$$

which leaves Q unchanged. There is another symmetry transformation which arises from projective transformations of ρ, z coordinates treating Q and U as scalar functions of ρ, z. This symmetry transformation is important for some purposes but will not be considered here.

The following commutation relations among these symmetry transformations can be easily derived:

$$\sigma\tilde{I}_2^{(s)} = \tilde{I}_2^{(s)}\sigma, \tag{107}$$

$$R(C)\tilde{I}_2^{(s)} = \tilde{I}_2^{(s)}R(C), \tag{108}$$

$$R(C)\cdot\sigma = \sigma\cdot R(\sigma(C)), \tag{109}$$

$$R(C)\hat{I}^{(q)}(k_0,X_0) = \hat{I}^{(q)}(\sigma(C)k_0C^{-1},CX_0))R(C), \tag{110}$$

$$\sigma\hat{I}^{(q)}(k_0,X_0) = \hat{I}^{(q)}(k_0^{-1},\sigma(k_0)^{-1}X_0)^{-1}\sigma, \tag{111}$$

$$\tilde{I}_2^{(s_1)}L(G(s)) = L(G(s+s_1)G(s_1)^{-1})R(\sigma(G(s_1)))\tilde{I}_2^{(s_1)}, \tag{112}$$

$$L(G(s))\tilde{I}_2^{(s_1)} = \tilde{I}_2^{(s_1)}L(G(s-s_1)G(-s_1)^{-1}) \times R(\sigma(G(-s_1))), \tag{112'}$$

$$R(C)L(G(s)) = L(\sigma(C)G\sigma(C)^{-1})R(C), \tag{113}$$

$$\sigma L(G(s)) = L(\sigma(G(s))\sigma, \tag{114}$$

$$\hat{I}^{(q)}L(G(s)) = L(\sigma(A(s))G(s)\sigma(A(s))^{-1}\sigma(C)^{-1}) \times R(C)\hat{I}^{(q)}, \tag{115}$$

$$\sigma(\mathrm{C}) = \lim_{s\to 0}\sigma(A(s))G(s)\sigma(A(s))^{-1}.$$

Now we may give the exact relations between the BZ one-soliton transformation $I_{\mathrm{BZ}}^{(s_1)}(X_0)$ and the NK transformations $\hat{I}^{(q)}(k_0,X_0)$:

$$\sigma I_{\mathrm{BZ}}^{(s_1)}(X_0)\sigma = R(\sigma(A(-s_1)^{-1})L(A(-s_1)A(s-s_1)^{-1}) \times \tilde{I}_2^{(-s_1)}\hat{I}^{(q)}(k_0,X_0)\tilde{I}_2^{(s_1)}. \tag{116}$$

Note that the BZ one-soliton transformation is singular at $s_1 = 0$ and $s = s_1$. The $\mathrm{SL}(n,R)$ factor and the gauge transformation removes the singularities so that $\tilde{I}_2^{(-s_1)}\hat{I}^{(q)}\tilde{I}_2^{(s_1)}$ is regular at these points. Aside from these factors, we may think of $\hat{I}^{(q)}$ as the basic one-soliton transformation while

$\tilde{I}_2^{(s_1)}$ merely translates the position of the pole in the s plane.

Because of Eq. (110), we have only to consider the case $k_0 = 1$. We shall therefore set $k_0 = 1$ and omit k_0 from the arguments of $\hat{I}^{(q)}$. As we have noted in the previous section, $I^{(q)}(X_0)$ is involutive

$$I^{(q)}(X_0)^2 = 1 . \tag{116'}$$

On the other hand, $\hat{I}^{(q)}(X_0)$ can be raised to any power. We shall first show that it is enough to consider the rank one case, i.e., $q = 1$. We shall omit the superscript q from $\hat{I}^{(q)}$ when the rank of X_0 is clear or is irrelevant.

Let us consider two successive rank one $\hat{I}$ transformations. Recall that $Q = \sigma(G)G^{-1}$ from Eq. (15). We shall choose a gauge so that G satisfies

$$(1 - P_0)GP_0 = 0 , \tag{117}$$

which implies

$$(1 - P_0)G^{-1}P_0 = 0 , \quad PG = GP_0 , \quad G^{-1}P = P_0G^{-1} . \tag{118}$$

In proving these, we used the identity

$$(X_0^TX_0)^{-1}(X_0^TGX_0)(X_0^TX_0)^{-1} = (X_0^TG^{-1}X_0)^{-1} , \tag{119}$$

which is valid in this gauge.

Equation (88) implies that we may choose

$$G_I = YG = (1 - P_0)\sigma(G)(1 - P_0) + (\tau/2)P_0\sigma(G)P_0 + P_0R_0(1 - P_0)\sigma(G)(1 - P_0) , \tag{120}$$

which again is in the gauge of Eq. (117). The NK transformation given by Eq. (82) can now be expressed as

$$U_{\hat{I}}G_{\hat{I}} = \sigma(A(s))(UG)\left(1 - P_0 + \frac{\gamma - 1}{\gamma + 1}P_0\right) . \tag{121}$$

Now consider two $\hat{I}$ transformations denoted as $\hat{I}_1$, $\hat{I}_2$ with the corresponding projection operators P_{01}, P_{02} satisfying

$$P_{01}P_{02} = P_{02}P_{01} = 0 \tag{122}$$

so that $P_{01} + P_{02}$ is a rank-2 projection operator.

From Eq. (120), we find that

$$G_{\hat{I}_1} = (1 - P_{01})G(1 - P_{01}) + (2/\tau)(1 - P_{01})\tilde{R}_0GP_{01} + (2/\tau)P_{01}GP_{01} . \tag{123}$$

In order to apply Eq. (121), $G_{\hat{I}_1}$ has to be chosen in the gauge

$$(1 - P_{02})G_{\hat{I}_1}P_{02} = (1 - P_{01} - P_{02})GP_{02} = 0 . \tag{124}$$

This can always be done and we find

$$\hat{I}(P_{02})\hat{I}(P_{01}) = \hat{I}(P_{01} + P_{02}) . \tag{125}$$

The iterations can be continued. We have to choose G to satisfy

$$\left(1 - \sum_{i=1}^{m} P_{0i}\right)GP_{0m} = 0 \quad \forall m = 1,...,n - 1 . \tag{126}$$

Let e_i be an orthonormal basis for R^n and $P_{0i} = e_ie_i^T$. In this basis, G satisfies Eq. (126) if and only if it is upper triangular. Carrying out $n - 1$ iterations we find

$$\begin{aligned} &\hat{I}(e_n)\cdots\hat{I}(e_1)(UG) = (\lambda/2s\tau)UG , \\ &\hat{I}(e_n)\cdots\hat{I}(e_1)(\tilde{U}G) = (2s\tau/\lambda)\tilde{U}G . \end{aligned} \tag{127}$$

The factor $(\lambda/2s\tau)$ can be renormalized away. It appears because we have not carried out the proper renormalization in the formulas (82) and (87), but this can always be done in the last step of successive transformations. With this in mind, we can write Eqs. (125) and (127) as

$$\hat{I}(e_i)\hat{I}(e_j) = \hat{I}(e_j)\hat{I}(e_i) = \hat{I}(e_ie_j) , \quad i \neq j , \tag{128}$$

$$\hat{I}(e_n)\cdots\hat{I}(e_1) = 1 . \tag{129}$$

Note that we have used e_i to represent P_{0i} and e_ie_j to represent $P_{0i} + P_{0j}$. In general we shall use $e_i\cdots e_j$ to represent the projection operator on to the subspace spanned by $e_i,...,e_j$. Combining Eqs. (111), (128), and (129) we find

$$\hat{I}(e_i)^{-1} = \sigma\hat{I}(e_i)\sigma = \hat{I}(*e_i) , \tag{130}$$

where

$$*e_i = e_1...e_{i-1}e_{i+1}...e_n . \tag{131}$$

It is clear that the transformations

$$(\hat{I}(e_i))^m , \quad m \in Z ,$$

are the basic ingredients of soliton transformations. Defining

$$I_m(e_i) = \sigma(\hat{I}(e_i))^m , \quad m \in Z , \tag{132}$$

we find easily that

$$I_m(e_i)^2 = 1 , \quad m \in Z . \tag{133}$$

The explicit formulas for $\hat{I}(e_i)^m$ and $I_m(e_i)$ are quite complicated. They are still under investigation. On the other hand, although the BZ soliton transformations have singularities on the s plane, their iterations for any number of times are relatively easy to write down. These formulas are derived in a subsequent work.[21]

VI. DISCUSSION

In this work, we first established that finding the metric tensor satisfying vacuum Einstein equations in $4 + K$ dimensions and admitting $2 + K$ commuting Killing vectors is equivalent to solving a generalized nonlinear sigma model. We then derived the formulas for BZ one-soliton transformations with proper normalization and with a formula for the conformal factor f appearing in Eq. (3). These formulas can be iterated an arbitrary number of times. The result is given in a subsequent work. From BZ one-soliton transformations, we derived a series of involutive transformations $I^{(q)}(X_0)$ generalizing the NK transformation in the four-dimensional case. The accompanying transformations $\hat{I}(e_i)$ are shown to form the basic ingredients of BZ soliton transformations. The spectral dependence of the BZ soliton transformations is generated by conjugating $\hat{I}(e_i)$ with the dual symmetry transformation $\tilde{I}_2^{(s)}$.

As we mentioned in Sec. III, given a solution of the sigma model, new solutions can be generated by the method of dressing proposed by Zakharov and Shabat.[12] Finding the dressing matrix is equivalent to solving the matrix Riemann–Hilbert problem. The general solution of the dressing matrix can be written in two parts. The first part arises from solving the regular Riemann–Hilbert problem, which amounts to solving a singular integral equation. The second part is the pure solitonic part, which arises from solving the singular Riemann–Hilbert problem and can be obtained by a

purely algebraic method. It is known that in the four-dimensional case, all solutions can be generated from the trivial solution in this way.[15] It is believed that this continues to be true in the higher-dimensional case. One interesting question is whether one can solve the singular integral equation and obtain the "continuum" part of the solution also by a purely algebraic method. The answer is believed to be yes but the actual construction of finite transformations is usually difficult.

The solution space decomposes into sectors labeled by soliton numbers. It is believed that an infinite dimensional Lie group acts on the vacuum sector and the corresponding Lie algebra is a Kac–Moody algebra.[22] Except in the four-dimensional case, the derivation is based on infinitesimal Riemann–Hilbert transformations, which is not practically useful since we do not know how to solve matrix singular integral equations in close form except in special cases. On the other hand, elements of the group are easily written down in terms of soliton transformations. For example, $I_m(e_i)\tilde{I}_2^{(s)}I_m(e_i)$ gives infinitely many (labeled by m) one parameter (s) families of group elements. Investigation of the group structure along this direction is in progress.

ACKNOWLEDGMENTS

I wish to thank S.-L. Lou for help in part of the calculations. This work is supported in part by a grant from the National Science Council, Taiwan, Republic of China.

[1]R. Sorkin, Phys. Rev. Lett. **51**, 87 (1983).
[2]D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 87 (1983).
[3]H.-M. Lee and S.-C. Lee, Phys. Lett. B **149**, 95 (1984).
[4]S.-C. Lee, Phys. Lett. B **149**, 98 (1984).
[5]S.-C. Lee and S.-L. Lou, Lett. Nuovo Cimento **44**, 69 (1985).
[6]S.-C. Lee, Class. Quant. Grav. **3**, 373 (1986); S.-C. Lee and S.-L. Lou, J. Math. Phys. **27**, 2751 (1986).
[7]S.-C. Lee, Lett. Nuovo Cimento **44**, 133 (1985).
[8]S.-C. Lee, Phys. Lett. B **158**, 413 (1985).
[9]D. Maison, J. Math. Phys. **20**, 871 (1979).
[10]V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978); **50**, 1 (1979).
[11]C. M. Cosgrove, *Proceedings of the The Second Marcel Grossmann Meetings on the Recent Developments of General Relativity*, Trieste, Italy, 5–11 July 1979, edited by R. Ruffini (North-Holland, Amsterdam, 1982); J. Math. Phys. **21**, 2417 (1980); **23**, 615 (1982).
[12]V. E. Zakharov and A. B. Shabat, Funkts. Anal. **13**, 13 (1979); V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978); V. E. Zakharov, "The inverse scattering method," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980); S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons* (Consultants Bureau, New York, 1984).
[13]V. Belinsky and R. Ruffini, Phys. Lett. B **89**, 195 (1980).
[14]J. Burzlaff and D. Maison, J. Math. Phys. **20**, 1352 (1979).
[15]I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 362, 1783 (1979); J. Math. Phys. **21**, 1126, 1418 (1980); **22**, 1051 (1981).
[16]G. Neugebauer and D. Kramer, Ann. Phys. (Leipzig) **24**, 62 (1969).
[17]S.-C. Lee, Phys. Lett. B **164**, 75 (1985).
[18]C.-C. Chiang, S.-C. Lee, G. Marmo, and S.-L. Lou, Phys. Rev. D **32**, 1364 (1985).
[19]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979); Nucl. Phys. B **164**, 52 (1980).
[20]G. Neugebauer, J. Phys. A **12**, 67 (1979).
[21]S.-C. Lee, J. Math. Phys. **28**, 901 (1987).
[22]Y.-S. Wu, Commun. Math. Phys. **90**, 461 (1983); L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981).

Soliton transformations for axially symmetric higher-dimensional gravity. II. Belinskii–Zakharov N-soliton transformations

S.-C. Lee
Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China

(Received 17 July 1986; accepted for publication 10 December 1986)

The solutions of vacuum Einstein equations in $4+K$ dimensions with $2+K$ commuting Killing vectors under the Abelian Kaluza–Klein ansatz are considered. This system admits Belinskii–Zakharov-type soliton transformations. The explicit formulas for general N-soliton transformations are obtained by iterating the formulas for the one-soliton transformation.

I. INTRODUCTION

In a previous work,[1] which we shall refer to as paper I in the following, we showed that the vacuum Einstein equations in $4+K$ dimensions under the Abelian Kaluza–Klein ansatz and the assumption of axial symmetry separate into two sets. The main set is equivalent to a generalized nonlinear sigma model[2] so that the well-known method of Belinskii–Zakharov (BZ)[3,4] can be applied to obtain the soliton transformations. The one-soliton transformation is worked out in detail in I and the results lead us to a series of new discrete symmetries[5] generalizing the Neugebauer–Kramer mapping.[6] In principle, the N-soliton transformations can be obtained by iterating the one-soliton transformation. It is sometimes desirable, however, to have explicit expressions for the general N-soliton transformations. In the four-dimensional case ($K=0$), this had been worked out by Belinskii–Zakharov[4] for N rank-1 soliton transformations. Here, we generalize their results to $4+K$ dimensions and for general N-soliton transformations of arbitrary ranks. These formulas can then be applied to flat background, to Weyl solutions[7] generalized to higher dimensions, or to other known solutions. This will be discussed in a subsequent work.

Our notations follow those of paper I. In Sec. II, we shall first recall the formulas for a one-soliton transformation and then derive the formulas for a two-soliton transformation. This is because two-soliton transformations can preserve both reality and asymptotic flatness[8] and hence are particularly useful. The derivation also shows that two successive one-soliton transformations commute (permutation theorem). Moreover, it prepares us for the derivation of general N-soliton transformations in Sec. III where we give the results and prove them by induction.

II. TWO-SOLITON TRANSFORMATIONS

The metric in $4+K$ dimensions under the Abelian Kaluza–Klein ansatz and the assumption of axial symmetry can be written as

$$d\bar{s}^2 = f(d\rho^2+dz^2) + Q_{\bar{a}\bar{b}}^{-1}\,dx^{\bar{a}}\,dx^{\bar{b}}, \quad \bar{a}=\bar{\mu},a, \tag{1}$$

where $Q_{\bar{a}\bar{b}}$ is a symmetric $n\times n$ matrix ($n=K+2$) given in (I.16). Here f and Q depends on ρ and z only. Moreover,

$$\det Q^{-1} = -\tau^2, \tag{2}$$

where $\Box\tau=0$. Instead of f, we shall often use the determinant $\bar{g}$ of the total metric tensor $\bar{g}_{AB}$, which is related to f by

$$\bar{g} = \tau^2 f^2. \tag{3}$$

Given a solution $Q,\bar{g}$, we can generate new solutions by solving the linearized equations

$$\begin{aligned}\partial_\xi U(\cdot;s) &= \tau^{-1}(1+\gamma^{-1})U(\cdot;s)J_\xi,\\ \partial_\eta U(\cdot;s) &= \tau^{-1}(1+\gamma)U(\cdot;s)J_\eta,\end{aligned} \tag{4}$$

where ξ,η are $z-i\rho$, $z+i\rho$, respectively, and γ is a given function of ξ,η,s defined in (I.29). The new solution obtained by a one-soliton transformation is given by $Q_1,\bar{g}_1,U_1(\cdot;s)$,

$$U_1(\cdot;s) = \left(\frac{\lambda-\lambda_1}{\lambda-\mu_1}\right)^{q_1/n} U(\cdot;s)\left(1-P_1+\frac{\lambda-\mu_1}{\lambda-\lambda_1}P_1\right), \tag{5}$$

$$Q_1 = \left(\frac{\lambda_1}{\mu_1}\right)^{q_1/n} Q\left(1-P_1+\frac{\mu_1}{\lambda_1}\right)P_1, \tag{6}$$

$$\begin{aligned}(-\bar{g}_1)^{1/2} \doteq{}& (-\bar{g})^{1/2}\tau^{q_1(n-q_1+2)/n}\lambda_1^{2q_1(n-q_1-1)/n}\\ &\times(\tau^2-\lambda_1^2)^{q_1(2q_1-n)/n}\\ &\times\det\left(\frac{1}{\tau^2-\lambda_1^2}X_1^T Q X_1\right).\end{aligned} \tag{7}$$

In these formulas, P_1 is a projection operator

$$P_1 = X_1(X_1^T Q X_1)^{-1}X_1^T Q, \tag{8}$$

with

$$X_1 = U(\cdot;s_1)^{-1}X_1^{(0)}. \tag{9}$$

Here $X_1^{(0)}$ is an arbitrary constant $n\times q_1$ matrix of rank q_1; and s_1 is a complex parameter and is related to λ_1 in the same way s is related to λ as given in Eq. (I.26). Finally, $\mu_1=\tau^2/\lambda_1$ is the value of λ_1 on the second sheet of s.

Note that P_1 can also be written as

$$\begin{aligned}P_1 = {}& U(\cdot;s_1)^{-1}X_1^{(0)}(X_1^{(0)\,T}Q^{(s_1)}X_1^{(0)})^{-1}\\ &\times X_1^{(0)\,T}Q^{(s_1)}U(\cdot;s_1),\end{aligned} \tag{10}$$

where

$$Q^{(s_1)} = \sigma(U(\cdot;s_1))QU(\cdot;s_1)^{-1}, \tag{11}$$

with σ being the involutive automorphism defined in (I.14).

Now we can iterate Eqs. (5)–(7) one more time to obtain a new solution $Q_2,\bar{g}_2$, and $U_2(\cdot;s)$ depending on the parameters $X_1^{(0)},X_2^{(0)},s_1,s_2$, where $X_2^{(0)}$ is an arbitrary $n\times q_2$ constant matrix of rank q_2.

The first step is to construct $Q_1^{(s_2)}$ and P_2 following Eqs. (10) and (11). We find

 0022-2488/87/040901-04$02.50

$$Q_1^{(s_2)} = \left(\frac{\lambda_1}{\mu_1}\right)^{q_1/n}\left(\frac{\lambda_2-\mu_1}{\lambda_2-\lambda_1}\right)^{2q_1/n}\sigma(U(\cdot;s_2))\tilde{Q}_1 U(\cdot;s_2)^{-1}, \tag{12}$$

where

$$\tilde{Q}_1 = Q - (\tau^2-\lambda_2^2)Q\left(\frac{1}{\tau^2-\lambda_1\lambda_2}X_1\right) \times\left(\frac{1}{\tau^2-\lambda_1^2}X_1^T Q X_1\right)^{-1}\left(X_1^T\frac{1}{\tau^2-\lambda_1\lambda_2}\right)Q \tag{13}$$

and

$$P_2 = \left(1-P_1+\frac{\lambda_2-\lambda_1}{\lambda_2-\mu_1}P_1\right)X_2(X_2^T\tilde{Q}_1X_2)^{-1} \times X_2^T\tilde{Q}_1\left(1-P_1+\frac{\lambda_2-\mu_1}{\lambda_2-\lambda_1}P_1\right) \tag{14}$$

with

$$X_2 = U(\cdot;s_2)^{-1}X_2^{(0)}. \tag{15}$$

Next we can construct $\bar{g}_2$ from Eqs. (7) and (12). This will involve the determinant

$$\det\left(\frac{1}{\tau^2-\lambda_1^2}X_1^T Q X_1\right)\det\left(\frac{1}{\tau^2-\lambda_2^2}X_2^T\tilde{Q}X_2\right). \tag{16}$$

The crucial point lies in observing that $X_2^T\tilde{Q}_1X_2$ arises naturally in the inverse of the matrix Σ_2,

$$\Sigma_2 = \begin{bmatrix} \frac{1}{\tau^2-\lambda_1^2}X_1^TQX_1, & \frac{1}{\tau^2-\lambda_1\lambda_2}X_1^TQX_2 \\ \frac{1}{\tau^2-\lambda_2\lambda_1}X_2^TQX_1, & \frac{1}{\tau^2-\lambda_2^2}X_2^TQX_2 \end{bmatrix}. \tag{17}$$

This can be seen from the following lemma.

Lemma: Let S be an invertible matrix and write

$$S = \begin{bmatrix} S_{11}, & S_{12} \\ S_{21}, & S_{22} \end{bmatrix},$$

where S_{11} is a submatrix of dimension $n_1\times n_1$ and S_{22} is of dimension $n_1\times n_2$. Assume S_{11} is invertible. Then

$$S^{-1} = \begin{bmatrix} (S_{11})^{-1}+(S_{11})^{-1}S_{12}(\tilde{S}_{22})^{-1}S_{21}(S_{11})^{-1}, & -(S_{11})^{-1}S_{12}(\tilde{S}_{22})^{-1} \\ -(\tilde{S}_{22})^{-1}S_{21}(S_{11})^{-1}, & (\tilde{S}_{22})^{-1} \end{bmatrix}, \tag{18}$$

where

$$\tilde{S}_{22} = S_{22} - S_{21}(S_{11})^{-1}S_{12}. \tag{19}$$

Moreover,

$$\det S = \det S_{11}\cdot\det\tilde{S}_{22}. \tag{20}$$

The proof of the first part of the lemma is by straightforward computation. For the second part, we observe that S_{21} can be made to vanish by multiplying on the left with the matrix

$$\begin{bmatrix} 1_{n_1\times n_1}, & 0 \\ -S_{21}(S_{11})^{-1}, & 1_{n_2\times n_2} \end{bmatrix},$$

which does not change its determinant.

Applying this lemma to Σ_2 and recall the definition of $\tilde{Q}_1$ in Eq. (13), we easily see that the determinant in Eq. (16) is just $\det\Sigma_2$. Putting everything together, we get

$$(-\bar{g}_2)^{1/2} \doteq (-\bar{g})^{1/2}\tau^{N_2(n-N_2+2)/n}\left(\frac{\tau^2-\lambda_1\lambda_2}{\lambda_1-\lambda_2}\right)^{2q_1q_2/n} \times\prod_{i=1}^{2}\left[\lambda_i^{2q_i(n-q_i-1)/n}(\tau^2-\lambda_i^2)^{q_i(2q_i-n)/n}\right] \cdot\det\Sigma_2, \tag{21}$$

where $N_2 = q_1+q_2$.

Now we are ready to compute $U_2(\cdot;s)$. For this, we have to compute

$$\left(1-P_1+\frac{\lambda-\mu_1}{\lambda-\lambda_1}P_1\right)\left(1-P_2+\frac{\lambda-\mu_2}{\lambda-\lambda_2}P_2\right). \tag{22}$$

Since the form of P_2 suggests that Σ_2^{-1} may play a role, it is easy to guess at the answer. We shall first introduce some notations. Let $H_2(\lambda)$, $D_2(\lambda)$ be the diagonal matrices

$$H_2(\lambda) = \begin{bmatrix} (\lambda-\lambda_1)1_{q_1\times q_1}, & \\ & (\lambda-\lambda_2)1_{q_2\times q_2} \end{bmatrix},$$

$$D_2(\lambda) = \begin{bmatrix} (\tau^2-\lambda\lambda_1)1_{q_1\times q_1}, & \\ & (\tau^2-\lambda\lambda_2)1_{q_2\times q_2} \end{bmatrix}. \tag{23}$$

We shall use $\hat{H}_2(\lambda)$ to denote the matrix obtained from $H_2(\lambda)$ by the replacement $\lambda_1\to\mu_1$, $\lambda_2\to\mu_2$. This notation will be consistently used for other matrices that depend on i. Similarly, we shall use $\tilde{H}_2(\lambda)$ to denote the matrix obtained from $H_2(\lambda)$ by the replacement $\lambda\to\mu=\tau^2/\lambda$. This will be consistently used for other matrices that depend on λ and is consistent with our notation that "tilde" denotes quantities on the second sheet of s. Now we define

$$\begin{aligned} \Gamma_2(\lambda) &= -H_2(0)\Sigma_2H_2(\lambda), \\ \tilde{\Gamma}_2(\lambda) &= -H_2(0)\Sigma_2\tilde{H}_2(\lambda) \end{aligned} \tag{24}$$

and note that

$$\hat{\Gamma}_2(\lambda) = -\hat{H}_2(0)\hat{\Sigma}_2\hat{H}_2(\lambda) = \Sigma_2D_2(\lambda). \tag{25}$$

Also define the $n\times(q_1+q_2)$ matrix as

$$Y_2 = [X_1,X_2]. \tag{26}$$

With these preparations, we can state the following theorem.

Theorem: The expression in Eq. (22) is equal to

$$1 - Y_2\Gamma_2(\lambda)^{-1}Y_2^TQ \tag{27}$$

so that

$$U_2(\cdot;s) = \prod_{i=1}^{2}\left(\frac{\lambda-\lambda_i}{\lambda-\mu_i}\right)^{q_i/n}U(\cdot;s)(1-Y_2\Gamma_2(\lambda)^{-1}Y_2^TQ), \tag{28}$$

$$Q_2 = \prod_{i=1}^{2} \left(\frac{\lambda_i}{\mu_i}\right)^{q_i/n} Q(1 - Y_2\Gamma_2(0)^{-1}Y_2^T Q). \tag{29}$$

Note that Q_2 is manifestly symmetric.

The proof of the theorem is by straightforward computation using Eq. (18).

Summarizing, Eqs. (28), (29), and (21) give the formulas for two-soliton transformation. It is clear that the final result is invariant under simultaneous interchange of s_1, s_2 and $X_1^{(0)}, X_2^{(0)}$. This is the permutation theorem for the BZ two-soliton transformation.

Let us also remark that if we choose $s_2 = s_1$ while letting $X_1^{(0)}, X_2^{(0)}$ independent so that the matrix $[X_1^{(0)}, X_2^{(0)}]$ is of rank $N_2 = q_1 + q_2$, then the two-soliton transformation with ranks q_1 and q_2 reduces to a one-soliton transformation of rank N_2. This is quite obvious for Eqs. (28) and (29), while for Eq. (21), a simple computation using the relation

$$\lambda_1 - \lambda_2 = \lambda_1^2(\tau^2 - \lambda_1^2)^{-1}(s_1 - s_2) + O((s_1 - s_2)^2) \tag{30}$$

proves the assertion. This shows that for multisoliton transformations, it is enough to consider only the rank-1 case. The general situation is obtained by letting some of the s_i's coincide.

We have presented our results in such a way that they are ready for generalization. Indeed, it is quite obvious how to generalize the definitions of the matrices X_2, Y_2, Σ_2, H_2, D_2, Γ_2, etc. to $X_m, Y_m, \ldots$, etc. for any positive integer m. Using these matrices, we shall give the formulas for the general N-soliton transformations and prove them by induction in the following section.

III. *N*-SOLITON TRANSFORMATIONS

Let us state our main result as a thoerem.

Theorem: Let Q, $\bar{g}$ be a solution and let $U(\cdot;s)$ be an integral of Eq. (4), properly normalized. Choose m complex constants $s_1, \ldots, s_m$ and m complex constant matrices $X_1^{(0)}, \ldots, X_m^{(0)}$ of ranks $q_1, \ldots, q_m$ and dimensions $n \times q_1, \ldots, n \times q_m$, respectively. Let

$$X_i = U(\cdot;s_i)^{-1}X_i^{(0)}, \quad Y_m = [X_1, \ldots, X_m], \tag{31}$$

then

$$Q_m = \prod_{i=1}^{m} \left(\frac{\lambda_i}{\mu_i}\right)^{q_i/n} Q(1 - Y_m\Gamma_m(0)^{-1}Y_m^T Q), \tag{32}$$

$$(-\bar{g}_m)^{1/2} \doteq (-\bar{g})^{1/2}\tau^{N_m(n-N_m+2)/n} \times \prod_{i<j}\left(\frac{\tau^2 - \lambda_i\lambda_j}{\lambda_i - \lambda_j}\right)^{2q_iq_j/n} \prod_{i=1}^{m} \lambda_i^{2q_i(n-q_i-1)/n} \times (\tau^2 - \lambda_i^2)^{q_i(2q_i-n)/n} \cdot \det \Sigma_m, \tag{33}$$

where $N_m = q_1 + \cdots + q_m$, $\Gamma_m(\lambda)$, and Σ_m are defined as in Eqs. (17) and (24), is also a solution. The corresponding generating function $U_m(\cdot;s)$ is given by

$$U_m(\cdot;s) = \prod_{i=1}^{m} \left(\frac{\lambda - \lambda_i}{\lambda - \mu_i}\right)^{q_i/n} U(\cdot;s) \times (1 - Y_m\Gamma_m(\lambda)^{-1}Y_m^T Q). \tag{34}$$

Before proving the theorem, let us also define the following useful quantities:

$$\tilde{Q}_m = Q - (\tau^2 - \lambda_{m+1}^2)QY_m D_m^{-1}(\lambda_{m+1}) \times \Sigma_m^{-1}D_m^{-1}(\lambda_{m+1})Y_m^T Q, \tag{35}$$

$$P_m = (1 - Y_{m-1}\sigma(\hat{\Gamma}_{m-1}(\lambda_m))Y_{m-1}^T Q) \times X_m(X_m^T\tilde{Q}_{m-1}X_m)^{-1} \times X_m^T\tilde{Q}_{m-1}(1 - Y_{m-1}\Gamma_{m-1}^{-1}(\lambda_m)Y_{m-1}^T Q), \tag{36}$$

$$\tilde{U}_m(\cdot;s) = \prod_{i=1}^{m} \left(\frac{\mu - \lambda_i}{\mu - \mu_i}\right)^{q_i/n} \times \tilde{U}(\cdot;s)(1 - Y_m\tilde{\Gamma}_m(\lambda)^{-1}Y_m^T Q). \tag{37}$$

Equations (35) and (36) are natural generalizations of Eqs. (13) and (14), respectively.

The theorem will be proved by induction and the main steps will be presented in a series of lemmas. Assuming the theorem is true for m, we shall construct Q_{m+1}, $\bar{g}_{m+1}$, and $U_{m+1}(\cdot;s)$ by doing a one-soliton transformation. As in the two-soliton case, we first construct

$$Q_m^{(s_{m+1})} = \sigma(U_m(\cdot;s))Q_m U_m(\cdot;s_{m+1})^{-1}. \tag{38}$$

This requires computation of the inverse of $(1 - Y_m\Gamma_m(\lambda)^{-1}Y_m^T Q)$. We have the following.

Lemma 1:

$$(1 - Y_m\Gamma_m(\lambda)^{-1}Y_m^T Q)(1 - Y_m\sigma(\hat{\Gamma}_m(\lambda))Y_m^T Q) = 1. \tag{39}$$

The proof relies on the following identities:

$$\Sigma_m^{-1}\hat{H}_m(\lambda) - H_m(\lambda)\Sigma_m^{-1} = \Sigma_m^{-1}H_m(0)^{-1}Y_m^T QY_m\Sigma_m^{-1}, \tag{40}$$

$$H_m(0)^{-1} = \hat{H}_m(\lambda)D_m(\lambda)^{-1}, \tag{41}$$

which are valid for all m. The identity (40) is easily verified by multiplying on both sides of each term with Σ_m and computing the ijth block while identity (41) follows from the definitions. A simple computation then shows that Lemma 1 is true.

Next we show the following.

Lemma 2:

$$\tilde{U}_m(\cdot;s) = \sigma(U_m(\cdot;s))Q_m \tag{42}$$

or, equivalently,

$$(1 - Y_m\hat{\Gamma}_m(\lambda)^{-1}Y_m^T Q)(1 - Y_m\Gamma_m(0)^{-1}Y_m^T Q) = 1 - Y_m\tilde{\Gamma}_m(\lambda)^{-1}Y_m^T Q. \tag{43}$$

The proof is again a simple computation using the transpose of the identity (40).

Substituting Eq. (42) into Eq. (38) and using Eq. (37), we can identify $\tilde{Q}_m$.

Lemma 3:

$$\tilde{Q}_m = Q(1 - Y_m\tilde{\Gamma}_m(\lambda_{m+1})^{-1}Y_m^T Q) \times (1 - Y_m\sigma(\hat{\Gamma}_m(\lambda_{m+1}))Y_m^T Q) \tag{44}$$

or, equivalently,

$$(1 - Y_m \tilde{\Gamma}_m(\lambda)^{-1} Y_m^T Q)(1 - Y_m \sigma(\hat{\Gamma}_m(\lambda)) Y_m^T Q)$$
$$= 1 - (\tau^2 - \lambda^2) Y_m D_m(\lambda)^{-1} \Sigma_m^{-1} D_m(\lambda)^{-1} Y_m^T Q. \tag{45}$$

This is also easily verified using the identity (40).

Now $Q_m^{(s_{m+1})}$ can be expressed in terms of $\tilde{Q}_m$. It is then easy to see that Eq. (36), when replacing m by $m+1$, gives the correct formula for P_{m+1}. With P_{m+1}, we can now compute $U_{m+1}(\cdot;s)$. As in the two-soliton case, we first establish the relations.

Lemma 4: The $(m+1)$, $(m+1)$ block of Σ_{m+1}^{-1} is

$$[\Sigma_{m+1}^{-1}]_{m+1,m+1} = (\tau^2 - \lambda_{m+1}^2)(X_{m+1}^T \tilde{Q}_m X_{m+1})^{-1} \tag{46}$$

and

$$\det \Sigma_{m+1} = \det \Sigma_m \cdot \det\left(\frac{1}{\tau^2 - \lambda_{m+1}^2} X_{m+1}^T \tilde{Q}_m X_{m+1}\right). \tag{47}$$

The proof follows from the definition of $\tilde{Q}_m$ in Eq. (35) and the lemma of the last section, i.e., Eqs. (18) and (19).

Now we can compute $\bar{g}_{m+1}$ using the induction hypothesis equation (33), the formula for one-soliton transformation equation (7), and the formula for $Q_m^{(s_{m+1})}$. The straightforward computation proves the theorem for $\bar{g}_{m+1}$.

Finally to complete the construction of $U_{m+1}(\cdot;s)$, we have the following.

Lemma 5:

$$(1 - Y_m \Gamma_m(\lambda)^{-1} Y_m^T Q)\left(1 - P_{m+1} + \frac{\lambda - \mu_{m+1}}{\lambda - \lambda_{m+1}} P_{m+1}\right)$$
$$= 1 - Y_{m+1} \Gamma_{m+1}(\lambda)^{-1} Y_{m+1}^T Q \tag{48}$$

or, equivalently,

$$1 - Y_m \Gamma_m(\lambda)^{-1} Y_m^T Q = \prod_{i=1}^{m} \left(1 - P_i + \frac{\lambda - \mu_i}{\lambda - \lambda_i} P_i\right) \quad \forall m. \tag{49}$$

The proof of Lemma 5 is quite straightforward. We expand both sides of the equation. When expanding the right-hand side, we use the lemma of the last section to separate Σ_{m+1}^{-1} into the $(m+1),(m+1)$ block, $(m+1),i$ blocks, $i,(m+1)$ blocks, and the rest. Then one collects terms and shows that the two sides are equal.

With this lemma, the theorem for U_{m+1}, Q_{m+1} is easily established.

IV. DISCUSSION

There are other ways to obtain the formulas for general N-soliton transformations than to iterate the one-soliton transformation. For example, one may use the method of soliton correlation matrix introduced by Harnad *et al.*[9] However, in those approaches, it is usually difficult to obtain the correct normalization and to obtain the formula for $\bar{g}_m$. What we have achieved is essentially to provide a series of identities so that the different approaches can be bridged.

In this respect, it is perhaps appropriate to mention that our $\Gamma(\lambda)^{-1}$ forms the backbone of the soliton correlation matrix of Ref. 9 consequently generalizing their construction for sigma models to the present case.[2]

We should also mention that a lot of work has been done in the four-dimensional case as referred to in I. In particular, BZ had worked out the general N successive rank-1 soliton transformations in four dimensions in their original works.[4]

As explained in I, our interest in the soliton transformation stems partly from the possibility of using them to generate solutions in the vacuum sector. This can be achieved by letting the s-plane poles in the BZ soliton transformations coincide and at the same time taking suitable limits of $X_i^{(0)}$ associated with each pole. An example of this in the four-dimensional case is given in Ref. 10 where a confluent double soliton transformation was carried out. The higher-dimensional case is much more complicated and is currently under investigation.

ACKNOWLEDGMENT

This work is supported in part by a grant from the National Science Council, Taiwan, Republic of China.

[1] S.-C. Lee, J. Math. Phys. **28**, 893 (1987).
[2] S.-C. Lee, Phys. Lett. B **158**, 413 (1985).
[3] V. E. Zakharov and A. B. Shabat, Funkts. Anal. **13**, 13 (1979); V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978); V.E. Zakharov, "The inverse scattering method," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980); S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons* (Consultants Bureau, New York, 1984).
[4] V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978); **50**, 1 (1979).
[5] S.-C. Lee. Phys. Lett. B **164**, 75 (1985).
[6] G. Neugebauer and D. Kramer, Ann. Phys. (Leipzig) **24**, 62 (1969).
[7] H. Weyl, Ann. Phys. (Leipzig) **54**, 117 (1917).
[8] C. M. Cosgrove, *Proceedings of the The Second Marcel Grossmann Meetings on the Recent Developments of General Relativity*, Trieste, Italy, 5–11 July 1979, edited by R. Ruffini (North-Holland, Amsterdam, 1982); J. Math. Phys. **21**, 2417 (1980); **23**, 615 (1982).
[9] J. Harnad, Y. Saint-Aubin, and S. Shnider, Commun. Math. Phys. **92**, 329 (1984).
[10] C. M. Cosgrove, J. Math. Phys. **21**, 2417 (1980).

Perfect fluids satisfying a less than extremely relativistic equation of state

Basilis C. Xanthopoulos
Department of Physics, University of Crete, Greece and Research Center of Crete, Iraklion, Greece

(Received 22 January 1986; accepted for publication 22 October 1986)

The Einstein perfect-fluid equations for a fluid with "energy density ϵ = pressure p + constant" equation of state are considered *ab initio* for space-times with two hypersurface orthogonal, spacelike, commuting Killing fields. Gauge conditions compatible with the field equations are imposed while the fluid is described by means of a stream potential. Contrary to the vacuum Einstein equations, the gauge functions cannot be specified in early stages of the analysis. Instead, the problem is described by a system of five (coupled) partial differential equations for the two gauge functions, the conformal factor of the two-dimensional geometry orthogonal to the two Killing fields, a scalar determining the norms of the Killing fields, and the stream potential of the fluid. It is remarkable that the last two scalars satisfy conjugate equations in the sense that each of them is a potential for the other. By using this conjugation, a subcase of the problem is eventually reduced to a system of two highly nonlinear equations for the two gauge functions. This system is explicitly solved, leading to three distinct families of solutions. To complete the solution for each family one has to solve a second-order, linear, hyperbolic equation and perform certain quadratures.

I. INTRODUCTION

One can possibly say with some confidence that in the last ten years the vacuum Einstein equations with two commuting Killing fields have been put under control. And one clear message from all this effort is that whatever progress can be made in studying the vacuum Einstein equations, it can also be readily achieved for the Einstein–Maxwell electrovacuum equations.

Another equally interesting generalization would be to treat the Einstein-hydrodynamic equations for the gravitational field coupled with a perfect fluid. Here the message is different, namely that the perfect fluid equations are much more difficult, and rumor has it that while the Einstein and the Maxwell equations are God-given, the perfect-fluid equations are man-made.

Two particular cases distinguish themselves as "easy" among the perfect-fluid solutions with two commuting Killing fields. The stationary axisymmetric dust (pressure $p=0$) solutions[1] and the perfect fluids with two spacelike Killing fields and a stiff (energy density $\epsilon=p$) equation of state.[2] In the latter case, in fact, the equations reduce[3,4] to the vacuum Einstein equations and to a linear second-order partial differential equation for the "stream potential" of the fluid; and the two problems can be solved independently. Except for these two cases, essentially all the other known perfect-fluid solutions[2,5]—and they are not that many—with two commuting Killing fields have been obtained by making the additional requirement that the space-time also admits a Killing tensor (or imposing some other algebraic condition on the curvature), thus obtaining an overdetermined system of equations. But then there is no control on the form of the equation of state that the fluid satisfies and, to no one's surprise, it usually turns out to be unphysical.

Our far-reaching objective is to change this unsatisfactory situation in perfect fluids. And as a small step forward in that direction, we report in this paper a method for studying and solving perfect fluids with two spacelike, hypersurface orthogonal Killing fields satisfying the simplest less-than-extremely-relativistic equation of state,

$$\epsilon = p + k, \quad k = \text{const}. \tag{1}$$

Three families of solutions, each enjoying the freedom of the solutions of a second-order linear hyperbolic equation, are obtained.

There is more in our objective than gaining experience and obtaining solutions with perfect fluids. Solutions with two spacelike and hypersurface orthogonal Killing fields describe inhomogeneous cosmological models as well as the interaction region resulting from the collision of two plane gravitational and hydrodynamic waves with linear polarizations.[6,7] Recently we have shown[4] that the collision of massless dust may result in the creation of a massive fluid which satisfies the extremely relativistic $\epsilon=p$ equation of state. What we would like in a next step, therefore, would be to consider a collision resulting in the creation of a massive fluid with the simplest, but not extremely relativistic, equation of state $\epsilon=p+$ const. Since we are mainly interested, at this stage, in learning how to deal with more involved, predetermined equations of state, we decided to put aside the additional complications arising from failing to assume that the two Killing fields are hypersurface orthogonal. We plan to return to this problem by considering particular solutions of the three families obtained in Sec. V and studying their physical interpretations.

II. THE HYDRODYNAMIC EQUATIONS

For space-times with two spacelike commuting and hypersurface orthogonal Killing fields the metric is of the form[8]

$$(ds)^2 = e^{2\nu}(dx^0)^2 - e^{2\psi}(dx^1)^2 - e^{2\mu_2}(dx^2)^2 - e^{2\mu_3}(dx^3)^2, \tag{2}$$

where $\partial/\partial x^1$ and $\partial/\partial x^2$ are the two Killing fields and ν, ψ, μ_2, and μ_3 are functions of x^0 and x^3 only. The metric satisfies

 0022-2488/87/040905-08$02.50

the Einstein and the hydrodynamic equations for a perfect fluid described by

$$T^{ij} = (\epsilon + p)u^i u^j - pg^{ij}, \tag{3}$$

where

$$g_{ij}u^i u^j = 1 \tag{4}$$

[signature $(+\ ---)$]. Compatibly with the existence of the two Killing fiels we assume that the only nonvanishing components of the four-velocity of the fluid are u^0 and u^3 and that these components as well as ϵ and p depend only on x^0 and x^3.

First we consider the hydrodynamic equations.[4] The conservation of the energy-momentum $T^{ij}{}_{;j} = 0$ gives

$$(\epsilon + p)u^i{}_{;j}u^j + u^i[u^j(\epsilon + p)_{,j} + (\epsilon + p)u^j{}_{;j}] - g^{ij}p_{,j} = 0, \tag{5}$$

which contracted with u_i becomes

$$u^j\epsilon_{,j} + (\epsilon + p)u^j{}_{;j} = 0. \tag{6}$$

Expressed in terms of ordinary derivatives the last equation gives

$$u^j\epsilon_{,j} + [(\epsilon + p)/\sqrt{-g}]\ [\sqrt{-g}u^j]_{,j} = 0, \tag{7}$$

where

$$\sqrt{-g} = e^{\nu+\psi+\mu_2+\mu_3}. \tag{8}$$

We introduce the orthonormal tetrad

$$\omega^{(0)} = e^{\nu}(dx^0), \quad \omega^{(1)} = e^{\psi}(dx^1), \quad \omega^{(2)} = e^{\mu_2}(dx^2), \quad \omega^{(3)} = e^{\mu_3}(dx^3), \tag{9}$$

and use the tetrad components $u_{(0)}$ and $u_{(3)}$ of the four-velocity, related to the tensor components by

$$u^{(0)} = u_{(0)} = e^{\nu}u^0 = e^{-\nu}u_0$$

and

$$u^{(3)} = -u_{(3)} = e^{\mu_3}u^3 = -e^{-\mu_3}u_3; \tag{10}$$

Eq. (4) then reads

$$u^2_{(0)} - u^2_{(3)} = 1. \tag{11}$$

We expand Eq. (7) and we express it in terms of the tetrad components of the four-velocity of the fluid. We get

$$e^{-\nu}u_{(0)}\epsilon_{,0} - e^{-\mu_3}u_{(3)}\epsilon_{,3} + (\epsilon + p)e^{-(\nu+\psi+\mu_2+\mu_3)} \times\{[e^{\psi+\mu_2+\mu_3}u_{(0)}]_{,0} - [e^{\nu+\psi+\mu_2}u_{(3)}]_{,3}\} = 0. \tag{12}$$

A. Assumption that the fluid statisfies an equation of state

Without specifying it, we now assume that the fluid satisfies an equation of state, i.e., a functional relation between ϵ and p. We shall be describing it by the relation

$$\epsilon + p = f(\epsilon) \tag{13}$$

for some unspecified function f. We shall also introduce

$$\ln f_1(\epsilon) = \int \frac{d\epsilon}{f(\epsilon)} \Leftrightarrow \frac{\dot{f}_1}{f_1} = \frac{1}{f}, \tag{14}$$

where, in this section, the dot denotes differentiation with respect to ϵ. Either of the functions f and f_1 determines uniquely the equation of state.

Equation (12) can now be written in the following two equivalent forms:

$$e^{\psi+\mu_2+\mu_3}u_{(0)}\epsilon_{,0} + f(\epsilon)[e^{\psi+\mu_2+\mu_3}u_{(0)}]_{,0} - e^{\nu+\psi+\mu_2}u_{(3)}\epsilon_{,3} - f(\epsilon)\ [e^{\nu+\psi+\mu_2}u_{(3)}]_{,3} = 0, \tag{15}$$

or

$$[e^{\psi+\mu_2+\mu_3}u_{(0)}f_1]_{,0} - [e^{\nu+\psi+\mu_2}u_{(3)}f_1]_{,3} = 0. \tag{16}$$

Clearly, Eq. (16) is solved by introducing the "stream" potential ϕ. Thus setting

$$e^{\psi+\mu_2+\mu_3}u_{(0)}f_1 = \phi_{,3}, \quad e^{\nu+\psi+\mu_2}u_{(3)}f_1 = \phi_{,0}, \tag{17}$$

Eq. (16) is satisfied by virtue of the existence of ϕ.

Next we turn to Eq. (5), which, in the presence of its contracted version (6), simplifies to

$$(\epsilon + p)u^j u_{i;j} = p_{,i} - u_i u^j p_{,j}; \tag{18}$$

expressed in tetrad components it also reads

$$u^{(b)}u_{(a)|(b)} = [1/(\epsilon + p)]\ [p_{,(a)} - u_{(a)}u^{(b)}p_{,(b)}], \tag{19}$$

where the bar denotes the intrinsic derivative, defined in Ref. 9, p. 37 (this book will be referred to hereafter as MT). Nontrivial equations are obtained for $a = 0$ and $a = 3$. Expanding Eq. (19) and using the expressions for the rotation coefficients listed in MT, p. 82, Eq. (91), we obtain, respectively,

$$e^{-\nu}u_{(0)}u_{(0),0} - e^{-\mu_3}u_{(3)}u_{(0),3} - e^{-\mu_3}\nu_{,3}u_{(0)}u_{(3)} + e^{-\nu}\mu_{3,0}u^2_{(3)} = [1/(\epsilon + p)]\ [-e^{-\nu}u^2_{(3)}p_{,0} + u_{(0)}u_{(3)}e^{-\mu_3}p_{,3}] \tag{20}$$

and

$$-e^{-\mu_3}u_{(3)}u_{(3),3} + e^{-\nu}u_{(0)}u_{(3),0} + e^{-\nu}\mu_{3,0}u_{(3)}u_{(0)} - e^{-\mu_3}\nu_{,3}u^2_{(0)} = [1/(\epsilon + p)]\ [e^{-\mu_3}u^2_{(0)}p_{,3} - u_{(3)}u_{(0)}e^{-\nu}p_{,0}]. \tag{21}$$

By subtracting Eqs. (20) and (21) and dividing by $u_{(0)} - u_{(3)}$ we obtain

$$e^{-\nu}u_{(0)}[\ln(u_{(0)} - u_{(3)})]_{,0} - e^{-\mu_3}u_{(3)}[\ln(u_{(0)} - u_{(3)})]_{,3} + e^{-\mu_3}\nu_{,3}u_{(0)} - e^{-\nu}\mu_{3,0}u_{(3)} = [1/(\epsilon + p)][e^{-\nu}u_{(3)}p_{,0} - u_{(0)}e^{-\mu_3}p_{,3}], \tag{22}$$

while by adding them and dividing by $u_{(0)} + u_{(3)}$ we obtain

$$e^{-\nu}u_{(0)}[\ln(u_{(0)} + u_{(3)})]_{,0} - e^{-\mu_3}u_{(3)}[\ln(u_{(0)} + u_{(3)})]_{,3} - e^{-\mu_3}u_{(0)}\nu_{,3} + e^{-\nu}\mu_{3,0}u_{(3)} = [1/(\epsilon + p)]\ [-e^{-\nu}p_{,0}u_{(3)} + e^{-\mu_3}p_{,3}u_{(0)}]. \tag{23}$$

Finally, by adding Eqs. (22) and (23) we obtain

$$e^{-\nu}u_{(0)}[\ln(u^2_{(0)} - u^2_{(3)})]_{,0} - e^{-\mu_3}u_{(0)}[\ln(u^2_{(0)} - u^2_{(3)})]_{,3} = 0, \tag{24}$$

i.e., an identity by virtue of Eq. (11). We have shown, therefore, that there are precisely two independent hydrodynamic

equations, namely, Eq. (16) and one of Eqs. (20) and (21)—we shall keep Eq. (20).

We would like to express Eq. (20) in terms of the stream potential ϕ introduced by Eqs. (17). We set

$$\mathbf{H} = e^{\mu_3} u_{(0)} f_1, \quad \Theta = e^{\mu_3} u_{(3)} f_1 , \tag{25}$$

which imply, by using Eq. (11), that

$$u_{(0)} = \mathbf{H}/\sqrt{\mathbf{H}^2 - \Theta^2}, \quad u_{(3)} = \Theta/\sqrt{\mathbf{H}^2 - \Theta^2} . \tag{26}$$

By differentiating the first of Eqs. (25) and the equation of state (13) we get

$$\begin{aligned} \frac{u_{(0),0}}{u_{(0)}} &= \frac{\mathbf{H}_{,0}}{\mathbf{H}} - \frac{\epsilon_{,0}}{f} - \mu_{3,0} , \\ \frac{u_{(0),3}}{u_{(0)}} &= \frac{\mathbf{H}_{,3}}{\mathbf{H}} - \frac{\epsilon_{,3}}{f} - \mu_{3,3} \end{aligned} \tag{27}$$

and

$$p_{,i} = (\dot{f} - 1)\epsilon_{,i}, \quad i = 0,3 . \tag{28}$$

Finally, by eliminating $u_{(0)}$ and $u_{(3)}$ and their derivatives from Eq. (20) we get

$$\begin{aligned} &[\Theta_{,0} - (e^{\nu - \mu_3}\mathbf{H})_{,3}] \\ &\quad + f^{-1}(\dot{f} - 2)\,[\Theta\epsilon_{,0} - e^{\nu - \mu_3}\mathbf{H}\epsilon_{,3}] = 0 . \end{aligned} \tag{29}$$

B. Assumption that the fluid satisfies the equation of state, $\epsilon = p + \text{const}$

When $\epsilon = p + k$, $k = \text{const}$, Eq. (13) gives that

$$f(\epsilon) = 2\epsilon - k \tag{30}$$

and therefore $\dot{f} - 2 = 0$. Thus the second term of Eq. (29) drops out and the hydrodynamic equation simplifies. In fact, it will be seen that it can be readily expressed solely in terms of the potential ϕ. Thus, by using Eqs. (17) and (25) we find

$$\mathbf{H} = e^{-\psi - \mu_2}\phi_{,3}, \quad \Theta = e^{\mu_3 - \nu - \psi - \mu_2}\phi_{,0} , \tag{31}$$

which, substituted in Eq. (29), give

$$[e^{\mu_3 - \mu_2 - \psi - \nu}\phi_{,0}]_{,0} - [e^{\nu - \psi - \mu_2 - \mu_3}\phi_{,3}]_{,3} = 0 . \tag{32}$$

We shall be referring to Eq. (32) as the stream potential equation. Were the geometry known, the solutions of the linear equation (32) for ϕ would determine all the quantities of the fluid. Thus Eq. (14) gives

$$f_1 = \sqrt{2\epsilon - k} = \sqrt{\epsilon + p} , \tag{33}$$

which, combined with Eqs. (11) and (17), gives the energy density

$$2\epsilon - k = \epsilon + p = e^{-2(\psi + \mu_2)}\,[\phi_{,3}^2 e^{-2\mu_3} - \phi_{,0}^2 e^{-2\nu}] . \tag{34}$$

And then, Eq. (17) is used to determine the four-velocity of the fluid,

$$\begin{aligned} u_{(0)} &= (e^{-\psi - \mu_2 - \mu_3}/\sqrt{2\epsilon - k})\,\phi_{,3} , \\ u_{(3)} &= (e^{-\psi - \nu - \mu_2}/\sqrt{2\epsilon - k})\,\phi_{,0} . \end{aligned} \tag{35}$$

III. THE EINSTEIN EQUATIONS

The Einstein equations are most easily written in the orthonormal tetrad (9). We use (i) the orthonormal components of the Ricci tensor obtained from Eqs. (10)–(17) of Ref. 8, (ii) Eqs. (10)–(13) of Ref. 4 giving the tetrad components of the Ricci and the energy-momentum tensor, and (iii) that $G_{(a)(b)} = -2\,T_{(a)(b)}$. The resulting equations are manipulated as described in Ref. 10. With the notation

$$\beta = \psi + \mu_2, \quad \chi = e^{\mu_2 - \psi} , \tag{36}$$

we get the Einstein equations [see Ref. 4, Eqs. (19)–(23)]

$$[e^{\mu_3 - \nu}(e^{\beta})_{,0}]_{,0} - [e^{\nu - \mu_3}(e^{\beta})_{,3}]_{,3} = -2ke^{\beta + \nu + \mu_3} , \tag{37}$$

$$[e^{\beta + \mu_3 - \nu}(\ln\chi)_{,0}]_{,0} - [e^{\beta + \nu - \mu_3}(\ln\chi)_{,3}]_{,3} = 0 , \tag{38}$$

$$\begin{aligned} &\beta_{,0}(\nu + \mu_3)_{,3} + \beta_{,3}(\nu + \mu_3)_{,0} \\ &\quad + [\beta_{,0}(\nu - \mu_3)_{,3} - \beta_{,3}(\nu - \mu_3)_{,0} - 2\beta_{,03} - \beta_{,0}\beta_{,3}] \\ &\quad = \frac{\chi_{,0}\chi_{,3}}{\chi^2} - 4(\epsilon + p)u_{(0)}u_{(3)}e^{\nu + \mu_3} , \end{aligned} \tag{39}$$

and

$$\begin{aligned} &2\beta_{,3}e^{\nu - \mu_3}(\nu + \mu_3)_{,3} + 2\beta_{,0}e^{\mu_3 - \nu}(\nu + \mu_3)_{,0} + e^{\nu - \mu_3}[\beta_{,3}^2 + 2\beta_{,3}(\nu - \mu_3)_{,3}] + e^{\mu_3 - \nu}[\beta_{,0}^2 + 2\beta_{,0}(\mu_3 - \nu)_{,0}] \\ &\quad = 2e^{-\beta}\{[e^{\nu - \mu_3}(e^{\beta})_{,3}]_{,3} + [e^{\mu_3 - \nu}(e^{\beta})_{,0}]_{,0}\} + (1/\chi^2)\,[e^{\nu - \mu_3}\chi_{,3}^2 + e^{\mu_3 - \nu}\chi_{,0}^2] - 4(\epsilon + p)\,(u_{(0)}^2 + u_{(3)}^2)e^{\nu + \mu_3} . \end{aligned} \tag{40}$$

Equations (32), (37), (38), (39), and (40) are the basic equations of the problem. Recall that k is a constant, which shall be assumed to be different from zero. The unknowns are ϕ, β, χ, ν, and μ_3 and we have the gauge freedom of choosing the coordinates x^0 and x^3.

In the vacuum, Einstein–Maxwell electrovacuum, and perfect-fluid solutions with $\epsilon = p$ equation of state, the right-hand side of Eq. (37) is zero. This permits the specification of the gauge—namely the determination of e^{β}—consistently with the field equations at the very beginning, independently of χ, $\nu + \mu_3$, and ϕ. The complications of the problem for $k \neq 0$ arise because Eq. (37) now couples β to $\nu + \mu_3$ and one has to consider the entire system of equations and determine β, $\nu + \mu_3$, χ, and ϕ simultaneously.

Following the literature[2,3] and compatibility with the field equations, we shall adopt the gauge condition

$$\nu = \mu_3; \tag{41}$$

it would correspond to adopting the Papapetrou[9,11] gauge in the vacuum and electrovacuum cases. For e^{β} *we shall assume* that it separates in the two coordinates $x^0 = t$ and $x^3 = z$:

$$e^{\beta} = \mathbf{A}(t)\mathbf{B}(z) , \tag{42}$$

where $\mathbf{A} = \mathbf{A}(t)$ and $\mathbf{B} = \mathbf{B}(z)$ are unspecified functions of the indicated variables. The assumption of the separability of e^{β} has been proved useful in the considerations of the vacuum and electrovacuum equations, leading to the standard forms of the Ernst[12] equation. The four Einstein equations

(37)–(40) and the stream potential equation then reduce to

$$2ke^{2\nu}=\frac{\mathbf{B}''}{\mathbf{B}}-\frac{\ddot{\mathbf{A}}}{\mathbf{A}}, \tag{43}$$

$$\mathbf{A}^{-1}[\mathbf{A}(\ln\chi)_{,t}]_{,t}-\mathbf{B}^{-1}[\mathbf{B}(\ln\chi)_{,z}]_{,z}=0, \tag{44}$$

$$\frac{2\dot{\mathbf{A}}}{\mathbf{A}}\nu_{,z}+\frac{2\mathbf{B}'}{\mathbf{B}}\nu_{,t}-\frac{\dot{\mathbf{A}}\mathbf{B}'}{\mathbf{AB}}=\frac{\chi_{,t}\chi_{,z}}{\chi^2}-4(\epsilon+p)u_{(t)}u_{(z)}e^{2\nu}, \tag{45}$$

$$4\frac{\mathbf{B}'}{\mathbf{B}}\nu_{,z}+4\frac{\dot{\mathbf{A}}}{\mathbf{A}}\nu_{,t}+\frac{\dot{\mathbf{A}}^2}{\mathbf{A}^2}+\frac{\mathbf{B}'^2}{\mathbf{B}^2}=2\left(\frac{\ddot{\mathbf{A}}}{\mathbf{A}}+\frac{\mathbf{B}''}{\mathbf{B}}\right)+\frac{1}{\chi^2}(\chi_{,t}^2+\chi_{,z}^2)-4(\epsilon+p)\,(u_{(t)}^2+u_{(z)}^2)e^{2\nu}, \tag{46}$$

$$\mathbf{A}(\mathbf{A}^{-1}\phi_{,t})_{,t}-\mathbf{B}(\mathbf{B}^{-1}\phi_{,z})_{,z}=0, \tag{47}$$

where now a dot or a prime denotes differentiation with respect to t or z, respectively. Moreover, in the gauge (41) and (42) the line element reads

$$(ds)^2=e^{2\nu}[(dt)^2-(dz)^2]-\mathbf{AB}[\chi^{-1}(dx^1)^2+\chi(dx^2)^2]. \tag{48}$$

In the present variables the fluid's quantities are determined from the stream potential by [Eqs. (34) and (35)]

$$2\epsilon-k=\frac{e^{-2\nu}}{\mathbf{A}^2\mathbf{B}^2}(\phi_{,z}^2-\phi_{,t}^2),\quad u_{(t)}=\frac{\phi_{,z}}{\sqrt{\phi_{,z}^2-\phi_{,t}^2}},\quad u_{(z)}=\phi_{,t}/\sqrt{\phi_{,z}^2-\phi_{,t}^2}. \tag{49}$$

Finally, we express the right-hand sides of Eqs. (45) and (46) in terms of ϕ. They become

$$\frac{2\dot{\mathbf{A}}}{\mathbf{A}}\nu_{,z}+\frac{2\mathbf{B}'}{\mathbf{B}}\nu_{,t}-\frac{\dot{\mathbf{A}}\mathbf{B}'}{\mathbf{AB}}=\frac{\chi_{,t}\chi_{,z}}{\chi^2}-\frac{4\phi_{,t}\phi_{,z}}{\mathbf{A}^2\mathbf{B}^2} \tag{50}$$

and

$$\frac{4\mathbf{B}'}{\mathbf{B}}\nu_{,z}+\frac{4\dot{\mathbf{A}}}{\mathbf{A}}\nu_{,t}+\left(\frac{\dot{\mathbf{A}}^2}{\mathbf{A}^2}+\frac{\mathbf{B}'^2}{\mathbf{B}^2}\right)-2\left(\frac{\ddot{\mathbf{A}}}{\mathbf{A}}+\frac{\mathbf{B}''}{\mathbf{B}}\right)=\frac{\chi_{,t}^2+\chi_{,z}^2}{\chi^2}-\frac{4(\phi_{,t}^2+\phi_{,z}^2)}{\mathbf{A}^2\mathbf{B}^2}. \tag{51}$$

To summarize, we have reduced the problem on hand to solving the system of the five equations (43), (44), (47), (50), and (51) for the unknowns χ, ϕ, and ν, which are functions of t and z, and for $\mathbf{A}=\mathbf{A}(t)$ and $\mathbf{B}=\mathbf{B}(z)$. The line element is given by the expression (48) while the quantities of the fluid are obtained from Eqs. (49). The difficulty of the problem is apparent: all five equations of the system should be considered simultaneously.

IV. CONJUGATE EQUATIONS

Multiplied by $\mathbf{AB}$, Eq. (44) can be written as a total divergence and it can be "solved" by introducing a potential function G,

$$\mathbf{AB}(\ln\chi)_{,t}=G_{,z},\quad \mathbf{AB}(\ln\chi)_{,z}=G_{,t}. \tag{52}$$

Then, the existence of χ implies the integrability condition

$$(G_{,t}/\mathbf{AB})_{,t}-(G_{,z}/\mathbf{AB})_{,z}=0, \tag{53}$$

which is precisely the stream potential equation (47).

Although they satisfy the same (linear, hyperbolic) equation, ϕ and G generally would be different solutions of this equation. *We shall restrict our considerations, from now on, only to perfect-fluid solutions for which G and ϕ are proportional,*

$$G=c\phi, \tag{54}$$

with a constant proportionality factor. By using Eqs. (52) and (54) we find that the right-hand sides of Eqs. (50) and (51) become

$$(c^2-4)\frac{\phi_{,t}\phi_{,z}}{\mathbf{A}^2\mathbf{B}^2}\quad\text{and}\quad\frac{(c^2-4)}{\mathbf{A}^2\mathbf{B}^2}(\phi_{,t}^2+\phi_{,z}^2), \tag{55}$$

respectively; and by choosing $c=\pm 2$, the unknowns χ and ϕ drop out from Eqs. (50) and (51) and the problem simplifies considerably. *We shall adopt the choice* $c=+2$. [$c=-2$ amounts to changing $\chi\to\chi^{-1}$ which, by the expression (48), corresponds to the mere interchange of the Killing fields $\partial/\partial x^1$ and $\partial/\partial x^2$.] The field equations now become

$$\frac{2\dot{\mathbf{A}}}{\mathbf{A}}\nu_{,z}+\frac{2\mathbf{B}'}{\mathbf{B}}\nu_{,t}=\frac{\dot{\mathbf{A}}\mathbf{B}'}{\mathbf{AB}}, \tag{56}$$

$$\frac{4\mathbf{B}'}{\mathbf{B}}\nu_{,z}+\frac{4\dot{\mathbf{A}}}{\mathbf{A}}\nu_{,t}=2\left(\frac{\ddot{\mathbf{A}}}{\mathbf{A}}+\frac{\mathbf{B}''}{\mathbf{B}}\right)-\left(\frac{\dot{\mathbf{A}}^2}{\mathbf{A}^2}+\frac{\mathbf{B}'^2}{\mathbf{B}^2}\right), \tag{57}$$

$$(\ln\chi)_{,t}=\frac{2}{\mathbf{AB}}\phi_{,z},\quad(\ln\chi)_{,z}=\frac{2}{\mathbf{AB}}\phi_{,t}, \tag{58}$$

supplemented by Eqs. (43) and (47). Note that the original system decouples, providing three equations—Eqs. (43), (56), and (57)—for the unknowns $\mathbf{A}(t)$, $\mathbf{B}(z)$, and $\nu(t,z)$.

The next step is to eliminate ν from Eqs. (56) and (57); it can be readily achieved since ν can be determined algebraically from Eq. (43). The price that we pay for the elimination of ν is that we increase the order of Eqs. (56) and (57) from second to third. We obtain

$$\frac{\dot{\mathbf{A}}}{\mathbf{A}}\left(\frac{\mathbf{B}''}{\mathbf{B}}\right)'-\frac{\mathbf{B}'}{\mathbf{B}}\left(\frac{\ddot{\mathbf{A}}}{\mathbf{A}}\right)^{\cdot}=\frac{\dot{\mathbf{A}}\mathbf{B}'}{\mathbf{AB}}\left(\frac{\mathbf{B}''}{\mathbf{B}}-\frac{\ddot{\mathbf{A}}}{\mathbf{A}}\right), \tag{59}$$

$$\frac{2\mathbf{B}'}{\mathbf{B}}\left(\frac{\mathbf{B}''}{\mathbf{B}}\right)'-\frac{2\dot{\mathbf{A}}}{\mathbf{A}}\left(\frac{\ddot{\mathbf{A}}}{\mathbf{A}}\right)^{\cdot}=\left(\frac{\mathbf{B}''}{\mathbf{B}}-\frac{\ddot{\mathbf{A}}}{\mathbf{A}}\right)\left(\frac{2\ddot{\mathbf{A}}}{\mathbf{A}}-\frac{\dot{\mathbf{A}}^2}{\mathbf{A}^2}+\frac{2\mathbf{B}''}{\mathbf{B}}-\frac{\mathbf{B}'^2}{\mathbf{B}^2}\right). \tag{60}$$

The strategy for solving the problem should now be obvious. First we should solve Eqs. (59) and (60) for the gauge functions $\mathbf{A}(t)$ and $\mathbf{B}(z)$. Then ν will be determined algebraically from Eq. (43). Next, we should solve, for the obtained $\mathbf{A}$ and $\mathbf{B}$, the linear equation (47) for the stream potential ϕ. And finally, we should determine χ from Eqs. (58) via quadratures. From the form of the linear equation (47) it is obvious that it admits separable solutions. The difficult step is, therefore, to solve Eqs. (59) and (60) for the two-gauge functions.

V. DETERMINATION OF THE GAUGE FUNCTIONS

We start with the remark that, since $\mathbf{A}=\mathbf{A}(t)$ and $\mathbf{B}=\mathbf{B}(z)$, Eqs. (59) and (60) are overdetermined. Thus, although they are quite involved, the chances that we might be able to solve them are quite good. But there also exists the chance that they may admit no nontrivial solution.

To simplify the notation we set

$$\mathbf{A}=\exp[P(t)], \quad \mathbf{B}=\exp[Q(z)]. \tag{61}$$

After some reductions we find that the variables in Eq. (59) separate, leading to the two ordinary differential equations,

$$\begin{aligned}&(\dot{P})^{-1}(\ddot{P}+\dot{P}^2)^{\cdot}-(\ddot{P}+\dot{P}^2)=\alpha,\\&(Q')^{-1}(Q''+Q'^2)'-(Q''+Q'^2)=\alpha,\end{aligned} \tag{62}$$

where α is an (undetermined) separation constant. In addition, Eq. (60) becomes

$$\begin{aligned}&2Q'(Q''+Q'^2)'-2\dot{P}(\ddot{P}+\dot{P}^2)^{\cdot}\\&\quad=(Q''+Q'^2-\ddot{P}-\dot{P}^2)(2\ddot{P}+\dot{P}^2+2Q''+Q'^2),\end{aligned} \tag{63}$$

or, after eliminating the third derivatives by virtue of Eqs. (62),

$$\begin{aligned}&(\ddot{P}+\dot{P}^2)(2\ddot{P}-\dot{P}^2)-(Q''+Q'^2)(2Q''-Q'^2)\\&\quad+\dot{P}^2Q''-\ddot{P}Q'^2+2\alpha(Q'^2-\dot{P}^2)=0.\end{aligned} \tag{64}$$

Since Eqs. (62) and (64) involve only the derivatives of P and Q we can reduce their order by setting

$$\dot{P}=x=x(t), \quad Q'=y=y(z). \tag{65}$$

Equations (62) then become

$$\begin{aligned}&(\dot{x}+x^2+\alpha)^{\cdot}-x(\dot{x}+x^2+\alpha)=0,\\&(y'+y^2+\alpha)'-y(y'+y^2+\alpha)=0,\end{aligned} \tag{66}$$

while Eq. (64) simplifies to

$$\begin{aligned}&2(\dot{x}-y')(\dot{x}+y')\\&\quad+(x^2-y^2)(\dot{x}+y'-x^2-y^2-2a)=0.\end{aligned} \tag{67}$$

In the notation of x and y Eq. (43) reads

$$2ke^{2\nu}=y'+y^2-\dot{x}-x^2. \tag{68}$$

The simplest solution of Eqs. (66) is $\dot{x}+x^2+\alpha=0$, $y'+y^2+\alpha=0$, which, however, gives $2ke^{2\nu}=0$; so it is unacceptable for our present considerations.

We transform Eqs. (66) to integrodifferential ones. They become

$$\dot{x}=-x^2-\alpha+\gamma_1 I, \quad y'=-y^2-\alpha+\gamma_2 J, \tag{69}$$

where

$$\begin{aligned}&I=I(t)=\exp\left[\int x(t)dt\right],\\&J=J(z)=\exp\left[\int y(z)dz\right],\end{aligned} \tag{70}$$

and γ_1 and γ_2 are integration constants. We can now eliminate the first derivatives of x and y from Eq. (67). A lot of simplifications occur. We finally obtain

$$\begin{aligned}&\gamma_1 I(3x^2+y^2+4\alpha)-\gamma_2 J(x^2+3y^2+4\alpha)\\&\quad=2(\gamma_1^2 I^2-\gamma_2^2 J^2).\end{aligned} \tag{71}$$

The aim now is to obtain additional integrability conditions from Eqs. (69) and (71). By differentiating Eq. (71) with respect to t and z and using Eqs. (69) to eliminate the first derivatives of x and y we obtain

$$\gamma_1 I(3x^2-y^2+2\alpha)-2\gamma_2 J(x^2+\alpha)=2\gamma_1 I(\gamma_1 I-\gamma_2 J), \tag{72a}$$

$$\gamma_2 J(3y^2-x^2+2\alpha)-2\gamma_1 I(y^2+\alpha)=2\gamma_2 J(\gamma_2 J-\gamma_1 I). \tag{72b}$$

The difference of Eqs. (72) gives, once more, the condition (71) while their sum gives

$$(\gamma_1 I-\gamma_2 J)\left[x^2-y^2-\tfrac{2}{3}(\gamma_1 I-\gamma_2 J)\right]=0. \tag{73}$$

For a nontrivial solution the terms in the squared brackets of Eq. (73) should vanish. Since $x=x(t)$ and $y=y(z)$ this requirement leads to the two conditions

$$x^2-\tfrac{2}{3}\gamma_1 I=y^2-\tfrac{2}{3}\gamma_2 J=\text{const}. \tag{74}$$

Compatibility of Eqs. (69), (70), and (74) requires that the constant in Eqs. (74) should be $-\alpha$. Thus we obtain

$$x^2+\alpha=\tfrac{2}{3}\gamma_1 I, \quad y^2+\alpha=\tfrac{2}{3}\gamma_2 J. \tag{75}$$

It is then straightforward to show that the two equations (75) imply Eqs. (69) and (71) and therefore they are equivalent to the problem on hand. Hence, we do not have to consider additional integrability conditions.

It is not difficult to integrate Eqs. (75). They lead to three distinct families of solutions, according to the sign of the parameter α. For negative α we obtain

$$\begin{aligned}&x=-2b\coth(bt+\lambda_1),\\&y=-2b\coth(bz+\lambda_2), \quad \alpha=-4b^2,\\&P=-2\ln\sinh(bt+\lambda_1),\\&Q=-2\ln\sinh(bz+\lambda_2),\\&\mathbf{A}=\sinh^{-2}(bt+\lambda_1), \quad \mathbf{B}=\sinh^{-2}(bz+\lambda_2),\end{aligned} \tag{76}$$

λ_1 and λ_2 are integration constants. Then from Eq. (68) we can evaluate $e^{2\nu}$ and write the metric. It reads

$$\begin{aligned}(ds)^2=3b^2k^{-1}&[\sinh^{-2}(bz+\lambda_2)-\sinh^{-2}(bt+\lambda_1)]\\&\times[(dt)^2-(dz)^2]-\sinh^{-2}(bt+\lambda_1)\\&\times\sinh^{-2}(bz+\lambda_2)\,[\chi^{-1}(dx^1)^2+\chi(dx^2)^2].\end{aligned} \tag{77}$$

Obviously, the parameters b, λ_1, and λ_2 can be absorbed in redefinitions of t and z. The parameter k of the equation of state (1) can also be absorbed but we shall retain it, keeping in mind that only its sign is relevant. Hence, without loss of generality, the family (76) for $\alpha<0$ is family (i):

$$\begin{aligned}&(x=-2\coth t,\ y=-2\coth z),\\&(P=-2\ln\sinh t,\ Q=-2\ln\sinh z),\\&(\mathbf{A}=\sinh^{-2}t,\ \mathbf{B}=\sinh^{-2}z),\end{aligned} \tag{78}$$

with line element

$$\begin{aligned}(ds)^2=(3/k)\,&(\sinh^{-2}z-\sinh^{-2}t)\,[(dt)^2-(dz)^2]\\&-\sinh^{-2}t\sinh^{-2}z[\chi^{-1}(dx^1)^2+\chi(dx^2)^2].\end{aligned} \tag{79}$$

Similarly we find, for $\alpha>0$, family (ii):

$$\begin{aligned}&(x=2\tan t,\ y=2\tan z),\\&(P=-2\ln\cos t,\ Q=-2\ln\cos z),\end{aligned} \tag{80}$$

$$(\mathbf{A}=\cos^{-2}t,\ \mathbf{B}=\cos^{-2}z)\,,$$

with line element

$$(ds)^2=(3/k)(\cos^{-2}z-\cos^{-2}t)\,[(dt)^2-(dz)^2]$$
$$-\cos^{-2}t\cos^{-2}z\,[\chi^{-1}(dx^1)^2+\chi(dx^2)^2]\,, \quad (81)$$

and for $\alpha=0$, family (iii):

$$(x=-2/t,\ y=-2/z)\,,$$
$$(P=-2\ln t,\ Q=-2\ln z)\,, \quad (82)$$
$$(\mathbf{A}=t^{-2},\ \mathbf{B}=z^{-2})\,,$$

with line element

$$(ds)^2=(3/k)(z^{-2}-t^{-2})\,[(dt)^2-(dz)^2]$$
$$-(tz)^{-2}\,[\chi^{-1}(dx^1)^2+\chi(dx^2)^2]\,. \quad (83)$$

VI. DESCRIPTION OF THE SPACE-TIME IN THE NEWMAN–PENROSE FORMALISM

The metric (48) is of the form

$$(ds)^2=U^2[(dt)^2-(dz)^2]-[V^2/(1-\mathscr{E}^2)]$$
$$\times[(1-\mathscr{E})^2(dx^1)^2+(1+\mathscr{E})^2(dx^2)^2] \quad (84)$$

considered in Ref. 10, Sec. 6 with

$$U=e^{\nu},\quad V=\sqrt{\mathbf{AB}},\quad \mathscr{E}=\frac{\chi-1}{\chi+1}\Leftrightarrow\chi=\frac{1+\mathscr{E}}{1-\mathscr{E}}\,. \quad (85)$$

Equation (124) of Ref. 10 gives a suitable null basis. Unfortunately, in Ref. 10 the Newman–Penrose quantities were only obtained for a particular choice of V which does not apply in nonvacuum or nonelectrovacuum cases that we are now considering. So we shall give the Newman–Penrose quantities for any value of V, but for real $\mathscr{E}$.

The only nonvanishing spin coefficients are

$$\sigma=\frac{\mathscr{E}_{,t}+\mathscr{E}_{,z}}{U\sqrt{2}(1-\mathscr{E}^2)},\quad \lambda=\frac{\mathscr{E}_{,z}-\mathscr{E}_{,t}}{U\sqrt{2}(1-\mathscr{E}^2)},$$
$$\rho=-\frac{V_{,t}+V_{,z}}{UV\sqrt{2}},\quad \mu=\frac{V_{,t}-V_{,z}}{UV\sqrt{2}}, \quad (86)$$
$$\varepsilon=\frac{U_{,t}+U_{,z}}{2U^2\sqrt{2}},\quad \gamma=\frac{U_{,z}-U_{,t}}{2U^2\sqrt{2}},$$

and the only nonvanishing Weyl and Ricci scalars are

$$\Psi_0=\frac{(\mathscr{E}_{,t}+\mathscr{E}_{,z})}{U^2(1-\mathscr{E}^2)}\left\{\left(\ln\frac{V}{U}\right)_{,t}+\left(\ln\frac{V}{U}\right)_{,z}+\frac{\mathscr{E}_{,tt}+\mathscr{E}_{,zz}+2\mathscr{E}_{,tz}}{2(\mathscr{E}_{,t}+\mathscr{E}_{,z})}+\frac{\mathscr{E}(\mathscr{E}_{,t}+\mathscr{E}_{,z})}{1-\mathscr{E}^2}\right\}, \quad (87a)$$

$$\Psi_4=\frac{(\mathscr{E}_{,t}-\mathscr{E}_{,z})}{U^2(1-\mathscr{E}^2)}\left\{\left(\ln\frac{V}{U}\right)_{,t}-\left(\ln\frac{V}{U}\right)_{,z}+\frac{\mathscr{E}_{,tt}+\mathscr{E}_{,zz}-2\mathscr{E}_{,tz}}{2(\mathscr{E}_{,t}-\mathscr{E}_{,z})}+\frac{\mathscr{E}(\mathscr{E}_{,t}-\mathscr{E}_{,z})}{1-\mathscr{E}^2}\right\}, \quad (87b)$$

$$\Psi_2=\frac{1}{6U^2}\left[\left(\ln\frac{V}{U}\right)_{,tt}-\left(\ln\frac{V}{U}\right)_{,zz}\right]+\frac{\mathscr{E}_{,t}^2-\mathscr{E}_{,z}^2}{3U^2(1-\mathscr{E}^2)}, \quad (87c)$$

$$\Phi_{00}=\frac{(V_{,t}+V_{,z})\,(U_{,t}+U_{,z})}{U^3V}-\frac{V_{,tt}+V_{,zz}+2V_{,tz}}{2U^2V}-\frac{(\mathscr{E}_{,t}+\mathscr{E}_{,z})^2}{2U^2(1-\mathscr{E}^2)^2}, \quad (87d)$$

$$\Phi_{22}=\frac{(V_{,t}-V_{,z})(U_{,t}-U_{,z})}{U^3V}-\frac{V_{,tt}+V_{,zz}-2V_{,tz}}{2U^2V}-\frac{(\mathscr{E}_{,t}-\mathscr{E}_{,z})^2}{2U^2(1-\mathscr{E}^2)^2}, \quad (87e)$$

$$\Phi_{11}=\frac{1}{4U^2}\left[(\ln U)_{,zz}-(\ln U)_{,tt}+(\ln V)_{,t}{}^2-(\ln V)_{,z}^2\right]+\frac{\mathscr{E}_{,z}^2-\mathscr{E}_{,t}^2}{4U^2(1-\mathscr{E}^2)^2}, \quad (87f)$$

$$\Phi_{20}=\frac{\mathscr{E}_{,zz}-\mathscr{E}_{,tt}}{2U^2(1-\mathscr{E}^2)}+\frac{\mathscr{E}(\mathscr{E}_{,z}^2-\mathscr{E}_{,t}^2)}{U^2(1-\mathscr{E}^2)^2}+\frac{\mathscr{E}_{,z}V_{,z}-\mathscr{E}_{,t}V_{,t}}{U^2V(1-\mathscr{E}^2)}, \quad (87g)$$

$$\Lambda=\frac{1}{12U^2}\left[(\ln U)_{,tt}-(\ln U)_{,zz}\right]+\frac{1}{9U^2V^{3/2}}\left[(V^{3/2})_{,tt}-(V^{3/2})_{,zz}\right]+\frac{\mathscr{E}_{,t}^2-\mathscr{E}_{,z}^2}{12U^2(1-\mathscr{E}^2)^2}. \quad (87h)$$

For the presently considered case

$$U=\frac{1}{\sqrt{2k}}\left(\frac{\mathbf{B}''}{\mathbf{B}}-\frac{\ddot{\mathbf{A}}}{\mathbf{A}}\right)^{1/2},\quad V=(\mathbf{AB})^{1/2}. \quad (88)$$

In all previous considerations of space-times with two spacelike Killing fields (vacuum, electrovacuum, $\epsilon=p$ fluids), V was fixed by gauge while the determination of U had to wait on the previous determination of $\mathscr{E}$ as a solution of the Ernst equation. In the presently considered problem, however, both U and V are determined from the gauge functions before the specification of $\mathscr{E}$. This situation opens the possibility of simplifying the expressions (87) for any $\mathscr{E}$. The simplifications which occur are really remarkable.

For the three families of solutions (78), (80), and (82) we find, after some lengthy reductions, that

$$(1/U^3V)\,(V_{,t}\pm V_{,z})\,(U_{,t}\pm U_{,z})-(1/2U^2V)\,(V_{,tt}+V_{,zz}\pm 2V_{,tz})=0\,, \quad (89a)$$

$$(\ln(V/U))_{,tt}-(\ln(V/U))_{,zz}=0\,, \quad (89b)$$

$$(\ln U)_{,zz}-(\ln U)_{,tt}+(\ln V)_{,t}{}^2-(\ln V)_{,z}{}^2=0\,, \quad (89c)$$

$$(1/12U^2)\,[(\ln U)_{,tt}-(\ln U)_{,zz}]+(1/9U^2V^{3/2})\,[(V^{3/2})_{,tt}-(V^{3/2})_{,zz}]=-(k/6)\,, \tag{89d}$$

and

$$\left(\ln\frac{V}{U}\right)_{,t}\pm\left(\ln\frac{V}{U}\right)_{,z}=\begin{cases}-\coth(t\pm z), & \text{for family (i)}\,,\\ -\cot(t\pm z), & \text{for family (ii)}\,,\\ -(t\pm z)^{-1}, & \text{for family (iii)}\,.\end{cases} \tag{90}$$

We conclude, therefore, that for the three families of solutions obtained in the present paper the expressions for the Weyl and Ricci scalars Ψ_2, Φ_{00}, Φ_{11}, Φ_{22}, and Λ simplify considerably to

$$\Psi_2=\frac{\mathscr{E}_{,t}^2-\mathscr{E}_{,z}^2}{3U^2(1-\mathscr{E}^2)^2},\quad \Phi_{00}=-\frac{(\mathscr{E}_{,t}+\mathscr{E}_{,z})^2}{2U^2(1-\mathscr{E}^2)^2},\quad \Phi_{22}=-\frac{(\mathscr{E}_{,t}-\mathscr{E}_{,z})^2}{2U^2(1-\mathscr{E}^2)^2}\,,$$

$$\Phi_{11}=\frac{(\mathscr{E}_{,z}^2-\mathscr{E}_{,t}^2)}{4U^2(1-\mathscr{E}^2)^2},\quad \Lambda=-\frac{k}{6}+\frac{\mathscr{E}_{,t}^2-\mathscr{E}_{,z}^2}{12U^2(1-\mathscr{E}^2)^2}\,, \tag{91}$$

while Ψ_0, Ψ_4, and Φ_{20} are given by Eqs. (87a), (87b), (87g), and (90).

VII. DISCUSSION

The most interesting conclusion arising from the present work is that the method of solving the Einstein perfect-fluid equations for $\epsilon=p+$ constant equation of state is completely different—almost the opposite, I would say—from the method used in solving the Einstein equations either in the vacuum case or when gravity is coupled to a perfect fluid with $\epsilon=p$ equation of state. The first step in both considerations is, of course, to simplify the metric according to the symmetries of the problem. Then, in the vacuum case, we proceed as follows: (i) we solve one of the field equations by imposing a suitable gauge condition, (ii) we solve a linear partial differential equation (the specialization of the Ernst equation to the case of hypersurface orthogonal Killing fields), and (iii) we determine, by straightforward quadratures, the conformal factor of the two-dimensional geometry orthogonal to the two Killing fields. Moreover, in the presence of a fluid with $\epsilon=p$ equation of state, steps (ii) and (iii) are performed twice, the second time for the stream potential of the fluid and its contribution to the conformal factor.

For a fluid with $\epsilon=p+$ constant equation of state, on the other hand, we had to devise a completely different strategy. In this case the main source of difficulty is that we cannot choose in early stages the gauge and simultaneously solve one of the Einstein equations, because the "gauge fixing" equation does not decouple from the remaining field equations. Instead, the strategy now is the following: (i) we use the gauge fixing equation to express algebraically the conformal factor in terms of the gauge functions **A** and **B**; (ii) we use the two linear equations for the conformal factor (the ones leading to the quadratures in the vacuum case) and we eliminate the conformal factor, obtaining a system of two coupled equations for the gauge functions **A** and **B**; and (iii) we solve the provious system for **A** and **B**. This task has to be done only once and it has been performed in the present paper; (iv) the conformal factor is then easily obtained; (v) we have to solve the linear partial differential equation, corresponding to step (ii) of the vacuum case, to determine the norms of the two Killing fields; and finally, (vi) the stream potential, which determines the characteristics of the fluid, is obtained by straightforward quadratures. Steps (v) and (vi) can be interchanged: we can solve a linear equation for the stream potential and determine the norms of the Killing fields by quadratures.

It should be emphasized that our considerations do not cover all the solutions of $\epsilon=p+k$ fluids with two hypersurface orthogonal Killing fields. Instead, the obtained families of solutions exhaust those fluids satisfying, in addition, the separability condition (42) for the gauge functions and the assumption (54) that the potential G and ϕ are proportional. The condition (54) means that while the fluid is freely distributed (subject, of course, to the field equations), the gravitational field is determined uniquely from the fluid, or vice versa. In the general case one expects that each one—the fluid and the gravitational field—would have infinite degrees of freedom, corresponding to the solutions of the two second-order, linear, hyperbolic equations (44) and (47).

That the $\epsilon=p+$ constant and the $\epsilon=p$ fluids are solved completely differently is reflected in that neither of the three families of solutions obtained in the present paper is connected continuously with solutions with $\epsilon=p$ fluids. Thus, in particular, we cannot set the constant k of the equation of state equal to zero and get solutions with $\epsilon=p$ equation of state.

Having determined the gauge functions, the eventual solution of the presently considered problem looks very similar to the solution of the vacuum Einstein equations with two hypersurface orthogonal Killing fields, the Weyl[2,13] solutions. In both cases one has to solve one second-order linear partial differential equation—elliptic for static solutions, hyperbolic for solutions with two spacelike Killing fields—equations which can be solved by separation of variables. In addition, to complete the metric, one has to perform certain quadratures. The details and the physical interpretation of each solution, of course, would depend on the particular solution of the linear equation that has been chosen, a problem that we do not address here. However, an important difference between the Weyl and the presently considered solutions should be stressed: although the Weyl solutions are all the vacuum solutions (with the appropriate symmetries), the presently considered solutions are only those fluid solutions satisfying the requirements (42) and (54) for the separability of the gauge function and the proportionality of the gravitational and the fluid potentials.

ACKNOWLEDGMENTS

I would like to thank Professor S. Chandrasekhar, of the University of Chicago, for many discussions concerning the collision of plane gravitational and hydrodynamic waves, from which I was motivated to undertake the present investigation. Also I would like to thank the referee for suggesting many improvements in the manuscript.

[1]J. Winicour, J. Math. Phys. **16**, 1806 (1975).
[2]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980).
[3]J. Wainwright, W. C. W. Ince, and B. J. Marshman, Gen. Relativ. Gravit. **10**, 259 (1979).
[4]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **402**, 37 (1985).
[5]D. Kramer, Class. Quantum Gravit. **2**, L135 (1985).
[6]P. Szekeres, Nature (London) **228**, 1183 (1970).
[7]K. Khan and R. Penrose, Nature (London) **229**, 185 (1971).
[8]S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984).
[9]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[10]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 223 (1985).
[11]A. Papapetrou, Ann. Phys. (Leipzig) **12**, 309 (1953).
[12]F. Ernst, Phys. Rev. **167**, 1175 (1968); **168**, 1415 (1968).
[13]H. Weyl, Ann. Phys. (Leipzig) **54**, 117 (1917); **59**, 185 (1919).

The anharmonic oscillator driven by Gaussian noise. Comparison between direct perturbation and Markov approximation methods

L. Pesquera and R. Blanco
Departamento de Física Teórica, Universidad de Santander, Avenida Los Castros s/n 39005, Santander, Spain

(Received 16 June 1986; accepted for publication 26 November 1986)

The anharmonic oscillator driven by Gaussian noise is studied in the limit of weak damping using the direct perturbation (DPM) and Markov approximation (MAM) methods. Mean values are obtained to first order in the anharmonic coupling constant g. From a careful treatment of the high-frequency behavior it is concluded that to first order in g the DPM takes high-frequency contributions into account whereas the MAM does not, while both agree if high-frequency contributions are not important. It is also shown that both methods give the same results to second order in g for the quartic anharmonic oscillator. The spectral density of the noise used in stochastic electrodynamics is considered as a particular example.

I. INTRODUCTION

In previous papers[1,2] the effect of a nonwhite Gaussian noise on the motion of a nonlinear oscillator was considered. The first reference used the direct perturbation method (DPM) taken over from classical perturbation theory.[3] The second reference used the Markov approximation method (MAM) based on an approximate Fokker–Planck equation in terms of the constants of the deterministic motion.[4,5]

The results in Refs. 1 and 2 referred explicitly to the quartic anharmonic oscillator in stochastic electrodynamics[6] (SED), meaning that the charge moves in a classical potential with x^2 and x^4 terms while being driven by a stochastic electric field with spectral density given by

$$S_{\mathscr{E}}(\omega) = \hbar|\omega|^3/3\pi c^3. \tag{1.1}$$

The fact that the DPM and the MAM should have wide applicability makes these results disturbing in a wider context because different answers were obtained.

In this paper we will therefore analyze the nonlinear system corresponding to a harmonic oscillator perturbed by a potential $gV(x)$:

$$\ddot{x} + \omega_0^2 x + \frac{g}{m} V'(x) + \tau\left(\omega_0^2 + \frac{g}{m} V''(x)\right)\dot{x} = \frac{e}{m}\mathscr{E}(t) = f(t). \tag{1.2}$$

Here $\tau = 2e^2/3mc^3$ and the equation corresponds to that for a charged particle in a stochastic electric field $\mathscr{E}(t)$, where the damping $\tau\ddot{x}$ has been approximated by[7] $\tau(d/dt)\ddot{x}_n$, where $\ddot{x}_n$ denotes the acceleration when no charge is present. Equation (1.2) is thus without runaway solutions.

The aim of the paper is to compare the two methods, and to obtain mean values in the stationary state, in particular that of the energy, to first order in the anharmonic coupling constant g. We will consider that both the form of $V(x)$ and the spectral density of $\mathscr{E}(t)$ are arbitrary functions. Contrary to earlier results,[1] we will show that the DPM and MAM coincide to order g when there are no high-frequency contributions from the spectral density of the random field $\mathscr{E}(t)$. We will also show that these methods give the same results to order g^2 for the quartic anharmonic oscillator.

II. THE DIRECT PERTURBATION METHOD (DPM)

Assuming the standard perturbation expansion[1,3]

$$x(t) = x_0(t) + gx_1(t) + g^2x_2(t) + \cdots, \tag{2.1}$$

one finds the following equations from (1.2):

$$\ddot{x}_0 + \omega_0^2 x_0 + \tau\omega_0^2\dot{x}_0 = f(t), \tag{2.2a}$$

$$\ddot{x}_1 + \omega_0^2 x_1 + \tau\omega_0^2\dot{x}_1 = -\left(1 + \tau\frac{d}{dt}\right)\frac{V'(x_0(t))}{m}, \tag{2.2b}$$

$$\ddot{x}_2 + \omega_0^2 x_2 + \tau\omega_0^2\dot{x}_2 = -\left(1 + \tau\frac{d}{dt}\right)\frac{V'(x_0(t))}{m}x_1(t), \tag{2.2c}$$

and so on.

A. Calculation of $\langle x^2\rangle$

To order g we have[1,3]

$$\langle x^2\rangle \simeq \langle x_0^2\rangle + 2g\langle x_0x_1\rangle, \tag{2.3}$$

where $\langle x_0^2\rangle$ corresponds to a linear oscillator and can be obtained by using standard Fourier techniques.[1,3] The result is

$$\langle x_0^2\rangle = \int_{-\infty}^{+\infty} \frac{dw}{|F(\omega)|^2} S_f(\omega), \tag{2.4}$$

where

$$F(\omega) = \omega_0^2 - \omega^2 + i\tau\omega_0^2\omega \tag{2.5}$$

and $S_f(\omega)$, the spectral density of f, is given by

$$S_f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\theta\, e^{i\omega\theta}\langle f(0)f(\theta)\rangle. \tag{2.6}$$

Because of the form of $|F(\omega)|^2$, the integrand in (2.4) is strongly peaked about ω_0. In Ref. 1 a delta function approximation was used to obtain

$$\langle x_0^2\rangle = \int_{-\infty}^{\infty} \frac{dw\, S_f(\omega)}{|F(\omega)|^2} \simeq \frac{\pi}{\tau}\frac{S_f(\omega_0)}{\omega_0^4}. \tag{2.7}$$

However, it is important to recognize that this result is valid only if there are no high-frequency contributions[8] (for frequencies $\gtrsim \tau^{-1}$) to $\langle x_0^2\rangle$. For example, if we consider

$$S_f(\omega) = N\tau\omega^4/(1 + \tau^2\omega^2), \tag{2.8}$$

 0022-2488/87/040913-08$02.50

then

$$\langle x_0^2\rangle \simeq N\pi + 2N\int_{\omega_1 \gg \omega_0}^{\infty} d\omega\, \frac{\tau\omega^4(1+\tau^2\omega^2)^{-1}}{(\omega_0^2-\omega^2)^2+\tau^2\omega_0^4\omega^2}. \tag{2.9}$$

For $\omega \gg \omega_0$, we may write

$$\frac{\tau\omega^4(1+\tau^2\omega^2)^{-1}}{(\omega_0^2-\omega^2)^2+\tau^2\omega_0^4\omega^2} \simeq \frac{\tau}{1+\tau^2\omega^2}, \tag{2.10}$$

so that the second term in (2.11) is approximately

$$2N\int_0^{\infty}\frac{dx}{1+x^2} = N\pi. \tag{2.11}$$

Thus the high-frequency contributions are equal to the contributions from frequencies $\omega \lesssim \omega_0$.

The problem is that with the delta function one is taking the limit $\tau \to 0$ through the integral sign. In this case there is no second term in (2.9). It can be shown that this is valid when $S_f(\omega) \leqslant N\omega^{2+r}$ for $\omega \gg \omega_0$, where $0<r<1$ and N is of order 1 with respect to τ. This is the case for the spectral density of SED, that is given, taking into account the approximation leading to (1.2), by[5]

$$S_f(\omega) = \frac{e^2}{m^2}S_{g^\tau}(\omega) = \frac{\tau\hbar|\omega|^3}{2\pi m(1+\tau^2\omega^2)}. \tag{2.12}$$

It is to be noted that if there exist high-frequency contributions, the electromagnetic model described by (1.2) is incorrect. However, equations like (1.2) arise in many contexts (e.g., the Van der Pol–Duffing equation) not associated with radiation damping, so our observations may be of more general interest.

We now consider $\langle x_0 x_1\rangle$. We first calculate it following the method used in Ref. 1 in order to compare our result with the one obtained in Ref. 1. From (2.2b) we have

$$x_1(t) = \int_{-\infty}^{+\infty} d\omega\, \frac{e^{i\omega t}}{F(\omega)}\int_{-\infty}^{+\infty} ds \times \frac{e^{-i\omega s}}{2\pi}\left(\frac{-1}{m}\right)\left(1+\tau\frac{d}{ds}\right)V'(x_0(s)), \tag{2.13}$$

so that

$$\langle x_0(t')x_1(t)\rangle = -\frac{1}{2\pi m}\int_{-\infty}^{+\infty} d\omega \frac{e^{i\omega t}(1+i\tau\omega)}{F(\omega)} \times \int_{-\infty}^{+\infty} ds\, e^{-i\omega s}\langle x_0(t')V'(x_0(s))\rangle. \tag{2.14}$$

The second integrand can be simplified by using Novikov's lemma which is valid for Gaussian processes[9,10]

$$\langle x_0(t')V'(x_0(s))\rangle = \int_{-\infty}^{+\infty} du \left\langle \frac{\delta V'(x_0(s))}{\delta x_0(u)}\right\rangle \langle x_0(u)x_0(t')\rangle, \tag{2.15}$$

so that

$$\langle x_0(t')x_1(t)\rangle = -\frac{1}{2\pi m}\langle V''(x_0)\rangle \int_{-\infty}^{+\infty} d\omega \frac{e^{i\omega t}(1+i\tau\omega)}{F(\omega)} \times \int_{-\infty}^{+\infty} ds\, e^{-i\omega s}\langle x_0(t')x_0(s)\rangle. \tag{2.16}$$

We next calculate the correlation function

$$\langle x_0(t')x_0(s)\rangle = \int_{-\infty}^{+\infty} d\omega \frac{S_f(\omega)e^{i\omega(t'-s)}}{|F(\omega)|^2}. \tag{2.17}$$

In Ref. 1 the same technique of approximating the integrand to a delta function was used. Clearly an alternative is to use the residue theorem, when $S_f(\omega)$ is analytic, applied to the poles

$$\omega_{\pm} = \pm(\omega_0^2 - \tau^2\omega_0^4/4)^{1/2} + i\tau\omega_0^2/2. \tag{2.18}$$

However, one must take care if $S_f(\omega)$ has poles of order τ^{-1}, as in the case of (2.8). An analysis of these high-frequency contributions will be made below. Assuming that these contributions are negligible, we have

$$\langle x_0(t')x_0(s)\rangle \simeq (\pi/\omega_0^3)e^{-\tau\omega_0^2|t'-s|/2}[(S_f(\omega_0)/\tau\omega_0)\cos\omega_0(t'-s) + \tfrac{1}{2}(S_f(\omega_0) - \omega_0 S_f'(\omega_0))\sin\omega_0|t'-s|]. \tag{2.19}$$

This expression coincides with Eq. (3.7) of Ref. 1 in the limit $\tau \to 0$ as it must (since in Ref. 1 the delta function approximation was used, valid for $\tau \to 0$). However (and here our present calculations differ from those of Ref. 1), the limit should not be taken first when substituting in Eq. (2.16) because the term in $\sin\omega_0|t'-s|$ contributes to the same order as $\cos\omega_0(t'-s)$.

Using

$$R(t-s) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega \frac{e^{i\omega(t-s)}(1+i\tau\omega)}{F(\omega)} \simeq H(t-s)\frac{e^{-\tau\omega_0^2(t-s)/2}}{\omega_0} \times[\sin\omega_0(t-s) + \tau\omega_0\cos\omega_0(t-s)], \tag{2.20}$$

where H is the Heaviside function, and introducing (2.19) in (2.16), we obtain ($\theta = t'-t \geqslant 0$)

$$\langle x_0(t')x_1(t)\rangle \simeq -\frac{\pi}{m\omega_0^5}\langle V''(x_0)\rangle e^{-\tau\omega_0^2\theta/2}\left[\cos\omega_0\theta\left(\frac{3}{4}\frac{S_f(\omega_0)}{\tau\omega_0} + \frac{1}{4}\frac{(S_f(\omega_0)-\omega_0 S_f'(\omega_0))}{\tau\omega_0}\right) - \frac{S_f(\omega_0)}{2\tau^2\omega_0^2}\sin\omega_0\theta\right]. \tag{2.21}$$

As noted earlier, Eq. (2.21) differs from Eq. (4.11) of Ref. 1 because of the inclusion of the $\sin\omega_0|t'-s|$ term. For equal times, it reduces to

$$\langle x_0 x_1\rangle \simeq -\frac{\pi}{m\omega_0^6}\frac{\langle V''(x_0)\rangle}{\tau}\left[S_f(\omega_0) - \frac{\omega_0}{4}S_f'(\omega_0)\right]. \tag{2.22}$$

Thus we finally have

$$\langle x^2\rangle \simeq \langle x_0^2\rangle - \frac{2\pi g}{\tau m\omega_0^6}\langle V''(x_0)\rangle \times\left[S_f(\omega_0) - \frac{\omega_0}{4}S_f'(\omega_0)\right]. \tag{2.23}$$

This result has been obtained by ignoring the high-fre-

quency contributions. We now analyze the validity of this approximation.

From (2.16) and (2.17) we obtain

$$\langle x_0 x_1\rangle = -\frac{2}{m}\langle V''(x_0)\rangle \int_0^\infty d\omega \times \frac{S_f(\omega)[\omega_0^2-\omega^2+\tau^2\omega_0^2\omega^2]}{[(\omega_0^2-\omega^2)^2+\tau^2\omega_0^4\omega^2]^2}. \tag{2.24}$$

This integral I is calculated in the Appendix. Spectral densities such that $S_f(\omega)<N\omega^{4+r}$ for $\omega\gg\omega_0$, where $0<r<1$ and N is of order 1 with respect to τ, are shown to have negligible high-frequency contributions to I. In this case I is given by (A5) with $S=S_f$, and we recover (2.23). Comparing with the discussion below Eq. (2.11), we see that high frequencies contribute to I for spectral densities that are more divergent than those which give the same type of high-frequency contributions to $\langle x_0^2\rangle$. For example if $S_f(\omega)$ is given by (2.8) the high-frequency contributions to $\langle x_0 x_1\rangle$ come only from $V''(x_0)$ and not from the integral I.

B. Calculation of $\langle v^2\rangle$

In order to calculate $\langle v^2\rangle$ we make $S=\omega^2 S_f$ in (A5) thus giving the result

$$\langle v^2\rangle \simeq \langle v_0^2\rangle - \frac{2\pi g}{\tau m\omega_0^4}\langle V''(x_0)\rangle \times\left[\frac{S_f(\omega_0)}{2}-\frac{\omega_0}{4}S_f'(\omega_0)\right]. \tag{2.25}$$

This can also be obtained by differentiating (2.21) with respect to t and t'. According to the analysis made in the Appendix, (2.25) is valid as long as there are no high-frequency contributions, i.e., if $S_f(\omega)<N\omega^{2+r}$ for some $0<r<1$ and for all $\omega\gg\omega_0$. As concerns $\langle v_0^2\rangle$, high-frequency contributions are negligible when $S_f(\omega)<N\omega^r$ for $\omega\gg\omega_0$. Then, these contributions to $\langle v_0^2\rangle$ appear for the Raleigh–Jeans radiation.[8] For the spectral density of SED, (2.12), it is clear that $\langle v_0^2\rangle$ is divergent. Hence one introduces the electrokinetic momentum[5(b)] p_{ek}.

From the definition of $\dot p_{ek}$,

$$\dot p_{ek} = -m\omega_0^2 x - gV'(x), \tag{2.26}$$

it can be seen that p_{ek} is equivalent to $m\dot x$, only more regular, so that $\langle p_{ek}^2\rangle/m^2$ will be the expression (2.25) when there are no high-frequency contributions to $\langle v^2\rangle$. However, it is also evident from (2.26) that $\langle p_{ek}^2\rangle$ can be finite only when $S_f(0)=0$. This is usually the case in the context of radiation theory (Rayleigh–Jeans radiation, zero-point field radiation of SED). We do not give here the explicit calculations, but it is straightforward to show that $\langle p_{ek}^2\rangle/m^2$ indeed coincides with (2.25) when there are neither high-frequency nor zero-frequency contributions.

C. Calculation of $\langle E\rangle$

The mean energy of the moving charge is given by

$$\langle E\rangle = \tfrac{1}{2}m\langle v^2\rangle + \tfrac{1}{2}m\omega_0^2\langle x^2\rangle + g\langle V(x)\rangle. \tag{2.27}$$

Thus if there are no high-frequency contributions, we get to order g from (2.7), (2.23), and (2.25),

$$\langle E\rangle \simeq \frac{m\pi S_f(\omega_0)}{\tau\omega_0^2} - \frac{\pi g}{\tau\omega_0^4}\langle V''(x_0)\rangle \times\left[\frac{3}{2}S_f(\omega_0)-\frac{\omega_0}{2}S_f'(\omega_0)\right]+g\langle V(x_0)\rangle. \tag{2.28}$$

At this point, it is interesting to consider the special case of SED. From the general program for work in this area,[6] one can ask when does (2.28) agree with quantum mechanics (QM). One sees that such an agreement can occur only if $S_f(\omega)$ behaves like ω^3 for low frequencies, for in that case the term in square brackets is zero. We will see later within the context of the MAM that this is due to the fact that the average action is independent of the system to first order in g only if $S_f(\omega)$ behaves like ω^3.

III. THE MARKOV APPROXIMATION METHOD (MAM)

The Markov approximation method (MAM) is based on the observation that the damping and stochastic forces are much smaller than the deterministic force due to the presence of the small parameter τ. As stated earlier this method uses an approximate Fokker–Planck equation expressed in terms of the constants of the deterministic motion. Details may be found elsewhere[2,4,5]; only the few equations needed in the development given here will be considered. In the case of the one-dimensional anharmonic oscillator the only constant of the motion is the energy. The properly normalized phase-space probability density is[2,5]

$$W_0(E) = C\exp\left[-\int^E dE'\frac{G^E(E')}{G^{EE}(E')}\right], \tag{3.1}$$

where

$$G^E = 2\pi\tau m\omega^3\sum_n n^4|x_n|^2, \tag{3.2a}$$

$$G^{EE} = 2\pi^2 m^2\omega\sum_n S_f(n\omega)n^2|x_n|^2, \tag{3.2b}$$

where ω is the frequency and the x_n are the Fourier components of the conservative deterministic motion.

To first order in g we have that

$$G^E/G^{EE}\simeq\tau\omega^2/\pi m S_f(\omega). \tag{3.3}$$

One may develop ω in powers of the coupling constant [formulas (9.192)–(9.197) of Ref. 11]:

$$\omega\simeq\omega_0\left(1+\frac{\omega_0}{2\pi}g\frac{d}{dE}\oint_E V\right). \tag{3.4}$$

Here

$$\oint_E V = \int_0^{2\pi/\omega_0} ds\, V\left(\left(\frac{2E}{m\omega_0^2}\right)^{1/2}\cos\omega_0 s\right) \tag{3.5}$$

is the integral of the potential over the unperturbed orbit corresponding to the energy E.

From (3.3) and (3.4) we obtain

$$\frac{G^E}{G^{EE}}\simeq\frac{\tau\omega_0^2}{\pi m S_f(\omega_0)}\left[1+\left(2-\frac{\omega_0 S_f'(\omega_0)}{S_f(\omega_0)}\right)\frac{g\omega_0}{2\pi}\frac{d}{dE}\oint_E V\right]. \tag{3.6}$$

This permits us to write the phase-space density in the form

$$W_0(E) = Ne^{-BE}\exp\left(-gC\oint_E V\right), \tag{3.7}$$

where [see (2.28)]

$$B = \tau\omega_0^2/\pi m S_f(\omega_0) = \langle E\rangle_{g=0}^{-1} \equiv \langle E\rangle_0^{-1}, \tag{3.8a}$$

$$C = (2 - (\omega_0 S_f'(\omega_0)/S_f(\omega_0)))\omega_0/2\pi\langle E\rangle_0. \tag{3.8b}$$

A. Calculation of $\langle x^2\rangle$

From (3.7) we find that

$$\langle x^2\rangle \simeq \frac{\int\int_{-\infty}^{+\infty} x^2 e^{-E_0/\langle E\rangle_0}(1 - gV/\langle E\rangle_0 - gC\oint_{E_0} V)dx\,dp}{\int\int_{-\infty}^{+\infty} e^{-E_0/\langle E\rangle_0}(1 - gV/\langle E\rangle_0 - gC\oint_{E_0} V)dx\,dp} \tag{3.9}$$

with $E_0 = \frac{1}{2}m\omega_0^2x^2 + (p^2/2m)$. We now calculate the various terms involved in (3.9).

First, let us define

$$\frac{\omega_0}{2\pi\langle E\rangle_0}\int\int_{-\infty}^{+\infty} e^{-E_0/\langle E\rangle_0} f(E_0)dx\,dp \equiv \langle f(E)\rangle_0. \tag{3.10}$$

Then

$$\int\int_{-\infty}^{+\infty} e^{-E_0/\langle E\rangle_0}\varphi(E_0)\left[\oint_{E_0} V\right]dx\,dp$$

$$= \int_0^{\infty}\int_0^{2\pi/\omega_0} e^{E_0/\langle E\rangle_0}\varphi(E_0)\left[\oint_{E_0} V\right]dE_0\,ds, \tag{3.11}$$

where s is the time along the path. Thus this integral reduces to

$$\int_0^{\infty} e^{-E_0/\langle E\rangle_0}\varphi(E_0)\left[\int_0^{2\pi/\omega_0} V\,ds\right]dE_0$$

$$= \langle V\varphi\rangle_0\left(\frac{2\pi}{\omega_0}\right)^2\langle E\rangle_0^{-1}. \tag{3.12}$$

We also have

$$m^2\omega_0^2\int\int_{-\infty}^{+\infty} x^2 e^{-E_0/\langle E\rangle_0}\left[\oint_{E_0} V\right]dx\,dp$$

$$= \int\int_{-\infty}^{+\infty} p^2 e^{-E_0/\langle E\rangle_0}\left[\oint_{E_0} V\right]dx\,dp, \tag{3.13}$$

so that

$$\int\int_{-\infty}^{+\infty} x^2 e^{-E_0/\langle E\rangle_0}\left[\oint_{E_0} V\right]dx\,dp$$

$$= \int\int_{-\infty}^{+\infty} \frac{E_0}{m\omega_0^2} e^{-E_0/\langle E\rangle_0}\left[\oint_{E_0} V\right]dx\,dp$$

$$= \frac{\langle E_0 V\rangle_0}{m\omega_0^2\langle E\rangle_0}\left(\frac{2\pi}{\omega_0}\right)^2. \tag{3.14}$$

So finally

$$\langle x^2\rangle \simeq \frac{\langle x^2\rangle_0 - (g/\langle E\rangle_0)\langle Vx^2\rangle_0 - gC(2\pi/m\omega_0^3)\langle E_0 V\rangle_0}{1 - (g/\langle E\rangle_0)\langle V\rangle_0 - gC(2\pi/\omega_0)\langle V\rangle_0}. \tag{3.15}$$

Now, for order 0, (x,p) is a Gaussian process, so that using functional techniques,[10] we have

$$\langle Vx^2\rangle_0 = \langle V\rangle_0\langle x^2\rangle_0 + \langle V''\rangle_0\langle x^2\rangle_0^2. \tag{3.16}$$

Also,

$$\langle x^2\rangle_0 = \frac{\langle E\rangle_0}{m\omega_0^2} = \frac{\pi S_f(\omega_0)}{\tau\omega_0^4}. \tag{3.17}$$

Thus we obtain the same Eq. (2.28) with $\langle x_0^2\rangle$ given by Eq. (2.7). However, since the expressions given by (2.7) and (2.28) are only valid if no high-frequency contributions are present, we may conclude that the results obtained by the MAM are correct only if no high-frequency contributions are present.

B. Calculation of $\langle v^2\rangle$

The only difference with the calculation of the previous section is the term $\langle v^2 V\rangle_0 = \langle v^2\rangle_0\langle V\rangle_0$ which appears in place of $\langle x^2 V\rangle_0$. Thus it is easy to see that the result here coincides with (2.25) with $\langle v_0^2\rangle = \pi S_f(\omega_0)/\tau\omega_0^2$. Again the MAM gives the correct result to first order in g if no high-frequency contributions are present. If we consider p_{ek} in order to define $\langle E\rangle$ the MAM gives the correct result for spectral densities satisfying $S_f(\omega) \leqslant N\omega^{2+r}$ for $\omega \gg \omega_0$ and $0 < r < 1$, as is the case for the spectral density of SED, (2.12). [Note that for the calculation of $\langle p_{ek}^2\rangle/m^2$ one must take $S = \omega_0^4 S_f/\omega^2$ in (A5).]

C. A special property of the ω^3 spectrum

Consider now the spectral density of SED (2.12). In this case we have seen that using p_{ek}, $\langle E\rangle \simeq \langle E\rangle_0 + g\langle V\rangle_0$ [cf. (2.28)]. Taking into account (2.12), (3.1), and (3.3) we have that

$$W_0 \simeq N\exp\left[-\int^E dE'\frac{2}{\hbar\omega(E')}\right] = Ne^{-2J/\hbar}. \tag{3.18}$$

Thus $J = \hbar/2$, independent of the potential. This result is correct to first order in g.

Note that this permits a simple calculation of the energy through the classical relation

$$E = \int dJ\,\omega(J) \tag{3.19}$$

combined with (3.4). We find again that

$$\langle E\rangle \simeq \hbar\omega_0/2 + g\langle V\rangle_0. \tag{3.20}$$

IV. COMPARISON BETWEEN DPM AND MAM TO SECOND ORDER

We have shown in previous sections that, if there are no high-frequency contributions, the DPM and MAM give the same mean values to first order in the anharmonic coupling constant g. In this section we compare the two methods taking into account the second order in g for the quartic anharmonic oscillator $V = mx^4$.

(A) We first consider the Markov approximation. The phase-space probability density is given by (3.1), where G^E/G^{EE} can be obtained from (3.2). To second order in g, G^E/G^{EE} contains contributions of the spectral density, S_f,

at frequencies given by the first harmonic $3\omega_0$ and not only at frequencies close to ω_0^2:

$$\frac{G^E}{G^{EE}}=\frac{\tau\omega(E)}{\pi m S_f(\omega(E))}\left[1+\frac{9}{16}\left(\frac{Eg}{m\omega_0^4}\right)^2\times\left(9-\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right)+O(g^3)\right]. \tag{4.1}$$

Using[2]

$$\omega(E)\simeq\omega_0\left[1+3\frac{gE}{m\omega_0^4}-\frac{69}{4}\left(\frac{gE}{m\omega_0^4}\right)^2\right] \tag{4.2}$$

in (4.1), we get the phase-space probability density to second order in g in the following form:

$$W_0(E)=Ne^{-E/\langle E\rangle_0}\left[1-\frac{\alpha gE^2}{2m\omega_0^4\langle E\rangle_0}+\frac{\alpha^2}{8}\left(\frac{gE^2}{m\omega_0^4\langle E\rangle_0}\right)^2-\frac{\beta}{3}\left(\frac{gE}{m\omega_0^4}\right)^2\frac{E}{\langle E\rangle_0}\right], \tag{4.3}$$

where $\langle E\rangle_0$ is given by (3.8a) and

$$\alpha=6-3\omega_0[S_f'(\omega_0)/S_f(\omega_0)], \tag{4.4a}$$

$$\beta=-\frac{327}{16}-\frac{3}{4}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}-\frac{9}{2}\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)}+9\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2-\frac{9}{16}\frac{S_f(3\omega_0)}{S_f(\omega_0)}. \tag{4.4b}$$

From (4.3) we obtain the mean energy to second order in g,

$$\langle E\rangle\simeq\frac{\pi m S_f(\omega_0)}{\tau\omega_0^2}\left[1+\frac{\pi g S_f(\omega_0)}{\tau\omega_0^6}\times\left(-15+6\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)+\left(\frac{\pi g S_f(\omega_0)}{\tau\omega_0^6}\right)^2\times\left(\frac{5493}{8}-\frac{819}{2}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}+36\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2+27\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)}+\frac{27}{8}\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right)\right]. \tag{4.5}$$

(B) As concerns the direct perturbation method, to second order in g we have

$$\langle x^2\rangle\simeq\langle x_0^2\rangle+2g\langle x_0x_1\rangle+g^2(\langle x_1^2\rangle+2\langle x_0x_2\rangle). \tag{4.6}$$

From (2.13) it is straightforward to get when $V=mx^4$,

$$\langle x_1^2\rangle=96\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}d\omega_1\int_{-\infty}^{+\infty}d\omega_2|\tilde R(\omega)|^2\tilde C(\omega_1)\times\tilde C(\omega_2)\tilde C(\omega-\omega_1-\omega_2)+144\langle x_0^2\rangle^2\int_{-\infty}^{+\infty}d\omega|\tilde R(\omega)|^2\tilde C(\omega), \tag{4.7}$$

where [see (2.17) and (2.20)]

$$\tilde C(\omega)=S_f(\omega)/[(\omega^2-\omega_0^2)^2+\tau^2\omega_0^4\omega^2], \tag{4.8a}$$

$$\tilde R(\omega)=(1+i\tau\omega)/(\omega_0^2-\omega^2+i\tau\omega_0^2\omega), \tag{4.8b}$$

and $\langle x_0^2\rangle$ is given by (2.4). In a similar way we get from (2.2)

$$\langle x_0x_2\rangle=288\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}d\omega_1\int_{-\infty}^{+\infty}d\omega_2\,\tilde R(\omega)\tilde R(\omega_1)\tilde C(\omega)\tilde C(\omega_2)\tilde C(\omega-\omega_1-\omega_2)+144\langle x_0^2\rangle^2\int_{-\infty}^{+\infty}d\omega\,\tilde R^2(\omega)\tilde C(\omega)+288\langle x_0^2\rangle\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}d\omega_1\,\tilde R(\omega)\tilde R(\omega_1)\tilde C(\omega)\tilde C(\omega_1). \tag{4.9}$$

If we consider spectral densities, S_f, such that there are no high-frequency contributions, $\langle x_0^2\rangle$ is given by (2.7), and the calculation of the integrals in (4.7) and (4.9) is straightforward, but very cumbersome. We give here just the results

$$\int_{-\infty}^{+\infty}d\omega|\tilde R(\omega)|^2\tilde C^3(\omega)\simeq\frac{1}{\omega_0^4}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[\frac{3}{16\tau^2\omega_0^2}+\frac{217}{256}-\frac{81}{128}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}+\frac{27}{128}\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)}+\frac{3}{64}\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2+\frac{3}{256}\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right], \tag{4.10a}$$

$$\int_{-\infty}^{+\infty}d\omega|\tilde R(\omega)|^2\tilde C(\omega)\simeq\frac{1}{\omega_0^4}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)\left[\frac{1}{2\tau^2\omega_0^2}+1-\frac{5}{16}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}+\frac{\omega_0^2}{16}\frac{S_f''(\omega_0)}{S_f(\omega_0)}\right], \tag{4.10b}$$

$$\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}d\omega_1\,\tilde R(\omega)\tilde R(\omega_1)\tilde C(\omega)\tilde C^2(\omega-\omega_1)\simeq\frac{1}{\omega_0^4}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[-\frac{1}{32\tau^2\omega_0^2}+\frac{69}{256}-\frac{15}{256}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}-\frac{1}{128}\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2-\frac{5}{256}\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)}+\frac{1}{256}\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right], \tag{4.10c}$$

$$\int_{-\infty}^{+\infty}d\omega\,\tilde R^2(\omega)\tilde C(\omega)\simeq\frac{1}{\omega_0^4}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)\left[-\frac{1}{4\tau^2\omega_0^2}+1-\frac{13}{32}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}+\frac{\omega_0^2}{32}\frac{S_f''(\omega_0)}{S_f(\omega_0)}\right]. \tag{4.10d}$$

As concerns the last integral in (4.9), it can be obtained from (A5). From (2.7), (4.7), (4.9), and (4.10) we get the second-order contribution to $\langle x^2\rangle$

$$\langle x_1^2\rangle + 2\langle x_0 x_2\rangle \simeq \frac{1}{\omega_0^4}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[\frac{9957}{8} - \frac{1089}{2}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)} + 36\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2 + 27\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)} + \frac{27}{8}\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right]. \tag{4.11}$$

It is straightforward to check that the same result is obtained by using the MAM from (4.3). We note that a previous calculation[12] for a quartic anharmonic asocillator in SED showed a divergence for the DPM to order g^2. We have shown that a correct evaluation of (4.7) and (4.9) lead to the convergent result (4.11).

As concerns $\langle v^2\rangle$, we have

$$\langle v^2\rangle \simeq \langle v_0^2\rangle + 2g\langle v_0 v_1\rangle + g^2(\langle v_1^2\rangle + 2\langle v_0 v_2\rangle). \tag{4.12}$$

It is easy to show that $\langle v_1^2\rangle$ and $\langle v_0 v_2\rangle$ are given by (4.7) and (4.9) if we introduce a factor ω^2 in the integrals. The calculation of the integrals appearing in these expressions leads to

$$\int_{-\infty}^{+\infty} d\omega\,\omega^2|\tilde{R}(\omega)|^2\tilde{C}^3(\omega) \simeq \frac{1}{\omega_0^2}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[\frac{3}{16\tau^2\omega_0^2} + \frac{153}{256} - \frac{69}{128}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)} + \frac{3}{64}\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2 + \frac{27}{128}\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)} + \frac{3}{256}\frac{S_f(3\omega_0}{S_f(\omega_0)}\right], \tag{4.13a}$$

$$\int_{-\infty}^{+\infty} d\omega\,\omega^2|\tilde{R}(\omega)|^2\tilde{C}(\omega) \simeq \frac{\pi S_f(\omega_0)}{\tau\omega_0^6}\left[\frac{1}{2\tau^2\omega_0^2} + \frac{1}{2} - \frac{\omega_0}{16}\frac{S_f'(\omega_0)}{S_f(\omega_0)} + \frac{\omega_0^2}{16}\frac{S_f''(\omega_0)}{S_f(\omega_0)}\right], \tag{4.13b}$$

$$\int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} d\omega_1\,\omega^2\tilde{R}(\omega)\tilde{R}(\omega_1)\tilde{C}(\omega)\tilde{C}^2(\omega-\omega_1) \simeq \frac{1}{\omega_0^2}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[-\frac{1}{32\tau^2\omega_0^2} + \frac{5}{256} - \frac{3\omega_0}{256}\frac{S_f'(\omega_0)}{S_f(\omega_0)} - \frac{1}{128}\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2 - \frac{5\omega_0^2}{256}\frac{S_f''(\omega_0)}{S_f(\omega_0)} + \frac{1}{256}\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right], \tag{4.13c}$$

$$\int_{-\infty}^{+\infty} d\omega\,\omega^2\tilde{R}^2(\omega)\tilde{C}(\omega) \simeq \frac{\pi S_f(\omega_0)}{\tau\omega_0^6}\left[-\frac{1}{4\tau^2\omega_0^2} + \frac{1}{4} - \frac{9\omega_0}{32}\frac{S_f'(\omega_0)}{S_f(\omega_0)} + \frac{\omega_0^2}{32}\frac{S_f''(\omega_0)}{S_f(\omega_0)}\right]. \tag{4.13d}$$

Using (2.7), (4.7), (4.9), (4.13), and (A5), we get

$$\langle v_1^2\rangle + 2\langle v_0 v_2\rangle \simeq \frac{1}{\omega_0^2}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[\frac{4005}{8} - \frac{729}{2}\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)} + 36\left(\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)^2 + 27\omega_0^2\frac{S_f''(\omega_0)}{S_f(\omega_0)} + \frac{27}{8}\frac{S_f(3\omega_0)}{S_f(\omega_0)}\right]. \tag{4.14}$$

It can be again checked that MAM leads to the same result.

Finally, in order to obtain $\langle E\rangle$ for the quartic anharmonic oscillator to second order in g, we need to evaluate $\langle x^4\rangle$ to first order in g, given by

$$\langle x^4\rangle \simeq \langle x_0^4\rangle + 4g\langle x_0^3 x_1\rangle. \tag{4.15}$$

Using (2.13), we obtain

$$\langle x_0^3 x_1\rangle = -36\langle x_0^2\rangle^2\int_{-\infty}^{+\infty} d\omega\,\tilde{R}(\omega)\tilde{C}(\omega) - 24\int_{-\infty}^{+\infty} d\omega\,\tilde{R}(\omega)\tilde{C}^3(\omega). \tag{4.16}$$

The integrals in (4.16) are given by

$$\int_{-\infty}^{+\infty} d\omega\,\tilde{R}(\omega)\tilde{C}(\omega) \simeq \frac{\pi S_f(\omega_0)}{\tau\omega_0^6}\left[1 - \frac{\omega_0}{4}\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right], \tag{4.17a}$$

$$\int_{-\infty}^{+\infty} d\omega\,\tilde{R}(\omega)\tilde{C}^3(\omega) \simeq \frac{1}{\omega_0^2}\left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^3\left[\frac{7}{16} - \frac{3\omega_0}{32}\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right]. \tag{4.17b}$$

Then, we get

$$\langle x^4\rangle \simeq \left(\frac{\pi S_f(\omega_0)}{\tau\omega_0^4}\right)^2\left[3 + \frac{\pi g S_f(\omega_0)}{\tau\omega_0^6}\times\left(-186 + 45\omega_0\frac{S_f'(\omega_0)}{S_f(\omega_0)}\right)\right]. \tag{4.18}$$

From (4.11), (4.14), and (4.18) it is verified that the DPM and MAM lead to the same result for $\langle E\rangle$ to second order in g given by (4.5).

V. CONCLUSIONS

In this paper we have compared two methods for solving for the motion of a charge in a random Gaussian electric

field, the direct perturbation method (DPM) and the Markov approximation method (MAM). We have shown that if there are no high-frequency contributions, the two methods coincide to first order in the anharmonic coupling constant g. If high-frequency contributions exist, the DPM includes them whereas the MAM always eliminates them. The MAM is then not valid when frequencies $\omega \gg \omega_0$ contribute, which happens for some divergent spectral densities.

Explicit results have been obtained for the mean values of $\langle x^2 \rangle$, $\langle v^2 \rangle$ and the energy to first order in g. These results are valid for arbitrary anharmonic potentials, V, and for random fields without high-frequency contributions.

We have also shown that the two methods give the same results to second order in g, for the quartic anharmonic oscillator. The agreement holds for any spectral density of the noise, such that there is no high-frequency contribution. It is reasonable to expect the same result for any perturbing potential.

Finally, we have shown that agreement of the average energy in stochastic electrodynamics (spectral density proportional to ω^3) with that of quantum mechanics, due to a precise cancellation of a term in the expression for $\langle E \rangle$ that in general would be proportional to V'', has its origin in the fact that the average action is $\hbar/2$ (to first order in the anharmonic coupling constant). However, this agreement with quantum mechanics does not continue to second order.[2] This result has been obtained using the MAM, but as we have shown the DPM leads to the same conclusion: no agreement is obtained with quantum mechanics to second order in g.

ACKNOWLEDGMENTS

Helpful discussions with Dr. E. Santos and interesting correspondence with Dr. S. M. Moore are gratefully acknowledged.

This work was supported by CAICYT Project No. 361/84 (Spain).

APPENDIX: ANALYSIS OF HIGH-FREQUENCY CONTRIBUTIONS

In this appendix we calculate the following integral:

$$I = \int_0^\infty d\omega\, S(\omega)\, \frac{\omega_0^2 - \omega^2 + \tau^2\omega_0^2\omega^2}{[(\omega_0^2 - \omega^2)^2 + \tau^2\omega_0^4\omega^2]^2}. \tag{A1}$$

The calculation of I to first order in $\tau\omega_0$ can be done by dividing the interval of integration into three subintervals: $J_1 \equiv (0, \omega_0(1 - (\tau\omega_0)^{1/4})^{1/2})$ for frequencies $\omega < \omega_0$, which never contribute to order $\tau\omega_0$; $J_2 \equiv (\omega_0(1 - (\tau\omega_0)^{1/4})^{1/2}, \omega_0(1 + (\tau\omega_0)^{1/4})^{1/2})$, which gives the contribution corresponding to $\omega \simeq \omega_0$; and $J_3 \equiv (\omega_0(1 + (\tau\omega_0)^{1/4})^{1/2}, \infty)$, which only contributes for noise with a spectral density giving high-frequency contributions.

By making the substitution $v = (\omega^2 - \omega_0^2)/\omega_0^2$, I_1, the integral corresponding to J_1, has the form

$$I_1 \simeq -\frac{1}{2\omega_0^5} \int_{-1}^{-(\tau\omega_0)^{1/4}} dv\, \frac{S(\omega_0(1+v)^{1/2})}{(1+v)^{1/2} v^3} \lesssim \frac{M}{\omega_0^5 (\tau\omega_0)^{3/4}}, \tag{A2}$$

where $M = \sup_{J_1} S(\omega)$ is of order τ. Thus $I_1 = O(\tau^{1/4})$, so it is ignorable, as was claimed.

For I_2, let us make the substitution $\tau\omega_0 u = (\omega^2 - \omega_0^2)/\omega_0^2$. Then

$$I_2 \simeq \int_{-(\tau\omega_0)^{3/4}}^{(\tau\omega_0)^{-3/4}} \frac{[S(\omega_0) + \tau\omega_0^2 u S'(\omega_0)/2][-u + \tau\omega_0]}{2\tau^2\omega_0^7 (1 + \tau\omega_0 u/2)(u^2+1)^2(1 + 2\tau\omega_0 u/(u^2+1))} \simeq \frac{1}{2}\int_{-\infty}^{\infty} du (\tau\omega_0^6)^{-1} \left\{ \frac{S(\omega_0)}{(u^2+1)^2}\left[1 + \frac{u^2}{2} + \frac{2u^2}{u^2+1}\right] - \frac{\omega_0 S'(\omega_0) u^2}{2(u^2+1)^2} \right\}. \tag{A3}$$

Using

$$\int_{-\infty}^{+\infty} \frac{du}{(u^2+1)^n} = \frac{(-1)^{n-1}}{(n-1)!}\pi \frac{d^{n-1}}{d\lambda^{n-1}} \lambda^{-1/2}\Big|_{\lambda=1}, \tag{A4}$$

we obtain

$$I_2 \simeq (\pi/2\tau\omega_0^6)[S(\omega_0) - (\omega_0/4)S'(\omega_0)]. \tag{A5}$$

For I_3, if $S(\omega)$ is bounded by some power of ω for all $\omega > \omega_0$, i.e., if $S(\omega) \leqslant N\tau\omega^n$, where N is a constant of order 1, then it is always possible to find $0 < k < 1$ so that the integral over the interval $(\omega_0(1 + (\tau\omega_0)^{1/4})^{1/2}, \omega_0(1 + (\tau\omega_0)^{-k})^{1/2})$ is ignorable. To see this, call it I_3^1 and make the variable change $\omega = \omega_0(1+v)^{1/2}$. Then

$$I_3^1 \lesssim \int_{(\tau\omega_0)^{1/4}}^{(\tau\omega_0)^{-k}} dv\, \frac{N\tau\omega_0^n (1+v)^{(n-1)/2}}{2\omega_0^5 v^3} \lesssim N\omega_0^{n-6}/2(\tau\omega_0)^{1/4 - [k(n+1)/2]}. \tag{A6}$$

Now call I_3^2 the part that is left. We have

$$I_3^2 \simeq -\int_{\omega_0(\tau\omega_0)^{-k/2}}^{\infty} d\omega\, \frac{S(\omega)}{\omega^6}. \tag{A7}$$

Thus if $S(\omega) \leqslant N\omega^{4+r}$ for $\omega \gg \omega_0$, where $0 < r < 1$ and N is of order 1, I_3^2 is ignorable.

Therefore for this kind of spectral density without high-frequency contributions, I is given by (A5).

[1] S. M. Moore and J. A. Ramírez, Nuovo Cimento B **64**, 275 (1981).
[2] L. Pesquera and P. Claverie, J. Math. Phys. **23**, 1315 (1982).

[3]T. T. Soong, *Random Differential Equations in Science and Engineering* (Academic, New York, 1973), Chap. 7.
[4]N. G. Van Kampen, Phys. Rep. C **24**, 171 (1976); H. Haken, Rev. Mod. Phys. **47**, 67 (1975).
[5](a) L. Pesquera, Ph. D. thesis, Universidad de Santander, Spain, 1980 (unpublished); (b) P. Claverie, in *Proceedings of the Einstein Centennial Symposium on Fundamental Physics*, edited by S. M. Moore, A. M. Rodríguez-Vargas, A. Rueda, and G. Violini (Univ. de los Andes, Bogotá, 1981).
[6]For an excellent review, see L. De la Peña, in *Stochastic Processes and their Application to Physics and Other Fields*, edited by B. Gómez, S. M. Moore, A. M. Rodríguez-Vargas, and A. Rueda (World Scientific, Singapore, 1982).
[7]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed.
[8]R. Blanco and L. Pesquera, Phys. Rev. D **33**, 421 (1986).
[9]E. A. Novikov, Sov. Phys. JETP **20**, 1290 (1965).
[10]P. Hänggi, in *Stochastic Processes Applied to Physics*, edited by L. Pesquera and M. A. Rodriguez (World Scientific, Singapore, 1985).
[11]L. Meirovitch, *Methods of Analytical Dynamics* (McGraw-Hill, New York, 1970).
[12]J. A. Ramírez, Ph. D. thesis, Universidad de los Andes, Colombia, 1981 (unpublished).

Coexistence of periodic points in reversible dynamical systems on a surface

Kiyotaka Tanikawa[a)]
Department of Astronomy, University of Texas, Austin, Texas 78712

Yoshihiro Yamaguchi
Ilya Prigogine Center for Studies in Statistical Mechanics, University of Texas, Austin, Texas 78712

(Received 20 May 1986; accepted for publication 3 December 1986)

Coexistence of periodic points is investigated in area-preserving reversible dynamical systems on a plane which are not necessarily close to an integrable one. It is shown that on the symmetry axis a higher cycle implies a lower cycle and any cycle implies an infinite number of higher cycles. The standard map is investigated as an example. It is shown that k-cycles for any $k \geqslant 2$ densely fill the phase plane of the map as the parameter goes to infinity. In addition, the number of these cycles are estimated.

I. INTRODUCTION

In one-dimensional dynamical systems on a line, a beautiful result is obtained by Sarkovskii[1] on the coexistence of cycles (or periodic points). Let n_1 and n_2 be two natural numbers. If an existence of an n_1-cycle implies the existence of an n_2-cycle, then we write $n_1 \prec n_2$. Sarkovskii's result is as follows.

Theorem: $3 \prec 5 \prec 7 \prec 9 \prec 11 \prec \cdots \prec 3\cdot 2 \prec 5\cdot 2 \prec 7\cdot 2 \prec \cdots \prec 3\cdot 2^2 \prec 5\cdot 2^2 \prec \cdots \prec 2^3 \prec 2^2 \prec 2 \prec 1$.

Sarkovskii obtained his result essentially by using the intermediate-value theorem. This suggests that a simple geometrical consideration in higher dimensions should be effective to a certain extent.

In the present paper, we consider the problem of the coexistence of cycles for two-dimensional area-preserving dynamical systems with reversibility which are not necessarily close to an integrable one. Corresponding to the intermediate-value theorem for one-dimensional systems, we utilize the Jordan curve theorem. An advantage of considering reversible systems[2,3] is that these systems have a symmetry axis and that cycles of a certain type, which are called symmetric cycles, have their points on the symmetry axis. In a sense, the problem of finding cycles reduces to one dimension. Frequently, these symmetric cycles play an important role when investigating the global structure of the systems. It is to be noticed here that Mather[4] has recently proved a theorem on the coexistence of periodic and quasiperiodic points in maps which satisfy the twist condition. His result heavily depends on the twist condition. Therefore his method seems not applicable to our problem.

In Sec. II, we state the problem and the results on the coexistence of cycles. We consider a surface map not necessarily defined for the whole plane $\mathbb{R}^2$. The domain of the surface map T^n shrinks as n increases. This situation is realized in the restricted three-body problem.[5] Our results are expressed in the form of three theorems. Theorems 1 and 2 are obtained essentially by the Jordan curve theorem. Theorem 3 utilizes the area-preserving property. As consequences, we shall show the coexistence of countably many cycles and uncountably many accumulation points of cycles.

In Sec. III, we investigate the standard map as an example of reversible systems. We shall show that k-cycles for any $k \geqslant 2$ densely fill the phase plane as the parameter K goes to infinity. In addition, the number of cycles will be estimated. It turns out that the number of two-cycles increases in proportion to K^2 as $K \to \infty$ and that the number of k-cycles for $k \geqslant 3$ increases at least in proportion to $K^{k/2}$ for an even k and in proportion to $K^{(k+1)/2}$ for an odd k.

II. STATEMENT OF THE PROBLEM AND THE RESULTS

Let W_1 be a closed region in the plane $\mathbb{R}^2$ surrounded by a piecewise smooth Jordan curve. Let us consider a map T: $W_1 \to \mathbb{R}^2$. We assume that T is (1) analytic in $W_1 - \partial W_1$ and homeomorphic in ∂W_1, (2) area preserving, and (3) reversible.

Reversibility is expressed as follows.[3] Let R be the reflection of points with respect to a symmetry axis of the plane. Then we have

$$T^n R = R T^{-n} \quad \text{for } n = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

with

$$R^2 = I. \tag{2}$$

The domains W_{n+1} of T^{n+1} are obtained by

$$W_{n+1} = T^{-n}(T^n W_n \cap W_1) \quad \text{for } n = 1,2,\ldots. \tag{3}$$

Let

$$W_\infty = \bigcap_{n=1}^{\infty} W_n \quad \text{and} \quad W = \bigcap_{n=1}^{\infty} T^n W_\infty. \tag{4}$$

Then W is the invariant set under T^n for an arbitrary integer n. By reversibility we have

$$T^n W_n = R W_n \quad \text{for } n = 1,2,\ldots. \tag{5}$$

Let

$$W_{-1} = T W_1. \tag{6}$$

Then W_{-1} is the domain of T^{-1}. The domains W_{-n-1} of T^{-n-1} are obtained successively by

$$W_{-n-1} = T^n(T^{-n} W_{-n} \cap W_{-1}) \quad \text{for } n = 1,2,\ldots. \tag{7}$$

Let

$$W_{-\infty} = \bigcap_{n=1}^{\infty} W_{-n}. \tag{8}$$

Then the invariant set W is given also by

[a)] Permanent address: International Latitude Observatory, Mizusawa, Iwate 023, Japan.

 0022-2488/87/040921-08$02.50

$$W=\bigcap_{n=1}^{\infty} T^{-n}W_{-\infty}. \tag{9}$$

With the aid of the equality $TW_{n+1}=W_n\cap TW_1$ which is readily verified, we obtain simple relations between W_n and W_{-n} as

$$W_{-n}=T^nW_n \quad \text{for } n=1,2,..., \tag{10}$$

which show in view of Eq. (5) that W_n and W_{-n} are symmetric to each other with respect to the symmetry axis.

In general, the sets W_n and hence T^nW_n and W_{-n} are composed of a finite number of connected components. However, we assume for simplicity of description that each W_n is composed of one arcwise connected component. This assumption corresponds to the fact that we consider only a component which contains $\mathbf{M}_n$ [see Eqs. (12) and (14)]. Following DeVogelaere,[3] let

$$M_n=T^nR, \tag{11}$$

and

$$\mathbf{M}_n=\{p|p=M_np\}, \quad n=0,\pm 1,\pm 2,.... \tag{12}$$

By definition, $\mathbf{M}_0$ is the symmetry axis. Let

$$\mathbf{M}_{n,m}=\mathbf{M}_n\cap\mathbf{M}_m \quad \text{for } n\neq m. \tag{13}$$

$\mathbf{M}_{n,m}$ is a set of symmetric cycles. In particular, $\mathbf{M}_{1,0}$ is the set of symmetric one-cycles (fixed points). For more details on $\mathbf{M}_n$ and $\mathbf{M}_{n,m}$, see DeVogelaere.[3] It is easily seen that

$$\mathbf{M}_n\subset T^nW_n=W_{-n} \quad \text{and} \quad \mathbf{M}_{-n}\subset W_n \quad \text{for } n=1,2,.... \tag{14}$$

If n points $p_1,p_2,...,p_n$ constitute an n-cycle, they are said to belong to this n-cycle. We denote intervals on the symmetry axis by (p,q), $[p,q]$, etc. for brevity where p and q are points on the symmetry axis.

Now let us state our main results. Proofs of the lemmas are given in the Appendix.

Lemma 1: Let $q_1,q_2\in\mathbf{M}_0$ belong to a $2n$-cycle. Then, the behavior of the symmetry axis around q_1 and q_2 under the operation of T^n is one of the cases illustrated in Fig. 1 or those symmetric to them with respect to the symmetry axis.

With the aid of Lemma 1 and the Jordan curve theorem, we can prove the following.

Theorem 1: If $q_1,q_2\in\mathbf{M}_0$ belong to a $2n$-cycle, and if $(q_1,q_2)\subset W_n$, then there is a fixed point of T^n in (q_1,q_2).

Proof: We assume that a fixed point of T^n does not exist and show that this leads to a contradiction. The interval (q_1,q_2) may contain other $2n$-cycle points r_i $(i=1,2,...,m)$. Without loss of generality, we can assume that there are no two points in r_i which belong to a single $2n$-cycle. Actually, if there exist these points, say r_k and r_l, then we only need to put $q_1=r_k$ and $q_2=r_l$, and start our discussion from the beginning. Thus we have $(q_1,q_2)\cap T^n(q_1,q_2)=\emptyset$. The curve $[q_1,q_2]\cup T^n[q_1,q_2]$ forms a Jordan curve, which we shall denote by C including its orientation. Here, we take the orientation of C as proceeding from q_1 to q_2 on the symmetry axis. Let D be the interior of C. Let C' and D' be the sets symmetric to C and D with respect to the symmetry axis. By reversibility, we have

$$T^nC'=C \quad \text{and} \quad T^nD'=D. \tag{15}$$

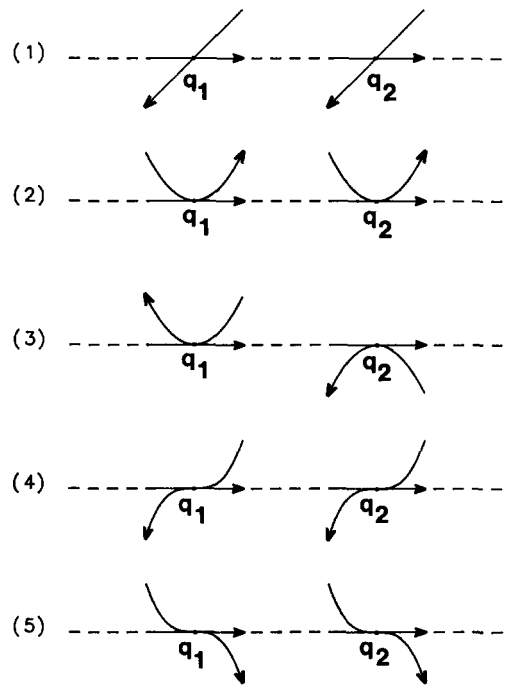


FIG. 1. Images of the symmetry axis near an even cycle on it. Two points q_1 and q_2 belong to a single $2n$-cycle. A portion of the symmetry axis around q_1 (indicated by an arrow centered at q_1) is mapped by T^n into the curve around q_2. Also, a portion of the symmetry axis around q_2 is mapped into the curve around q_1.

Let T^nr_k $(1\leqslant k\leqslant m)$ in $\{T^nr_i\}$ be the nearest point on the symmetry axis to q_1 from the left, and let $\gamma=(T^nr_k,q_1)$. Since $\gamma\cap C=\emptyset$, we have by the Jordan curve theorem

$$\gamma\subset D \quad \text{or} \quad \gamma\cap D=\emptyset. \tag{16}$$

On the other hand, since γ is an interval on the symmetry axis, we have

$$\gamma\subset D\Leftrightarrow\gamma\subset D', \quad \gamma\cap D=\emptyset\Leftrightarrow\gamma\cap D'=\emptyset. \tag{17}$$

Now, let us follow the behavior of γ under T^n for each case in Fig. 1. The methods are almost identical, so we describe the first case and omit the remaining cases. Here D is either on the right or left of C (see, e.g., Kodaira[6]). If D is on the right of C, then we have $\gamma\subset D$ and $T^n\gamma\cap D=\emptyset$. This contradicts Eqs. (15) and (17). If D is on the left of C, then we have $T^n\gamma\subset D$ and $\gamma\cap D=\emptyset$. This also contradicts Eqs. (15) and (17). Thus a contradiction is derived assuming nonexistence of a fixed point of T^n. Q.E.D.

Lemma 2: Take a $p\in T^nW_n$ such that $p\neq M_np$. Then there exists a point of $\mathbf{M}_n$ on an arbitrary arc in T^nW_n connecting p and M_np.

Lemma 3: $\mathbf{M}_n$ is locally an analytical arc in T^nW_n.

For the proof of Lemma 3, see Finn.[7] Using Lemmas 2 and 3, we can prove the following two lemmas.

Lemma 4: $\mathbf{M}_n$ divides T^nW_n into two.

Lemma 5: $\mathbf{M}_n$ is a single arc.

From Lemmas 4 and 5, we obtain the following.

Theorem 2: $\mathbf{M}_n$ is an arc dividing T^nW_n into two.

We obtain the following theorem by the area-preserving property of the map.

Theorem 3: If $q_1,q_2\in\mathbf{M}_0\cap W$ belong to different cycles

and if $(q_1,q_2)\subset W$, then there exists a $q\in(q_1,q_2)$ such that q belongs to a higher cycle than do q_1 and q_2.

Proof: Let m and n be arbitrary positive integers and $k=[m,n]$ the least common multiple. Let $q_1,q_2\in\mathbf{M}_0\cap W$ belong to n- and m-cycles, respectively. Without loss of generality, we can regard that there is no cycle lower than n or m in (q_1,q_2). For, if there exists such a point q, then it is enough to put $q_1=q$. Denote the interval (q_1,q_2) by γ. By our assumption just above, γ and $T^k\gamma$ have no common points, $\bar{\gamma}\cup T^k\bar{\gamma}$ forms a Jordan curve. Let us denote the interior of $\bar{\gamma}\cup T^k\bar{\gamma}$ by D and the area of D by $m(D)$. Let us consider the sequence $T^{jk}D(j=0,1,2,\ldots)$. It is evident that one of $D,T^kD,T^{2k}D,\ldots,T^{lk}D(l=m(W)/2(m(D))+1)$ has common points with the symmetry axis, i.e., one of $T^k\gamma,T^{2k}\gamma,\ldots,T^{(l+1)k}\gamma$ intersects the symmetry axis. This means that γ contains a cycle. Q.E.D.

From the above theorems, we obtain several consequences. For simplicity, we restrict ourselves to the interior of the invariant set W.

Corollary 1: An odd cycle on $\mathbf{M}_0$ implies a fixed point.

Proof: The method of proof is similar for any odd cycle, so we consider a three-cycle. Let p_1, p_2, and p_3 constitute a three-cycle. Suppose $p_1=Tp_3=T^2p_2=T^3p_1$ and $p_1\in\mathbf{M}_0$. Then, $p_1=Tp_3=TRp_2=M_1p_2$, that is, there is a point $p\in\mathbf{M}_1$ on an arbitrary arc connecting p_1 and p_2 by Lemma 2. On the other hand, $p_3=Tp_2=TRp_3=M_1p_3$, i.e., $p_3\in\mathbf{M}_1$. Therefore $\mathbf{M}_1\cap\mathbf{M}_0=\mathbf{M}_{1,0}\neq\emptyset$. In fact, p_2 and p_3 are at the opposite sides with respect to $\mathbf{M}_0$ and p can be chosen to be on the same side with p_2, hence $\mathbf{M}_1$, which is an arc passing through p and p_3, intersects $\mathbf{M}_0$. Q.E.D.

Directly from Theorem 1 and Corollary 1, we obtain the following.

Corollary 2: An n-cycle implies a fixed point for $n>1$. The following results are evident from Theorem 3.

Corollary 3: There exist denumerably many cycles between two arbitrary cycles on $\mathbf{M}_0$.

Corollary 4: There exist uncountably many accumulation points of cycles between two arbitrary cycles on $\mathbf{M}_0$.

Similar results hold on $\mathbf{M}_1$, the complementary symmetry curve. In the case where the surface itself is compact and T can be operated arbitrarily many times on this surface, the above results hold on the whole surface.

III. AN EXAMPLE: THE STANDARD MAP AS $K\to\infty$

Let us consider the standard map as an example. In particular, we consider the problem of the coexistence of cycles as the parameter K goes to infinity. Our main purpose is to prove the following two theorems.

Theorem 4: For any integer $k>1$, k-cycles densely fill the phase plane of the standard map as $K\to\infty$.

Theorem 5: The number of two-cycles increases in proportion to K^2 as $K\to\infty$. For $k>2$, the number of k-cycles increases at least in proportion to $K^{k/2}$ if k is even and in proportion to $K^{(k+1)/2}$ if k is odd as $K\to\infty$.

We shall discuss in detail the case of two-cycles in Sec. III A. Section III B deals with the case of k-cycles for $k>2$.

The standard map has been extensively investigated.[8–10] The map is given by

$$\begin{aligned} T:\ & x_{n+1}=x_n+h_n \pmod 1,\\ & \theta_{n+1}=\theta_n+x_{n+1} \pmod 1, \end{aligned} \tag{18}$$

where

$$h_n=h(\theta_n)=-(K/2\pi)\sin 2\pi\theta_n. \tag{19}$$

The phase plane of the map is a torus. We take the square $-\frac{1}{2}\leqslant x\leqslant\frac{1}{2}$, $-\frac{1}{2}\leqslant\theta\leqslant\frac{1}{2}$ as the fundamental domain of T where opposite sides must be identified. The standard map is reversible.[11,12] The main symmetry axes (or the set $\mathbf{M}_0$ of Sec. II) are the x axis and the lines $\theta=\pm\frac{1}{2}$, while the complementary axes (or the set $\mathbf{M}_1$ of Sec. II) are the lines $x=2\theta$ and $x+1=2\theta$. As is easily confirmed, the θ axis and the lines $x=\pm\frac{1}{2}$ have a similar property to that of the symmetry axes, that is, if their iterates intersect themselves, then the points of intersection are cycles.

A. Two-cycles

Let three points (x_1,θ_1), (x_2,θ_2), (x_3,θ_3) be related by $(x_3,\theta_3)=T(x_2,\theta_2)=T^2(x_1,\theta_1)$. We obtain from Eq. (18)

$$x_3=x_1+h_1+h_2,\quad \theta_3=\theta_1+h_1+h_2+x_1+x_2.$$

In order that these three points constitute a two-cycle, it is necessary and sufficient that they satisfy

$$h_1+h_2=m \quad\text{and}\quad x_1+x_2=n,$$

for some integers m and n. With the aid of Eq. (18), these conditions are rewritten as

$$h(\theta_1)+h(\theta_1+h(\theta_1)/2+n/2)=m, \tag{20}$$

and

$$\begin{aligned} &\theta_2=\theta_1+h(\theta_1)/2+n/2,\\ &x_1=-h(\theta_1)/2+n/2,\\ &x_2=h(\theta_1)/2+n/2, \end{aligned} \tag{21}$$

with integers m and n.

First, consider the case $m=n=0$. Equation (20) reduces to

$$2\theta_1+l=(K/4\pi)\sin 2\pi\theta_1, \tag{22}$$

where l is an integer. Clearly, two-cycles of this type are on the complementary symmetry axes $x=2\theta$ and $x+1=2\theta$. If the value of K is increased, two new two-cycles are born every time the graph $f_1(\theta)=(K/4\pi)\sin 2\pi\theta$ crosses the graph $f_2(\theta)=2\theta+l$. The value of K at which new two-cycles are born is given approximately by $K=4\pi(l+\frac{1}{2})-4/\pi(l+\frac{1}{2})$ when K is sufficiently large. Therefore the number of two-cycles increases in proportion to K. Two-cycles densely fill the complementary axes as $K\to\infty$. In fact, for any $\epsilon>0$, and a large $N>0$, the slope $(K/2)\cos 2\pi\theta$ of $f_1(\theta)$ $(0\leqslant\theta\leqslant\frac{1}{2})$ becomes larger than N except for $\frac{1}{4}-\epsilon\leqslant\theta\leqslant\frac{1}{4}+\epsilon$ if we take K sufficiently large. For this value of K, we can find at least one two-cycle in any interval of length $\delta\approx 1/(N-2)$ contained in $0\leqslant\theta\leqslant\frac{1}{4}-\epsilon$ or $\frac{1}{4}+\epsilon\leqslant\theta\leqslant\frac{1}{2}$.

Next, consider the case $m=0$, $n\neq 0$. It is easily seen that the case of an even n reduces to the case $n=0$ and the case of an odd n reduces to the case $n=1$. Therefore it suffices to consider the case $n=1$. In this case, Eq. (20) becomes

$$(K/4\pi)\sin 2\pi\theta_1 = l, \tag{23}$$

where l is an integer. Clearly, two-cycles are on the line $x = \frac{1}{2}$ (or $x = -\frac{1}{2}$). The situation is similar to that in the preceding paragraph. Two new two-cycles are born each time when the value of K passes through $4\pi l (l = 1,2,\cdots)$. Thus the number of two-cycles increases in proportion to K. Two-cycles densely fill the line $x = \pm\frac{1}{2}$ as $K \to \infty$.

Finally, consider the case $m \neq 0$. The case of an even n reduces to the case $n = 0$ and the case of an odd n reduces to the case $n = 1$. Therefore it suffices to consider the cases $n = 0$ and $n = 1$. Let us consider the case $n = 0$ in detail. Equation (20) becomes

$$(K/2\pi)\sin 2\pi\theta_1 + (K/2\pi)\sin 2\pi(\theta_1 - (K/4\pi)\sin 2\pi\theta_1) = m, \tag{24}$$

where m is an integer. Two-cycles of this type are on the curve $x = (K/4\pi)\sin 2\pi\theta$ and their iterates on the curve $x = -(K/4\pi)\sin 2\pi\theta$ because of Eq. (21).

Let us estimate the number of two-cycles of this type. Let

$$\begin{aligned} f_1(\theta) &= (K/2\pi)\sin 2\pi\theta, \\ f_2(\theta) &= (K/2\pi)\sin 2\pi g(\theta), \end{aligned} \tag{25}$$

where

$$g(\theta) = \theta - (K/4\pi)\sin 2\pi\theta. \tag{26}$$

For a large $K > 0$, the graph $g(\theta)$ has one negative peak of amplitude nearly equal to $K/4\pi$ at $\theta \approx \frac{1}{4}$ for $0 \leqslant \theta \leqslant \frac{1}{2}$. Therefore $f_2(\theta)$ oscillates approximately $2[K/4\pi]$ times with amplitude $K/2\pi$, where [ ] is the Gauss notation. The graph of $f_1 + f_2$ oscillates $2[K/4\pi]$ times with amplitude $K/2\pi$, the centers of oscillations being on f_1. In each oscillation of amplitude $K/2\pi$, the graph $f_1 + f_2$ crosses the graph $f =$ integer $4[K/4\pi]$ times. Consequently, $f_1 + f_2$ intersects $f =$ integer $8[K/4\pi]^2$ times for $0 \leqslant \theta \leqslant \frac{1}{2}$ (see Fig. 2). As is easily verified, each point of intersection uniquely corresponds to a two-cycle. Thus the number of two-cycles increases in proportion to K^2.

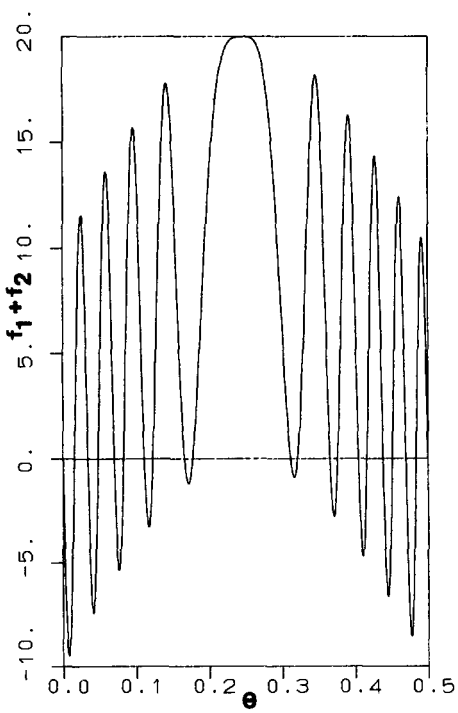


FIG. 2. The graph of $f_1(\theta) + f_2(\theta) = (K/2\pi)\sin 2\pi\theta + (K/2\pi)\sin 2\pi(\theta - (K/4\pi)\sin 2\pi\theta)$ with $K/2\pi = 10$. Two-cycles are obtained as points of intersection of $f = f_1 + f_2$ with f an integer.

Now, two-cycles obtained from Eq. (24) with $m = \pm 1, \pm 2, \ldots$ densely fill the phase plane as $K \to \infty$. First, observe that these cycles densely fill the curve $x = (K/4\pi)\sin 2\pi\theta$ as $K \to \infty$. In fact, two neighboring peaks of f_2, in which there are $2[K/2\pi]$ cycles, become arbitrarily close to each other everywhere in $-\frac{1}{2} \leqslant \theta \leqslant \frac{1}{2}$ as $K \to \infty$ since the slope of $g(\theta)$ becomes arbitrarily large except at $\theta = \pm\frac{1}{4}$. On the other hand, the curve $x = (K/4\pi)\sin 2\pi\theta$ densely fills the phase plane as $K \to \infty$ since the slope of this curve becomes arbitrarily large except at $\theta = \pm\frac{1}{4}$, and hence the distance between two neighboring branches of the curve in the fundamental domain $-\frac{1}{2} \leqslant x \leqslant \frac{1}{2}$, $-\frac{1}{2} \leqslant \theta \leqslant \frac{1}{2}$ becomes arbitrarily small.

In the case $n = 1$, Eq. (20) becomes

$$(K/2\pi)\sin 2\pi\theta_1 - (K/2\pi)\sin 2\pi(\theta_1 - (K/4\pi)\sin 2\pi\theta_1) = m.$$

A similar discussion applies to this case. We obtain two-cycles different from those in the previous case. These cycles are on the curve $x = (K/4\pi)\sin 2\pi\theta + \frac{1}{2}$.

B. k-cycles ($k > 2$)

Let us extend our discussion to the case of k-cycles for any $k > 2$. We consider a particular type of k-cycles which have at least one of their points on the θ axis. The conditions for a k-cycle are from Eq. (18)

$$h_1 + h_2 + \cdots + h_k = m, \quad x_1 + x_2 + \cdots + x_k = n, \tag{27}$$

with m and n integers. Let $(x_1,\theta_1), (x_2,\theta_2), \ldots, (x_k,\theta_k)$ constitute a k-cycle and let $x_2 = 0$ for convenience. Then we obtain

$$x_1 = -h_1 = (K/2\pi)\sin 2\pi\theta_1 \tag{28}$$

and

$$\theta_2 = \theta_1, \theta_j = \theta_{j-1} + \sum_{i=2}^{j-1} h(\theta_i), \quad j = 3,4,\ldots,k. \tag{29}$$

The conditions (27) for a k-cycle can be simplified with the aid of its property similar to symmetric periodic points. In fact, we have $\theta_{k-i+1} = \theta_{2+i}, x_{k-i+1} + x_{3+i} = 0$, $i = 1,2,\ldots,k/2 - 1$, and $x_3 + x_1 = 0$, $x_{k/2+2} = 0 \pmod 1$, for an even k, and $\theta_{k-i+1} = \theta_{2+i}, x_{k-i+1} + x_{3+i} = 0, i = 1,2,\ldots,(k-3)/2$, and $x_3 + x_1 = 0 \pmod 1$, for an odd k. Owing to these relations, the second equation of Eq. (27) is automatically satisfied. In addition, if we let

$$P_k(\theta_1) = h(\theta_1) + h(\theta_2) + \cdots + h(\theta_k), \tag{30}$$

we obtain recursion formulas for P_j as

$$P_2(\theta) = 2h(\theta),$$
$$P_j(\theta) = 2P_{j-1}(\theta) - P_{j-2}(\theta) \quad \text{for an even } j \geq 4, \tag{31}$$
$$P_j(\theta) = P_{j-1}(\theta) + h\left(\theta + \sum_{i=1}^{(j-1)/2} P_{2i}(\theta)/2\right)$$
$$\text{for an odd } j \geq 3.$$

Here k-cycles are obtained as the points of intersection of the graphs $P = P_k(\theta)$ and $P =$ integer. Denote the number of these intersection points by I_k.

Now, let us estimate the number N_k of k-cycles for a large $K > 0$. It should be noted that N_k is not less than $I_k/2$. In fact, the point (x_1, θ_1) is on the curve given by Eq. (28) for $-\frac{1}{2} < \theta < \frac{1}{2}$. Even cycles have two points on this curve and odd cycles have one point on it. Hence $N_k \geq I_k/2$. From now on, we consider the region $0 \leq \theta \leq \frac{1}{2}$ as in the previous subsection. First, consider three-cycles. Here $P_3(\theta)$ is written fully as

$$-P_3(\theta) = (K/\pi)\sin 2\pi\theta + (K/2\pi)\sin 2\pi(\theta - (K/2\pi)\sin 2\pi\theta).$$

The function $-P_3(\theta)$ is essentially of the same form with the left-hand members of Eq. (24), and it oscillates $2[K/2\pi]$ times. In addition, individual oscillations have the amplitude $K/2\pi$ with their centers of oscillation on the curve $P = (K/\pi)\sin 2\pi\theta$. We obtain an estimate $I_3 \approx 8[K/2\pi]^2$. Thus N_3 is proportional to $[K/2\pi]^2$, and N_4 is also proportional to $[K/2\pi]^2$.

Next, consider five-cycles. We have

$$P_5(\theta) = P_4(\theta) - (K/2\pi) \times \sin 2\pi(\theta + P_2(\theta)/2 + P_4(\theta)/2).$$

One oscillation of $P_4(\theta)/2$ with amplitude $K/2\pi$ corresponds to $4[K/2\pi]$ oscillations of $\sin 2\pi(P_4(\theta)/2)$. Since the numbers of oscillations of $P_4(\theta)/2$ and $\theta + P_2(\theta)/2 + P_4(\theta)/2$ are approximately equal in $0 \leq \theta \leq \frac{1}{2}$ for a large $K > 0$, $P_5(\theta)$ has approximately $8[K/2\pi]^2$ oscillations of amplitude $K/2\pi$ with their centers of oscillation on the curve $P = P_4(\theta)$. The number N_5 is proportional to $[K/2\pi]^3$.

The argument in the preceding paragraph can be repeated for $P_k(\theta)$, $k \geq 6$. We obtain the result as stated in Theorem 5. The graphs of P_3, P_4, and P_5 are plotted for a moderate value of K in Fig. 3.

Now, it is evident that k-cycles for $k > 2$ densely fill the phase plane as $K \to \infty$. In fact, the argument is identical for three-cycles with that in Sec. III A. From the construction, it is clear that k-cycles for $k > 3$ are more numerous than three-cycles, hence the assertion follows.

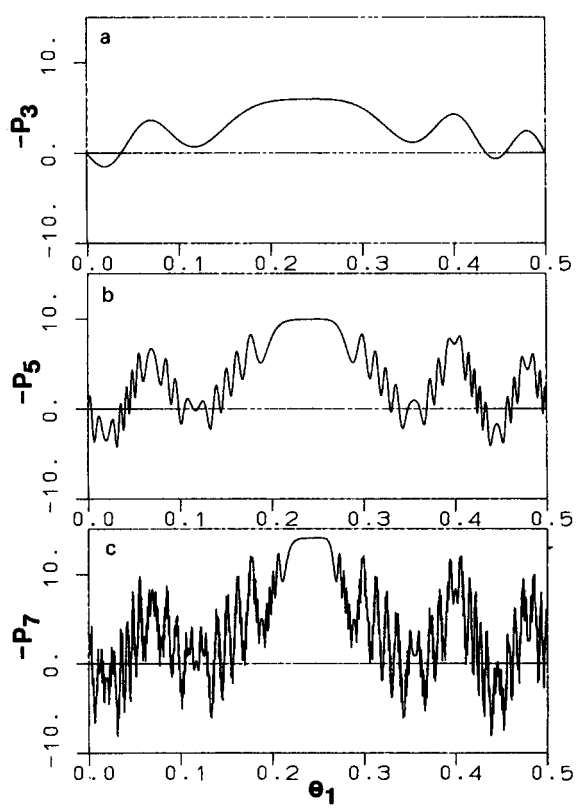


FIG. 3. The graph of $-P_k(\theta)$ for $k = 3$, 5, and 7 with $K/2\pi = 2$. The k-cycles are obtained as points of intersection of $f = P_k(\theta)$ with f an integer.

ACKNOWLEDGMENTS

One of the authors (K. T.) would like to thank Professor W. H. Jefferys for useful discussions and for his kind hospitality while he was staying at the University of Texas. He also would like to thank Dr. J. K. Park for discussions. The other author (Y. Y.) would like to thank Professor I. Prigogine for his kind encouragement.

APPENDIX: PROOFS OF LEMMAS

Proof of Lemma 1: Take local rectangular coordinate systems (x,y) and (x',y') around q_1 and q_2, respectively, where x and x' axes are along the symmetry axis. Let A_1 and A_2 be the linearizations of T^n around q_1 and q_2, and let

$$A_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{A1}$$

By reversibility, we have $RA_2RA_1 = I$, i.e.,

$$A_2 = \begin{bmatrix} d & b \\ c & a \end{bmatrix}. \tag{A2}$$

Let $(x,0)$ and $(x',0)$ be points in the small neighborhood of q_1 and q_2 on the symmetry axis. Let (x_1', y_1') and (x_1, y_1) be their images under A_1 and A_2. We have

$$x_1' = ax, \quad x_1 = dx', \quad y_1' = cx, \quad y_1 = cx'. \tag{A3}$$

If $c \neq 0$, then $y_1' = (c/a)x_1'$ and $y_1 = (c/d)x_1$, which mean transversal crossings of the images of the symmetry axis (see the first figure of Fig. 1). The direction of crossings is determined by the sign of c.

If $c = 0$, the images of the symmetry axis are tangent to the symmetry axis. To get more information at the tangency, let us return to T^n and take nonlinear terms into account. Let us expand T^n as power series in local coordinates around q_1 and q_2 up to and including the k th order terms in x and y. We have

$$x_1' = ax + by + \sum_{i=2}^{k} e_i x^i + yF_1(x,y),$$
$$y_1' = dy + \sum_{i=2}^{k} f_i x^i + yG_1(x,y), \tag{A4}$$

around q_1 and

$$x_1 = ax' + by' + \sum_{i=2}^{k} g_i x'^i + y' F_2(x',y'),$$
$$y_1 = dy' + \sum_{i=2}^{k} h_i x'^i + y' G_2(x',y'), \tag{A5}$$

around q_2, where F_1, G_1, F_2, and G_2 are functions of degree 1 to $k-1$ with respect to their arguments. Here, it is to be understood that points (x,y) and (x_1,y_1) are in the neighborhood of q_1, and points (x',y') and (x_1',y_1') are in the neighborhood of q_2.

Now, take a point $p_0(x,0)$ on the symmetry axis near q_1. Then we have the image point $p_1(x_1',y_1')$ under T^n as

$$x_1' = ax + \sum_{i=2}^{k} e_i x^i, \quad y_1' = \sum_{i=2}^{k} f_i x^i. \tag{A6}$$

By reversibility, we have $Rp_0 = T^n Rp_1$, i.e., $(x,0) = T^n(x_1', -y_1')$ or by Eq. (A5)

$$x = d\left(ax + \sum_{i=2}^{k} e_i x^i\right) + b\left(-\sum_{i=2}^{k} f_i x^i\right) + \sum_{i=2}^{k} g_i \left(ax + \sum_{j=2}^{k} e_j x^j\right)^i - \sum_{i=2}^{k} f_i x^i F_3(x), \tag{A7}$$

$$0 = a\left(-\sum_{i=2}^{k} f_i x^i\right) + \sum_{i=2}^{k} h_i \left(ax + \sum_{j=2}^{k} e_j x^j\right)^i - \sum_{i=2}^{k} f_i x^i G_3(x),$$

where F_3 and G_3 are functions of degree 1 or greater with respect to x. From Eq. (A7), we obtain

$$f_{m+1} = a^m h_{m+1} \quad (1 \leqslant m \leqslant k-1), \tag{A8}$$

if $f_i = 0$ (hence $h_i = 0$) for $i = 2,...,m$. When $f_i = 0$ for $i = 2,...,m$, the transformation of y in Eqs. (A4) and (A5) reduces to

$$y_1' = dy + a^{m-1} h_m x^m + y F_4(x,y),$$
$$y_1 = ay' + h_m x'^m + y' G_4(x',y'), \tag{A9}$$

where F_4 and G_4 are functions of x and y of degree 1 or greater. In particular, if we transform the points $(x,0)$ and $(x',0)$ on the symmetry axis near q_1 and q_2, we obtain

$$x_1' = ax + \cdots, \quad y_1' = a^m h_m x^m + \cdots,$$
$$x_1 = dx' + \cdots, \quad y_1 = h_m x'^m + \cdots, \tag{A10}$$

from which we obtain the assertion of Lemma 1. Q.E.D.

Proof of Lemma 2: The proof is analogous to that of Theorem 1. First, consider the case $p \in T^n W_n - \partial(T^n W_n)$, hence $M_n p \in T^n W_n - \partial(T^n W_n)$. Let Γ be an arbitrary open arc in $T^n W_n$ connecting p and $M_n p$. We can choose an $\epsilon > 0$ such that $U_\epsilon(p) \cap \mathbf{M}_n = \emptyset$ and $U_\epsilon(M_n p) \cap \mathbf{M}_n = \emptyset$. Our assertion is not affected by deforming Γ in $U_\epsilon(p)$ and $U_\epsilon(M_n p)$. Let us deform Γ such that Γ becomes parallel to the symmetry axis in the $\epsilon/2$ neighborhoods of p and $M_n p$, but remains unchanged outside $U_\epsilon(p)$ and $U_\epsilon(M_n p)$. Further, let us extend Γ beyond p and $M_n p$ parallel to the symmetry axis in both the $\epsilon/2$ neighborhoods. In particular, let us denote by γ_1 and γ_2 the two partial arcs in extended arcs contained in $U_{\epsilon/2}(p)$ and $U_{\epsilon/2}(M_n p)$.

Now, we shall show that the images of γ_1 and γ_2 under M_n behave as illustrated in Fig. 1. Take a point $q \in T^n W_n$. Let the coordinates of eight points p, q, $M_n p$, $M_n q$, $RM_n p$, $RM_n q$, Rp, and Rq be (x_1,y_1), (x_1',y_1'), (x_2,y_2), (x_2',y_2'), (x_3,y_3), (x_3',y_3'), (x_4,y_4), and (x_4',y_4') in the global coordinate system where the x axis is along the symmetry axis. By reversibility, we have

$$x_1 = x_4, \quad x_1' = x_4', \quad x_2 = x_3, \quad x_2' = x_3',$$
$$y_1 = -y_4, \quad y_1' = -y_4', \quad y_2 = -y_3, \quad y_2' = -y_3'. \tag{A11}$$

Consider a local coordinate system whose axes are parallel to the global ones and whose origin is at p. Let (x_1'',y_1'') be the coordinates of q in this system. Similarly, consider local coordinate systems with origins, respectively, at $M_n p$, $RM_n p$, and Rp, and let (x_2'',y_2''), (x_3'',y_3''), and (x_4'',y_4'') be the coordinates of $M_n q$, $RM_n q$, and Rq in the corresponding systems. Then, we have

$$x_1'' = x_4'', \quad x_2'' = x_3'',$$
$$y_1'' = -y_4'', \quad y_2'' = -y_3''. \tag{A12}$$

Let R^* be the reflection with respect to each local x axis. Equation (A12) shows that a formal reversibility $T^n R^* = R^* T^{-n}$ holds in the local coordinate systems. From this property, it is easily seen that the situations illustrated in Fig. 1 are realized.

Let us assume $\Gamma \cap \mathbf{M}_n = \emptyset$ and show that this leads to a contradiction. We must examine each case in Fig. 1. We describe, however, only the first case for brevity. Without loss of generality, we can assume that $r \in \Gamma$ implies $M_n r \notin \Gamma$. In fact, suppose that there exists an $r \in \Gamma$ for which $M_n r \in \Gamma$. In this case, we need only to let $p = r$ and start our discussion from the beginning. Therefore $\Gamma \cap M_n \Gamma = \emptyset$, and $\overline{\Gamma} \cup M_n \overline{\Gamma}$ forms a Jordan curve. In addition, $\overline{\Gamma} \cup M_n \overline{\Gamma}$ is invariant under M_n. Hence the region surrounded by $\overline{\Gamma} \cup M_n \overline{\Gamma}$ is also invariant under M_n. Let $\gamma = \gamma_1 - \overline{\Gamma}$. Then the arc γ or its partial arc plays a similar role as γ does in the proof of Theorem 1 and a contradiction is also derived.

If $p \in \partial(T^n W_n)$, take $U_\epsilon(p)$ and $U_\epsilon(M_n p)$ such that $U_\epsilon(p) \cap \mathbf{M}_n = \emptyset$ and $U_\epsilon(M_n p) \cap \mathbf{M}_n = \emptyset$. If we take a sufficiently small $\delta < \epsilon$, then $M_n q \in U_\epsilon(M_n p)$ for any $q \in U_\delta(p)$. Let Γ be an arc connecting p and $M_n p$. Let us extend Γ beyond both ends until it reaches q and $M_n q$ in $U_\delta(p)$ and $U_\epsilon(M_n p)$. To prove our assertion, we only need to repeat the preceding argument with this extended Γ. Q.E.D.

Proof of Lemma 4: Put $D = T^n W_n - \mathbf{M}_n$. If $p,q \in D$ can be connected by an arc in D, we express it as $p \equiv q$. The relation $\equiv$ is evidently an equivalent relation. Let us denote by $[p]$ the equivalence class to which $p \in D$ belongs. Take a $p \in D$. Then it is easily shown with the aid of Lemma 2 that $[p] \neq [M_n p]$. Thus D consists of at least two arcwise connected components. We shall show in the following that D consists of at most two arcwise connected components, which proves our assertion.

To begin with, $M_n[p] = [M_n p]$ and $M_n[M_n p] = [p]$. In fact, take a $q \in [p]$ and let $\gamma \subset D$ be an arc connecting p and q. We have $M_n \gamma \subset [M_n p]$, since $M_n \gamma$ is arcwise connected. Hence $M_n q \in [M_n p]$, which means $M_n[p] \subset [M_n p]$. We obtain $[M_n p] \subset M_n[p]$ in a similar manner. The second equality is obtained similarly.

Next, we shall show that $[p]$ is open in $T^n W_n$. Take a $q \in [p]$. We have $q \neq M_n q$ by definition. If we take a sufficiently small $\epsilon > 0$, we have $U_\epsilon(q) \cap \mathbf{M}_n = \emptyset$ by the continuity of

M_n where $U_\epsilon(q)$ is the ϵ neighborhood of q in T^nW_n. Therefore $U_\epsilon(q) \subset D$. Any $r \in U_\epsilon(q)$ can be arcwise connected to q in D, which in turn means that r can be arcwise connected to p in D. Hence $U_\epsilon(q) \subset [p]$, i.e., $[p]$ is open in T^nW_n.

Next, we shall show that $\overline{[p]} - [p] = \overline{[M_np]} - [M_np] \subset \mathbf{M}_n$. Take a $q \in \overline{[p]} - [p]$. If $q \notin \mathbf{M}_n$, then $q \in [q]$. However, this contradicts that $[q]$ is open in T^nW_n. Therefore $q \in \mathbf{M}_n$. Take an arbitrary sequence $\{q_i\} \subset [p]$ which converges to q. We have $M_nq_i \in [M_np]$ and $M_nq_i \to M_nq = q$ as $i \to \infty$, which means $q \in \overline{[M_np]} - [M_np]$. Hence $\overline{[p]} - [p] \subset \overline{[M_np]} - [M_np]$. Then $\overline{[M_np]} - [M_np] \subset \overline{[p]} - [p]$ can be shown similarly.

According to Lemma 3, T^nW_n is always locally divided into two by $\mathbf{M}_n$. In other words, if we take a small $\epsilon > 0$, then the ϵ neighborhood $U_\epsilon(q)$ of $q \in \overline{[p]} - [p]$ in T^nW_n consists of points of $[p]$, $[M_np]$, and $\overline{[p]} - [p]$, i.e., $U_\epsilon(q) \subset \overline{[p]} \cup \overline{[M_np]}$.

Assume that there exists a $q \in D$ such that $q \notin [p] \cup [M_np]$, hence $[q] \neq [p]$ and $[q] \neq [M_np]$. Then q can be arcwise connected to p by an arc γ in T^nW_n. Consider an open subarc γ_t of γ in which one of the ends is q and the other $t \in \gamma$. If t is sufficiently close to p, we have $\gamma_t \cap [p] \neq \emptyset$. On the other hand, if t is sufficiently close to q, we have $\gamma_t \cap [p] = \gamma_t \cap [M_np] = \emptyset$. Therefore there exists a point $s \in \gamma$ such that $\gamma_s \cap [p] = \gamma_s \cap [M_np] \neq \emptyset$ and $\gamma_t \cap [p] \neq \emptyset$ or $\gamma_t \cap [M_np] \neq \emptyset$ for any t which is closer to p along γ than s is. Evidently, $s \in \overline{[p]} - [p]$. Consider $U_\epsilon(s)$ in T^nW_n for any $\epsilon > 0$. It contains a point $t \in \gamma$ which is closer to q along γ than s is and $t \notin \overline{[p]} \cup \overline{[M_np]}$. This contradicts the property stated in the preceding paragraph. Therefore there does not exist a $q \in D$ such that $q \notin [p] \cup [M_np]$. Q.E.D.

Proof of Lemma 5: It is easily seen by Lemma 2 that there are just two points of $\mathbf{M}_n$ on $J = \partial(T^nM_n)$. Let us denote them by m_1 and m_2. We shall show that $\mathbf{M}_n$ is an arc connecting m_1 and m_2, which proves the lemma. We use the following proposition.

Proposition: $\mathbf{M}_n$ is a single arc in each neighborhood $U_{\delta_0}(m_1)$ and $U_{\delta_0}(m_2)$ for a sufficiently small $\delta_0 > 0$.

Proof: The proof is identical for m_1 and m_2, so we consider the case for m_1. Let us consider a family of circles with center at m_1 and denote each of the circles by C_α where α is the radius. Let $C_\alpha^+ = C_\alpha \cap T^nW_n$. Since J is a piecewise smooth Jordan curve, there is a $\delta > 0$ such that C_α^+ is a single arc for $\alpha \leqslant \delta$. Let us fix such a δ. It is easily seen that we have at least one point of $\mathbf{M}_n$ on each C_α^+ for $0 < \alpha \leqslant \delta$. Let us denote by p_α and p'_α the two end points of C_α^+. We can find an $\epsilon = \epsilon(\alpha) > 0$ such that $(U_\epsilon(p_\alpha) \cup U_\epsilon(p'_\alpha)) \cap \mathbf{M}_n = \emptyset$ since $p_\alpha, p'_\alpha \notin \mathbf{M}_n$. Consequently, if we let $\gamma_\alpha = C_\alpha^+ - U_\epsilon(p_\alpha) - U_\epsilon(p'_\alpha)$, we have $\gamma_\alpha \cap \mathbf{M}_n = C_\alpha^+ \cap \mathbf{M}_n$. Let

$$D_k = \{p \in \gamma_\alpha \mid \delta/k \geqslant \alpha \geqslant \delta/(k+1)\}, \quad k = 1,2,\ldots .$$

Take a point $q \in \mathbf{M}_n \cap D_k$. Then, $\mathbf{M}_n$ is a single arc in a sufficiently small neighborhood of q by Lemma 3, and $D_k \subset (T^nW_n)^i$ is closed, so we can find an $\epsilon_k > 0$ not depending on the position in D_k such that $\mathbf{M}_n$ is a single arc in the ϵ_k neighborhood $U_{\epsilon_k}(p)$ of any $p \in D_k \cap \mathbf{M}_n$. Let us divide our plane by an infinite number of squares with sides equal to $\epsilon_k/2$. Take ϵ_k neighborhoods at each vertex. Then D_k is covered by a finite number of these neighborhoods.

Now, take a $p_1 \in \gamma_\delta \cap \mathbf{M}_n$ and let Γ_1 be the arc of $\mathbf{M}_n$ in $U_{\epsilon_1}(p_1)$. Then we have $d(\Gamma_1, m_1) \leqslant \delta$. With a suitable choice of p_1, we can make $d(\Gamma_1, m_1) < \delta$. This can be easily shown, so we omit the proof. The arc Γ_1 can be extended in D_1, and gets out of D_1 through either γ_δ or $\gamma_{\delta/2}$ eventually. In fact, if it can be extended endlessly in D_1, it passes through some of the ϵ_1 neighborhoods more than once. Let p_1^2 be the intersection point of Γ_1 with γ_δ if it gets out of D_1 through γ_δ, and let p_2 denote the intersection point of Γ_1 with $\gamma_{\delta/2}$ if it gets out of D_1 through $\gamma_{\delta/2}$. We extend Γ_1 farther if it gets out of D_1 through $\gamma_{\delta/2}$ until it necessarily gets out of $D_1 \cup D_2$. It gets out of $D_1 \cap D_2$ through either γ_δ or $\gamma_{\delta/3}$. Let p_1^2 denote the intersection point of Γ_1 with γ_δ in the former case, and p_3 denote the intersection point of Γ_1 with $\gamma_{\delta/3}$ in the latter case.

If we continue the above process, we obtain either an infinite sequence $p_1, p_2, \ldots$, or a finite sequence $p_1, p_2, \ldots, p_{j_0}$ with some j_0. Then Γ_1 reaches m_1 in the former case. Consider the latter case. In this case, Γ_1 starting from p_1 gets out of $\cup_{k=1}^{j_0-1} D_k$ through γ_δ. Let p_1^2 be the intersection point of Γ_1 with γ_δ. We have $d(\Gamma_1, m_1) = \delta/j_0$ in the ϵ neighborhood of p_{j_0}. As in the case of p_1, we can take another point $p'_{j_0} \in \mathbf{M}_n \cap \gamma_{\delta/j_0}$ such that the arc Γ_2 of $\mathbf{M}_n$ through p'_{j_0} goes into D_{j_0}. Let us extend Γ_2. If Γ_2 intersects $\gamma_{\delta/(j_0+i)}$ for $i = 1,2,\ldots$, let p_{j_0+i} denote its intersection points. If p_{j_0+i} is defined for all $i \geqslant 1$, then Γ_2 reaches m_1. Suppose there exists an i_0 such that p_{j_0+i} exist only for $1 \leqslant i \leqslant i_0$. Then Γ_2 cannot stay in $\cup_{k=1}^{j_0+i_0-1} D_k$ indefinitely but it gets out there through γ_δ. Let p_1^3 denote its intersection point with γ_δ.

We repeat the above process. Suppose that $\Gamma_1, \Gamma_2, \ldots, \Gamma_{k_0}$ do not reach m_1. Then we have $p_1^1 = p_1, p_1^2, \ldots, p_1^{k_0+1}$ on γ_δ. These p_1^i all belong to $\mathbf{M}_n$ and locally to different arcs of $\mathbf{M}_n$. Thus k_0 must be finite, since otherwise more than one arc of $\mathbf{M}_n$ enter into one of the ϵ_1 neighborhoods which cover γ_δ. Therefore Γ_i starting from $\gamma_{\delta/i}$ for some i reaches m_1.

It is easy to show that the number of arcs of $\mathbf{M}_n$ starting from m_1 is 1, so we omit the proof. Let us denote the arc by Γ, where Γ gets out of $U_{\delta/i}(m_1)$ once and it can be extended outside $U_{\delta/i}(m_1)$ and may reenter $U_{\delta/i}(m_1)$. However, clearly it cannot approach infinitely close to m_1. In other words, there exists a $\delta^* > 0$ such that $\mathbf{M}_n$ is a single arc in $U_{\delta^*}(m_1)$.

We obtain a similar neighborhood $U_{\delta^{**}}(m_2)$ for the case of m_2. Let $\delta_0 = \min\{\delta^*, \delta^{**}\}$, then our assertion is proved. Q.E.D.

Now, return to the proof of Lemma 5. Take a positive $\epsilon_0 < \delta_0$ and let

$$G = \{p \mid d(p,q) < \epsilon_0,\ p \in T^nW_n,\ q \in J\} - U_{\delta_0}(m_1) - U_{\delta_0}(m_2),$$

$$D = T^nW_n - G - U_{\delta_0}(m_1) - U_{\delta_0}(m_2).$$

For a sufficiently small ϵ_0, we can make $G \cap \mathbf{M}_n = \emptyset$. Therefore Γ getting out of $U_{\delta_0}(m_1)$ enters D. By a similar reasoning as in the proof of the proposition, Γ necessarily gets out of D. It cannot reenter $U_{\delta_0}(m_1)$, so it enters $U_{\delta_0}(m_2)$. On the

other hand, an arc of $\mathbf{M}_n$ which can enter $U_{\delta_0}(m_2)$ is the one which reaches m_2. Thus Γ starting from m_1 reaches m_2.

[1]A. N. Sarkovskii, Ukrainian Math. J. **16**, 61 (1964).
[2]G. D. Birkhoff, *Dynamical Systems* (Am. Math. Soc., New York, 1927).
[3]R. DeVogelaere, *Contributions to the Theory of Nonlinear Oscillations*, edited by S. Lefschetz (Princeton U. P., Princeton, NJ, 1958), Vol. IV, p. 53.
[4]J. N. Mather, Topology **21**, 457 (1982).
[5]K. Tanikawa, Celestial Mech. **29**, 367 (1983).
[6]K. Kodaira, *Introduction to Complex Analysis* (Cambridge U. P., Cambridge, England, 1984).
[7]J. M. Finn, Ph. D. thesis, University of Maryland, 1974, Appendix B.
[8]B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[9]J. M. Greene, J. Math. Phys. **20**, 1183 (1979).
[10]A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer, New York, 1983).
[11]J. M. Greene, R. S. Mackay, F. Vivaldi, and M. J. Feigenbaum, Physica D **3**, 468 (1981).
[12]S. J. Shenker and L. P. Kadanoff, J. Stat. Phys. **27**, 631 (1982).

Mass corrections to periodic solitons

H. W. Braden
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, United Kingdom and Institute of Field Physics, Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27514

(Received 9 December 1985; accepted for publication 18 November 1986)

The quantum mass corrections to static periodic solitons are considered and expressed in terms of a functional quantity, the discriminant of the Hill's operator appearing. Several general expansions for this and hence the mass corrections are given. In particular, the connection with integrable systems and inverse scattering is clearly seen; this enables one to obtain simple expressions for these corrections.

I. INTRODUCTION

A frequent question arising in field theories is the effect of quantization in the presence of a background field. Recurring background fields that have attracted much attention over the last decade are those which correspond classically to homotopically nontrivial configurations. These include "kink"-type solutions as well as the more restrictive class of soliton solutions. Calculating the quantum corrections to these background fields usually proceeds via the semiclassical approximation .[1,2] In the process of evaluating these corrections for two-dimensional soliton supporting theories a remarkable observation was made: the semiclassical approximation is in fact exact.[1] Indeed very elegant closed expressions could be found for the mass corrections of static solitons.[3] This led to several fruitful connections being established between completely integrable systems, inverse scattering, and field theory. The chiral σ model, which possesses both an infinite number of (nonlocal) conserved quantities that survive quantization and an infinite-dimensional loop algebra of symmetry transformation acting on the solution space of these models, has received much investigation along such lines.[4]

In the present paper we are going to evaluate the semiclassical corrections to static two-dimensional solitons, which are periodic in the spatial dimension. This analysis has the merit of enabling us to isolate and examine several of the connections with inverse scattering and integrable systems. Much of the analysis carries through to the nonstatic situation. By letting the period tend to infinity we obtain the usual soliton results. Further the setting has some physical interest of its own[5] in that the $R\times S^1$ topology has a phase structure for the twisted field solitons. (We shall not discuss this latter aspect here, nor the connection with field theory at finite temperature obtained by inverting the roles of space and time.) The nub of the problem in evaluating the quantum corrections lies in finding tractable expressions for a certain functional determinant. Before outlining what the difficulties are we must begin by describing the functional determinant and its regularization.

The evaluation of one loop quantum corrections reduces to the study of a functional determinant, or a related zeta function.[6–8] Following Faddeev[9] the quantum mass correction δM for a time independent classical field $\phi_{\rm cl}(\lambda)$ in a potential $U[\phi]$ may be expressed in the following manner:

$$\delta M=\frac{1}{2}\mathrm{tr}'(\sqrt{K}-\sqrt{K_0})-\frac{1}{8\pi}\int dx\,[U''[\phi_{\rm cl}]-U''[\phi_{\rm vac}]]\int\frac{dk}{(k^2+m^2)^{1/2}} \tag{1a}$$

$$=-\tfrac{1}{4}\mathrm{tr}'\left[(\sqrt{K}-K_0)^2K_0^{-1/2}\right]\leqslant 0, \tag{1b}$$

$$K_0=-\frac{d^2}{dx^2}+U''[\phi_{\rm vac}], \tag{2a}$$

$$K=-\frac{d^2}{dx^2}+U''[\phi_{\rm cl}]=K_0+\tilde{U},$$

$$\tilde{U}=U''[\phi_{\rm cl}]-U''[\phi_{\rm vac}],\quad a=\int dx\,\tilde{U}. \tag{2b}$$

Here and throughout tr' denotes the trace over the eigenfunctions orthogonal to the zero modes, these being taken care of in the standard fashion. The first term in (1a) contains the vacuum subtraction removing the quadratic divergences from δM. The second term corresponds to the mass counterterm removing the logarithmic divergences. As this latter term depends only on the short distance behavior it has the same form for the $R\times R$ and $R\times S^1$ topology. Equations (1a) and (1b) are finite, though due care must be taken with the noncommutativity of K, K_0 (Ref. 6).

It is convenient to consider the following regularization of Eqs. (1a) and (1b):

$$M(s)=-\tfrac{1}{2}\mathrm{tr}'\left[s(a/L)K_0^{-1+s}-(K^s-K_0^s)\right]. \tag{3}$$

Then $\delta M=M(\frac{1}{2})$. [The factor L^{-1} appears when changing the integral in (1a) into a sum.] Equation (3) is readily seen to agree with (1a) and (1b) up to terms of order $s-\frac{1}{2}$, and by using (3) we are led to investigate the following zeta functions:

$$\zeta(s)=\frac{1}{2i\pi}\int d\lambda\,\lambda^s(K-\lambda)^{-1}. \tag{4}$$

The operators K and K_0 that appear throughout are Hill's operators; that is, considered as operators on R, they have periodic potentials arising from the spatial periodicity. Such operators appear naturally when considering quantum mechanics on a lattice and they have been well studied.[10,11] The zeta functions (4) for Hill's operators have been investigated[12,13] for s integral, where they are finite. We, however, are interested in the behavior as $s\to\frac{1}{2}$ (with the renormalizing terms indicated above). By studying these we shall de-

 0022-2488/87/040929-06$02.50

rive expressions for the mass corrections.

In Sec. II of this paper we introduce the discriminant $\Delta(\lambda)$ of a Hill's operator. This function provides a suitable resolvent for the zeta functions (4) and we establish a general expression for the mass correction $M(s)$ [Eq. (3)] in terms of it. The problem is then reduced to finding expansions for the discriminant. We end this section by giving a general expansion for $M(s)$ in terms of the Fourier coefficients of the potential $\tilde{U}$. For all but the simplest potentials this rapidly becomes intractable and so in Sec. III we use a different approach. This revolves around the fact that a resolvent can be constructed from a subset of the associated Hill's equation's periodic eigenvalues. We express the mass correction then in terms of this subset. Of course, for a general potential this subset need neither be finite nor proper, yet for the potentials that arise the subset is both finite and known. This section leads us into several connections with integrable systems. It is perhaps worth bearing in mind that, for both the sine–Gordon and ϕ^4 theories, the Hill's operators that appear are the periodic analog of reflectionless potentials; they are in fact the $n=1$ and $n=2$ forms of Lame's equation (32), which is mentioned later. We conclude Sec. III by rederiving the usual R^2 soliton (kink) mass corrections from this subset of periodic eigenvalues.

II. THE ZETA FUNCTION OF HILL'S EQUATION

The purpose of this section is to derive several general expressions for the quantum mass corrections (1). These will be in terms of the discriminant of the Hill's operator, and we begin by first recalling some general properties of Hill's equation and its discriminant.[10–15]

Let Q be the Hill's operator

$$Q \equiv -\frac{d^2}{dz^2} + q(z), \tag{5}$$

with the potential such that $q(z+L)=q(z)$ is at least three times continuously differentiable. Let $y_{1,2}(z,\lambda)$ be solutions of the Hill's equation

$$Qy_i = \lambda y_i \tag{6}$$

satisfying the boundary conditions $y_1(0,\lambda)=y_2'(0,\lambda)=1$, $y_1'(0,\lambda)=y_2(0,\lambda)=0$. The monodromy matrix relates the solutions $y_i(z+L,\lambda)$ to $y_i(z,\lambda)$. That is, we have

$$\begin{pmatrix} y_1(L) \\ y_2(L) \end{pmatrix} = \begin{pmatrix} y_1(L) & y_1'(L) \\ y_2(L) & y_2'(L) \end{pmatrix} \begin{pmatrix} y_1(0) \\ y_2(0) \end{pmatrix} = M_\lambda \begin{pmatrix} y_1(0) \\ y_2(0) \end{pmatrix}. \tag{7}$$

The discriminant $\Delta(\lambda)$ of Hill's equation (6) is defined to be the trace of this matrix

$$\Delta(\lambda) = \mathrm{Tr}\, M_\lambda = y_1(L) + y_2'(L). \tag{8}$$

The determinant of M_λ is just the Wronskian of the two solutions; that is, 1. The characteristic equation for the eigenvalues of M_λ is then

$$p^2 - \Delta(\lambda)p + 1 = 0, \tag{9a}$$

$$p_\pm = \frac{\Delta(\lambda) \pm [\Delta^2(\lambda)-4]^{1/2}}{2}, \quad p_+ p_- = 1. \tag{9b}$$

Setting $p_\pm = e^{\pm i\alpha L}$, Floquet's theorem[10,11] tells us that when these roots are distinct (6) has two linearly independent solutions f_1, f_2 such that

$$f_1(z) = e^{i\alpha z} p_1(z), \quad f_2(z) = e^{i\alpha z} p_2(z), \tag{10}$$

where p_1 and p_2 are periodic functions with period L.

From (9) and (10) we see the Hill's equation has stable solutions provided $|\Delta(\lambda)|<2$. The solutions of $|\Delta(\lambda)|=2$ give the intervals of stability and instability. Let λ_i be the solutions of this equation (see Fig. 1). Then for $(\lambda_{2i-1},\lambda_{2i})$ $(i=0,1,2,\ldots)$ we have an interval of instability; no solution here is bounded. The complementary intervals $(\lambda_{2i},\lambda_{2i+1})$ $(i=0,1,2,\ldots)$ are the intervals of stability: here every solution is bounded, but none is of period L or $2L$. These intervals of stability are just the conduction and valence bands for the Schrödinger equation (6). The λ_i represent the periodic spectrum; the solutions have period L for λ_0, λ_{4i-1}, λ_{4i} $(i=1,2,\ldots)$; the solutions have period $2L$ for λ_{4i+1}, λ_{4i+2} $(i=0,1,2,\ldots)$. In general, the solutions at the end points are unstable. This is always true for λ_0. Solutions are stable if an interval of instability collapses—that is, if we have a double root $\lambda_{2i-1}=\lambda_{2i}$. We call the simple roots the simple periodic spectrum. Obviously these come in pairs apart from λ_0.

We may now proceed to construct resolvents—depending on the boundary conditions—for the functions (3) and (4) out of the monodromy matrix M_λ. Obviously $\Delta(\lambda)\pm 2 = (\mathrm{Tr}[M_\lambda \pm 1]) = 0$ has the antiperiodic and periodic spectrum as roots. Dirichlet boundary conditions are given by the roots of $(M_\mu)_{21} = y_2(L,\mu) = 0$. The roots μ_i of this latter equation, known as the auxiliary spectrum, interlace the periodic spectrum with $\mu_i \in [\lambda_{2i-1},\lambda_{2i}]$ $(i=1,2,\ldots)$. In what follows we shall concentrate on the former boundary conditions even though similar results are available for the auxiliary spectrum.

The first step in constructing a resolvent is summarized in the following.

Proposition 1:

$$\mathrm{tr}'(K^s - K_0^s) = \frac{1}{2i\pi}\int_\Gamma d\lambda\, \lambda^s \frac{d}{d\lambda} \ln\left[\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}\right]. \tag{11}$$

The proof of this follows from the fact that $\Delta(\lambda)$ is an integral function of order $\frac{1}{2}$ (see Refs. 12 and 16). [$\Delta(\lambda)$ is an entire function of the complex variable λ and there exist constants c_1, c_2 such that (i) $|\Delta(\lambda)|\exp -c_1|\lambda|^{1/2}$ is bounded for all λ; and (ii) $|\Delta(\lambda)|\exp -c_2|\lambda|^{1/2}$ diverges for λ real $\to -\infty$.] So too is $\Delta(\lambda)+2$ and thus it possesses a product expansion of the form

$$c_1 \prod_{n=0}\left(1-\frac{\lambda}{\lambda_{4n+1}}\right)\left(1-\frac{\lambda}{\lambda_{4n+2}}\right).$$

Here the trace in (11) is over the antiperiodic spectrum; with suitable modifications it can be made over the periodic spectrum, which includes the zero mode $\lambda=0$. The contour Γ encloses the real positive λ axis passing between 0 and the first pole at λ_1. Note that (11) naturally accounts for the

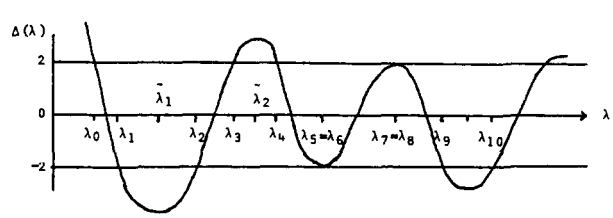


FIG. 1. The discriminant $\Delta(\lambda)$.

right multiplicity of periodic solutions, there being two when an interval of instability collapses and we have a double root.

Here $\Delta_0(\lambda)$, which appears in (11), is the discriminant corresponding to the simple harmonic oscillator vacuum solution

$$\left[-\frac{d^2}{dz^2}+U''(\phi_{\rm vac})\right]y=\lambda y.$$

By redefining λ we can make this vacuum subtraction simply

$$-\frac{d^2}{dz^2}y=\lambda y,\quad \sqrt{\lambda}=\frac{2n+1}{L}\pi.$$

With such a shift the potential corresponding to $\Delta(\lambda)$ becomes

$$q(z)=U''(\phi_{\rm cl})-U''(\phi_{\rm vac}). \tag{12}$$

We remark that in the usual accounts of Hill's equation the normalization for $q(z)$ is taken to be $\int_0^L q(z)dz=0$. We are unable to impose this, so caution is necessary in taking formulas directly from the standard accounts. Let

$$a=\int_0^L q(z)dz. \tag{13}$$

After such a shift in the vacuum energy we may write (11) as

$$\mathrm{tr}(K^s-K_0^s)=\frac{1}{2i\pi}\int_\Gamma d\lambda\,[\lambda+U''(\phi_{\rm vac})]^s\frac{d}{d\lambda}\ln\left[\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}\right], \tag{14}$$

where $\Delta_0(\lambda)$ now corresponds to the shifted simple harmonic oscillator. Similarly,

$$\mathrm{tr}\left(s\frac{a}{L}K_0^{s-1}\right)=\frac{s}{2i\pi}\int_\Gamma d\lambda\,[\lambda+U''(\phi_{\rm vac})]^{s-1}\times\frac{a}{L}\frac{d}{d\lambda}\ln[\Delta_0(\lambda)+2]. \tag{15}$$

Using the regularization described in the Introduction, Eq. (3) becomes, upon substitution of (14) and (15),

$$\begin{aligned}M(s)=&\frac{1}{4i\pi}\int_\Gamma d\lambda\frac{d}{d\lambda}[\lambda+U''(\phi_{\rm vac})]^s\ln\left[\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}\right]\\&-\frac{s}{4i\pi}\int_\Gamma d\lambda\,[\lambda+U''(\phi_{\rm vac})]^{s-1}\\&\times\left\{\ln\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}+\frac{a}{L}\frac{d}{d\lambda}\ln[\Delta_0(\lambda)+2]\right\}\\=&-\frac{s}{4i\pi}\int_\Gamma d\lambda\,[\lambda+U''(\phi_{\rm vac})]^{s-1}\\&\times\left\{\ln\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}+\frac{a}{L}\frac{d}{d\lambda}\ln[\Delta_0(\lambda)+2]\right\}.\end{aligned} \tag{16}$$

Now the integrand in (16) has a branch cut along the negative λ axis to $-U''(\phi_{\rm vac})$. Using standard asymptotic[10,11] estimates we find

$$\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}=1+\tan\frac{\sqrt{\lambda}L}{2}\frac{a}{2\sqrt{\lambda}}+O(\lambda^{-1}). \tag{17}$$

Further, the properties already mentioned of $\Delta(\lambda)$ show us that the quantity

$$f(\lambda)=\left\{\ln\frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2}+\frac{a}{L}\frac{d}{d\lambda}\ln[\Delta_0(\lambda)+2]\right\} \tag{18}$$

is continuous across the branch cut. Finally, as may be shown directly or from a more general analysis (to be given in the next section),

$$\frac{d}{d\lambda}\ln[\Delta_0(\lambda)+2]=-\frac{L}{2\sqrt{\lambda}}\tan\frac{\sqrt{\lambda}L}{2}. \tag{19}$$

The mass counterterm then cancels the $\lambda^{-1/2}$ term in (17). The integrand of (17) for $s<1$ is seen to become vanishingly small on a large circle. We can then change the contour of integration to encircle the λ plane, avoiding the cut by going above and just below it. Doing this we pick up a discontinuity across the cut coming from

$$(-\lambda+i0)^s=\lambda^s e^{i\pi s},\quad (-\lambda-i0)^s=\lambda^s e^{-i\pi s}.$$

Changing the contour we get

$$2M(s)=-s\frac{\sin\pi(s-1)}{\pi}\times\int_{-\infty}^{-U''(\phi_{\rm vac})}d\lambda\,[-\lambda-U''(\phi_{\rm vac})]^{s-1}f(\lambda).$$

Thus we have established the following.

Lemma 2:

$$M(s)=\frac{s\sin\pi s}{2\pi}\int_{U''(\phi_{\rm vac})}^{\infty}d\lambda\,[\lambda-U''(\phi_{\rm vac})]^{s-1}f(-\lambda). \tag{20}$$

The asymptotic estimates (17) show this quantity is finite at $s=\frac{1}{2}$. Some comments are in order. First, when s is integral, the rhs of (14) has a pole term which is picked up by the sin πs term. Thus the usual trace formulas[12,16,17] reduce to the expression of residues. Unfortunately, for s half-integral we have to subtract off the pole term—here the mass counter term—and we are still left with the calculation of a finite piece; there appears no analog of the trace formulas as yet to simplify this. Second, care is needed when changing the contour of integration. If we had not been dealing with subtracted quantity we would have had contributions from the large circles.

Equation (20) now gives us a closed expression for the mass corrections. To proceed further we need some expansions for $\Delta(-\lambda)+2$. Fortunately the shift in integration means we can look at asymptotic expansions to $f(-\lambda)$. The heart of the problem becomes finding tractable expansions for

$$\ln\frac{\Delta(-\lambda)+2}{\Delta_0(-\lambda)+2}.$$

We consider two approaches to this. The first is general to all Hill's equations with a sufficiently smooth potential. The second, making contact with integrable systems, will be considered in the next section.

After some straightforward changes in the usual proofs[10] we have the following.

Lemma 3: If the potential $q(x)$ possesses a Fourier expansion

$$q(x) = \sum a_n \exp \frac{i2\pi nx}{L}, \tag{21}$$

then

$$\Delta(\lambda) + 2 = 4\cos^2(\sqrt{\lambda}\, L/2) D_1(\lambda), \tag{22}$$

where

$$D_1(\lambda) = \| \delta_{n,m} - (a_{n-m})/[\lambda - ((\pi/L)(2n+1))^2] \|_{-\infty}^{\infty}. \tag{23}$$

Using Lemma 3 we then have

$$\ln \frac{\Delta(-\lambda)+2}{\Delta_0(-\lambda)+2} = \ln \det |1 + W|$$

$$= \mathrm{Tr} \ln[1+W] = \mathrm{Tr} \sum_{p=1}^{\infty} \frac{-1}{p} (-W)^p \tag{24}$$

and

$$(W)_{n,m} = a_{n-m}/[\lambda + ((\pi/L)[2n+1])^2]. \tag{25}$$

For a potential with few Fourier coefficients this expansion can be used. We note again the mass renormalization (19) cancels the first term in this expansion

$$\mathrm{Tr}\, W = \frac{a}{L} \sum_n \frac{1}{\lambda + ((\pi/L)(2n+1))^2} = \frac{a}{2\sqrt{\lambda}} \tanh \frac{\sqrt{\lambda}\, L}{2}.$$

This gives us an expansion for the mass correction

$$\delta M = \frac{1}{4\pi} \int_{U''(\phi_{\mathrm{vac}})}^{\infty} d\lambda\, [\lambda - U''(\phi_{\mathrm{vac}})]^{-1/2}$$

$$\times \mathrm{Tr} \sum_{p=2}^{\infty} \frac{1}{p} (-W)^p. \tag{26}$$

To conclude this section it is perhaps worth noting that we may well be able to solve the Schrödinger equation (5) and construct the discriminant $\Delta(\lambda)$ explicitly (8) in terms of elementary solutions. In practice we are hindered from using this—even for quantum mechanics—because the eigenvalue λ is seldom simply related to the parameters that characterize the elementary solutions. This point is clearly seen in Ref. 18.

III. INVERSE SCATTERING AND MASS CORRECTIONS

In the previous section we expressed in terms of the discriminant of the associated Hill's operator the regularized functional determinant that appears in the semiclassical approximation. The finiteness of this expression was made manifest and a general expansion (26) provided in terms of the Fourier coefficients of the potential. For a general potential, however, whose Fourier expansion passes beyond just a few coefficients, this expression rapidly becomes intractable. Indeed the potentials associated with soliton and kink-type fields have an infinite Fourier series. To deal with these latter theories we now describe an alternate expansion for the discriminant which takes a particularly simple form for soliton related background fields. Here we shall make connection with the inverse scattering approach.

There is one further property of the discriminant $\Delta(\lambda)$ of which we have not made use. This is the remarkable result of Hochstadt: the simple periodic spectrum determines both the double periodic spectrum and the nontrivial roots of $\Delta'(\lambda) = 0$ (Ref. 14). Physically this is saying that given the edges of the various conduction bands for the Schrödinger equation (6) we can construct the periodic spectrum. Furthermore, we have a simple representation for this.

Suppose there are $n+1$ bands with edges given by $\lambda_0^0, \ldots, \lambda_{2n}^0$. The nontrivial roots $\tilde{\lambda}$ of $\Delta'(\lambda) = 0$ are determined by the conditions

$$0 = \int_{\lambda_{2j-1}^0}^{\lambda_{2j}^0} \prod_{k=1}^{n} (\mu - \tilde{\lambda}_k) \frac{d\mu}{l(\mu)}, \tag{27}$$

where $l(\mu)$ in 27 is determined by the simple periodic spectrum

$$l(\lambda) = \left[- \prod_{i=0}^{2n} (\lambda - \lambda_i^0) \right]^{1/2}. \tag{28}$$

Having determined the nontrivial roots $\tilde{\lambda}_j$ the discriminant may be expressed[12] in the following form:

$$\Delta(\lambda) = 2\cos \psi(\lambda) L, \tag{29}$$

with

$$\psi(\lambda) = \frac{i}{2} \int_{\lambda_0}^{\lambda} \prod_{j=1}^{n} (\mu - \tilde{\lambda}_j) \frac{d\mu}{l(\mu)}, \quad \psi(\lambda_0) = 0. \tag{30}$$

Hochstadt's result then says $\psi(\lambda) = \pm m\pi/L$ if $\lambda = \lambda_{2m-1}$ or $\lambda = \lambda_{2m}$. That is, (27) and (29) show the simple periodic spectrum determines both the nontrivial roots of $\Delta'(\lambda) = 0$ as well as the double periodic spectrum.

We remark in passing that for the simplest example, the simple harmonic oscillator with potential "a," $l(\lambda) = (a-\lambda)^{1/2}$, and $\Delta(\lambda) = 2\cos(\lambda - a)^{1/2} L$. For more complicated potentials these expressions have also been calculated.[18,19]

At this stage then we have replaced the calculation of $\Delta(\lambda)$ by the determination of the simple periodic spectrum. Using (29) we then have

$$\ln \frac{\Delta(\lambda)+2}{\Delta_0(\lambda)+2} = 2 \ln \frac{\sin \psi(\lambda)(L/2)}{\sin(\lambda^{1/2} L/2)} \tag{31}$$

and so on.

So far we have not mentioned integrable systems or solitons, and there is no reason as yet to consider determining the simple periodic spectrum at all straightforward. The connection with integrable systems is the following: just as the KdV equation is associated with a Schrödinger equation and an isospectral deformation in the usual theory of integrable systems, it is also true that the KdV equation—and indeed any member of the KdV hierarchy—preserves the periodic spectrum of Hill's equation.[12,16,20,21] The hyperelliptic curves (27) and (30) reflect this underlying integrable structure just as solving by quadratures reduces to evaluating hyperelliptic integrals. Next, just as solitons are associated with particularly simple scattering problems (Schrödinger equations), their periodic analogs are also associated with a particularly simple subset of Hill's operators: the finite zone Hill's operators. These are Hill's operators whose simple periodic spectrum is finite in number. Physically they are the one-dimensional lattice potentials with a finite number of conduction and valence bands; for $n+1$ bands there correspond $2n+1$ simple eigenvalues. One Hill's equation

classically studied with these properties is Lame's equation[22-24]

$$\left[-\frac{d^2}{dz^2}+n(n+1)k^2\,\mathrm{sn}^2(z,k)\right]y=(\lambda+a_n)y. \quad (32)$$

Here a_n is a constant and $\mathrm{sn}(z,k)$ an elliptic function. A theorem of Borg[25] in fact tells us that there is one band if and only if the potential q is constant; a theorem of Hochstadt tells us there are two bands if and only if the potential is given by the $n=1$ form of Lame's equation.[15] Actually the manifold of potentials[9] with fixed simple spectrum has been described in Ref. 12. This work generalizes the results of Borg and Hochstadt and shows that most finite gap operators are not just of Lame type; the general shape of the potentials q for $n\geqslant 2$ is still unclear (see Ref. 12 for further comment).

Soliton associated potentials then are associated with finite zoned Hill's operators. Although for a general potential of this class an expression for the simple periodic spectrum may prove intractable, in particular cases this may be explicitly calculated. As we mentioned in the Introduction the $n=1$, 2 forms of Lame's equation occur for physically interesting theories and fortuitously these examples fall into the latter group. Indeed, for general n Lame's equation has $2n+1$ known simple eigenvalues and the solutions corresponding to these are the Lame polynomials.[23,26] The quantum mechanical version of the $n=1$ theory is treated in Ref. 18 where the quantities described thus far are explicitly calculated.

We must now show how these connections with integrable systems and finite zone Hill's operators yield the usual soliton mass corrections in terms of the asymptotic scattering behavior of the Schrödinger equation, i.e., the phase shift. If we have the phase shift $\delta(\lambda)$ defined as the difference between the eigenvalues in the presence of a potential and those in its absence, we have, from our definition $\psi(\lambda)$, (29), and Hochstadt's result,

$$\delta(\lambda)=\psi(\lambda)-\sqrt{\lambda} \quad (33a)$$

$$=\int_\lambda^\infty d\lambda\left(\frac{i}{2}\frac{\Pi_{j=1}^n(\lambda-\tilde{\lambda}_j)}{l(\lambda)}-\frac{1}{\sqrt{\lambda}}\right). \quad (33b)$$

For a finite zone operator this expression is calculable. Indeed in the $L\to\infty$ limit when Lame's equation reduces to the usual reflectionless potential

$$q(z)=-n(n+1)/\cosh^2 z+a_n', \quad (34)$$

the expressions (33) agree (up to conventions) with the usual phase shifts. It is worth emphasizing that (33) is L dependent through the L dependence of the simple eigenvalues λ_i^0. This can be made manifest and the L dependence can be treated systematically. Further, there are useful asymptotic expansions for $\psi(\lambda)$ that arise because of its connection with an integrable system. Using the results of Ref. 12 we may derive simply the asymptotic result

$$\psi(\lambda)\sim\sum_{j=0}^\infty\frac{H_{j-1}}{2^j}\left[\frac{1}{\lambda^{j-(1/2)}}-\frac{1}{\lambda_0^{j-(1/2)}}\right]. \quad (35)$$

Here the H_{j-1} are functions of q,q', etc. The factors 2^j depend on authors; the lowest terms are

$$H_{-1}=-1,\quad \frac{H_0}{2}=\frac{1}{2L}\int_0^1 q\,dx,$$
$$\frac{H_1}{4}=\frac{1}{8L}\int_0^L q^2(x)dx, \quad (36)$$
$$\frac{H_2}{8}=\frac{1}{32L}\int_0^L[2q^3+(q')^2]dx.$$

For finite zone operators these lead to attractive trace formulas[12,13] but we shall not use these here.

We conclude by deriving the usual R^2 mass corrections for the ϕ^4 static kink solution. Although we are not strictly dealing with a soliton, the small oscillation problem about the periodic classical solution is in fact the $n=2$ form of Lame's equation (32) and so the semiclassical corrections may be treated within our framework. Choosing our constants so that the $L\to\infty$ limit of (32) yields

$$K=-\frac{d^2}{dz^2}+4-\frac{6}{\cosh^2 y}, \quad (37a)$$

$$K_0=-\frac{d^2}{dz^2}+4, \quad (37b)$$

we find from (33) the L independent piece to the phase shift

$$\delta(\lambda)=-2\tan^{-1}\left[\frac{3\sqrt{\lambda}}{(2-\lambda)}\right]. \quad (38)$$

In this limit the λ_i^0 correspond to the bound states of the potential (37a). Similarly expanding our general expression (20) we find

$$\frac{\delta M}{m}=-\frac{1}{4\pi}\int_4^\infty d\lambda\,[\lambda-4]^{-1/2}\times\left\{2\tanh^{-1}\frac{3\sqrt{\lambda}}{2+\lambda}+\frac{a}{2\sqrt{\lambda}}\right\}$$
$$=-(3/\pi)[1-\pi/6\sqrt{3}]. \quad (39)$$

The result is the standard one,[3] though derived by a different route. The expansion (35) and (36) corresponds to the Taylor expansion of the inverse tangent (38). It is interesting to note how the bound states of the potential (37a) corresponds to the coalescing of the simple periodic spectrum of the associated Hill's equation. This observation, which is in fact general, is just reexpressing Levinson's theorem.

IV. CONCLUSION

In this paper we have examined the semiclassical quantum corrections for static field configurations on $S^1\times R$. In Sec. II we found a general expression (20) for the quantum mass corrections in terms of the discriminant of the associated Hill's operator; even for the nonstatic case this quantity will appear. The evaluation of quantum corrections then reduces to finding suitable expansions of this discriminant; one in terms of the Fourier components of the potential was given. Several advantages appear by bringing the discriminant to the foreground. Section III noted several of the connections between this quantity and the theory of integrable systems. There we saw how the discriminant could be con-

structed from a subset of the Hill's operators spectrum; this is the analog of the bound state information needed in the usual inverse scattering analysis. For certain Hill's operators, this subset is in principle both known and finite, and we then constructed the quantum mass corrections in terms of this. For both ϕ^4 and the sine–Gordon theories the operators that arise are the finite zone operators of Lame type with these properties and as an example we found the usual ϕ^4 mass corrections as the $L \to \infty$ limit of our analysis. A particular advantage of this analysis is that the L dependence becomes amenable to study. Last, by focusing on the discriminant we can isolate what features are general to any Hill's operator by virtue of its connection with an integrable system and those that are peculiar to its finite zone character. Though trace formulas exist for the operators under discussion when the eigenvalues appear to integral powers no similar formulas exist as yet for half-integral powers: such generalized trace formulas would lead to new evaluations of the mass correction presented here.

[1]R. F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **11**, 3424 (1975).
[2]S. Coleman, in *The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1979).
[3]K. Cahill, A. Comtet, and R. J. Glauber, Phys. Lett. B **64**, 283 (1976).
[4]H. J. de Vega, H. Eichenherr, and J. M. Maillet, Commun. Math. Phys. **92**, 507 (1984).
[5]S. J. Avis and C. J. Isham, Proc. R. Soc. London Ser. A **363**, 581 (1978).
[6]L. A. Dikii, Uspehi Mat. Nauk **13**, 111 (1958) (in Russian); Math. Soc. Transl. Ser. 2 **18**, 81 (1961).
[7]B. M. Levitan and I. S. Sargsjan, "Introduction to spectral theory: self-adjoint ordinary differential operators," in *Translations of Mathematical Monographs*, Vol. 39 (Am. Math. Soc., Providence, RI, 1975).
[8]S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[9]L. D. Faddeev and V. E. Korepin, Theor. Math. Phys. USSR **25**, 147 (1975).
[10]W. Magnus and S. Winkler, *Hill's Equation* (Wiley–Interscience, NY, 1966).
[11]M. S. P. Eastham, *The Spectral Theory of Periodic Differential Equations* (Scottish Academic, Edinburgh, 1973).
[12]H. P. McKean and P. van Moerbeke, Invent. Math. **30**, 217 (1975).
[13]E. Trubowitz, Commun. Pure Appl. Math. **30**, 321 (1977).
[14]H. Hochstadt, Math. Z. **82**, 237 (1963).
[15]H. Hochstadt, Arch. Rational Mech. Anal. **19**, 353 (1965).
[16]H. P. McKean and E. Trubowitz, Commun. Pure Appl. Math. **29**, 143 (1976).
[17]H. Flaschka, Arch. Rational Mech. Anal. **59**, 293 (1975).
[18]H. W. Braden, J. Phys. A **18**, 2127 (1985).
[19]H. W. Braden, Ph. D. dissertation, Cambridge University, 1983.
[20]H. P. McKean, Commun. Pure Appl. Math. **34**, 197 (1981).
[21]P. Lax, Commun. Pure Appl. Math. **28**, 141 (1978).
[22]C. Hermite, *Euvres* (Gauthier-Villars, Paris, 1912), Vol. iii, pp. 118–22, 266–83, 374–79, 475–8; Vol. iv, pp. 8–18.
[23]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, England, 1978), 4th edition.
[24]P. Appel and E. Lacour, *Principes Fonctions Elliptiques et Applications* (Gautheir-Villars, Paris 1922), pp. 378–381.
[25]G. Borg, Acta Math. **78**, 1 (1945).
[26]E. L. Ince, Proc. R. Soc. Edinburgh **60**, 47, 83 (1939–1940).

Symmetry reduction and simple supersymmetric models

Martin Légaré[a)]
Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 30 July 1986; accepted for publication 18 November 1986)

The reduction by the subgroups of the conformal group space-time of the Wess–Zumino model and the four-dimensional simple ($N = 1$) supersymmetric Yang–Mills theories is considered. It is shown for both models that the translations are the only subgroups of the Poincaré group that leave a supersymmetric residual system on the reduced space if we require compatibility of the reduction with all Poincaré supersymmetry transformations.

I. INTRODUCTION

In previous papers, a study of the geometric formulation of invariant gauge fields under smooth group actions[1,2] has been used to reduce gauge[3,4] and matter-coupled gauge[5,6] systems under subgroups of the conformal group of space-time. Hamiltonian interpretations of the residual systems have been given and a number of invariant solutions found. As a next step in these investigations, we address ourselves to the symmetry reduction of supersymmetric models.

The (manifest) supersymmetries of the S matrix in more than $1 + 1$ dimensions have been classified by Nahm.[7] He has also obtained the representations of the flat space-time supersymmetries and presented examples with supermultiplets containing a vector as their highest spin component. In this article, we start instead from a definite representation of a supersymmetry to generate supersymmetric models in lower dimensions by symmetry reduction. This approach has been successfully applied in flat space-time by Brink, Schwarz, and Scherk,[8] who derived the extended supersymmetric Yang–Mills models in $3 + 1$ dimensions from higher dimensional $N = 1$ supersymmetric Yang–Mills models.

In the following, we examine the Wess–Zumino model[9] and the four-dimensional $N = 1$ supersymmetric Yang–Mills system.[10,11] Our purpose is to identify group actions leading to supersymmetric reduced systems. First, we present in Sec. II the invariance conditions for the fields of the (on-shell) supersymmetric multiplet involved in each model. In Sec. III, we show that the translations are the only subgroups (compatible with all Poincaré supersymmetry transformations) of the invariance group of the massive Wess–Zumino model that leave a supersymmetric residual system on the reduced space. Next, we devote Sec. IV to the investigation of the $N = 1$ supersymmetric Yang–Mills theories. We look for subgroups of the conformal group [$C(3,1)$] leading to reduced supersymmetry. It is found that the translations fulfill this requirement among the subgroups of the Poincaré group [$P(3,1)$]. We also give a condition for the subgroups of $C(3,1)$. In the Appendix, a reduction of a supersymmetric Yang–Mills system without residual supersymmetry is discussed.

II. INVARIANCE CONDITIONS

The action of a symmetry group G on Minkowski space (M) generates a number of strata on which the invariance conditions are solved, each stratum being the set of orbits with conjugate isotropy subgroups (G_0). For tensor and spinor fields, the invariance conditions under the action of the subgroups of the conformal group are, respectively, given in Refs. 12 and 13. However, for a multiplet of Dirac spinor fields transforming under a representation ($\tilde{D}$) of the gauge group (H), we have the following expression[5]:

$$\Psi(gx) = D({}^{g}J(x)) \otimes \tilde{D}(\rho^{-1}(g,x))\Psi(x), \tag{2.1}$$

for every element $g \in G \subset C(3,1)$, where $x \in M$, D is the representation of the Jacobian (J) of the group transformation g [which is for an orthonormal frame the product of a conformal factor and the direct sum of the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ basic representations of SL(2,C)], and ρ: $G \times M \to H$, is the "transformation function"[1] which characterizes the group action on the gauge bundle. In order to determine the invariant fields, we have first to solve the isotropy condition, i.e., the set of linear algebraic equations derived from the invariance condition at the isotropy point x_0 of each orbit for each element (g_0) of the isotropy subgroup at this point [$G_0(x_0)$]. The isotropy condition for the spinor fields multiplet derived from (2.1) is

$$\Psi(x_0) = D({}^{g_0}J(x_0)) \otimes \tilde{D}(\lambda(g_0))\Psi(x_0), \tag{2.2}$$

where $\lambda(g_0) = \rho^{-1}(g_0, x_0)$, corresponds to the homomorphism of G_0 into the gauge group classifying the gauge bundles admitting a lift of the group action (see Ref. 1). We obtain the invariant field on each stratum by acting with G on the solution to the isotropy condition (2.2) and by smoothly (C^∞) joining (if possible) on the stratum the resulting fields on each orbit (for details and examples, see Refs. 12 and 13).

As for the invariant gauge fields, they satisfy, for every $g \in G$, the condition[1]

$$(f_g^* \omega)(x) = \mathrm{Ad}\,\rho^{-1}(g,x)\omega(x) + \rho^{-1}(g,x)d\rho(g,x), \tag{2.3}$$

where ω is a gauge field (one-form connection) and Ad denotes the adjoint representation of the gauge group. Their isotropy condition is simply

$$(f_{g_0}^* \omega)(x_0) = \mathrm{Ad}\,\lambda(g_0)\omega(x_0). \tag{2.4}$$

Let us note that the compatibility between the invariance conditions and the gauge transformations requires that

[a)] Address after 1 September 1986: Centre de Recherches Mathématiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Canada, H3C 3J7.

 0022-2488/87/040935-05$02.50

a gauge field ω', obtained from a gauge transformation $h(x)$ on ω, obeys the invariance condition (2.3) with a transformation function $\rho'(g,x)$ given by

$$\rho'^{-1}(g,x) = h^{-1}(x)\rho^{-1}(g,x)h(x). \tag{2.5}$$

III. WESS–ZUMINO MODEL

The (on-shell) Wess–Zumino model is formed of one complex scalar field (ϕ) and one anticommuting Majorana spinor field (Weyl component: ψ) with the following Lagrangian density, written in the Weyl spinor notation (see Ref. 14):

$$\mathscr{L} = \tfrac{1}{2}(\partial_\mu\phi\,\partial^\mu\phi^* - m^2\phi^*\phi) + \tfrac{1}{2}(i\psi\sigma^\mu\partial_\mu\bar{\psi} + i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi - m\psi\psi - m\bar{\psi}\bar{\psi}). \tag{3.1}$$

The (on-shell) supersymmetry transformations leaving the corresponding action invariant are

$$\delta_\epsilon\phi = 2\epsilon\psi \tag{3.2}$$

and

$$\delta_\epsilon\psi = -(i\sigma^\mu\bar{\epsilon}\,\partial_\mu\phi + m\epsilon\phi^*), \tag{3.3}$$

where ϵ and $\bar{\epsilon}$ are defined as the SL(2,C) components of the infinitesimal anticommuting Majorana spinor parameter.

Since the reduced systems are required to be supersymmtric, the fields (ϕ' and ψ') resulting from any supersymmetry transformation must also satisfy the invariance conditions imposed on the original fields ϕ and ψ:

$$\phi'(gx) = \phi'(x) \tag{3.4}$$

and

$$\psi'(gx) = D^{(1/2,0)}([{}^gJ(x)])\psi'(x), \tag{3.5}$$

where $\phi' = \phi + \delta_\epsilon\phi$, $\psi' = \psi + \delta_\epsilon\psi$, and $D^{(1/2,0)}$ stands for the $(\frac{1}{2}, 0)$ basic representation of SL(2,C). The invariance conditions and all the supersymmetry transformations are then compatible if

$$(\delta_\epsilon\phi)(gx) = (\delta_\epsilon\phi)(x) \tag{3.6}$$

and

$$(\delta_\epsilon\psi)(gx) = D([{}^gJ(x)])(\delta_\epsilon\psi)(x), \tag{3.7}$$

for all $g\in G$ and all ϵ.

In addition to the invariance condition

$$\psi(gx) = D^{(1/2,0)}([{}^gJ(x)])\psi(x), \tag{3.8}$$

we find from the transformation (3.2) and the condition (3.6) that the spinor ψ is subjected to the constraint

$$\psi(gx) = \psi(x). \tag{3.9}$$

Moreover, the invariance condition for massive ϕ, the transformation (3.3) and the condition (3.7) imply that

$$\phi^*(x)(\mathbf{1}_2 - D^{(1/2,0)}([{}^gJ(x)])) = 0, \tag{3.10}$$

for all $g\in G$. Assuming a nontrivial invariant scalar field ϕ, we deduce from (3.10) that the spinor representation of the Jacobian must be the identity at any point of the stratum. Although the invariance group of the model is the Poincaré group, only the translations (neglecting the discrete subgroups) satisfy the above constraint. These subgroups also ensure the compatibility between the conditions (3.8) and (3.9). Consequently, the translations are the only (continuous) space-time transformations allowing reductions of the massive Wess–Zumino model with residual supersymmetry. For example, the invariance under two-dimensional space translations yields a supersymmetric system composed of a noninteracting Dirac spinor field with a scalar field in 1 + 1 dimensions. The residual supersymmetry transformations are easily obtained by substituting the invariant fields in the formulas (3.2) and (3.3).

IV. SIMPLE SUPERSYMMETRIC YANG–MILLS SYSTEMS

A simple ($N = 1$) supersymmetric Yang–Mills theory (on-shell) in 3 + 1 dimensions consists of a Yang–Mills field minimally coupled to a multiplet of anticommuting Majorana spinor fields transforming under the adjoint representation of the gauge group. Its Lagrangian density, expressed in the Weyl spinor notation (see Ref. 14), is

$$\mathscr{L} = -\tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu a} + (i/2)[\bar{\psi}^a\bar{\sigma}^\mu(D_\mu\psi)^a + \psi^a\sigma^\mu(D_\mu\bar{\psi})^a], \tag{4.1}$$

where $F_{\mu\nu}$ represents the field strength corresponding to the gauge field ω_μ, D_μ is the covariant derivative with respect to the gauge field, and ψ is the $(\frac{1}{2},0)$ Weyl component of the Majorana spinor field. The corresponding action is left invariant under the following (on-shell) supersymmetry transformations, which form a representation of the super-Poincaré algebra:

$$\delta_\epsilon\omega^a_\mu = i\bar{\mathscr{E}}\gamma_\mu\Psi^a = i(\bar{\epsilon}\bar{\sigma}_\mu\psi^a + \epsilon\sigma_\mu\bar{\psi}^a) \tag{4.2}$$

and

$$\delta_\epsilon\psi^a = -\tfrac{1}{2}F^a_{\mu\nu}\sigma^{\mu\nu}\epsilon, \tag{4.3}$$

where ϵ and $\bar{\epsilon}$ are the SL(2,C) components of the anticommuting Majorana four-spinor parameter $\mathscr{E}$, and Ψ is the four-component Majorana spinor field with values in the Lie algebra of the gauge group. (Use of the four-spinor notation will sometimes be made for brevity.[14])

As in Sec. III, a reduced supersymmetry is ensured if the invariance conditions are also imposed on the gauge and spinor fields coming from any supersymmetry transformations (4.2) and (4.3). This means that

$$(f_g^*\omega')(x) = \mathrm{Ad}\,\rho^{-1}(g,x)\omega'(x) + \rho^{-1}(g,x)d\rho(g,x) \tag{4.4}$$

and

$$\psi'(gx) = D^{(1/2,0)}([{}^gJ(x)]) \otimes \mathrm{Ad}\,\rho^{-1}(g,x)\psi'(x), \tag{4.5}$$

where $\omega' = \omega + \delta_\epsilon\omega$ and $\psi' = \psi + \delta_\epsilon\psi$. Let us remark that a supersymmetric residual system is also produced if the fields ω' and ψ' are gauge equivalent to invariant fields. We discuss this case at the end of this section.

From Eqs. (4.4) and (4.5) and the invariance conditions for the gauge field (ω) and the spinor field (ψ), we get

$$(f_g^*\delta_\epsilon\omega)(x) = \mathrm{Ad}\,\rho^{-1}(g,x)\delta_\epsilon\omega(x) \tag{4.6}$$

and

$$(\delta_\epsilon\psi)(gx) = D^{(1/2,0)}([{}^gJ(x)]) \otimes \mathrm{Ad}\,\rho^{-1}(g,x)\delta_\epsilon\psi(x). \tag{4.7}$$

These conditions hold for every Majorana spinor ϵ (i.e., for every Poincaré supersymmetry transformation) if we have

$$\bar{\sigma}_\mu \psi(gx)\,{}^{g}J(x)^{\mu}{}_{\nu} = \bar{\sigma}_\nu \,\mathrm{Ad}\,\rho^{-1}(g,x)\psi(x) \tag{4.8}$$

and

$$\sigma^{\mu\nu}F_{\mu\nu}(gx) = D^{(1/2,0)}({}^{g}J(x))\sigma^{\mu\nu}\,\mathrm{Ad}\,\rho^{-1}(g,x)F_{\mu\nu}(x). \tag{4.9}$$

At the isotropy point x_0, the above expressions simplify to

$$\bar{\sigma}_\mu \psi(x_0)\,{}^{g_0}J(x_0)^{\mu}{}_{\nu} = \bar{\sigma}_\nu \,\mathrm{Ad}\,\lambda(g_0)\psi(x_0) \tag{4.10}$$

and

$$\sigma^{\mu\nu}F_{\mu\nu}(x_0) = D^{(1/2,0)}({}^{g_0}J(x_0))\,\sigma^{\mu\nu}\,\mathrm{Ad}\,\lambda(g_0)F_{\mu\nu}(x_0), \tag{4.11}$$

where $g_0 \in G_0$. Substituting, respectively, the isotropy conditions for the spinor field and the field strength in Eqs. (4.10) and (4.11), we find the constraints

$$[\bar{\sigma}^{\mu}\,{}^{g_0}J(x_0)^{\mu}{}_{\nu} - \bar{\sigma}_\nu D^{-1}({}^{g_0}J(x_0))]\psi(x_0) = 0 \tag{4.12}$$

and

$$[\sigma^{\mu\nu} - D({}^{g_0}J(x_0))\sigma^{\lambda\rho}\,{}^{g_0}J(x_0)^{\mu}{}_{\lambda}\,{}^{g_0}J(x_0)^{\nu}{}_{\rho}]F_{\mu\nu}(x_0) = 0, \tag{4.13}$$

where $\psi(x_0)$ and $F_{\mu\nu}(x_0)$ are assumed to be nontrivial.

We next look for the symmetry subgroups of $P(3,1)$ obeying (4.8) and (4.9) with the invariant fields ψ and $F_{\mu\nu}$. Choosing an orthonormal frame on Minkowski space, the Jacobian matrix ${}^{g_0}J(x_0)$ induces a homomorphism of the isotropy subgroup G_0 into the Lorentz group SO(3, 1). Therefore Eq. (4.12) can be written as

$$[D^{(1/2,0)}({}^{g_0}\Lambda^{-1})^{\dagger} - \mathbb{1}_2]\bar{\sigma}_\nu D^{(1/2,0)}({}^{g_0}\Lambda^{-1})\psi(x_0) = 0, \tag{4.14}$$

with ${}^{g_0}\Lambda \in SO(3,1)$. However, no one-parameter subgroup of SO(3,1) satisfies (4.14). Since condition (4.14) is preserved by conjugations under SO(3,1), this can be shown by considering only the one-parameter subgroups of SO(3,1) up to conjugations by SO(3,1): $\cos\phi L_3 + \sin\phi K_3$ $(0 \leqslant \phi < \pi)$, and $L_2 + K_1$ (Ref. 15), where L_i stands for the generator of the rotations around the x^i axis, and K_i, for the generator of the boosts along the x^i axis. But no nonvanishing $\psi(x_0)$ fulfills this condition for each of these one-parameter subgroups of SO(3,1). Hence the image of the Jacobian homomorphism at x_0 is the identity element (since we ignore the discrete subgroups), which means that $G_0(x_0)$ is included in the kernel of $J(x_0)$ $(G_0(x_0) \subseteq \mathrm{Ker}\, J(x_0))$. For the action of any subgroup of $P(3,1)$ on Minkowski space, the kernel of $J(x_0) \subseteq T(4)$ (the pure translations in four dimensions), and thus $G_0 \subseteq T(4)$. This implies that $G_0(x_0)$ is the identity element. Consequently, the homomorphism λ is trivial and from Corollary 3 of Ref. 1, the transformation function equals the identity $(\rho^{-1}(g,x) = e)$.

Returning to Eq. (4.8), the value of ρ enables us to write

$$\bar{\sigma}_\mu \psi(gx)\,{}^{g}\Lambda^{\mu}{}_{\nu} = \bar{\sigma}_\nu \psi(x). \tag{4.15}$$

With use of the invariance condition (2.1) for the spinor field, this relation becomes

$$[D^{(1/2,0)}({}^{g}\Lambda^{-1})^{\dagger} - \mathbb{1}_2]\bar{\sigma}_\nu \psi(x) = 0, \tag{4.16}$$

for every value of the index ν and at every point x of the stratum. But there exists no nontrivial spinor field $\psi(x)$ that obeys this equation for any of the one-parameter subgroups of SO(3,1). As in the isotropy case, one can arrive at this result by looking at the above-mentioned one-parameter subgroups of SO(3,1). We are thus left with the translations. The compatibility condition (4.9) for the invariant field strength $F_{\mu\nu}$ is also satisfied for these subgroups. Hence, only the reductions by translations permit a supersymmetry preserving the set of invariant fields. An example of such reductions is detailed in Brink *et al.*[8] for two-dimensional space translations where a Yang–Mills system interacting with a complex scalar field and a Dirac spinor field in $1+1$ dimensions is derived.

Let us make a brief comment on the possible reductions for the subgroups of the conformal group $C(3,1)$. With a choice of orthonormal frames, the Jacobian at x_0 is a homomorphism of G_0 into the direct product of the Lorentz group with the group of dilatations:

$${}^{g_0}J(x_0): g_0 \in G_0(x_0) \mapsto e^{\lambda_{g_0}(x_0)}\,{}^{g_0}\Lambda(x_0) \in \mathbb{R}^+ \times SO(3,1), \tag{4.17}$$

where $\lambda_{g_0}(x_0) \in \mathbb{R}$ and ${}^{g_0}\Lambda \in SO(3,1)$. The isotropy condition (4.10) then reads

$$[e^{\lambda_{g_0}(x_0)}\bar{\sigma}^{\mu}\,{}^{g_0}\Lambda(x_0)^{\mu}{}_{\nu} - e^{-(3/2)\lambda_{g_0}(x_0)}\bar{\sigma}_\nu D^{(1/2,0)}({}^{g_0}\Lambda^{-1}(x_0))]\psi(x_0) = 0. \tag{4.18}$$

It turns out that all the one-parameter subgroups of $\mathbb{R}^+ \times SO(3,1)$ force the spinor field at the isotropy point to vanish. Thus we conclude that $G_0(x_0) \subseteq \mathrm{Ker}\, J(x_0)$.

Finally, in the hope of finding new subgroups leading to supersymmetric systems, we relax the compatibility conditions (4.4) and (4.5) by replacing the transformation function ρ by a gauge equivalent one [see Eq. (2.5)]:

$$(f_g^{*}\omega')(x) = \mathrm{Ad}\,\rho'^{-1}(g,x)\omega'(x) + \rho'^{-1}(g,x)d\rho'(g,x) \tag{4.19}$$

and

$$\Psi'(gx) = D({}^{g}J(x)) \otimes \mathrm{Ad}\,\rho'^{-1}(g,x)\Psi'(x). \tag{4.20}$$

It corresponds to allow each invariant field transformed by a supersymmetry to be equivalent up to a gauge transformation (which can be a function of the Majorana spinor parameter $\mathscr{E}$) to a field of the set of invariant fields under the symmetry group G. On the other hand, these conditions involve explicitly the transformation function ρ.

For $G \subset P(3,1)$, the following expressions in four-spinor notation are deduced from (4.19) and (4.20) at the isotropy point (for every value of the Majorana spinor $\mathscr{E}$):

$$X_\mu(x_0,\mathscr{E})\,{}^{g_0}\Lambda^{\mu}{}_{\nu} = \mathrm{Ad}\,\lambda(g_0)X_\nu(x_0,\mathscr{E}) \tag{4.21}$$

and

$$D^{-1}({}^{g_0}\Lambda)\Phi(x_0,\mathscr{E}) = \mathrm{Ad}\,\lambda(g_0)\Phi(x_0,\mathscr{E}), \tag{4.22}$$

with the definitions

$$X_\mu(x_0,\mathscr{E}) \equiv \mathrm{Ad}\,h(x_0,\mathscr{E})[\bar{\mathscr{E}}\gamma_\mu\Psi(x_0) - \omega_\mu(x_0)] \tag{4.23}$$

and

$$\Phi(x_0,\mathscr{E}) \equiv \mathrm{Ad}\,h(x_0,\mathscr{E})[F^{\mu\nu}(x_0)\sigma_{\mu\nu}\mathscr{E} - \Psi(x_0)]. \tag{4.24}$$

If $G_0 = \{e\}$, then $\rho' = \rho = \{e\}$, and therefore, no symmetry groups other than the translations are admissible. For $G_0 \neq \{e\}$, Eqs. (4.21) and (4.22) for $X_\mu(x_0,\mathscr{E})$ and

$\Phi(x_0,\mathscr{E})$ are, respectively, equivalent to the isotropy conditions for the gauge field ω_μ and the spinor field Ψ, and thus the solution X_μ has the same form as $\omega_\mu(x_0)$, and the solution Φ, the same as Ψ. Since $\Psi(x_0)$ and $\omega(x_0)$ span different subspaces of the gauge algebra (when $G_0 \neq \{e\}$), there exists no Ad $h(x_0,\mathscr{E})$ automorphism of the gauge algebra such that (4.23) and (4.24) are satisfied, and therefore no new supersymmetric residual system is obtained. In the Appendix, we look at the SO(3) invariant gauge and spinor fields of the SU(3) supersymmetric Yang–Mills system as an example of the above discussion.

V. SUMMARY

We have investigated in this paper the reduction by symmetry of two simple supersymmetric models. We have found that the massive Wess–Zumino model can be reduced to supersymmetric systems by the translation subgroups of the Poincaré group by looking at the compatibility between all the supersymmetry transformations and the invariance conditions. The residual supersymmetry is derived by substitution of the invariant fields in the supersymmetry transformations of the model. We have also studied the $N=1$ four-dimensional supersymmetric Yang–Mills system. It has been shown that no residual supersymmetry is left for any subgroup of the Poincaré group except the pure translations (if we demand compatibility of the reduction with all Poincaré supersymmetry transformations), and that a reduced supersymmetry requires at least a trivial Jacobian homomorphism of the isotropy subgroup for any subgroup of the conformal group. No other subgroup of $P(3,1)$ produces such reductions even if we relax the compatibility conditions by allowing each invariant field tranformed by supersymmetry to be equivalent up to a gauge transformation to a field of the set of invariant fields under the symmetry group chosen. Finally, the SO(3) invariant gauge and spinor fields of the $N=1$ supersymmetric Yang–Mills system with gauge group SU(3) are discussed in the Appendix as an example of a reduction that leaves no supersymmetry.

In future work, it would be interesting to study the symmetry reduction of higher-dimensional supersymmetric systems to four-dimensional supersymmetric models. For instance, this method can be applied to derive the SO(8) supergravity model from simple supergravity in 11 dimensions.[16] Further results have been published recently by Manton[17] who considered nontrivial coset spaces as extra-dimensional compact manifolds.

ACKNOWLEDGMENTS

The author would like to thank C. J. C. Burges, J. Harnad, and L. Vinet for useful discussions.

The author would also like to acknowledge a fellowship from the Natural Sciences and Engineering Research Council of Canada. This work was supported in part by funds provided by the U. S. Department of Energy (D. O. E.) under Contract No. DE-AC02-76ER03069.

APPENDIX: SO(3) INVARIANCE OF THE SU(3) ON-SHELL SIMPLE SUPER-YANG–MILLS MULTIPLET

In the following, we give an example of a reduction leading to a residual system which is not supersymmetric. For simplicity, we look at the $N=1$ supersymmetric Yang–Mills system with gauge group SU(3) reduced by symmetry SO(3). The SO(3) reduction of the same model with gauge group SU(2) leaves only a Yang–Mills field and a multiplet of scalar fields. We use for that purpose the Gell-Mann representation of the generators λ_a of SU(3), with $a=1,\ldots,8$, and their adjoint representation $(\lambda_a^{\text{Ad}})_b{}^c = -(f_a)_b{}^c$ $(a,b,c=1,\ldots,8)$, where the quantities $f_{ab}{}^c$ are the structure constants of the Lie algebra su(3). The orbits generated under the action of SO(3) are two-dimensional spheres that form a stratum parametrized by the time $t\in\mathbb{R}$ and the radii $r\in\mathbb{R}^+$ of the spheres, leaving the origin as a singular orbit.

The invariant gauge field is obtained either by the invariance condition (2.3) or equivalently by the Theorem 2 of Ref. 5. In the former case, the transformation function can be written as an extension of the homomorphism λ of the isotropy group SO(2) into SU(3). For any isotropy point x_0 along the z axis, except the origin, the isotropy subgroup consists of the rotations $e^{\chi L_3}$. Accordingly, we choose the following homomorphism (up to a $\mathbb{Z}_2$ factor) of SO(3) for the transformation function (ρ^{-1}):

$$\rho^{-1}: g = e^{\phi L_3}e^{\theta L_2}e^{\chi L_3} \in \mathrm{SO}(3) \mapsto \begin{bmatrix} e^{\phi\tau_3}e^{\theta\tau_2}e^{\chi\tau_3} & 0 \\ 0 & 1 \end{bmatrix} \in \mathrm{SU}(2)\subset\mathrm{SU}(3), \tag{A1}$$

where ϕ, θ, and χ are the rotation angles, $\{L_i\}$ are the generators of SO(3), and $\{\tau_i \equiv \sigma_i/2i\}$ are the generators of SU(2). Let us mention that the isomorphism of SO(3) onto an SO(3) subgroup of SU(3) implies a zero invariant spinor field and that the trivial homomorphism $[\lambda(G_0)=\{e\}]$ permits only a pure (zero curvature) invariant Yang–Mills field. [As for the other homomorphisms λ, they can not be extended to global homomorphisms (ρ^{-1}) including the origin, and we ignore them.]

The isotropy condition for the gauge field (2.4) is explicitly

$$\omega_\mu(x_0)^{g_0}J(x_0)^\mu{}_\nu = \mathrm{Ad}\,\lambda(g_0)\omega_\nu(x_0), \tag{A2}$$

for every $g_0\in G_0$. Its solution at the isotropy point $x_0=(0,0,0,z\neq 0)$ is given in matrix form by

$$[\omega(x_0)]_\mu{}^a = \begin{bmatrix} 0 & 0 & C & 0 & 0 & 0 & 0 & E \\ A & B & 0 & 0 & 0 & 0 & 0 & 0 \\ -B & A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & D & 0 & 0 & 0 & 0 & F \end{bmatrix}, \tag{A3}$$

where $A,\ldots,F$ are constants. Inserting this expression in the invariance condition (2.3) and letting the parameters $A,\ldots,F$ vary smoothly (C^∞) on the stratum, we obtain the most general SO(3) invariant SU(3) gauge field (for λ):

$$\begin{aligned}\omega_\mu(x) = {} & \delta_{\mu 0}(Cx^i\lambda'_i + E\lambda'_8) \\ & + \delta_{\mu i}[A\lambda'_i + (D-A)(x^ix^j/r^2)\lambda'_j \\ & - Be^{ijk}(x^j/r)\lambda'_k + Fx^i\lambda'_8],\end{aligned} \tag{A4}$$

where $A,\ldots,F$ are functions of r and t, and $\lambda'_a \equiv \lambda_a/2i$ $(i,j,k=1,2,3)$.

For the spinor field, the isotropy condition (2.2) reads

$$\Psi(x_0) = [D^{(1/2,0)}({}^{g_0}J(x_0))] \otimes \mathrm{Ad}\,\lambda(g_0)\Psi(x_0), \tag{A5}$$

where $\Psi \in \mathbb{C}^{4\times 8}$ is a Dirac spinor field. At the isotropy point $x_0 = (0, 0, 0, z)$, the solution to (A4) has the form

$$\Psi(x_0) = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} \lambda'_4 + \begin{bmatrix} -i\alpha \\ i\beta \\ -i\gamma \\ i\delta \end{bmatrix} \gamma'_5 + \begin{bmatrix} \alpha' \\ \beta' \\ \gamma' \\ \delta' \end{bmatrix} \gamma'_6 + \begin{bmatrix} i\alpha' \\ -i\beta' \\ i\gamma' \\ i\delta' \end{bmatrix} \lambda'_7, \tag{A6}$$

where $\alpha,\ldots,\delta$ and $\alpha',\ldots,\delta'$ are constants. Imposing the Majorana constraint requires that

$$\alpha = -\delta^*, \quad \beta = \gamma^*, \quad \alpha' = -\delta'^*, \quad \beta' = \gamma'^*. \tag{A7}$$

We determine the SO(3) invariant Majorana spinor field with values in the gauge algebra su(3) by substituting (A6) and (A7) in the invariance condition (2.1) and by allowing α, β, α', and β' to depend smoothly on the parameters r and t of the orbits. For simplicity, we write

$$\Psi^{\mathrm{Maj}}(t, r, \theta, \phi) = \begin{bmatrix} e^{\phi\tau_3} e^{\theta\tau_2} & O_2 \\ O_2 & e^{\phi\tau_3} e^{\theta\tau_2} \end{bmatrix} \otimes \mathrm{Ad}\left(e^{\phi\lambda'_3} e^{\theta\lambda'_2}\right) \Psi^{\mathrm{Maj}}(t,r), \tag{A8}$$

where $\Psi^{\mathrm{Maj}}(t,r)$ stands for the spinor (A6) with the restrictions (A7). Let us point out that the SO(3) invariant spinor $\Psi^{\mathrm{Maj}}(t, r, \theta, \phi)$ belongs to the subspace of su(3) spanned by $\{\lambda_4, \lambda_5, \lambda_6, \lambda_7\}$, and that the SO(3) invariant gauge field (A3) takes values in the orthogonal subspace (with respect to the Killing form) generated by $\{\lambda_1, \lambda_2, \lambda_3, \lambda_8\}$. From this fact, it is easily seen by introducing the invariant fields (A3) and (A8) in the supersymmetry transformations (4.2) and (4.3) that no supersymmetry is left.

However, as mentioned in Sec. IV, the compatibility conditions can be weakened by demanding instead a residual supersymmetry up to a gauge transformation. But even in that case, the conditions (4.1) and (4.2) are still too stringent. The relation (4.21) for $X_\mu(x_0, \mathscr{E})$ is equivalent to the isotropy condition (A2) for $\omega_\mu(x_0)$, and therefore the solution X_μ has the form (A3), except that the parameters can depend on the Majorana spinor $\mathscr{E}$. It follows that the expression (4.23) cannot be satisfied even if we add a gauge transformation since $\mathrm{Ad}\, h(x_0, \mathscr{E})$ is an automorphism of the gauge algebra.

[1] J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980).
[2] P. Forgacs and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[3] M. Légaré and H. Harnard, J. Math. Phys. **25**, 1542 (1984).
[4] J.-P. Antoine and M. Jacques, Class. Quantum Gravit. **1**, 431 (1984).
[5] J. Doneux, Y. Saint-Aubin, and L. Vinet, Phys. Rev. D **25**, 484 (1982).
[6] M. Légaré and J. Harnad, J. Math. Phys. **27**, 620 (1986).
[7] W. Nahm, Nucl. Phys. B **135**, 149 (1978).
[8] L. Brink, J. H. Schwarz, and J. Scherk, Nucl. Phys. B **121**, 77 (1977).
[9] J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[10] S. Ferrara and B. Zumino, Nucl. Phys. B **79**, 413 (1974).
[11] A. Salam and J. Strathdee, Phys. Lett. B **51**, 353 (1974).
[12] J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[13] J. Beckers, J. Harnad, and P. Jasselette, J. Math. Phys. **21**, 2491 (1980).
[14] M. Sohnius, Phys. Rep. **128**, 39 (1985).
[15] J. Patera, P. Winternitz, and H. Zassenhauss, J. Math. Phys. **16**, 1597 (1975).
[16] E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1979).
[17] N. S. Manton, Ann. Phys. (NY) **167**, 328 (1986).

On gauge fields with sources

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D. F., Mexico

(Received 5 February 1986; accepted for publication 10 September 1986)

It is shown that in an algebraically special space-time that admits a congruence of null strings, the Yang–Mills equations with sources reduce to a pair of nonlinear first-order differential equations for two matrices, provided that the gauge field is aligned with the congruence. In the case where the current is tangent to the null strings, the gauge field is determined by a matrix potential that has to satisfy a second-order differential equation with quadratic nonlinearities. As an example of this case, the Yang–Mills–Weyl equations are reduced, assuming that the multiplet of Weyl neutrino fields are also aligned with the congruence, and a reduced form of the Einstein–Yang–Mills–Weyl equations is also given.

I. INTRODUCTION

In a previous paper[1] it was shown that, for an arbitrary gauge group, the source-free Yang–Mills equations reduce to a single nonlinear second-order differential equation for a matrix potential, provided that the field strength vanishes along a family of totally null two-dimensional surfaces on which the conformal curvature is zero. This restriction, which is the key in the process of integration of the equations, concerns the gauge field as well as the space-time. The latter, owing to the assumption that its conformal curvature vanishes on totally null two-dimensional subspaces, is algebraically special. The fact that these subspaces are tangent to a family, or congruence, of surfaces constitutes a further restriction.

The existence of a congruence of two-dimensional, totally null, geodesic surfaces ("null strings"), suitably correlated with the curvature of (the complex extension of) the space-time, is an essential ingredient in the treatment followed in the study of the self-dual gauge fields as well as in certain approaches used to obtain algebraically special solutions of Einstein's equations or of the equations for other fields in such spaces (see, e.g., Ref. 1 and the references cited therein, see also Ref. 2).

In the present paper the Yang–Mills equations with sources are reduced, with the restriction on the field strength mentioned above. In the more general case we show that the gauge field is expressed in terms of two matrices that must satisfy one pair of nonlinear first-order differential equations. In the special case where the source of the field is also aligned with the congruence of null strings, the field is given in terms of a single matrix potential that obeys a nonlinear second-order differential equation. As an example of this special case, the simultaneous reduction of the equations for a gauge field coupled to a multiplet of spin-$\frac{1}{2}$ massless fields (Weyl neutrino fields) is given. Finally, we show how one can take into account the interaction of these fields with the gravitational field, by reducing the coupled Einstein–Yang–Mills–Weyl equations.

II. PRELIMINARIES

At a point of (the complex extension of) the space-time, each totally null two-dimensional subspace is determined by a one-index spinor, l_A or $l_{\dot A}$ (depending on whether the bivector corresponding to the subspace is self-dual or anti-self-dual). The subspaces defined by a spinor field, e.g., $l_{\dot A}$, are tangent to a family of two-dimensional surfaces provided that

$$l^{\dot A} l^{\dot B} \nabla_{C\dot A} l_{\dot B} = 0. \tag{2.1}$$

If $l_{\dot A}$ satisfies the additional condition

$$l^{\dot A} l^{\dot B} l^{\dot C} C_{\dot A\dot B\dot C\dot D} = 0, \tag{2.2}$$

where $C_{\dot A\dot B\dot C\dot D}$ denotes the spinor components of the anti-self-dual part of the conformal curvature (i.e., the conformal curvature restricted to the subspaces defined by $l_{\dot A}$ vanishes), then the metric of the space-time can be written as[3–5]

$$g = 2\phi^{-2}\, dq^A \underset{s}{\otimes} (dp_A + Q_{AB}\, dq^B), \tag{2.3}$$

where q^A and p^A are (complex) coordinate functions, ϕ is a function that satisfies the equation

$$l^{\dot A} \nabla_{B\dot C} l_{\dot A} = l_{\dot C} l^{\dot A}\, \partial_{B\dot A} \ln \phi, \tag{2.4}$$

and the functions Q_{AB} are such that $Q_{AB} = Q_{BA}$. The tangent vectors

$$\partial_{A\dot 1} = \sqrt{2}\,\frac{\partial}{\partial p^A} \equiv \sqrt{2}\,\partial_A,$$
$$\partial_{A\dot 2} = \sqrt{2}\phi^2\left(\frac{\partial}{\partial q^A} + Q_{AB}\,\partial^B\right) \equiv \sqrt{2}\phi^2 D_A \tag{2.5}$$

form a null tetrad (with $\partial_{A\dot B}\cdot\partial_{C\dot D} = -2\epsilon_{AC}\epsilon_{\dot B\dot D}$). (The spinor indices are raised and lowered according to $\psi_A = \epsilon_{AB}\psi^B$, $\psi^B = \epsilon^{AB}\psi_A$, and similarly for dotted indices.) The differential operators defined in (2.5) satisfy the commutation relations

$$\partial^A \partial_A = 0, \tag{2.6a}$$

 0022-2488/87/040940-05$02.50

$$\partial^A D_B = D_B \partial^A + (\partial^A Q_{BC})\partial^C, \tag{2.6b}$$

$$D^A D_A = (D^A Q_{AB})\partial^B. \tag{2.6c}$$

III. REDUCTION OF THE YANG–MILLS EQUATIONS

The Yang–Mills equations with sources, written in spinor notation, are

$$\nabla^A{}_{\dot R} F_{AB} + [A^A{}_{\dot R}, F_{AB}] = j_{B\dot R}, \tag{3.1a}$$

$$\nabla_R{}^{\dot A} F_{\dot A\dot B} + [A_R{}^{\dot A}, F_{\dot A\dot B}] = j_{R\dot B}, \tag{3.1b}$$

with

$$F_{AB} = \nabla_{(A}{}^{\dot R} A_{B)\dot R} + A_{(A}{}^{\dot R} A_{B)\dot R}, \tag{3.2a}$$

$$F_{\dot A\dot B} = \nabla^R{}_{(\dot A} A_{|R|\dot B)} + A^R{}_{(\dot A} A_{|R|\dot B)}. \tag{3.2b}$$

(Round brackets denote symmetrization on the indices enclosed. The indices between bars are excluded from the symmetrization.) Matrices F_{AB} and $F_{\dot A\dot B}$ are matrices that represent the spinor components of the self-dual and the anti-self-dual parts of the field strength, respectively. Here, $F_{AB} = F_{BA}$ and $F_{\dot A\dot B} = F_{\dot B\dot A}$. $A_{A\dot B} \equiv A(\partial_{A\dot B})$, where A denotes the matrix-valued potential one-form of the gauge field, $\nabla_{A\dot B}$ is the covariant derivative with respect to the Levi-Civita connection along $\partial_{A\dot B}$, and $j_{A\dot B}$ denotes the spinor components of the matrix-valued current, which is produced by a matter field coupled to the gauge field. If the gauge group consists of unitary matrices then, for a real tangent vector v, $A(v)$ is skew Hermitian, therefore if the tetrad $\partial_{A\dot B}$ satisfies $\overline{\partial_{A\dot B}} = \partial_{B\dot A}$, then $(A_{A\dot B})^\dagger = -A_{B\dot A}$, $(F_{AB})^\dagger = -F_{\dot A\dot B}$, and $(j_{A\dot B})^\dagger = -j_{B\dot A}$, where † denotes the Hermitian adjoint.

In the forthcoming we will reduce Eqs. (3.1) assuming that the space-time admits a nontrivial solution, $l_{\dot A}$, of (2.1) and (2.2) and that the field strength satisfies the alignment condition

$$l^{\dot A} l^{\dot B} F_{\dot A\dot B} = 0. \tag{3.3}$$

With these assumptions, the metric of the space-time is given by (2.3), and with respect to the basis (2.5), condition (3.3) amounts to $F_{\dot 1\dot 1} = 0$. Then, expressing all the components with respect to the tetrad (2.5), from Eq. (3.2b) it follows that[1]

$$A_{R\dot 1} = \sqrt{2} M^{-1} \partial_R M, \tag{3.4}$$

where M is a nonsingular matrix. In other words, the condition $F_{\dot 1\dot 1} = 0$ means that the curvature of the gauge field, represented by the field strength, restricted to the null strings, vanishes. Since $A_{R\dot 1} = A(\partial_{R\dot 1})$ and $\partial_{R\dot 1}$ are tangent to the null strings, the matrices $A_{R\dot 1}$ determine the parallel transport (in the "internal space") along the directions tangent to the null strings, therefore, there exists a guage in which $A_{R\dot 1}$ is equal to zero. The matrix M represents the transformation to the gauge in which $A_{R\dot 1}$ vanishes. However, the fact that the tangent directions to the null strings cannot be all real implies that M does not necessarily belong to the gauge group.

Using (3.4) and the connection coefficients[4,5] for the tetrad (2.5), one finds that the Yang–Mills equations (3.1) are given explicitly by

$$\sqrt{2}\partial^A (M\phi^{-2} F_{AB} M^{-1}) = M\phi^{-2} j_{B\dot 1} M^{-1}, \tag{3.5a}$$

$$\sqrt{2}\{D^A (M\phi^{-2} F_{AB} M^{-1}) - (\partial_B Q^{AR}) M\phi^{-2} F_{AR} M^{-1}\} + [M(\phi^{-2} A^A{}_{\dot 2} - \sqrt{2} M^{-1} D^A M) M^{-1}, M\phi^{-2} F_{AB} M^{-1}] = M\phi^{-4} j_{B\dot 2} M^{-1}, \tag{3.5b}$$

$$\sqrt{2}\partial_R (M\phi^{-2} F_{\dot 1\dot 2} M^{-1}) = M\phi^{-2} j_{R\dot 1} M^{-1}, \tag{3.6a}$$

$$\sqrt{2}\partial_R (M\phi^{-4} F_{\dot 2\dot 2} M^{-1}) - \sqrt{2} D_R (M\phi^{-2} F_{\dot 1\dot 2} M^{-1}) - [M(\phi^{-2} A_{R\dot 2} - \sqrt{2} M^{-1} D_R M) M^{-1}, M\phi^{-2} F_{\dot 1\dot 2} M^{-1}] = M\phi^{-4} j_{R\dot 2} M^{-1}, \tag{3.6b}$$

with

$$M\phi^{-2} F_{AB} M^{-1} = \sqrt{2}\,\partial_{(A}\{M(\phi^{-2} A_{B)\dot 2} - \sqrt{2} M^{-1} D_{B)} M) M^{-1}\}, \tag{3.7}$$

$$2M\phi^{-2} F_{\dot 1\dot 2} M^{-1} = \sqrt{2}\,\partial^R \{M(\phi^{-2} A_{R\dot 2} - \sqrt{2} M^{-1} D_R M) M^{-1}\}, \tag{3.8a}$$

$$M\phi^{-4} F_{\dot 2\dot 2} M^{-1} = \sqrt{2} D^R \{M(\phi^{-2} A_{R\dot 2} - \sqrt{2} M^{-1} D_R M) M^{-1}\} + M(\phi^{-2} A^R{}_{\dot 2} - \sqrt{2} M^{-1} D^R M) \times (\phi^{-2} A_{R\dot 2} - \sqrt{2} M^{-1} D_R M) M^{-1}. \tag{3.8b}$$

The integrability condition of Eq. (3.6a) is obtained by applying ∂^R on both sides of the equation; this gives [see (2.6a)] $\partial^R (M\phi^{-2} j_{R\dot 1} M^{-1}) = 0$. Therefore, under the present assumptions, the source of the gauge field must be such that, locally,

$$M\phi^{-2} j_{R\dot 1} M^{-1} = 2\sqrt{2}\,\partial_R \Lambda, \tag{3.9}$$

where Λ is a matrix made out of the matter field coupled to the gauge field. The matrix Λ is gauge invariant and conformal invariant.

From the Yang–Mills equations (3.1) it follows that the current is "conserved" in the sense that

$$\nabla^{A\dot B} j_{A\dot B} + [A^{A\dot B}, j_{A\dot B}] = 0.$$

Using (3.4), this equation amounts to

$$D^A(\phi^{-2} j_{A\dot 1}) - (\partial_S Q^{SA})\phi^{-2} j_{A\dot 1} + (1/\sqrt{2})[\phi^{-2} A^A{}_{\dot 2}, \phi^{-2} j_{A\dot 1}] - M^{-1}\partial^A (M\phi^{-4} j_{A\dot 2} M^{-1}) M = 0.$$

Then, substituting (3.9) and using the commutation relations (2.6), one finds

$$\partial^A D_A \Lambda - [M((1/\sqrt{2})\phi^{-2} A^A{}_{\dot 2} - M^{-1} D^A M) M^{-1}, \partial_A \Lambda] + (1/2\sqrt{2})\partial^A (M\phi^{-4} j_{A\dot 2} M^{-1}) = 0. \tag{3.10}$$

By substituting Eq. (3.9) into (3.6a) it follows that $M\phi^{-2} F_{\dot 1\dot 2} M^{-1} = 2(\Lambda + \epsilon)$, where ϵ is a matrix that depends on q^A only. If we replace $\Lambda + \epsilon$ by Λ, then the expression for $j_{R\dot 1}$ is left unchanged [see Eq. (3.9)] and

$$M\phi^{-2} F_{\dot 1\dot 2} M^{-1} = 2\Lambda. \tag{3.11}$$

Using Eqs. (3.8a) and (3.11), Eq. (3.10) can be rewritten as

$$\partial^A\{D_A \Lambda + [M((1/\sqrt{2})\phi^{-2} A_{A\dot 2} - M^{-1} D_A M) M^{-1}, \Lambda] + (1/2\sqrt{2}) M\phi^{-4} j_{A\dot 2} M^{-1}\} = 0, \tag{3.12}$$

which implies the existence of a matrix Ξ such that

$$M\phi^{-4}j_{A\dot{2}}M^{-1} = 2\sqrt{2}\{-D_A\Lambda - [M((1/\sqrt{2})\phi^{-2}A_{A\dot{2}} - M^{-1}D_AM)M^{-1},\Lambda] + \partial_A\Xi\}. \quad (3.13)$$

The matrix Ξ is also a gauge- and conformal-invariant quantity made out of the field that acts as a source of the gauge field (see Sec. IV).

We shall now introduce a pair of matrices C_A such that

$$\partial^A C_A = 2\Lambda, \quad (3.14)$$

then from (3.8a) and (3.11) it follows that, locally,

$$M(\phi^{-2}A_{R\dot{2}} - \sqrt{2}M^{-1}D_RM)M^{-1} = \sqrt{2}(C_R + \partial_R H),$$

where H is some matrix. By replacing $C_A + \partial_A H$ by C_A, we still have $\partial^A C_A = 2\Lambda$ and

$$A_{R\dot{2}} = \sqrt{2}\phi^2\{M^{-1}D_RM + M^{-1}C_RM\}. \quad (3.15)$$

Substituting (3.11), (3.13), and (3.15) into Eq. (3.6b) we get $\partial_R(M\phi^{-4}F_{\dot{2}\dot{2}}M^{-1}) = 2\partial_R\Xi$, hence $M\phi^{-4}F_{\dot{2}\dot{2}}M^{-1} = 2(\Xi + \delta)$, where δ is a matrix which depends on q^A only. If we now replace $\Xi + \delta$ by Ξ, then the expression for $j_{A\dot{2}}$ is unaltered [see Eq. (3.13)] and

$$M\phi^{-4}F_{\dot{2}\dot{2}}M^{-1} = 2\Xi. \quad (3.16)$$

On the other hand, from Eqs. (3.8b) and (3.15) we have $M\phi^{-4}F_{\dot{2}\dot{2}}M^{-1} = 2(D^RC_R + C^RC_R)$. Therefore, the matrices C_A, defined by Eq. (3.14), must fulfill the condition

$$D^RC_R + C^RC_R = \Xi. \quad (3.17)$$

The components F_{AB} can now be obtained by substituting (3.15) into (3.7), which yields

$$M\phi^{-2}F_{AB}M^{-1} = 2\,\partial_{(A}C_{B)}. \quad (3.18)$$

Then the remaining Yang–Mills equations [Eqs. (3.5)] are satisfied as a consequence of (3.9), (3.13)–(3.15), and (3.17).

In the case where the current is also aligned with the congruence of null strings, in the sense that the current is tangent to those surfaces, with respect to the tetrad (2.5) one has $j_{A\dot{1}} = 0$. Then, from Eq. (3.9), one finds that $\Lambda = \epsilon(q^A)$ and from (3.14) it follows that $C_A = -\epsilon p_A + \partial_A H$, which substituted into Eq. (3.17) leads to

$$D^A\partial_A H + (\partial^A H - \epsilon p^A)(\partial_A H - \epsilon p_A) + p^A\frac{\partial\epsilon}{\partial q^A} = \Xi. \quad (3.19)$$

In this case, Eqs. (3.15) and (3.18) take the form

$$A_{R\dot{2}} = \sqrt{2}\phi^2\{M^{-1}D_RM + M^{-1}(\partial_R H - \epsilon p_R)M\}, \quad (3.20a)$$

$$F_{AB} = 2\phi^2M^{-1}(\partial_A\,\partial_B H)M. \quad (3.20b)$$

Finally, when the gauge field is source-free, from Eqs. (3.13) and (3.20a) we get

$$\Xi = p^A\frac{\partial\epsilon}{\partial q^A} + [H,\epsilon] + \delta.$$

Hence Eq. (3.19) reduces to

$$D^A\partial_A H + \partial^A H\,\partial_A H + [p^A\partial_A H - H,\epsilon] = \delta, \quad (3.21)$$

which was previously given in Ref. 1.

IV. REDUCTION OF THE YANG–MILLS–WEYL EQUATIONS

The components of a multiplet of Weyl neutrino fields $\psi_A{}^i$, where $A = 1,2$ is a spinor index and $i = 1,2,...,n$ labels the components with respect to a basis in the "internal space," interacting with a gauge field obey the equations

$$\nabla^{A\dot{B}}\psi_A{}^i + A^{A\dot{B}i}{}_j\psi_A{}^j = 0, \quad (4.1)$$

where $A^{A\dot{B}i}{}_j$ denotes the elements of the matrix $A^{A\dot{B}}$, or, in matrix form regarding ψ_A as a column,

$$\nabla^{A\dot{B}}\psi_A + A^{A\dot{B}}\psi_A = 0. \quad (4.2)$$

The "adjoint" components of the multiplet, $\psi_{\dot{A}i}$, fulfill

$$\nabla^{A\dot{B}}\psi_{\dot{B}} - \psi_{\dot{B}}A^{A\dot{B}} = 0, \quad (4.3)$$

where $\psi_{\dot{B}}$ is the row $\psi_{\dot{B}} = (\psi_{\dot{B}1},...,\psi_{\dot{B}n})$.

We shall impose the alignment conditions $l^A\psi_A = 0$, and that for the gauge field given in (3.3), with l_A being a solution of Eqs. (2.1) and (2.2). Expressing again the components with respect to the basis (2.5), condition $l^A\psi_A = 0$ amounts to $\psi_{\dot{1}} = 0$. Then, using Eq. (3.4) and the connection coefficients[4,5] for the tetrad (2.5), from Eq. (4.3) one gets $\partial^A(\phi^{-5/2}\psi_{\dot{2}}M^{-1}) = 0$, which implies that

$$\psi_{\dot{2}} = \sqrt{2}\phi^{5/2}\zeta M, \quad (4.4)$$

where $\zeta = (\zeta_1,...,\zeta_n)$ depends only on q^A.

Similarly, from Eq. (4.2), with $\dot{B} = \dot{2}$, it follows that $\partial^A(M\phi^{-3/2}\psi_A) = 0$. Therefore, there exists locally a column η such that

$$\psi_A = \phi^{3/2}M^{-1}\partial_A\eta. \quad (4.5)$$

And from (4.2) with $\dot{B} = \dot{1}$, using (3.15) one finds that

$$D^A(\phi^{-3/2}\psi_A) - \tfrac{1}{2}\partial_CQ^{CA}\phi^{-3/2}\psi_A + (M^{-1}D^AM + M^{-1}C^AM)\phi^{-3/2}\psi_A = 0.$$

Hence, substituting the expression given in (4.5) one has

$$D^A\partial_A\eta - \tfrac{1}{2}(\partial_CQ^{CA})\partial_A\eta + C^A\partial_A\eta = 0. \quad (4.6)$$

The matrix-valued current produced by a multiplet of Weyl neutrino fields is given by $j_{A\dot{B}m}{}^l = ik\psi_A{}^l\psi_{\dot{B}m}$, where k is a (real) coupling constant, or in abbreviated form, $j_{A\dot{B}} = ik\psi_A\otimes\psi_{\dot{B}}$. The coupling constant k can be determined, e.g., from the Lagrangian density for the composite system. In this way the relative normalization of the energy-momentum tensors of these fields is also fixed (see Sec. V). In the present case, from (4.4) and (4.5) we have

$$\begin{aligned} j_{A\dot{1}} &= 0, \\ j_{A\dot{2}} &= \sqrt{2}ik\phi^4M^{-1}(\partial_A\eta\otimes\zeta)M. \end{aligned} \quad (4.7)$$

Thus we are in the special case considered in the previous section, where $\Lambda = \epsilon(q^A)$. Then, using (3.20a), a comparison of (3.13) with (4.7) leads to

$$\Xi = p^A\frac{\partial\epsilon}{\partial q^A} + [H,\epsilon] + \delta + \frac{1}{2}ik\eta\otimes\zeta,$$

where $\delta = \delta(q^A)$, and from Eq. (3.19) we get

$$D^A\partial_A H + \partial^A H\,\partial_A H + [p^A\partial_A H - H,\epsilon] = \delta + \tfrac{1}{2}ik\eta\otimes\zeta. \quad (4.8)$$

The fields are given in Eqs. (3.4), (3.20), and [see Eqs. (3.11) and (3.16)]

$$F_{\dot{1}\dot{2}} = 2\phi^2M^{-1}\epsilon M,$$

$$F_{\dot{2}\dot{2}} = 2\phi^4M^{-1}\left\{p^A\frac{\partial\epsilon}{\partial q^A} + [H,\epsilon] + \delta + \frac{1}{2}ik\eta\otimes\zeta\right\}. \quad (4.9)$$

Finally, substituting $C_A = -\epsilon p_A + \partial_A H$ into (4.6) we get

$$D^A\partial_A\eta - \tfrac{1}{2}(\partial_C Q^{CA})\partial_A\eta + (\partial^A H - \epsilon p^A)\partial_A\eta = 0. \quad (4.10)$$

The pair of coupled equations (4.8) and (4.10) for the unknowns H and η, with ϵ, δ, and ζ being arbitrary functions of q^A only, is equivalent, under the imposed restrictions, to the full system of the coupled Yang–Mills–Weyl equations. Here we are not considering the gravitational effects produced by these fields, i.e., they are taken as test fields on a background space-time that admits a congruence of null strings. In Sec. V we shall show how the interaction with the gravitational field can be included.

V. REDUCTION OF THE EINSTEIN–YANG–MILLS–WEYL EQUATIONS

We shall take the energy-momentum tensor of the multiplet of Weyl neutrino fields as given by (cf., Ref. IV)

$$\begin{aligned} T_{AB\dot C\dot D} &= (i\hbar/16)\{(\nabla_{B\dot D}\psi_{\dot C} - \psi_{\dot C}A_{B\dot D})\psi_A \\ &\quad + (\nabla_{A\dot C}\psi_{\dot D} - \psi_{\dot D}A_{A\dot C})\psi_B \\ &\quad - \psi_{\dot C}(\nabla_{B\dot D}\psi_A + A_{B\dot D}\psi_A) \\ &\quad - \psi_{\dot D}(\nabla_{A\dot C}\psi_B + A_{A\dot C}\psi_B)\}, \end{aligned} \quad (5.1)$$

and that of the gauge field as

$$T_{AB\dot C\dot D} = -(1/\pi)\mathrm{Tr}\, F_{AB}F_{\dot C\dot D}, \quad (5.2)$$

where Tr denotes the trace. Both expressions are gauge-invariant. This choice amounts to taking the coupling constant k introduced in Sec. IV as $k = \pi\hbar/4$. By substituting the expressions given by Eqs. (3.4), (3.20), (4.4), (4.5), and (4.9) one finds that the components of the total energy-momentum tensor, with respect to the tetrad (2.5), are

$$\begin{aligned} T_{AB\dot1\dot1} &= 0, \\ T_{AB\dot1\dot2} &= (1/8\pi)\phi^4\,\partial_A\,\partial_B T_0, \\ T_{AB\dot2\dot2} &= -(1/8\pi)\phi^6\,\partial_{(A}T_{B)}, \end{aligned}$$

where

$$T_0 = -\pi i\hbar\zeta\eta - 32\,\mathrm{Tr}\,\epsilon H, \quad (5.3a)$$

$$\begin{aligned} T_A &= -2\pi i\hbar\left\{\left(\frac{\partial\zeta}{\partial q^A}\right)\eta - \zeta D_A\eta + 2\zeta\epsilon\eta p_A - 2\zeta(\partial_A H)\eta\right\} \\ &\quad + 32\,\mathrm{Tr}\left\{\left(p^B\frac{\partial\epsilon}{\partial q^B} + [H,\epsilon] + \delta\right)\partial_A H - H\frac{\partial\epsilon}{\partial q^A}\right\}. \end{aligned} \quad (5.3b)$$

We shall now outline the derivation of a reduced form of the equations for the system considered in Sec. IV coupled with the gravitational field, using the results given in Ref. 4.

With the alignment conditions $l^A\psi_A = 0$ and (3.3), where $l_{\dot A}$ defines a congruence of null strings [i.e., $l_{\dot A}$ satisfies Eq. (2.1)], and from the Einstein equations, it follows that $l^{\dot C}l^{\dot D}C_{AB\dot C\dot D} = 0$, where the $C_{AB\dot C\dot D}$ are the spinor components of the traceless part of the Ricci tensor, then Eq. (2.2) is automatically satisfied.[6]

Condition $T_{AB\dot1\dot1} = 0$ implies $\phi = J_A p^A + \kappa$, where J_A and κ depend on q^A only.[3–5] Choosing a coordinate system q^A, p^A such that when $J_A = 0$ (case I), κ is equal to 1, and in the case where $J_A \neq 0$ (case II), J_A and κ are constant, one finds that[4,5]

$$\text{I:}\quad Q_{AB} = -\partial_{(A}B_{B)} - \tfrac{2}{3}L_{(A}p_{B)}, \quad (5.4)$$

where $L_A = L_A(q^B)$ and B_A denotes the most general solution of

$$\phi^{-1}\partial_A(\phi^2 B^A) = 4T_0, \quad (5.5)$$

and

$$\text{II:}\quad Q_{AB} = -\phi^3\partial_{(A}B_{B)} + \mu\phi^3 K_A K_B, \quad (5.6)$$

where $\mu = \mu(q^A)$, K_A is a pair of constants such that $K^A J_A = 1$ and B_A denotes the most general solution of (5.5). Then, using Eqs. (3.33) of Ref. 4 and Eqs. (5.3) above, one finds that the functions B_A, which essentially determine the metric, must satisfy the further condition

$$\begin{aligned} \text{I:}\quad &\partial_A\left\{D^C B_C - \frac{1}{2}\partial^B B^C\partial_{(B}B_{C)} - L^C B_C + \frac{1}{18}(L_B p^B)^2 - \frac{1}{6}\frac{\partial L_C}{\partial q^B}p^B p^C - N_B p^B - 32\,\mathrm{Tr}\left(p^B\frac{\partial\epsilon}{\partial q^B} + \delta\right)H\right\} \\ &= 4\pi i\hbar\zeta\{D_A\eta - \tfrac{1}{2}L_A\eta + (\partial_A H - \epsilon p_A)\eta\} + 64\,\mathrm{Tr}\,\epsilon\{D_A H - L_A H + \tfrac{1}{2}[\partial_A H, H] - \tfrac{1}{2}\delta p_A\}, \end{aligned} \quad (5.7)$$

where $N_A = N_A(q^B)$, and

$$\begin{aligned} \text{II:}\quad &\partial_A\left\{\phi D^C B_C - \tfrac{1}{2}\phi^4\partial^B B^C\partial_{(B}B_{C)} + (\phi J^C B_C)^2 + \mu\phi^3 K^C B_C\right. \\ &\left. + \frac{1}{2}K^B p_B[K^D p_D J^C - (\phi+\kappa)K^C]\frac{\partial\mu}{\partial q^C} - N_B p^B - 32\,\mathrm{Tr}\left(p^B\frac{\partial\epsilon}{\partial q^B} + \delta\right)H\right\} \\ &= 4\pi i\hbar\zeta\{D_A\eta + J^B\partial_{(A}\phi^2 B_{B)}\eta + \tfrac{3}{2}\mu\phi^2 K_A\eta + (\partial_A H - \epsilon p_A)\eta\} \\ &\quad + 64\,\mathrm{Tr}\epsilon\{D_A H + 2J^B\partial_{(A}\phi^2 B_{B)}H + 3\mu\phi^2 K_A H + \tfrac{1}{2}[\partial_A H, H] - \tfrac{1}{2}\delta p_A\} + 3\mu\phi^2 B_A. \end{aligned} \quad (5.8)$$

Consistency of Eqs. (5.7) and (5.8) requires that ∂^A, when applied to the right-hand side of these equations, yields zero. In fact, using Eqs. (4.8) and (4.10) it can be verified that this consistency condition is fulfilled. Hence, the right-hand side of (5.7) and (5.8) must be, locally, of the form $\partial_A S$, where S is a function. If one can write Eqs. (4.8) and (4.10) in the form $\partial^A R_A = 0$ (as a "continuity equation"), which implies the existence of χ (a matrix or a column, respectively) such that $R_A = \partial_A\chi$, then the function S mentioned above can be written in terms of χ and ∂_A can be eliminated from both sides of Eqs. (5.7) and (5.8). Thus, B_A would be determined by a first-order equation and, since the R_A's would involve only the first derivatives of H or η, Eqs. (4.8) and (4.10), written as $R_A = \partial_A\chi$, are also of first

order. Therefore, finding such potentials χ would allow us, introducing the χ's as new unknowns, to reduce the differential order of Eqs. (4.8), (4.10), and (5.7) or (5.8) by one unit. Unfortunately, the fact that H and η are matrices complicates the determination of the χ's. However, they can be readily obtained when the gauge field is absent[4] and, in case I, when the Weyl field is absent.[1]

VI. FINAL REMARKS

In order to obtain real fields, the solutions of Eqs. (3.17), (4.6), (5.7), and (5.8) must satisfy further conditions corresponding to the assumed alignments of the Yang–Mills and Weyl fields and to the algebraic degeneracy of the curvature of the space-time, e.g., for a real gauge field, the self-dual part of the field strength given by Eq. (3.18) must satisfy a condition of the form $l^A l^B F_{AB} = 0$, which is an additional constraint on C_A and where l_A must satisfy equations analogous to (2.1) and (2.2). The necessity of considering these further conditions is a consequence of the use of complex coordinates and complex tetrads in the derivation followed here. Nevertheless, as compared with the original equations, taking into account their nonlinearity, differential order, and the number of unknowns, the results presented here represent a considerable simplification. An example of the usefulness of this approach is provided by the study of linear perturbations of source-free gauge fields given in Ref. 7, where one is able to eliminate all reference to complex coordinates in the final covariant expressions.

ACKNOWLEDGMENT

The author acknowledges partial support from the Sistema Nacional de Investigadores (Mexico).

[1] G. F. Torres del Castillo, J. Math. Phys. **26**, 836 (1985).
[2] M. S. Hickman and C.B. G. McIntosh, "Complex relativity and real solutions III: Real type N solutions from complex $N \otimes N$ ones," Gen. Relativ. Gravit. **18**, 107 (1986).
[3] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976). See also J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).
[4] G. F. Torres del Castillo, J. Math. Phys. **24**, 590 (1983).
[5] We employ the same notation as in Ref. 4, only changing undotted indices to dotted ones and vice versa.
[6] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[7] G. F. Torres del Castillo, Int. J. Theor. Phys. **25**, 883 (1986).

Integrals for two-loop calculations in massless QCD

G. Kramer and B. Lampe
II. Institut für Theoretische Physik der Universität Hamburg, West Germany

(Received 16 June 1986; accepted for publication 26 November 1986)

Tools that have been developed for the extension of the Sterman–Weinberg formula in two-loop order are presented. They are the essential ingredients for any perturbative two-loop calculation in a massless theory. Introducing dimensional regularization, this paper deals with poles up to fourth order in the dimensional parameters. The virtual two-loop integrations as well as real ones over phase space are studied.

I. INTRODUCTION

In this work we evaluate various integrals that appeared in the course of the calculation of e^+e^- jet cross sections to order α_s^2. This calculation[1] was done in the framework of massless perturbative QCD. Dimensional regularization has been used throughout for the regularization of the ultraviolet and the infrared divergencies. (Ultraviolet and infrared singularities can in principle be disentangled by examining the low energy and the high energy behavior of the integrals. We do not distinguish them in the following, because the ultraviolet singularities are removed by counterterms which are known *a priori*.) In the course of the calculation of the two jet cross sections,[1] which is the one most involved, poles of up to fourth order in the dimensional parameter appeared. [If n is the dimension we define $\varepsilon = (4-n)/2$.] Thus this is a true two-loop calculation because one-loop results contain at most poles ε^{-2}. So this work extends earlier papers on the techniques[2] of dimensionally regularized massless theories, which all stop at the next to leading order level. We hope that the techniques we present will be of help in other two-loop QCD calculations for ep and $p\bar{p}$ processes where at the moment the ε^{-2} level is arrived at.[3]

Some of the methods are described only for the scalar integrals. However, they can all be extended to the case where numerators (coming from the Dirac traces) are involved. We will present also results for this case, e.g., Eq. (33). The calculation of the total hadronic coss section in e^+e^- annihilation can be reduced to the calculation of the photon two-point function via the optical theorem. Though it is a genuine two-loop problem fortunately only poles up to ε^{-2} appeared here and could be handled with the help of the Gegenbauer technique.[4] For completeness we will review some features of this technique in the beginning of Sec. II. However, for the three-, four-, and five-point functions with one particle off shell, which are needed for jet cross sections, this method is of no advantage as compared to the usual technique of introducing Feynman parameters.[5] Then we give an elegant way for calculating the ladder diagram [Fig. 2(a)] and use Feynman parameters to calculate the crossed diagram [Fig. 2(b)]. In Sec. III we discuss box diagrams, where a one-loop virtual integration has to be done and also one real gluon has to be integrated out. This means that one integrates over those parts of phase space where the gluon is infrared or collinear with one of the other partons. This adds poles in ε to the poles coming from the virtual integrations.

Sections IV and V give two prescriptions how to handle the tree-level diagrams, where a two-gluon bremsstrahlung has to be integrated out. Again poles of up to fourth order in ε emerge.

In physical applications[1] these cancel against the infrared singularities from the virtual integrations.

II. INTEGRALS FOR VIRTUAL TWO-LOOP DIAGRAMS

The Gegenbauer expansion technique in momentum space[6] proves useful for diagrams with only two external legs, e.g., for the vacuum polarization of the photon.[4] In Refs. 4 and 7 extensive use is made of it. Here we describe some of its features for the sake of completeness.

The Gegenbauer polynomials C_j^λ, $j=0,1,\ldots$, $\lambda > -\frac{1}{2}$ form a complete set of functions in the interval $(-1,1)$ (Ref. 8). They are a generalization of the Legendre polynomials ($\lambda = \frac{1}{2}$) and the Chebyshev polynomials ($\lambda = 1$). The method consists in expanding propagators into Gegenbauer polynomials [cf. Eq. (2) of Ref. 6]. Then one does the angular integrations with the help of the orthogonality relation [Eq. (A1) of Ref. 7] and is left with radial integrations in the form of powers of the momenta. In simple cases such as the example in Ref. 6 this leads to infinite series which can be evaluated with the help of standard summation formulas.[9] The method works for integrals that depend only on one momentum, i.e., self-energy graphs. They must fulfill an additional property, namely there must exist a parametrization of integration momenta in such a way that not more than two Gegenbauer polynomials appear within one angular integration. For integrals of products of three Gegenbauer polynomials no simple closed form exists.

In the case of the vacuum polarization of the photon in order α_s^2 (Ref. 4) all graphs are calculable with the Gegenbauer technique with the exception of the graph of Fig. 1, which is "irreducible."

Next we consider the two-parton contributions of the two-jet cross section (Fig. 2). These diagrams can also be thought to define the (singular) electromagnetic form factor of massless quarks.[10] After the Dirac traces are done the dot products that emerge are transformed into sums of squares of momenta in such a way that as many of them as possible cancel against denominator factors, e.g.,

$$kp = \tfrac{1}{2}k^2 - \tfrac{1}{2}(k-p)^2, \tag{1}$$

for $p^2 = 0$. In general one is left with relatively complicated scalar integrals, which we will calculate in the following, and simple tensorial integrals. For the tensorial as well as for

 0022-2488/87/040945-12$02.50

FIG. 1. The contribution to the vacuum polarization of the photon in $O(\alpha_s^2)$. The symbols in brackets always denote the momenta of the particles.

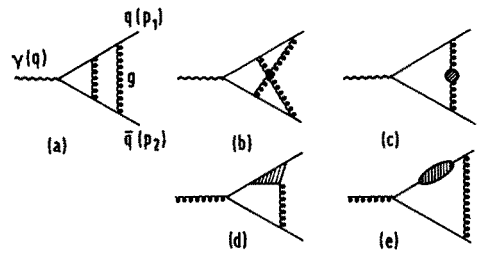


FIG. 2. Two parton contributions to the two-jet cross section in $O(\alpha_s^2)$. A shaded region always means a one-loop insertion.

some of the scalar integrals [Figs. 2(c)–2(e)] the strategy is to introduce Feynman parameters with the help of

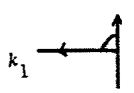
$$\frac{1}{k^2(k-p)^2}=\int_0^1 \frac{dx}{[(k-px)^2]^2} \tag{2}$$

which holds for $p^2=0$, i.e., one tries to preserve the massless structure of the theory. In general, one ends up with standard integrals given in Appendix A.

Figures 2(a) and 2(b) cannot be met with this general strategy. However, planar diagrams like Fig. 2(a) can be solved with the help of "partial integration." This method has been described in Ref. 11. Figure 2(a) can be reduced to the scalar integral

$$L_s=\int \frac{d^nk\,d^nl}{k^2l^2(l+p_1)^2(l-p_2)^2(k+l+p_1)^2(k+l-p_2)^2} \tag{3}$$

and some tensorial integrals which can be calculated by using the standard strategy (2). [The integrations are done in Euclidean space. Continuation to Minkowski space is done in the result of the integration in such a way that the propagators are the usual causal propagators $(k^2+i\mu)^{-1}$.] To calculate L_s by the method of Ref. 11 we need some notation. Let —✕— mean an additional factor of k^2 in the numerator and —●— an additional factor of k^2 in the denominator. Define



$$\equiv 2k_1(k_1+k_2) \tag{4}$$

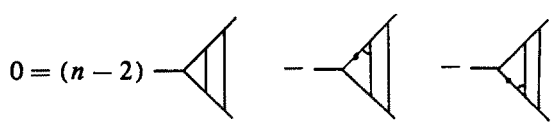
$$=(k_1+k_2)^2+k_1^2-k_2^2 \tag{5}$$

(6)

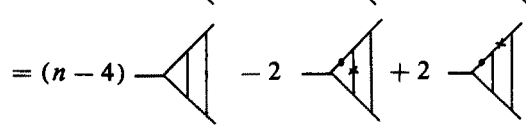
Because of the translational invariance of the integral one has the identity

$$\int d^nk\,d^nl\,\frac{\partial}{\partial k_\mu}\,\frac{(k+l-p_2)_\mu}{k^2l^2(l+p_1)^2(l-p_2)^2(k+l+p_1)^2(k+l-p_2)^2}=0. \tag{7}$$

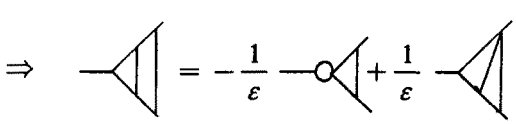
After differentiation one receives



$$0=(n-2)\qquad -\qquad -$$

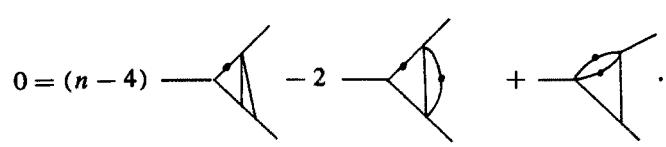


$$=(n-4)\qquad -2\qquad +2 \tag{8}$$



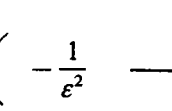
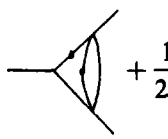
$$\Rightarrow\qquad =-\frac{1}{\varepsilon}\qquad +\frac{1}{\varepsilon}\qquad . \tag{9}$$

The first integral on the right-hand side is very simple and can be done with the standard formulas of Appendix A. The second integral can be simplified in the same way as the original one:



$$0=(n-4)\qquad -2\qquad + \qquad . \tag{10}$$

The result is

$$L_s = -\frac{1}{\varepsilon}$$  $$-\frac{1}{\varepsilon^2}$$  $$+\frac{1}{2\varepsilon}2$$  . (11)

With the help of Appendix A one gets

$$L_s = \pi^{4-2\varepsilon}(q^2)^{-2-2\varepsilon}\Gamma(1-\varepsilon)\Gamma(1+\varepsilon)\Gamma(1+2\varepsilon)\left\{\frac{1}{4\varepsilon^4}+\frac{\zeta_2}{2\varepsilon^2}+\frac{11\zeta_3}{2\varepsilon}+\frac{3}{4}\zeta_4\right\}. \tag{12}$$

Using Feynman parameters Gonsalves[10] has also arrived at this result. However, our method is much more elegant. The momentum dependence of L_s could have been derived from a simple dimensional analysis. The appearance of

$$\zeta_s = \sum_{k=1}^{\infty} k^{-s} \tag{13}$$

in connection with poles ε^{s-4} is a characteristic feature of the result (12).

The crossed diagram [Fig. 2(b)] cannot be done with this method because it is not planar. [To see this try to apply (8) to it!] The scalar crossed integral

$$K_s = \int \frac{d^n k\, d^n l}{k^2 l^2 (l-p_1)^2 (k-l)^2 (k-l-p_2)^2 (k-p_1-p_2)^2} \tag{14}$$

can, however, be calculated with the help of Feynman parameters,

$$K_s = \int_0^1 dx \int_0^1 dy \int \frac{d^n k\, d^n l}{k^2 (l-p_1 x)^4 (k-l-p_2 y)^4 (k-p_1-p_2)^2}. \tag{15}$$

Any further Feynman parameter is accompanied by a "mass term" in the denominator. One gets

$$K_s = \pi^{2-\varepsilon}(q^2)^{-1-\varepsilon}(3+\varepsilon)(2+\varepsilon)\int_0^1 dx \int_0^1 dy \int_0^1 dz \int_0^1 du\, u^{1+\varepsilon}(1-u)\int \frac{d^n k}{(k^2+C)^{4+\varepsilon}}, \tag{16}$$

$$C = z(1-z)q^2(1-u) + u(1-u)(qz - xp_1 - yp_2)^2. \tag{17}$$

The k integration can be done with Eq. (A1).

The y integration is then straightforward:

$$K_s = -\pi^{4-2\varepsilon}(q^2)^{-2-2\varepsilon}\frac{2}{\varepsilon}\frac{\Gamma^2(1-\varepsilon)\Gamma(1+2\varepsilon)}{\Gamma(1-2\varepsilon)}\int_0^1 du\, u^{\varepsilon}(1-u)^{-1-2\varepsilon}$$
$$\times \int_0^1 dx \int_0^1 \frac{dz}{z-x} z^{-1-2\varepsilon}(1-z)^{-1-2\varepsilon}\left\{\left(1-u\frac{z-x}{z}\right)^{-1-2\varepsilon} - \left(1-u\frac{x-z}{1-z}\right)^{-1-2\varepsilon}\right\}. \tag{18}$$

The divergence for $x=z$ is an artifact of the y integration and disappears in the difference. The u integration leads to hypergeometric functions. If one wants their series representation to converge one has to distinguish the regions $x<z$ and $x>z$. Then

$$K_s = \pi^{4-2\varepsilon}(q^2)^{-2-2\varepsilon}\Gamma(1-\varepsilon)\Gamma(1+2\varepsilon)\Gamma(1+\varepsilon)\hat{I}\cdot\varepsilon^{-2}, \tag{19}$$

$$\hat{I} = \int_0^1 dz \int_0^z dx\, z^{-1-2\varepsilon}(1-z)^{-1-2\varepsilon}F\left(1-\frac{x}{z}\right)(z-x)^{-1}$$
$$+\int_0^1 dz \int_0^z dx\, x^{-1-2\varepsilon}(1-z)^{-1-2\varepsilon}G\left(1-\frac{z}{x}\right)(z-x)^{-1}$$
$$-\int_0^z dz \int_z^1 dx\, z^{-1-2\varepsilon}(1-z)^{-1-2\varepsilon}F\left(1-\frac{1-x}{1-z}\right)(z-x)^{-1}$$
$$-\int_0^1 dz \int_z^1 dx\, (1-x)^{-1-2\varepsilon}z^{-1-2\varepsilon}G\left(1-\frac{1-z}{1-x}\right)(z-x)^{-1}, \tag{20}$$

with

$$F(a) = {}_2F_1(1+2\varepsilon, 1+\varepsilon, 1-\varepsilon, a), \tag{21}$$

$$G(a) = {}_2F_1(1+2\varepsilon, -2\varepsilon, 1-\varepsilon, a). \tag{22}$$

We change from x to $1-x$ and from z to $1-z$ in the last two terms of $\hat{I}$ and get

$$\hat{I} = 2(I_0 + I_1 + I_2), \tag{23}$$

$$I_0 = \int_0^1 dz \int_0^z dx\, z^{-1-2\varepsilon}[(1-z)^{-1-2\varepsilon} - (1-x)^{-1-2\varepsilon}](z-x)^{-1}$$
$$= \Gamma^{-1}(1+2\varepsilon)\sum_0 \frac{\Gamma(k+1+2\varepsilon)}{k!(k-2\varepsilon)}(\psi(k+1)-\psi(1)) \tag{24}$$

is the contribution from the first term in the series expansion of the hypergeometric functions. Here I_1 and I_2 contain the integration of all other terms. Because of absolute convergence of the hypergeometric series one can exchange summation and integration. The sums that finally appear can be done with the help of standard formulas.[9] The final result is

$$K_s = \pi^{4-2\varepsilon}(q^2)^{-2-2\varepsilon}\Gamma(1-\varepsilon)\Gamma(1+\varepsilon)\Gamma(1+2\varepsilon) \times\left\{\frac{1}{\varepsilon^4}-\frac{9\zeta_2}{\varepsilon^2}-\frac{25\zeta_3}{\varepsilon}+\frac{15}{2}\zeta_4\right\}. \tag{25}$$

From completeness we give also the results for the scalar integrals of Figs. 2(c)–2(e),

$$\int\frac{d^nk\,d^nl}{k^2(k-p_1)^4(k-p_1-p_2)^2l^2(k-p_1-l)^2} = v_0\left(-\frac{3}{4\varepsilon^2}-3+\frac{3}{2}\zeta_2\right), \tag{26}$$

$$\int\frac{d^nk\,d^nl}{k^2l^2(k-l)^2(l-p_1)^2(k-p_1)^2(k-p_1-p_2)^2} = v_0\left(\frac{3}{4\varepsilon^3}-\frac{3}{2\varepsilon^2}+\frac{3-9/2\zeta_2}{\varepsilon}-6-\frac{15}{2}\zeta_3+9\zeta_4\right), \tag{27}$$

$$\int\frac{d^nk\,d^nl}{k^2l^2(k-p_1)^2(k+l-p_1)^2(k+p_2)^2} = v_0\left(\frac{3}{2\varepsilon^2}+\frac{3}{2\varepsilon}+\frac{9}{2}-6\zeta_2\right). \tag{28}$$

Here

$$v_0 = \pi^{4-2\varepsilon}(q^2)^{-2-2\varepsilon}\Gamma(1-\varepsilon)\Gamma(1+\varepsilon)\Gamma(1+2\varepsilon). \tag{29}$$

In the case of Fig. 2(b) one cannot reduce all the tensor structure from the Dirac trace to the scalar integral K_s and the standard integrals of Appendix A. In addition one needs

$$R_l = \int\frac{d^nk\,d^nl(2kp_2)^l}{k^2l^2(l-p_1)^2(k-l)^2(k-l-p_2)^2(k-p_1-p_2)^2}, \tag{30}$$

for $l = 1,2,3$.

The denominator in Eq. (14) is invariant under the two transformations $(p_1\leftrightarrow p_2, l\leftrightarrow k-l)$ and $(k\leftrightarrow p_1+p_2-k, l\leftrightarrow p_1-l)$. Using the second invariance one can derive

$$R_1 = K_s/2, \tag{31}$$

$$R_3 = -K_s/2+3R_2/2. \tag{32}$$

One can calculate R_2 by the same methods as K_s:

$$R_2 = \pi^{4-2\varepsilon}(q^2)^{-2-2\varepsilon}\frac{\Gamma^3(1-\varepsilon)\Gamma(1+2\varepsilon)}{\Gamma(1-3\varepsilon)}\left\{\frac{1}{2\varepsilon^4}+\frac{1}{\varepsilon^3} +\frac{4}{\varepsilon^2}-\frac{5\zeta_2}{2\varepsilon^2}+\left(14-4\zeta_2-\frac{17}{2}\zeta_3\right)\frac{1}{\varepsilon} +46-12\zeta_2-\frac{8}{5}\zeta_3-\frac{85}{4}\zeta_4\right\}. \tag{33}$$

III. DISCUSSION OF BOX DIAGRAMS

In Sec. II the "two-loop" calculation consisted in doing two virtual integrations. Here we consider diagrams like Fig. 3, where one loop integration has to be done and one real particle (namely the gluon) has to be integrated "out" subsequently (i.e., a phase space integration over infrared and collinear regions has to be carried out). As an example we consider the scalar box integral with three particles on mass shell $p_1^2 = p_2^2 = p_3^2 = 0$. The case with all four particles on shell is contained in it. If all incoming and outgoing particles are off shell the integral can be done in four dimensions,

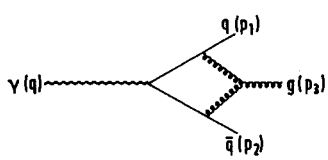


FIG. 3. A typical box diagram.

because no infrared singularity exists. (Box integrals are ultraviolet finite anyhow.) Four-dimensional box integrals are calculated in the third paper of Ref. 2. To calculate

$$I_{\text{Box}} = \int\frac{d^nk}{k^2(k+p_2)^2(k-p_3)^2(k-p_1-p_3)^2} \tag{34}$$

we follow the general strategy described in Sec. II. Namely we introduce Feynman parameters avoiding mass terms in the denominator

$$I_{\text{Box}} = \int_0^1 dx\int_0^1 dy\int\frac{d^nk}{(k+p_2x)^4(k-p_3-p_1y)^4}. \tag{35}$$

After a shift of variables one can use the standard formula (A2) for doing the virtual integration. The Feynman parameter integrations lead to hypergeometric functions

$$I_{\text{Box}} = -2(q^2)^{-2-\varepsilon}\pi^{2-\varepsilon}\frac{\Gamma^2(1-\varepsilon)\Gamma(1+\varepsilon)}{\varepsilon^2\Gamma(1-2\varepsilon)}\frac{1}{y_{13}y_{23}} \times\left\{{}_2F_1\left(1,-\varepsilon,1-\varepsilon,-\frac{y_{12}}{y_{13}y_{23}}\right) -y_{13}^{-\varepsilon}\,{}_2F_1\left(1,-\varepsilon,1-\varepsilon,-\frac{y_{12}}{y_{23}}\right) -y_{23}^{-\varepsilon}\,{}_2F_1\left(1,-\varepsilon,1-\varepsilon,-\frac{y_{12}}{y_{13}}\right)\right\}, \tag{36}$$

where the $y_{ij} = 2p_i\,p_j/q^2$ are dot products normalized to the energy and $y_{12}+y_{23}+y_{13} = 1$ is energy conservation. The ε^{-2} pole is due to the infrared singularity from the virtual integration. The q^2 dependence could also have been derived from a dimensional analysis of Eq. (34). Note also the singularity $y_{13}^{-1}y_{23}^{-1}$, which can produce another ε^{-2} pole after the integration over phase space. The hypergeometric functions

in Eq. (36) have the power series expansion

$$ {}_2F_1(1,-\varepsilon,1-\varepsilon,z)=1-\varepsilon\sum_{j=1}^{\infty}\frac{z^j}{j-\varepsilon}. \tag{37}$$

The series in Eq. (37) is the expansion of a logarithm generalized to $n=4-2\varepsilon$ dimensions. If expanded in a power series in ε we obtain

$$ \sum_{j=1}^{\infty}\frac{z^j}{j-\varepsilon}=-\ln(1-z)+\sum_{j=1}^{\infty}\varepsilon^j L_{j+1}(z), \tag{38}$$

where L_n is the generalized Euler dilogarithm

$$ L_n(z)=\sum_{j=1}^{\infty}\frac{z^j}{j^n}. \tag{39}$$

For further integration over phase space the expansion (38) is not useful, because additional singularities may make the inclusion of L_n, $n>2$, necessary and at the integration boundaries (39) may not be convergent.

So we first transform the hypergeometric functions in Eq. (36) according to

$$ {}_2F_1\left(1,-\varepsilon,1-\varepsilon,-\frac{a}{b}\right)=\left(\frac{b}{a+b}\right)^{-\varepsilon}E\left(\frac{a}{a+b}\right), \tag{40}$$

with $E(z)={}_2F_1(-\varepsilon,-\varepsilon,1-\varepsilon,z)$. The hypergeometric function E has an everywhere convergent series expansion

$$ E(z)=1+\frac{\varepsilon^2}{\Gamma(1-\varepsilon)}\sum_{k=1}^{\infty}\frac{\Gamma(k-\varepsilon)}{k!(k-\varepsilon)}z^k, \tag{41}$$

$$\begin{aligned} I_{\rm Box}=&-2(q^2)^{-2-\varepsilon}\pi^{2-\varepsilon}\frac{\Gamma^2(1-\varepsilon)\Gamma(1+\varepsilon)}{\varepsilon^2\Gamma(1-2\varepsilon)}\\ &\times y_{13}^{-1-\varepsilon}y_{23}^{-1-\varepsilon}\left\{(y_{13}y_{23}+y_{12})^{\varepsilon}E\left(\frac{y_{12}}{y_{13}y_{23}+y_{12}}\right)\right.\\ &-(y_{12}+y_{23})^{\varepsilon}E\left(\frac{y_{12}}{y_{12}+y_{23}}\right)\\ &\left.-(y_{12}+y_{13})^{\varepsilon}E\left(\frac{y_{12}}{y_{12}+y_{13}}\right)\right\}. \end{aligned} \tag{42}$$

One now wants to integrate Eq. (42) over the three-particle phase space drawn in Fig. 4 (Ref. 12). One can divide it into a "two-jet" region where (42) has singularities and a "three-jet" region where (42) is finite. In the finite region $E(z)=1+\varepsilon^2L_2(z)$ and the integration of the L_2 term can be done in four dimensions.

The singular region is defined $y_{13}<y$ or $y_{23}<y$. Equation (42) has no singularity for $y_{12}\to 0$. So this region which physically is two-jet can be formally added to the finite three-jet region.

Here $y\ll 1$ is the invariant mass cut defining a jet.[12] It is both an energy and an angle cut. Here $\{y_{13}<y, y_{23}>y\}$ is the collinear region $1\|3$, $\{y_{13}>y, y_{23}<y\}$ is the collinear region $2\|3$, and $\{y_{13}<y, y_{23}<y\}$ is the infrared region $3\to 0$. Because of the symmetry the two-jet region of three-particle phase space can be written as

$$ \int_0^z dy_{13}\,y_{13}^{-\varepsilon}\left\{2\int_0^{1-y_{13}}-\int_0^{y}\right\}dy_{23}\,y_{23}^{-\varepsilon}(1-y_{13}-y_{23})^{-\varepsilon}. \tag{43}$$

As long as one neglects contributions of order y one can restrict oneself to the y_{13} pole in (42). So the two-jet contribution from the scalar box integral is

$$\begin{aligned} B_I:=&\int_0^y dy_{13}\,y_{13}^{-1-\varepsilon}\left\{2\int_0^1-\int_0^y\right\}\\ &\times dy_{23}\,y_{23}^{-\varepsilon}(1-y_{23})^{-\varepsilon}\lim_{y_{13}\to 0}y_{13}I_{\rm Box}, \end{aligned} \tag{44}$$

which can be calculated straightforwardly. The result has an ε^{-4} pole with a structure similar to the two-loop virtual results in (12) or (25). In addition one has terms $\ln^k y/\varepsilon^{4-k}$, $k=0,1,2,3,4$,

$$\begin{aligned} B_I=&\pi^{2-\varepsilon}(q^2)^{-2-2\varepsilon}\frac{\Gamma^2(1-\varepsilon)\Gamma(1+\varepsilon)}{\Gamma(1-2\varepsilon)}\\ &\times\left\{\frac{1}{2\varepsilon^4}-\frac{3\zeta_2}{2\varepsilon^2}-\frac{2\zeta_3}{\varepsilon}-\frac{25}{8}\zeta_4\right.\\ &+\frac{4\zeta_2\ln y}{\varepsilon}+4\zeta_3\ln y-6\zeta_2\ln^2 y\\ &\left.-\frac{2\ln^2 y}{\varepsilon^2}+\frac{4\ln^3 y}{\varepsilon}-\frac{14}{3}\ln^4 y\right\}. \end{aligned} \tag{45}$$

IV. TREE DIAGRAMS

In the case of a diagram like Fig. 5 one has to do no virtual integrations. However, two gluons have to be integrated "out" here to calculate its two-jet contribution—the three- and four-jet contributions from such diagrams have been discussed extensively in the literature.[12–17] In the four-jet case no ε singularities appear. In the three-jet case one has to deal with ε^{-2} singularities coming from an integration measure very similar to (43).

There exist two topologically distinct two-jet configurations here:

+ permutations + infrared configurations. (46)

Because of the pole structure of the matrix element squared only the permutations $1\leftrightarrow 2$ and/or $3\leftrightarrow 4$ are necessary. All other configurations which are physically two-jet lead to the $O(y)$ contributions. However, because of the appearance of terms $\sim y/\varepsilon$ they cannot be simply added to the three- or

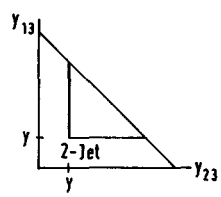


FIG. 4. Three-particle phase space for $e^+e^-\to q\bar{q}g$. It is divided into a two-jet region which contains the singularities and a three-jet region free of singularities.

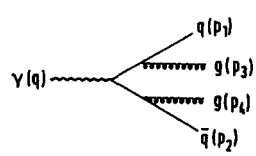


FIG. 5. A typical tree level diagram in $O(\alpha_s^2)$.

four-jet contributions; but must in principle be cancelled independently. In Sec. V we describe a method where such terms are avoided. That method will also provide us with all finite logarithms necessary for the two-jet cross section, whereas the procedure described here is only sufficient to cancel the singularities $\sim \ln^{k-m} y/\varepsilon^{4-k}$, $k=0,1,2,3$, $m=0,1,2$, $k \geqslant m$.

We have used the representation of four-particle phase space given in Ref. 14. In the two-jet limit (46) the four-particle phase space reduces to

$$2\int_0^y dy_{24}\, y_{24}^{-\varepsilon}\int_0^y dy_{13}\, y_{13}^{-\varepsilon}\int_y^1 dy_{134}\, y_{134}^{-\varepsilon}(1-y_{134})^{-\varepsilon}\int_0^1 du\, u^{-\varepsilon}(1-u)^{-\varepsilon}\int_0^\pi \frac{d\theta'}{N'_\theta}\sin^{-2\varepsilon}\theta'$$
$$+\int_0^y dy_{134}\, y_{134}^{1-2\varepsilon}\int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}\int_0^\pi \frac{d\theta'}{N_{\theta'}}\sin^{-2\varepsilon}\theta'$$
$$\times\left\{2\int_0^1 dy_{24}\int_0^1 du-\int_0^y dy_{24}\int_0^y dy_{23}-\int_0^y dy_{24}\int_y^1 dy_{23}-\int_0^y dy_{24}\int_0^y dy_{23}\right\} y_{24}^{-\varepsilon}(1-y_{24})^{1-2\varepsilon}u^{-\varepsilon}(1-u)^{-\varepsilon}. \quad (47)$$

Here $y_{ijk} := y_{ij}+y_{jk}+y_{ik}$, $u := y_{23}/(y_{12}+y_{23})$, $z = y_{13}/y_{134}y_{123}$, and $N_{\theta'} = 2^{2\varepsilon}\pi\Gamma(1-2\varepsilon)\Gamma^{-2}(1-\varepsilon)$ is the normalization of the θ' integration. The first term in (47) operates on the $y_{13}y_{24}$ poles of the matrix elements squared. In Fig. 6 we show how it appears as part of the total phase space. The second term in (47) corresponds to that part of four-particle phase space, where either $y_{134}<y$ or $y_{234}<y$. To visualize this region one should take Fig. 4 with y_{13} replaced by y_{134} and y_{23} replaced by y_{234}. Because of the symmetry of the matrix elements squared in $1\leftrightarrow 2$ one can take twice the contribution from the region $y_{134}<y$ and has to subtract overlap contributions. This is the content of the curly bracket in (47). (For an extensive discussion of this see Ref. 1.) Then the second term of the expression (47) operates on the y_{134}-double poles of the matrix elements squared. These double poles have a structure much more complicated than any pole structure which appeared so far. We will not write down the full expression here but only describe the calculation of two characteristic integrals. These integrals stem from poles in y_{14} in the matrix element squared,

$$\lim_{y_{134}\to 0} y_{14}/y_{134} = u(1-z)+zuy_{24} - 2\cos\theta'(u(1-u)z(1-z)y_{24})^{1/2} \quad (48)$$

has a relatively complicated structure (see Appendix B).

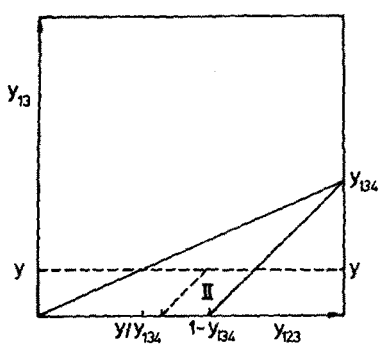


FIG. 6. Four-particle phase space for a fixed value of $y_{134}>y$. Region II is the two-jet region ($y_{13}<y$, $y_{24}<y$, $y_{134}>y$), which is the first part of Eq. (47).

We want to calculate

$$J^{\gamma\delta}_{\alpha\beta} := \int_0^1 dy\, y^{\alpha-\varepsilon}(1-y)^{\beta-2\varepsilon}\int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon} \times\int_0^1 du\, u^{\delta-\varepsilon}(1-u)^{\gamma-\varepsilon}\frac{I}{N_{\theta'}} \quad (49)$$

to order ε, where

$$I=\int_0^\pi \frac{d\theta'\sin^{-2\varepsilon}\theta'}{ay+b-2\sqrt{aby}\cos\theta'} \quad (50)$$

with $a=uz$, $b=(1-u)(1-z)$. The θ' integration can be done

$$\frac{I}{N_{\theta'}} = (ay+b)^{-1}\,{}_2F_1\left(\frac{1}{2},1,1-\varepsilon,\frac{4aby}{(ay+b)^2}\right). \quad (51)$$

The hypergeometric function can be transformed by a "quadratic transformation" [18] so that

$$I/N_{\theta'} = (s_+/2)^{2\varepsilon}r_-^{-1-2\varepsilon}\,{}_2F_1(-\varepsilon,-2\varepsilon,1-\varepsilon,s_-/s_+), \quad (52)$$

where $s_\pm = r_+ \pm r_-$, $r_\pm = |ay\pm b|$. We can set ${}_2F_1(-\varepsilon,-2\varepsilon,1-\varepsilon,s_-/s_+)=1$ because terms of order $\varepsilon^2 s_-/s_+$ give a contribution $O(\varepsilon)$ in the full result. The reason for that is the s_-/s_+ cancels all singularities which appear (u^{-1},y^{-1}) with the exception of the y_{134} singularity. So the u, y, and z integrations are finite. One has an ε^{-1} from the y_{134} integration but this pole is removed by the ε^2 factor.

Next the z integration is done. One divides the region of integration into two regions: (1) $z<\gamma$ and (2) $z>\gamma$ [$\gamma := v/\beta$, $\beta := 1-u(1-y)$, $v := 1-u$]. Then one has

$$J^{\gamma\delta}_{\alpha\beta} = \int_0^1 dy\, y^{\alpha-\varepsilon}(1-y)^{\beta-2\varepsilon} \times\int_0^1 du\, u^{\delta-\varepsilon}(1-u)^{\gamma-\varepsilon}(M_a+M_b), \quad (53)$$

where

$$\beta M_a = \beta\int_0^\gamma dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}(v-\beta z)^{-1-\varepsilon}v^{2\varepsilon}(1-z)^{2\varepsilon}$$
$$= \gamma^{-\varepsilon}\frac{\Gamma(1-\varepsilon)\Gamma(-2\varepsilon)}{\Gamma(1-3\varepsilon)}\,{}_2F_1(-\varepsilon,1-\varepsilon,1-3\varepsilon,\gamma)$$
$$= [\beta M_b]_{\gamma\to 1-\gamma}. \quad (54)$$

Expanding the hypergeometric function in Eq. (54) in powers of ε and $1-\gamma$ one gets[19]

$$\beta(M_a+M_b) = \left[\frac{\Gamma(1-\varepsilon)\Gamma(-2\varepsilon)}{\Gamma(1-3\varepsilon)}+\frac{\Gamma(\varepsilon)\Gamma(-2\varepsilon)}{\Gamma(-\varepsilon)}\right] \times\left(\frac{1-\gamma}{\gamma}\right)^{-\varepsilon}+\frac{\Gamma(1-\varepsilon)\Gamma(-\varepsilon)}{\Gamma(1-2\varepsilon)}+O(uy\varepsilon^2). \tag{55}$$

The terms of order $uy\varepsilon^2$ do not contribute [see the remarks after Eq. (52)]. We conclude

$$J^{\gamma\delta}_{\alpha\beta} = \frac{\Gamma(1-\varepsilon)\Gamma(-\varepsilon)}{\Gamma(1-2\varepsilon)}F^{\gamma\delta}_{1\alpha\beta} + \left[\frac{\Gamma(1-\varepsilon)\Gamma(-2\varepsilon)}{\Gamma(1-3\varepsilon)}+\frac{\Gamma(\varepsilon)\Gamma(-2\varepsilon)}{\Gamma(-\varepsilon)}\right]F^{\gamma\delta}_{2\alpha\beta}, \tag{56}$$

where

$$F^{\gamma\delta}_{n\alpha\beta} = \int_0^1 dy\int_0^1 dy\, y^{\alpha-n\varepsilon}(1-y)^{\beta-2\varepsilon} \times u^{\delta-n\varepsilon}(1-u)^{\gamma-(2-n)\varepsilon}(1-u(1-y))^{-1} \tag{57}$$

can be calculated by expanding $(1-u(1-y))^{-1}$ near 1:

$$F^{\gamma\delta}_{n\alpha\beta} = \frac{\Gamma(\beta+1-2\varepsilon)\Gamma(\alpha+\gamma+1-2\varepsilon)}{\Gamma(\alpha+\beta+1-(n+2)\varepsilon)}\sum_{k=0}^{\infty}\frac{\Gamma(k+\alpha+1-n\varepsilon)\Gamma(k+\delta+1-n\varepsilon)}{k!\Gamma(k+\gamma+\delta+\alpha+2-(n+2)\varepsilon)(k+\alpha+\beta+1-(n+2)\epsilon)}. \tag{58}$$

For all combinations of $\alpha,\beta,\gamma,\delta$ that appear in the problem the sum in Eq. (57) is convergent.

As an example we now calculate $J^{1\,-1}_{-1\,0}$. We have

$$F^{1\,-1}_{2\,-1\,0} = \Gamma^4(1-2\varepsilon)/4\varepsilon^2\Gamma^2(1-4\varepsilon), \tag{59}$$

$$F^{1\,-1}_{1\,-1\,0} = \frac{\Gamma^2(1-2\varepsilon)\Gamma^2(-\varepsilon)}{\Gamma^2(1-3\varepsilon)}+\frac{\Gamma^2(1-2\varepsilon)}{\Gamma(-3\varepsilon)} \times\sum_{k=1}^{\infty}\frac{\Gamma^2(k-\varepsilon)}{k!\Gamma(k+1-3\varepsilon)(k-3\varepsilon)}. \tag{60}$$

One has

$$\sum_{k=1}^{\infty}\frac{\Gamma^2(k-\varepsilon)}{k!\Gamma(k+1-3\varepsilon)(k-3\varepsilon)} = \sum_{k=1}^{\infty}\frac{\Gamma(k+\varepsilon)}{k!(k-2\varepsilon)(k-4\varepsilon)}+O(\varepsilon^2), \tag{61}$$

and the sum on the right-hand side of (60) can be done with (6.6.2) of Ref. 9. In total one gets

$$J^{1\,-1}_{-1\,0} = -\frac{1}{\varepsilon^3}+\frac{11\zeta_2}{2\varepsilon}+18\zeta_3+\frac{31}{2}\zeta_4\varepsilon+O(\varepsilon^2). \tag{62}$$

An integral closely related to $J^{\gamma\delta}_{\alpha\beta}$ is

$$K^{\gamma\delta}_{\alpha\beta} := \int_0^1 dy\, y^{\alpha-\varepsilon}(1-y)^{\beta-2\varepsilon} \times\int_0^1 dz\, z^{-1-\varepsilon}(1-z)^{-\varepsilon}\int_0^1 du\, u^{\delta-\varepsilon}(1-u)^{\gamma-\varepsilon} \times\int_0^{\pi}\frac{d\theta'}{N_{\theta'}}\frac{-2\sqrt{aby}\cos\theta'}{ay+b-2\sqrt{aby}\cos\theta'}. \tag{63}$$

Because of $\sqrt{aby}$ in the numerator it is less divergent than $J^{\gamma\delta}_{\alpha\beta}$. One can see this by writing [see (53)]

$$K^{\gamma\delta}_{\alpha\beta} = \int_0^1 dy\, y^{\alpha-\varepsilon}(1-y)^{\beta-2\varepsilon} \times\int_0^1 du\, u^{\delta-\varepsilon}(1-u)^{\gamma-\varepsilon}(N_a+N_b). \tag{64}$$

One finds

$$N_a+N_b = -2uy(N_a+N_b)-2\ln\gamma+\varepsilon\ln^2\gamma + 2\varepsilon L_2(1-\gamma)+O(\varepsilon^2 uy). \tag{65}$$

We write

$$-\ln\gamma+\epsilon\ln^2\gamma+2\varepsilon L_2(1-\gamma) = 2[(1-\gamma)/(1-\varepsilon)]\,{}_2F_1(1,1,2-\varepsilon,1-\gamma)+O(\varepsilon^2 uy). \tag{66}$$

So $K^{\gamma\delta}_{\alpha\beta}$ is reduced to $J^{\gamma\delta+1}_{\alpha+1\beta}$ modulo a third type of integral

$$L^{\gamma\delta}_{\alpha\beta} = \int_0^1\int_0^1\frac{dy\,du}{1-\varepsilon} \times y^{\alpha+1-\varepsilon}(1-y)^{\beta-2\varepsilon}u^{\delta+1-\varepsilon}(1-u)^{\gamma-\varepsilon} \times\frac{{}_2F_1(1,1,2-\varepsilon,uy/[1-u(1-y)])}{1-u(1-y)}, \tag{67}$$

which is finite for all combinations of $\alpha,\beta,\gamma,\delta$ that appear. Introducing γ instead of u as integration variable we can do the y integration. This leads to another hypergeometric function. Expanding the hypergeometric functions into series we can calculate $L^{\gamma\delta}_{\alpha\beta}$ in a standard way.

Instead of calculating all $J^{\gamma\delta}_{\alpha\beta}$ and $K^{\gamma\delta}_{\alpha\beta}$ in the 13-system (48) some of the integrals can be done more elegantly in other phase-space systems [e.g., (B20)–(B24)]. However, the expression $y_{12}/y_{13}\,y_{24}\,y_{14}\,y_{23}$ keeps its complexity in *any* system and can only be integrated using the methods described in Chap. 4.

V. THE PARTIAL FRACTIONING APPROACH

To get all finite contributions of the tree diagrams to the two- and three-jet cross section a representation of the matrix element squared has proved frutiful, which has singularities only when a certain y_{ij}, say y_{13}, is zero.[20,21] In contrast to Eq. (47) the stucture of the poles in ε can be derived from considering the region $y_{13}<y$ solely. Finite contributions from other regions can be easily included by numerical integrations.[21] In the three- and four-jet case they can even be calculated so as to include contributions of order y (Ref. 21).

We will give here some typical integrals which appear in the course of the analytical calculation of the singular contributions to the two-jet cross section. [In the three-jet case there is essentially one complicated integral, see Eq. (A1) of Ref. 14.]

One can write the matrix element squared in the form

$$\mathrm{ME} = (A/y_{13}) + (1-2) + (3-4) + (1-2,3-4). \tag{68}$$

Because of symmetry

$$(\text{full phase space})(A/y_{13}) = \tfrac{1}{4}(\text{full phase space})\mathrm{ME}. \tag{69}$$

So it is enough to consider the term A/y_{13}. The most interesting region is the region $y_{134}<y$, which is part of the two-jet region [cf. Eq. (B19)]. In the whole of Sec. V we shall restrict our attention to this region. Let us first consider an integral which originates from the pole component of the term A/y_{13} in Eq. (68). (By the pole component we mean the limit of A for y_{13} going to zero):

$$I_p = (\text{phase space})_{y_{134}<y} \times [y_{13}y_{134}(y_{13}+y_{23})(y_{13}+y_{24})]^{-1}. \tag{70}$$

Here

$$(\text{phase space})_{y_{134}<y} = \int_0^y dy_{134}\, y_{134}^{1-2\varepsilon} \int_0^1 dy_{123}\, y_{123}^{1-2\varepsilon}(1-y_{123})^{-\varepsilon} \times \int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon} \int_0^1 dv\, v^{-\varepsilon}(1-v)^{-\varepsilon} \tag{71}$$

[see (47)]. From

$$y_{123} = 1 - y_{24} + O(y_{134}), \tag{72}$$

$$z = y_{13}/y_{134}y_{123}, \tag{73}$$

$$u = 1 - v = y_{23}/(y_{123}-y_{13}), \tag{74}$$

one gets

$$I_p = \int_0^y dy_{134}\, y_{134}^{-1-2\varepsilon} \int_0^1 dz\, z^{-1-\varepsilon}(1-z)^{-\varepsilon} \times \int_0^1 dv\, v^{-\varepsilon}(1-v)^{-\varepsilon}(1-v(1-zy_{134}))^{-1} \times \int_0^1 dy_{123}\, y_{123}^{-2\varepsilon}(1-y_{123})^{-\varepsilon}(1-y_{123}(1-zy_{134}))^{-1}. \tag{75}$$

If one neglects terms of order y one can use the appropriate formula

$$\int_0^1 dx\, x^{k-a\varepsilon}(1-x)^{-\varepsilon}(1-x(1-zy_{134}))^{-1} = \Gamma(k+1-a\varepsilon)\Gamma(-\varepsilon)\Gamma^{-1}(k+1-(a+1)\varepsilon) \times \Gamma(1-\varepsilon)\Gamma(\varepsilon)z^{-\varepsilon}y_{134}^{-\varepsilon} + O(y_{134}), \tag{76}$$

which can be derived from 3.197(3) and 9.131 of Ref. 19. The final formula for J_p thus contains terms $\sim y^{-3\varepsilon}, y^{-4\varepsilon}$, which in Sec. IV appeared only after subtraction of doubly counted infrared regions [see (47)]:

$$I_p = (y^{-2\varepsilon}/-2\varepsilon)CB_1^2 + 2\Gamma(1-\varepsilon)\Gamma(\varepsilon)(C+B_1)(y^{-3\varepsilon}/-3\varepsilon)B_2 + 4\Gamma^2(1-\varepsilon)\Gamma^2(\varepsilon)(y^{-4\varepsilon}/-4\varepsilon)B_3, \tag{77}$$

$$B_n = \Gamma(1-\varepsilon)\Gamma(-n\varepsilon)/\Gamma(1-(n+1)\epsilon), \tag{78}$$

$$C = \Gamma(1-2\varepsilon)\Gamma(-\varepsilon)/\Gamma(1-3\varepsilon). \tag{79}$$

A remark is in order: in the scheme described in Sec. IV a denominator $\sim y_{13}y_{14}y_{23}y_{24}$ emerged. This made the θ' integration in the region $y_{134}<y$ so tedious. In the partial fractioning scheme it turns out that one has only three of those four y_{ij} appearing at a time, e.g., $y_{13}(y_{13}+y_{23})(y_{13}+y_{24})$ (I), as in Eq. (70), or $y_{13}(y_{13}+y_{14})(y_{14}+y_{24})$ (II). To avoid complicated θ' dependences in the case II we change the coordinate frame. Exchanging the role of particles 2 and 4 the θ' integration is trivial again and Eq. (76) can be applied. One should note that the structure of invariants is a little different in the two systems. In contrast to Eq. (48) one has here

$$\lim_{y_{134}\to 0} y_{12} = [y_{123}/(1-zy_{123})(1-u)(1-z) + uzy_{24} - 2\cos\theta'(u(1-u)z(1-z)y_{24})^{1/2}], \quad \text{with } u = y_{34}/(y_{134}-y_{13}). \tag{80}$$

Now we turn to some integrals which emerge in the nonpole part of Eq. (68). A typical example is

$$I_v := (\text{phase space})_{y_{134}<y} \times [y_{12}/y_{123}y_{134}(y_{13}+y_{14})(y_{13}+y_{24})]. \tag{81}$$

To avoid θ' dependences in the denominator we work in the system of Eq. (80). The θ' dependence in the numerator is removed by antisymmetry. One gets

$$I_v = \int_0^y dy_{134}\, y_{134}^{-1-2\varepsilon} \int_0^1 dy_{123}\, y_{123}^{1-2\varepsilon}(1-y_{123})^{-\varepsilon} \times \int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}H_v, \tag{82}$$

where

$$H_v := \int_0^1 \frac{dy\, u^{-\varepsilon}(1-u)^{1-\varepsilon}}{(1-u(1-zy_{123}))(1-y_{123}(1-zy_{134}))} + 2\int_0^1 \frac{du\, u^{-\varepsilon}(1-u)^{-\varepsilon}(2u-1)}{(1-zy_{123})(1-u(1-zy_{123}))}. \tag{83}$$

The second integral in Eq. (83) can be done after partial fractioning

$$\frac{1}{(1-zy_{123})(1-u(1-zy_{123}))} = \frac{1}{1-zy_{123}} + \frac{u}{1-u(1-zy_{123})}. \tag{84}$$

One is led to one integral of the type

$$K(k,l,m,n) := \int_0^1 dy\, y^{k-2\varepsilon}(1-y)^{l-\varepsilon} \times \int_0^1 dz\, z^{m-\varepsilon}(1-z)^{n-\varepsilon}(1-zy)^{-1} \tag{85}$$

and one integral of the type

$$I(k,l,m,n,a,b) := \frac{\Gamma(a+b-2\varepsilon)}{\Gamma(a-\varepsilon)\Gamma(b-\varepsilon)}\int_0^1 dy\, y^{k-2\varepsilon}(1-y)^{l-\varepsilon}\int_0^1 dz\, z^{m-\varepsilon}(1-z)^{n-\varepsilon}$$
$$\times\int_0^1 du\, u^{a-1-\varepsilon}(1-u)^{b-1-\varepsilon}(1-u(1-zy))^{-1}. \tag{86}$$

Here K can be calculated by expanding $(1-zy)^{-1}$ near 1, and I can be calculated by doing the u integration with the help of the same formulas that led to Eq. (76). In contrast to Eq. (76), in Eq. (86) no approximation can be made. Instead one introduces (absolutely convergent) hypergeometric series and gets

$$I(k,l,m,n,a,b) = [(a+b-1-2\varepsilon)/\Gamma(a-\varepsilon)]\Gamma(1+\varepsilon-b)\Gamma(l+1-\varepsilon)\Gamma(n+1-\varepsilon)(\sigma_2-\sigma_1), \tag{87}$$

where

$$\sigma_2 = \sum_{j=\infty}^{\infty} \frac{\Gamma(j+a+b-1-2\varepsilon)\Gamma(j+k+b-3\varepsilon)\Gamma(j+m+b-2\varepsilon)}{j!\Gamma(j+k+l+b+1-4\varepsilon)\Gamma(j+m+n+b+1-3\varepsilon)} \tag{88}$$

and

$$\sigma_1 = \sum_{j=0}^{\infty}\frac{\Gamma(j+a-\varepsilon)\Gamma(j+k+1-2\varepsilon)\Gamma(j+m+1-\varepsilon)}{\Gamma(j+2+\varepsilon-b)\Gamma(j+k+l+2-3\varepsilon)\Gamma(j+m+n+2-2\varepsilon)}. \tag{89}$$

Here $\sigma_2-\sigma_1$ converges for all values of k, l, m, n, a, and b that appear and can be calculated by a suitable expansion in ε.

Now we come to the first integral in Eq. (83). Call it H_1:

$$H_1 = \frac{\Gamma(2-\varepsilon)\Gamma(1-\varepsilon)}{\Gamma(3-2\varepsilon)}\,{}_2F_1(1,1-\varepsilon,3-2\varepsilon,1-zy_{123})$$
$$\times(1-y_{123}(1-zy_{134}))^{-1}. \tag{90}$$

Define H_l to be the limit of H_1 for $y_{123}\to 0$, so that in H_l Eq. (76) can be used and in $H_k = H_1 - H_l$ one can put $(1-y_{123}(1-zy_{134}))^{-1} = (1-y_{123})^{-1}$. Then for H_k one needs $I(1,-1,0,0,1,2)$ and for H_l

$$H_l := \int_0^1 dz\, z^{-2\varepsilon}(1-z)^{-\varepsilon}\,{}_2F_1(1,1-\varepsilon,3-2\varepsilon,1-z). \tag{91}$$

Both can be evaluated by standard techniques.[9,21,22]

In Sec. IV it was very simple to figure out whether a term contributed $O(y)$ in the region $y_{134}<y$. One simply had to put in the phase space variables z, u, and y_{24} and count the power of y_{134} that emerged. Only double poles in y_{134} had to be taken into account [see Eq. (71)]. Here we have an integrand which cannot be neglected even though formally it is of order y_{134}^{-1}. The integrand is $y_{12}y_{123}y_{124}/(y_{13}+y_{24})(y_{13}+y_{14})(y_{134}+y_{23})(y_{14}+y_{24})$. The contribution arises for $y_{24}\to 0$ as can be seen by partial fractioning

$$\frac{1}{(y_{13}+y_{24})(y_{14}+y_{24})} = \frac{1}{y_{13}-y_{14}}\left(\frac{1}{y_{24}+y_{14}}-\frac{1}{y_{24}+y_{13}}\right). \tag{92}$$

Working in the system of Eq. (80) one has

$$I_N = \int_0^y dy_{134}\, y_{134}^{-1-2\varepsilon}\int_0^1 dy_{123}\, y_{123}^{2\varepsilon}(1-y_{123})^{-\varepsilon}$$
$$\times\int_0^1 dv\, v^{2-\varepsilon}(1-v)^{-\varepsilon}$$
$$\times\int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}(z-v(1-z))^{-1}$$
$$\times(z+v(1-z))^{-1}(1-v(1-zy_{134}))^{-1}$$
$$\times\left[\frac{1}{1-y_{123}+v(1-z)y_{134}}-\frac{1}{1-y_{123}+zy_{134}}\right]. \tag{93}$$

In Eq. (93) the limit $y_{24}\to 0$ has been taken wherever it is allowed.

The y_{123} integration can be done with the help of Eq. (76). [One should convince oneself that it is allowed to replace $1-x(1-zy_{134})$ by $1-x+zy_{134}$ in the denominator of Eq. (76).] One gets

$$I_N = \Gamma(1-\varepsilon)\Gamma(\varepsilon)\int_0^y dy_{134}\, y_{134}^{-1-2\varepsilon}$$
$$\times\int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}\int_0^1 dv\, v^{2-\varepsilon}(1-v)^{-\varepsilon}$$
$$\times(v^{-\varepsilon}(1-z)^{-\varepsilon}-z^{-\varepsilon})(1-v(1-zy_{134}))^{-1}$$
$$\times(z+v(1-z))^{-1}(z-v(1-z))^{-1}. \tag{94}$$

Define I_N^l to be the limit of I_N for $v\to 1$, so that in $I_N^k = I_N - I_N^l$ one can put $(1-v(1-zy_{134}))^{-1} = (1-v)^{-1}$,

$$I_N^l = \Gamma(1-\varepsilon)\Gamma(\varepsilon)\int_0^y dy_{134}\, y_{134}^{-1-2\varepsilon}$$
$$\times\int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}(2z-1)^{-1}$$
$$\times((1-z)^{-\varepsilon}-z^{-\varepsilon})\int_0^1 dv\, v^{-\varepsilon}(1-v)^{-\varepsilon}$$
$$\times(1-v(1-zy_{134}))^{-1}. \tag{95}$$

In I_N^l the v integration can be done with the help of Eq. (76). One gets

$$I_N = \Gamma^2(1-\varepsilon)\Gamma^2(\varepsilon)(y^{-4\varepsilon}/-4\varepsilon)Z_0 + \Gamma(1-\varepsilon)\Gamma(\varepsilon)(y^{-3\varepsilon}/-3\varepsilon)Z_1, \tag{96}$$

where

$$Z_0 = \int_0^1 dz\, z^{-2\varepsilon}(1-z)^{-\varepsilon}\frac{(1-z)^{-\varepsilon}-z^{-\varepsilon}}{2z-1} \tag{97}$$

and

$$Z_1 = \int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}\int_0^1 dv\, v^{2-\varepsilon}(1-v)^{-1-\varepsilon} \times \frac{v^{-\varepsilon}(1-z)^{-\varepsilon}-z^{-\varepsilon}}{(1-(1-z)(1+v))(1-(1-z)(1-v))}. \tag{98}$$

Here Z_0 may be calculated by expanding in ε. This is not possible for Z_1, because Z_1 has a singularity for $v\to 1$. By successive partial fractioning one can isolate this singularity into an elementary integral

$$\frac{1}{(1-z_1(1+v))(1-z_1(1-v))} = \frac{1}{2z(1-z_1(1+v))} + \frac{1}{2z(1-z_1(1-v))}, \tag{99}$$

$$\frac{1}{(1-v)(1-z_1(1+v))} = \frac{1}{(1-2z_1)(1-v)} - \frac{z}{(1-2z_1)(1-z_1(1+v))}, \tag{100}$$

with $z_1 = 1-z$. The result is

$$Z_0 = \tfrac{3}{2}\zeta_2\varepsilon + (\tfrac{7}{2}\zeta_3 + 6\zeta_3\ln 2)\varepsilon^2 + (16\zeta_2\ln^2 2 - \tfrac{2}{3}\ln^4 2 + \tfrac{21}{8}\zeta_4 - 16L_4(\tfrac{1}{2}))\varepsilon^3 + O(\varepsilon^4), \tag{101}$$

$$Z_1 = -\tfrac{3}{2}\zeta_2 + (-\tfrac{21}{8}\zeta_3 - \tfrac{3}{2}\zeta_2 - \tfrac{9}{2}\zeta_2\ln 2)\varepsilon + (-9\zeta_2\ln 2 - 6\zeta_2\ln^2 2 + \tfrac{1}{4}\ln^4 2 - \tfrac{9}{8}\zeta_3 - \tfrac{33}{4}\zeta_4 - \tfrac{3}{2}\zeta_2 + 6L_4(\tfrac{1}{2}))\varepsilon^2 + O(\varepsilon^3).$$

Here[22]

$$L_4(\tfrac{1}{2}) = 0.517\,84. \tag{102}$$

VI. SUMMARY

We have described the calculation of various integrals which appear in two-loop calculations. All integrals can be calculated by analytical means. Properties of the hypergeometric functions and of the gamma function and its derivatives are the essential ingredients of the calculations.

ACKNOWLEDGMENTS

We acknowledge discussions with J. Bartels, W. Hollik, and G. Schierholz. We thank T. T. Wu for collaboration in the early stages of this work which led to the simple method in calculating the ladder diagram.

One of the authors (B.L.) was supported by Bundesministerium für Forschung und Technologie, Bonn, West Germany.

APPENDIX A: STANDARD VIRTUAL INTEGRALS

With the help of Ref. 5,

$$\int \frac{d^n k}{(2\pi)^n}\frac{(k^2)^r}{(k^2-C)^m} = \frac{i(-1)^{r-m}}{(16\pi^2)^{n/4}} C^{r-m+n/2} \times \frac{\Gamma(r+(m/2))\Gamma(m-r-(n/2))}{\Gamma(n/2)\Gamma(m)}, \tag{A1}$$

one gets

$$I := \int \frac{d^n k}{(k^2)^\alpha((k-q)^2)^\beta} = i(-q^2)^{-\varepsilon}\pi^{2-\varepsilon}(q^2)^{2-\alpha-\beta} \times \frac{\Gamma(2-\varepsilon-\beta)\Gamma(2-\varepsilon-\alpha)}{\Gamma(4-2\varepsilon-\alpha-\beta)}\frac{\Gamma(\alpha+\beta+\varepsilon-2)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{A2}$$

$$I_\mu := \int \frac{d^n k\, k_\mu}{(k^2)^\alpha((k-q)^2)^\beta} = q_\mu \frac{2-\varepsilon-\alpha}{4-2\varepsilon-\alpha-\beta} I, \tag{A3}$$

$$I_{\mu\nu} := \int \frac{d^n k\, k_\mu k_\nu}{(k^2)^\alpha((k-q)^2)^\beta} = g_{\mu\nu}(q^2)^{3-\alpha-\beta}V_1(\alpha,\beta) + q_\mu q_\nu (q^2)^{2-\alpha-\beta}V_2(\alpha,\beta), \tag{A4}$$

$$I_{\mu\nu\rho} := \int \frac{d^n k\, k_\mu k_\nu k_\rho}{(k^2)^\alpha((k-q)^2)^\beta} = Z_{\mu\nu\rho}(q)(q^2)^{3-\alpha-\beta}W_1(\alpha,\beta) + q_\mu q_\nu q_\rho (q^2)^{2-\alpha-\beta}W_2(\alpha,\beta), \tag{A5}$$

$$I_{\mu\nu\rho\sigma} := \int \frac{d^n k\, k_\mu k_\nu k_\rho k_\sigma}{(k^2)^\alpha((k-q)^2)^\beta} = X_{\mu\nu\rho\sigma}(q^2)^{4-\alpha-\beta}W_5(\alpha,\beta) + Y_{\mu\nu\rho\sigma}(q)(q^2)^{3-\alpha-\beta} \times W_4(\alpha,\beta) + q_\mu q_\nu q_\rho q_\sigma (q^2)^{2-\alpha-\beta}W_3(\alpha,\beta), \tag{A6}$$

$$V_1(\alpha,\beta) = i\pi^{2-\varepsilon}(-q^2)^{-\varepsilon}\frac{\Gamma(3-\alpha-\varepsilon)\Gamma(3-\beta-\varepsilon)}{2\Gamma(6-\alpha-\beta-2\varepsilon)} \times \frac{\Gamma(\alpha+\beta+\varepsilon-3)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{A7}$$

$$V_2(\alpha,\beta) = i\pi^{2-\varepsilon}(-q^2)^{-\varepsilon}\frac{\Gamma(4-\alpha-\varepsilon)\Gamma(2-\beta-\varepsilon)}{\Gamma(6-\alpha-\beta-2\varepsilon)} \times \frac{\Gamma(\alpha+\beta+\varepsilon-2)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{A8}$$

$$W_1(\alpha,\beta) = [(3-\alpha-\varepsilon)/(6-\alpha-\beta-2\varepsilon)]V_1(\alpha,\beta), \tag{A9}$$

$$W_2(\alpha,\beta) = [(4-\alpha-\epsilon)/(6-\alpha-\beta-2\varepsilon)]V_2(\alpha,\beta), \tag{A10}$$

$$W_{34}(\alpha,\beta) = [(5-\alpha-\varepsilon)/(7-\alpha-\beta-2\varepsilon)]W_2(\alpha,\beta), \tag{A11}$$

$$W_4(\alpha,\beta) = [(4-\alpha-\varepsilon)/(7-\alpha-\beta-2\varepsilon)]W_1(\alpha,\beta), \tag{A12}$$

$$W_5(\alpha,\beta) = i\pi^{2-\varepsilon}(-q^2)^{-\varepsilon}\frac{\Gamma(4-\alpha-\varepsilon)\Gamma(4-\beta-\varepsilon)}{4\Gamma(8-\alpha-\beta-2\varepsilon)} \times \frac{\Gamma(\alpha+\beta+\varepsilon-4)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{A13}$$

$$Z_{\mu\nu\rho}(q) = q_\mu g_{\nu\rho} + q_\nu g_{\mu\rho} + q_\rho g_{\mu\nu}, \tag{A14}$$

$$Y_{\mu\nu\rho\sigma}(q) = q_\mu q_\nu g_{\rho\sigma} + \text{permutations}, \tag{A15}$$

$$X_{\mu\nu\rho\sigma} = g_{\mu\nu}g_{\rho\sigma} + \text{permutations}. \tag{A16}$$

APPENDIX B: PHASE-SPACE FORMULAS

The phase space for j massless final state particles in n dimensions is

$$\mathrm{PS}^{(j)} = (2\pi)^n \int \prod_{i=1}^{j} \frac{d^n p_i}{(2\pi)^{n-1}} \delta^+(p_i^2)\delta^n\left(q - \sum_{i=1}^{j} p_i\right). \tag{B1}$$

For $j=3$ and q^2-channel processes it can be fully expressed by the invariants y_{13}, y_{23}:

$$\mathrm{PS}^{(3)} = \frac{q^2(4\pi/q^2)^{2\varepsilon}}{128\pi^3\Gamma(2-2\varepsilon)} \int_0^1 dy_{13}\, y_{13}^{-\varepsilon} \times \int_0^{1-y_{23}} dy_{23}\, y_{23}^{-\varepsilon}(1-y_{13}-y_{23})^{-\varepsilon}. \tag{B2}$$

For $j=4$, two angle variables θ,θ' are needed. They are defined as follows. One chooses a system, where $\vec{p}_1 + \vec{p}_3 = 0$ and where $\vec{p}_2 \| \vec{e}_z$ (Ref. 14),

$$p_1 = \sqrt{y_{13}}/2(1,\ldots,\sin\theta\cos\theta',\cos\theta)\sqrt{q^2}, \tag{B3}$$

$$p_2 = (y_{123}-y_{13})/2\sqrt{y_{13}}(1,\ldots,0,1)\sqrt{q^2}, \tag{B4}$$

$$p_3 = \sqrt{y_{13}}/2(1,\ldots,-\sin\theta\cos\theta',-\cos\theta)\sqrt{q^2}, \tag{B5}$$

$$p_4 = (y_{134}-y_{13})/2\sqrt{y_{13}}(1,\ldots,\sin\beta,\cos\beta)\sqrt{q^2}. \tag{B6}$$

Setting $v = \frac{1}{2}(1-\cos\theta)$ one gets

$$\mathrm{PS}^{(4)} = \frac{q^4(4\pi/q^2)^{3\varepsilon}}{2048\pi^5\Gamma(2-2\varepsilon)\Gamma(1-\varepsilon)} \int dy_{123}\, dy_{134}\, dy_{13}(y_{134}y_{123}-y_{13})^{-\varepsilon} \times (y_{13}+1-y_{123}-y_{134})^{-\varepsilon} y_{13}^{-\varepsilon}\theta(y_{13})\theta(y_{134}y_{123}-y_{13}) \times \theta(y_{13}+1-y_{123}-y_{134}) \int_0^1 dv\, v^{-\varepsilon}(1-v)^{-\varepsilon} \int_0^\pi \frac{d\theta'}{N_{\theta'}} \sin^{-2\varepsilon}\theta', \tag{B7}$$

where $N_{\theta'}$ is defined after Eq. (47).

For integrations over all of phase space a representation of $\mathrm{PS}^{(4)}$ is useful, where all integrations are between 0 and 1:

$$\mathrm{PS}^{(4)} = \frac{q^4(4\pi/q^2)^{3\varepsilon}}{2048\pi^5\Gamma(2-2\varepsilon)\Gamma(1-\varepsilon)} \int_0^1 dy_{134}\, y_{134}^{1-2\varepsilon}(1-y_{134})^{2-3\varepsilon} \times \int_0^1 ds\, s^{1-2\varepsilon}(1-s)^{-\varepsilon} \int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon}(1-zy_{134})^{-2+2\varepsilon} \int_0^1 dv\, v^{-\varepsilon}(1-v)^{-\varepsilon} \int_0^\pi \frac{d\theta'}{N_{\theta'}} \sin^{-2\varepsilon}\theta'. \tag{B8}$$

Here $z = y_{13}/y_{134}y_{123}$ and $s = y_{123}(1-zy_{134})/(1-y_{134})$. The invariants y_{ij} may be expressed with the help of variables appearing in (B8):

$$y_{12} = (1-y_{134})sv, \tag{B9}$$

$$y_{23} = (1-y_{134})s(1-v), \tag{B10}$$

$$y_{14} = y_{134}(1-zy_{123})[v(1-\gamma)+\gamma(1-v) - 2\cos\theta'(v(1-v)\gamma(1-\gamma))^{1/2}], \tag{B11}$$

$$y_{34} = y_{134}(1-zy_{123})[(1-v)(1-\gamma)+v\gamma + 2\cos\theta'(v(1-v)\gamma(1-\gamma))^{1/2}], \tag{B12}$$

$$y_{24} = y_{134}s, \tag{B13}$$

where $\gamma := zy_{24}/(1-zy_{123})(1-zy_{134})$.

In the main text we are concentrating on the region $y_{134} < y$. There the invariants may be approximated by

$$y_{12} = y_{123}v, \tag{B14}$$

$$y_{23} = y_{123}(1-v), \tag{B15}$$

$$y_{24} = 1-y_{123}, \tag{B16}$$

$$y_{14} = y_{134}[v(1-z)+zy_{24}(1-v) - 2\cos\theta'(v(1-v)z(1-z)y_{24})^{1/2}], \tag{B17}$$

$$y_{34} = y_{134}[(1-v)(1-z)+vzy_{24} + 2\cos\theta'(v(1-v)z(1-z)y_{24})^{1/2}]. \tag{B18}$$

The phase space in this limit is

$$\mathrm{PS}^{(4)}_{y_{134}<y} = \int_0^y dy_{134}\, y_{134}^{1-2\varepsilon} \times \int_0^1 dy_{24}\, y_{24}^{-\varepsilon}(1-y_{24})^{1-2\varepsilon} \times \int_0^1 dz\, z^{-\varepsilon}(1-z)^{-\varepsilon} \int_0^1 dv\, v^{-\varepsilon}(1-v)^{-\varepsilon} \times \int_0^\pi \frac{d\theta'}{N_{\theta'}} \sin^{-2\varepsilon}\theta'. \tag{B19}$$

If one exchanges the role of particles 2 and 4 in (B7) and evaluates the limit $y_{134}\to 0$ one gets back (B19). However, v now has a different meaning and the structure of the invariants differs from (B14)–(B18) (apart from $2\leftrightarrow 4$ interchange):

$$y_{14} = y_{134}(1-zy_{123})v, \tag{B20}$$

$$y_{34} = y_{134}(1-zy_{123})(1-v), \tag{B21}$$

$$y_{24} = 1 - y_{123}, \tag{B22}$$

$$y_{12} = [y_{123}/(1 - zy_{123})][v(1 - z) + z(1 - v)y_{24} - 2\cos\theta'(v(1 - v)z(1 - z)y_{24})^{1/2}], \tag{B23}$$

$$y_{23} = [y_{123}/(1 - zy_{123})][(1 - v)(1 - z) + zvy_{24} + 2\cos\theta'(v(1 - z)z(1 - v)y_{24})^{1/2}]. \tag{B24}$$

APPENDIX C: MISCELLANEOUS SERIES

In this appendix we have selected some series that have proved useful in two-loop calculations and that are not standard. Most of them are taken from Ref. 9. A few appear in Ref. 23 in the form of integrals:

$$\sum_{k=0}^{\infty} \frac{\Gamma(k+c)}{k!q(k)} = \Gamma(c)\Gamma(1-c) = \sum_{i=1}^{n} \frac{\Gamma(-a_i)}{\Gamma(1-c-a_i)q'(a_i)}, \quad c<1, \tag{C1}$$

where $q(k)$ is a polynomial of degree n in k and q' its derivative. The roots a_i of $q(k)$ must be simpler,

$$\sum_{k=1}^{\infty} \frac{\Gamma(k+b)}{\Gamma(k+a)k} = \frac{\Gamma(b)}{\Gamma(a)}(\chi(a) - \chi(a-b)), \quad a>b, \tag{C2}$$

where ψ is the logarithmic derivative of the Γ function

$$\sum_{k=0}^{\infty} \frac{\Gamma(k+c)}{k!(ka+b)^2} = \frac{\Gamma(c)\Gamma(1-c)\Gamma(b/a)}{a^2\Gamma(b/a-c+1)}\left(\psi\left(\frac{b}{a}-c+1\right) - \psi\left(\frac{b}{a}\right)\right), \tag{C3}$$

$$\sum_{k=0}^{\infty} \psi^{(n)}(k+1)x^k = \frac{(-1)^{n+1}}{n!}\frac{L_{n+1}(1) - L_{n+1}(x)}{1-x}, \quad -1\leqslant x\leqslant 1, \quad n = 1,2,\ldots, \tag{C4}$$

and the L_n are Eulers n logarithms,

$$\sum_{k=1}^{\infty} \frac{x^k(\psi(k) - \psi(1))}{k} = \frac{\ln^2(1-x)}{2}, \quad -1\leqslant x<1, \tag{C5}$$

$$\sum_{k=1}^{\infty} \frac{\psi'(k)}{k} = 2\zeta_3, \tag{C6}$$

$$\sum_{k=1}^{\infty} \frac{\psi(k) - \psi(1)}{k^2} = \zeta_3, \tag{C7}$$

$$\sum_{k=1}^{\infty} \frac{(\psi(k+1) - \psi(1))^2}{k^2} = \frac{17\zeta_4}{4}, \tag{C8}$$

$$\sum_{k=1}^{\infty} \frac{\psi(k) - \psi(1)}{k^3} = \frac{\zeta_4}{4}, \tag{C9}$$

$$\sum_{j=1}^{\infty} \frac{(-1)^j\psi'(j)}{j^2} = -\frac{85}{16}\zeta_4 + 4L_4\left(\frac{1}{2}\right) + \frac{7}{2}\zeta_3\ln 2 \tag{C10}$$

$$-\zeta_2\ln^2 2 + \frac{1}{6}\ln^4 2, \tag{C11}$$

$$\sum_{j=1}^{\infty} \frac{(-1)^j}{j^3}(\psi(j) - \psi(1)) = -\frac{15}{8}\zeta_4 + 2L_4\left(\frac{1}{2}\right) + \frac{7}{4}\zeta_3\ln 2 - \frac{\zeta_2\ln^2 2}{2} + \frac{1}{12}\ln^4 2. \tag{C12}$$

Sums with a denominator $(k+a)(k+b)\cdots$ can be calculated by partial fractioning, e.g., $\Sigma_1\psi_k'/k(k+1) = 1$. Further sums can be found in Ref. 22.

[1] B. Lampe and G. Kramer, DESY 85-030, 1985; G. Kramer and B. Lampe, DESY 86-103, 1986; for earlier work see also Refs. 7, 13, and 14.
[2] P. Cvitanović and T. Kinoshita, Phys. Ref. D **10**, 3978, 3991, 4007 (1974); M. A. Nowak, M. Praszalowicz, and W. Slominski, München preprint MPI-PAE/PTh 14/84, 1984; G. 't Hooft and M. Veltman, Nucl. Phys. B **153**, 365 (1979).
[3] R. K. Ellis and J. C. Sexton, Nucl. Phys. B **269**, 445 (1986).
[4] W. Celmaster and R. J. Gonsalves, Phys. Rev. D **21**, 3112 (1980); K. G. Chetyrkin, A. L. Kataev, and F. V. Tachov, Phys. Lett. B **85**, 277 (1979).
[5] W. J. Marciano, Phys. Rev. D **12**, 3861 (1975); G. 't Hooft and M. Veltman, Nucl. Phys. B **44**, 189 (1972).
[6] A. E. Terrano, Phys. Lett. B **93**, 424 (1980).
[7] B. Lampe and G. Kramer, Phys. Scr. **28**, 585 (1983).
[8] L. Durand, P. M. Fishbane, and L. M. Simmons, J. Math. Phys. **17**, 1973 (1976).
[9] A. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[10] R. J. Gonsalves, Phys. Rev. D **28**, 1542 (1983); see also F. A. Berends, D. Danckaert, P. de Gansmaecker, R. Gastmans, R. Kleiss, W. Troost, and T. T. Wu, in *Electroweak Effects at High Energies*, edited by H. B. Newman (Plenum, New York, 1985); W. L. Van Neerven, Nucl. Phys. B **268**, 453 (1986).
[11] K. G. Chetyrkin, Nucl. Phys. B **192**, 159 (1981).
[12] G. Kramer, *Springer Tracts in Modern Physics*, Vol. 102 (Springer, Berlin, 1984), p. 56.
[13] K. Fabricius, G. Kramer, G. Schierholz, and I. Schmitt, Z. Phys. C **11**, 315 (1982).
[14] R. K. Ellis, D. A. Ross, and A. E. Terrano, Nucl. Phys. B **178**, 421 (1981).
[15] B. Lampe and G. Kramer, Commun. Math. Phys. **97**, 257 (1985); J. G. Körner, G. Kramer, B. Lampe, and G. Schuler, Z. Phys. C **32**, 181 (1986).
[16] A. Ali, J. G. Körner, Z. Kunszt, E. Pietarinen, G. Kramer, G. Schierholz, and J. Willrodt, Nucl. Phys. B **167**, 454 (1980).
[17] T. D. Gottschalk and M. P. Shatz, CALT-68-1172, 1984.
[18] A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.
[19] J. S. Gradstheyen and J. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965); W. Gröbner and N. Hofreiter, *Integraltafeln* (New York, 1973), Vol. 2.
[20] T. D. Gottschalk and M. P. Shatz, CALT-68-1173, 1968.
[21] G. Kramer and B. Lampe, DESY 86-119, 1986.
[22] R. Gastmans and W. Troost, Simon Stevin, AQJPAM **55**, 205 (1981).
[23] A. Devoto and D. W. Duke, Tallahassee preprint FSU-HEP-831003, 1983.

Generalized coupling coefficients for O(6)⊃O(5) and O(5)⊃O(3) in Bose–Fermi symmetries for odd–odd nuclei

P. Van Isacker
School of Mathematical and Physical Sciences, University of Sussex, Brighton BN1 9QH, England

(Received 30 July 1986; accepted for publication 10 December 1986)

Generalized coupling coefficients associated with the reductions O(6)⊃O(5) and O(5)⊃O(3) are derived in cases of interest for the $U_\nu^B(6)\times U_\pi^B(6)\times U_\nu^F(12)\times U_\pi^F(4)$ scheme for odd–odd nuclei.

I. INTRODUCTION

With the success of the interacting boson model (IBM) of Arima and Iachello,[1] interest has grown in the application of group-theoretical methods in nuclear physics. Originally this approach was confined to the study of even–even nuclei and, based on group-theoretical techniques, an overview was given[2–5] of the predictions of the IBM with regard to such nuclei. At a later stage these ideas were extended to odd-A nuclei,[6] which are approximated as a system of bosons and one fermion. A group-theoretical treatment of such systems leads in a natural way to Bose–Fermi symmetries for odd-A nuclei,[7] which have been studied extensively in situations appropriate in various regions of the mass table.[8–13] Parallel to this development for odd-A nuclei, even–even nuclei were studied further in detail and, specifically, neutrons and protons were considered separately, unlike in the original version of the IBM, where no distinction is made between them. Again this neutron–proton version of the model—usually referred to as the IBM-2—has been studied in a group-theoretical context, and some of its symmetry properties are discussed in Ref. 14. In a recent development[15–17] a description of odd–odd nuclei was proposed in the framework of the IBM. In essence it represents a merger of the interacting boson–fermion model (IBFM) for odd-A nuclei and the IBM-2, in that the bosons and fermions as well as the neutrons and protons are treated explicitly. The relevance of such schemes for odd–odd nuclei is at present under investigation.

Although all models mentioned above have an identical conceptual structure, they have become increasingly more complex. This is especially true for the Bose–Fermi symmetries for odd–odd nuclei. Nevertheless, to assess the relevance of any of such schemes it is necessary to carry out explicit calculations to obtain its predictions. The most straightforward way to accomplish this is by finding the realization of the states in terms of bosons and fermions for the neutrons and protons. This in turn reduces to the calculation of generalized coupling coefficients for the various groups appearing in the symmetry under consideration.

In this paper the $U_\nu^B(6)\times U_\nu^F(12)\times U_\pi^B(6)\times U_\pi^F(4)$ dynamical symmetry of Ref. 16 will be analyzed. Two plausible group chains will be studied, one of which was applied in Ref. 16 to the excitation-energy properties of ^{198}Au and another one which conceivably might be of use for the interpretation of neutron transfer data in the Pt–Au mass region. The properties of both group chains will be shown to depend on O(6)⊃O(5) and O(5)⊃O(3) generalized coupling coefficients. A general algorithm to calculate these coefficients will be given and some explicit expressions, of interest in practical applications, will be derived.

Notation: All groups with a boson (fermion) realization are characterized by a superscript B (F); all groups with a neutron (proton) realization are characterized by a subscript ν (π). Groups with a mixed realization (i.e., formed by adding different types of generators) are denoted by a sum in the superscript and/or subscript. Irreducible representations (irreps) of unitary groups [mostly U(6)] are labeled as $[N_1, N_2, N_3, \ldots]$, irreps of O(6) as $\langle\sigma_1, \sigma_2, \sigma_3\rangle$, and irreps of O(5) as (τ_1, τ_2). If some label is suppressed, it is zero. For instance $\langle\sigma\rangle \equiv \langle\sigma, 0, 0\rangle$, $(\tau) \equiv (\tau, 0)$, etc.

II. BOSE–FERMI SYMMETRIES FOR ODD–ODD NUCLEI

The Bose–Fermi symmetry considered in Ref. 16 is appropriate for situations in which the neutrons occupy orbits with angular momenta $j = \frac{1}{2}, \frac{3}{2}$, and $\frac{5}{2}$, and the protons have $j = \frac{3}{2}$. The associated group chain is of the form

$$
\begin{aligned}
&U_\nu^B(6)\times U_\pi^B(6)\times U_\nu^F(12)\times U_\pi^F(4)\supset U^{B}_{\nu+\pi}(6)\times U_\nu^F(6)\times SU_\nu^F(2)\times SU_\pi^F(4)\supset U^{B+F}_{\nu+\pi}(6)\times SU_\nu^F(2)\\
&\quad\times O_\pi^F(6)\supset \overline{O}^{B+F}_{\nu+\pi}(6)\times SU_\nu^F(2)\times O_\pi^F(6)\supset O^{B+F}_{\nu+\pi}(6)\times SU_\nu^F(2)\supset O^{B+F}_{\nu+\pi}(5)\times SU_\nu^F(2)\supset O^{B+F}_{\nu+\pi}(3)\\
&\quad\times SU_\nu^F(2)\supset O(3)\supset O(2),
\end{aligned}
\qquad (2.1)
$$

$$
\begin{array}{llllllllllllll}
[N_\nu] & [N_\pi] & [1] & [1] & [\tilde{N}_1,\tilde{N}_2] & [1] & \tfrac{1}{2} & [1] & [N_1,N_2,N_3] & \tfrac{1}{2} & \langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle & \langle\bar{\sigma}_1,\bar{\sigma}_2,\bar{\sigma}_3\rangle & \tfrac{1}{2} & \langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle\\
\langle\sigma_1,\sigma_2,\sigma_3\rangle & \tfrac{1}{2} & (\tau_1,\tau_2) & \tfrac{1}{2} & J & & L & M & & & & & &
\end{array}
$$

where below each group the associated quantum numbers are indicated. The group $U_\nu^B(6)\times U_\pi^B(6)\times U_\nu^F(12)\times U_\pi^F(4)$ serves as a dynamical group[18] in the sense that one single representation of it contains the states of a specific odd–odd nucleus. Hence,

 0022-2488/87/040957-07$02.50

N_ν and N_π are fixed and are equal to the number of neutron and proton bosons. All other quantum numbers are obtained by standard reduction and multiplication rules. Note that in general the $O^{B+F}_{\nu+\pi}(5) \supset O^{B+F}_{\nu+\pi}(3)$ reduction from (τ_1,τ_2) to J is not multiplicity-free and that an additional index is needed besides J.[7,19] However, in the cases of practical interest considered below, this multiplicity problem does not arise and the additional index will be omitted in the $O^{B+F}_{\nu+\pi}(5) \supset O^{B+F}_{\nu+\pi}(3)$ reduction as well as in other, similar cases. The quadratic Casimir operator of $U^B_{\nu+\pi}(6)$ is proportional to a symmetry term between the neutron and proton bosons.[14] It brings the symmetric irrep $[N_\nu + N_\pi] \equiv [N]$ lower in energy by about 2 to 3 MeV compared with the next irrep $[N-1,1]$ of $U^B_{\nu+\pi}(6)$. For applications in odd–odd nuclei one may thus assume $[\tilde{N}_1,\tilde{N}_2] = [N]$. A wave function associated with the group chain (2.1) will be denoted as

$$|[N_1,N_2]\langle\bar{\sigma}_1,\bar{\sigma}_2\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle, \tag{2.2}$$

where the quantum numbers that are the same for all states of a given odd–odd nucleus (such as N_ν, N_π, etc.) are omitted. This wave function is written explicitly as follows:

$$\begin{aligned}
&|[N_1,N_2]\langle\bar{\sigma},\bar{\sigma}_2\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle \\
&= \sum_j \sum_{\{-\}} \sum_{\{\nu\}} \sum_{\{\pi\}} \sum_{\{B\}} \sum_{\{F\}} (-1)^{L_B+J+1/2}((2j+1)(2J+1)(2\bar{L}+1)(2\bar{J}+1))^{1/2} \begin{Bmatrix} L_B & L_F & \bar{L} \\ \frac{1}{2} & \bar{J} & j \end{Bmatrix} \begin{Bmatrix} \bar{L} & \frac{3}{2} & J \\ L & \frac{1}{2} & \bar{J} \end{Bmatrix} \\
&\times \left\langle \begin{matrix} [N_\nu] & [N_\pi] \\ \langle\sigma_\nu\rangle & \langle\sigma_\pi\rangle \end{matrix} \middle| \begin{matrix} [N] \\ \langle\sigma_B\rangle \end{matrix} \right\rangle \left\langle \begin{matrix} \langle\sigma_\nu\rangle\langle\sigma_\pi\rangle \\ (\tau_\nu)\ (\tau_\pi) \end{matrix} \middle| \begin{matrix} \langle\sigma_B\rangle \\ (\tau_B) \end{matrix} \right\rangle \left\langle \begin{matrix} (\tau_\nu)\ (\tau_\pi) \\ \alpha_\nu L_\nu \alpha_\pi L_\pi \end{matrix} \middle| \begin{matrix} (\tau_B) \\ \alpha_B L_B \end{matrix} \right\rangle \left\langle \begin{matrix} [N][1] \\ \langle\sigma_B\rangle\langle 1\rangle \end{matrix} \middle| \begin{matrix} [N_1,N_2] \\ \langle\bar{\sigma}_1,\bar{\sigma}_2\rangle \end{matrix} \right\rangle \left\langle \begin{matrix} \langle\sigma_B\rangle\langle 1\rangle \\ (\tau_B)(\tau_F) \end{matrix} \middle| \langle\bar{\sigma}_1,\bar{\sigma}_2\rangle \right\rangle \\
&\times \left\langle \begin{matrix} (\tau_B)\ (\tau_F) \\ \alpha_B L_B\ L_F \end{matrix} \middle| \begin{matrix} (\bar{\tau}_1,\bar{\tau}_2) \\ \bar{L} \end{matrix} \right\rangle \left\langle \begin{matrix} \langle\bar{\sigma}_1,\bar{\sigma}_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle \\ (\bar{\tau}_1,\bar{\tau}_2)\ (\frac{1}{2},\frac{1}{2}) \end{matrix} \middle| \begin{matrix} \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2) \end{matrix} \right\rangle \left\langle \begin{matrix} (\bar{\tau}_1,\bar{\tau}_2)(\frac{1}{2},\frac{1}{2}) \\ \bar{L}\qquad \frac{3}{2} \end{matrix} \middle| \begin{matrix} (\tau_1,\tau_2) \\ J \end{matrix} \right\rangle \\
&\times ((B^\dagger_\nu([N_\nu]\langle\sigma_\nu\rangle(\tau_\nu)\alpha_\nu L_\nu) \times B^\dagger_\pi([N_\pi]\langle\sigma_\pi\rangle(\tau_\pi)\alpha_\pi L_\pi))^{(L_B)} \times a^\dagger_\nu(j))^{(\bar{J})} \times a^\dagger_\pi(\tfrac{3}{2}))^{(L)}_M|0\rangle,
\end{aligned} \tag{2.3}$$

where the summation runs over $j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$,

$$\{-\} \equiv \{\bar{\tau}_1,\bar{\tau}_2,\bar{L},\bar{J}\}, \quad \{\nu\} \equiv \{\sigma_\nu,\tau_\nu,\alpha_\nu,L_\nu\}, \quad \{\pi\} \equiv \{\sigma_\pi,\tau_\pi,\alpha_\pi,L_\pi\}, \quad \{B\} \equiv \{\sigma_B,\tau_B,\alpha_B,L_B\}, \quad \{F\} \equiv \{\tau_F,L_F\}.$$

The operator $B^\dagger_\nu$ ($B^\dagger_\pi$) creates a many-neutron (proton) boson state with $O(6) \supset O(5) \supset O(3)$ symmetry; the operator $a^\dagger_\nu$ ($a^\dagger_\pi$) creates a neutron (proton) fermion. The coefficients $\langle :: | : \rangle$ are generalized coupling coefficients (GCC) or isoscalar factors.[18] The first three GCC's in Eq. (2.3) are associated with the reduction $U(6) \supset O(6)$, $O(6) \supset O(5)$, and $O(5) \supset O(3)$, respectively, and are defined in Ref. 14. The next three GCC's are associated with the same reductions, and general expressions for them are given in Ref. 11. The last two GCC's in Eq. (2.3) are $O(6) \supset O(5)$ and $O(5) \supset O(3)$ coupling coefficients and are unknown. The problem of finding an explicit realization of the wave function (2.2) is thus reduced to the calculation of these two GCC's. The expression (2.3) can be simplified considerably as follows:

$$\begin{aligned}
&|[N_1,N_2]\langle\bar{\sigma},\bar{\sigma}_2\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle \\
&= \sum_{\{-\}} (-1)^{J-\bar{J}}((2J+1)(2\bar{J}+1))^{1/2} \begin{Bmatrix} \bar{L} & \frac{3}{2} & J \\ L & \frac{1}{2} & \bar{J} \end{Bmatrix} \left\langle \begin{matrix} \langle\bar{\sigma}_1,\bar{\sigma}_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle \\ (\bar{\tau}_1,\bar{\tau}_2)\ (\frac{1}{2},\frac{1}{2}) \end{matrix} \middle| \begin{matrix} \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2) \end{matrix} \right\rangle \left\langle \begin{matrix} (\bar{\tau}_1,\bar{\tau}_2)(\frac{1}{2},\frac{1}{2}) \\ \bar{L}\qquad \frac{3}{2} \end{matrix} \middle| \begin{matrix} (\tau_1,\tau_2) \\ J \end{matrix} \right\rangle \\
&\times (A^\dagger_\nu([N,N_2]\langle\bar{\sigma}_1,\bar{\sigma}_2\rangle(\bar{\tau}_1,\bar{\tau}_2)\bar{L}\bar{J}) \times a^\dagger_\pi(\tfrac{3}{2}))^{(L)}_M|0\rangle.
\end{aligned} \tag{2.4}$$

The operator $A^\dagger_\nu$ creates a state of an odd-neutron nucleus with quantum numbers defined according to the O(6) limit of $U^B(6) \times U^F(12)$.[11]

From Eqs. (2.3) and (2.4) it is clear that in the group chain (2.1), first the neutron is coupled to the bosons and then the proton is coupled to the bosons. This coupling order is appropriate for calculating *proton*-transfer strengths. For *neutron* transfer a different coupling order should be imposed which is associated with the group chain

$$\begin{aligned}
&U^B_\nu(6) \times U^B_\pi(6) \times U^F_\nu(12) \times U^F_\pi(4) \supset U^B_{\nu+\pi}(6) \times U^F_\nu(6) \times SU^F_\nu(2) \times SU^F_\pi(4) \supset O^B_{\nu+\pi}(6) \times O^F_\nu(6) \times SU^F_\nu(2) \\
&\quad [N_\nu] \quad [N_\pi] \quad [1] \quad [1] \quad [N] \quad [1] \quad \tfrac{1}{2} \quad [1] \quad \langle\sigma_B\rangle \quad \langle 1\rangle \quad \tfrac{1}{2} \\
&\times O^F_\pi(6) \supset \tilde{O}^{B+F}_{\nu+\pi}(6) \times O^F_\nu(6) \times SU^F_\nu(2) \supset O^{B+F}_{\nu+\pi}(6) \times SU^F_\nu(2) \supset O^{B+F}_{\nu+\pi}(5) \times SU^F_\nu(2) \supset O^{B+F}_{\nu+\pi}(3) \\
&\quad \langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle \quad \langle\tilde{\sigma}_1,\frac{1}{2},\frac{1}{2}\rangle \quad \langle 1\rangle \quad \tfrac{1}{2} \quad \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \quad \tfrac{1}{2} \quad (\tau_1,\tau_2) \quad \tfrac{1}{2} \quad J \\
&\times SU^F_\nu(2) \supset O(3) \supset O(2). \\
&\quad \tfrac{1}{2} \quad L \quad M
\end{aligned} \tag{2.5}$$

In the same way as for chain (2.1) an abbreviated notation can be introduced for the wave function, i.e.,

$$|\langle\sigma_B\rangle\langle\tilde{\sigma}_1,\tfrac{1}{2},\tfrac{1}{2}\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle, \tag{2.6}$$

and this wave function is written explicitly as follows:

$$|\langle\sigma_B\rangle\langle\tilde\sigma_1,\tfrac12,\tfrac12\rangle\langle\sigma_1,\sigma_2,\tfrac12\rangle(\tau_1,\tau_2)JLM\rangle$$

$$=\sum_j\sum_{\{-\}}\sum_{\{\nu\}}\sum_{\{\pi\}}\sum_{\{B\}}\sum_{\{F\}}(-1)^{\tilde J+L+3/2}((2j+1)(2J+1))^{1/2}\begin{Bmatrix}\tilde J & L_F & J\\ \frac12 & L & j\end{Bmatrix}\left\langle\begin{matrix}\langle\sigma_\nu\rangle\langle\sigma_\pi\rangle\\(\tau_\nu)\,(\tau_\pi)\end{matrix}\middle|\begin{matrix}\langle\sigma_B\rangle\\(\tau_B)\end{matrix}\right\rangle\left\langle\begin{matrix}(\tau_\nu)\ (\tau_\pi)\\ \alpha_\nu L_\nu\alpha_\pi L_\pi\end{matrix}\middle|\begin{matrix}(\tau_B)\\ \alpha_B L_B\end{matrix}\right\rangle$$
$$\times\left\langle\begin{matrix}\langle\sigma_B\rangle\langle\frac12,\frac12,\frac12\rangle\\(\tau_B)\ (\frac12,\frac12)\end{matrix}\middle|\begin{matrix}\langle\tilde\sigma_1,\frac12,\frac12\rangle\\(\tilde\tau_1,\frac12)\end{matrix}\right\rangle\left\langle\begin{matrix}(\tau_B)\ (\frac12,\frac12)\\ \alpha_B L_B\ \ \frac32\end{matrix}\middle|\begin{matrix}(\tilde\tau_1,\frac12)\\ \tilde J\end{matrix}\right\rangle\left\langle\begin{matrix}\langle\tilde\sigma_1,\frac12,\frac12\rangle\langle1\rangle\\(\tilde\tau,\frac12)(\tau_F)\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac12\rangle\\(\tau_1,\tau_2)\end{matrix}\right\rangle\left\langle\begin{matrix}(\tilde\tau_1,\frac12)(\tau_F)\\ \tilde J\quad L_F\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$$
$$\times(((B^\dagger_\nu([N_\nu]\langle\sigma_\nu\rangle(\tau_\nu)\alpha_\nu L_\nu)\times B^\dagger_\pi([N_\pi]\langle\sigma_\pi\rangle(\tau_\pi)\alpha_\pi L_\pi))^{(L_B)}\times a^\dagger_\pi(\tfrac32))^{(\tilde J)}\times a^\dagger_\nu(j))^{(L)}_M|0\rangle. \tag{2.7}$$

Again, all GCC's are known (see Refs. 9 and 11) except for the last two, which are associated with the reductions $O(6)\supset O(5)$ and $O(5)\supset O(3)$, respectively. As above, expression (2.7) can be simplified as follows:

$$|\langle\sigma_B\rangle\langle\tilde\sigma_1,\tfrac12,\tfrac12\rangle\langle\sigma_1,\sigma_2,\tfrac12\rangle(\tau_1,\tau_2)JLM\rangle$$
$$=\sum_j\sum_{\{-\}}\sum_{\{F\}}(-1)^{\tilde J+L+3/2}\times((2j+1)(2J+1))^{1/2}\begin{Bmatrix}\tilde J & L_F & J\\ \frac12 & L & j\end{Bmatrix}$$
$$\times\left\langle\begin{matrix}\langle\tilde\sigma_1,\frac12,\frac12\rangle\langle1\rangle\\(\tilde\tau,\frac12)(\tau_F)\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac12\rangle\\(\tau_1,\tau_2)\end{matrix}\right\rangle\left\langle\begin{matrix}(\tilde\tau_1,\frac12)(\tau_F)\\ \tilde J\quad L_F\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$$
$$\times(A^\dagger_\pi(\langle\sigma_B\rangle\langle\tilde\sigma_1,\tfrac12,\tfrac12\rangle(\tilde\tau_1,\tfrac12)\tilde J)\times a^\dagger_\nu(j))^{(L)}_M. \tag{2.8}$$

The operator $A^\dagger_\pi$ creates a state of an odd-proton nucleus with quantum numbers defined according to the Spin(6) limit of $U^B(6)\times U^F(4)$.[9]

In Secs. III–VI expressions for the GCC's appearing in Eqs. (2.4) and (2.8) will be derived.

III. THE O(5)⊃O(3) GCC $\left\langle\begin{matrix}(\bar\tau_1,\bar\tau_2)(\frac12,\frac12)\\ L\quad\frac32\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$

These $O(5)\supset O(3)$ GCC's are defined as a transformation from the coupled basis $O_a(5)\times O_b(5)\supset O_{a+b}(5)\supset O_{a+b}(3)$ to the uncoupled basis $O_a(5)\times O_b(5)\supset O_a(3)\times O_b(3)\supset O_{a+b}(3)$:

$$|(\bar\tau,\bar\tau_2),(\tfrac12,\tfrac12);(\tau_1,\tau_2)JM\rangle$$
$$=\sum_{\bar L}\left\langle\begin{matrix}(\bar\tau_1,\bar\tau_2)(\frac12,\frac12)\\ L\quad\frac32\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle|(\bar\tau_1,\bar\tau_2)L,(\tfrac12,\tfrac12)\tfrac32;JM\rangle. \tag{3.1}$$

The $O(5)\supset O(3)$ GCC's are thus obtained by diagonalizing $C_2(O_{a+b}(5))$ [the quadratic Casimir operator of $O_{a+b}(5)$] in the uncoupled basis. Denoting the generators of $O_a(5)$ and $O_b(5)$ by $T^{(k)}_a$ and $T^{(k)}_b$ $(k=1,3)$, the operator $C_2(O_{a+b}(5))$ can be written as

$$C_2(O_{a+b}(5))=2\sum_{k=1,3}(T^{(k)}_a+T^{(k)}_b)\cdot(T^{(k)}_a+T^{(k)}_b)$$
$$=C_2(O_a(5))+C_2(O_b(5))$$
$$+4\sum_{k=1,3}T^{(k)}_a\cdot T^{(k)}_b, \tag{3.2}$$

and hence, the matrix elements of $C_2(O_{a+b}(5))$ are given by

$$\langle(\bar\tau_1,\bar\tau_2)L,(\tfrac12,\tfrac12)\tfrac32;J\,|C_2(O_{a+b}(5))|(\bar\tau_1,\bar\tau_2)L',(\tfrac12,\tfrac12)\tfrac32;J\rangle$$
$$=\left(\bar\tau_1(\bar\tau_1+3)+\bar\tau_2(\bar\tau_2+1)+\frac52\right)\delta_{LL'}$$
$$+4\sum_{k=1,3}(-1)^{3/2+J+L'}\begin{Bmatrix}L & \frac32 & J\\ \frac32 & L' & k\end{Bmatrix}$$
$$\times\langle(\bar\tau_1,\bar\tau_2)L\,\|T^{(k)}_a\|(\bar\tau_1,\bar\tau_2)L'\rangle$$
$$\times\langle(\tfrac12,\tfrac12)\tfrac32\|T^{(k)}_b\|(\tfrac12,\tfrac12)\tfrac32\rangle. \tag{3.3}$$

The expressions for the matrix elements of $T^{(k)}_a$ and $T^{(k)}_b$ in Eq. (3.3) are still needed. Using the realization[9]

$$T^{(k)}_b=[(-1)^{(k+1)/2}/\sqrt2](a^\dagger(\tfrac32)\tilde a(\tfrac32))^{(k)}, \tag{3.4a}$$
$$|(\tfrac12,\tfrac12)\tfrac32\rangle=a^\dagger(\tfrac32)|0\rangle, \tag{3.4b}$$

one finds

$$\langle(\tfrac12,\tfrac12)\tfrac32\|T^{(k)}_b\|(\tfrac12,\tfrac12)\tfrac32\rangle=(-1)^{(k-1)/2}\sqrt{(2k+1)/2}. \tag{3.5}$$

For the matrix element of $T^{(k)}_a$ two cases occur in the summation in Eq. (2.4): (i) $\bar\tau_2=0$, and (ii) $\bar\tau_2=1$. The matrix elements $\langle(\bar\tau)L\,\|T^{(k)}_a\|(\bar\tau)L'\rangle$ can be computed from a realization in terms of d-bosons.[2] The matrix elements $\langle(\bar\tau,1)L\,\|T^{(k)}_a\|(\bar\tau,1)L'\rangle$ can be reduced to the first case by realization of $O(5)$ in terms of two kinds of d-bosons (say ν and π):

$$|(\bar\tau,1)LM\rangle=\sum_{\bar L}\left\langle\begin{matrix}(\bar\tau)(1)\\ \bar L\quad 2\end{matrix}\middle|\begin{matrix}(\bar\tau,1)\\ L\end{matrix}\right\rangle(B^\dagger_\nu((\bar\tau)\bar L)\times d^\dagger_\pi)^{(L)}_M|0\rangle. \tag{3.6}$$

The matrix elements then become

$$\langle(\bar\tau,1)L\,\|T^{(k)}_a\|(\bar\tau,1)L'\rangle$$
$$\equiv\langle(\bar\tau,1)L\,\|T^{(k)}_{a\nu}+T^{(k)}_{a\pi}\|(\bar\tau,1)L'\rangle$$
$$=\sum_{\bar L,\bar L'}\left\langle\begin{matrix}(\bar\tau)(1)\\ \bar L\quad 2\end{matrix}\middle|\begin{matrix}(\bar\tau,1)\\ L\end{matrix}\right\rangle\left\langle\begin{matrix}(\bar\tau)(1)\\ \bar L'\quad 2\end{matrix}\middle|\begin{matrix}(\bar\tau,1)\\ L'\end{matrix}\right\rangle$$
$$\times\Big[(-1)^{\bar L+L'+k}((2L+1)(2L'+1))^{1/2}$$
$$\times\langle(\bar\tau)\bar L\,\|T^{(k)}_{a\nu}\|(\bar\tau)\bar L'\rangle\begin{Bmatrix}\bar L & L & 2\\ L' & \bar L' & k\end{Bmatrix}$$
$$+(-1)^{\bar L+L+k}((2L+1)(2L'+1))^{1/2}$$
$$\times\langle d_\pi\|T^{(k)}_{a\pi}\|d_\pi\rangle\begin{Bmatrix}2 & L & \bar L\\ L' & 2 & k\end{Bmatrix}\delta_{\bar L\bar L'}\Big], \tag{3.7}$$

where the $O(5)\supset O(3)$ GCC's can be taken from Ref. 11. From Eqs. (3.5) and (3.7) the matrix element (3.3) is obtained, which in turn determines the $O(5)\supset O(3)$ GCC's. The coefficients that are needed in the calculation of properties of odd–odd nuclei are given in Tables I–III.

TABLE I. The $O(5)\supset O(3)$ GCC's $\left\langle \begin{matrix}(2)&(\frac{1}{2},\frac{1}{2})\\ L&\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$.

(τ_1,τ_2)	J	$L=2$	$L=4$
$(\frac{3}{2},\frac{1}{2})$	$\frac{1}{2}$	1	
	$\frac{5}{2}$	$\sqrt{\frac{25}{49}}$	$-\sqrt{\frac{24}{49}}$
	$\frac{7}{2}$	$\sqrt{\frac{4}{49}}$	$\sqrt{\frac{45}{49}}$
$(\frac{5}{2},\frac{1}{2})$	$\frac{3}{2}$	1	
	$\frac{5}{2}$	$\sqrt{\frac{24}{49}}$	$\sqrt{\frac{25}{49}}$
	$\frac{7}{2}$	$\sqrt{\frac{45}{49}}$	$-\sqrt{\frac{4}{49}}$
	$\frac{9}{2}$		1
	$\frac{11}{2}$		1

TABLE III. Some $O(5)\supset O(3)$ GCC's $\left\langle \begin{matrix}(2,1)&(\frac{1}{2},\frac{1}{2})\\ L&\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$.

(τ_1,τ_2)	J	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$
$(\frac{3}{2},\frac{1}{2})$	$\frac{1}{2}$	$-\sqrt{\frac{3}{5}}$	$\sqrt{\frac{3}{5}}$			
	$\frac{5}{2}$	$\sqrt{\frac{2}{21}}$	$-\sqrt{\frac{5}{147}}$	$\sqrt{\frac{8}{21}}$	$-\sqrt{\frac{24}{49}}$	
	$\frac{7}{2}$		$-\sqrt{\frac{27}{245}}$	$-\sqrt{\frac{4}{35}}$	$-\sqrt{\frac{36}{245}}$	$\sqrt{\frac{32}{35}}$
$(\frac{3}{2},\frac{3}{2})$	$\frac{3}{2}$	$\sqrt{\frac{8}{35}}$	$\sqrt{\frac{3}{7}}$	$\sqrt{\frac{12}{35}}$		
	$\frac{5}{2}$	$\sqrt{\frac{7}{15}}$	$\sqrt{\frac{4}{21}}$	$-\sqrt{\frac{49}{120}}$	$-\sqrt{\frac{15}{56}}$	
	$\frac{9}{2}$			$\sqrt{\frac{1}{56}}$	$\sqrt{\frac{99}{280}}$	$\sqrt{\frac{32}{35}}$

IV. THE O(6)⊃O(5) GCC's $\left\langle \begin{matrix}\langle\bar\sigma_1,\bar\sigma_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\\ (\bar\tau_1,\bar\tau_2)\quad(\frac{1}{2},\frac{1}{2})\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (\tau_1,\tau_2)\end{matrix}\right\rangle$

These $O(6)\supset O(5)$ GCC's appear in the transformation from the coupled basis $O_a(6)\times O_b(6)\supset O_{a+b}(6)\supset O_{a+b}(5)\supset O_{a+b}(3)$ to the uncoupled basis $O_a(6)\times O_b(6)\supset O_a(5)\times O_b(5)\supset O_a(3)\times O_b(3)\supset O_{a+b}(3)$:

$$
\begin{aligned}
&|\langle\bar\sigma_1,\bar\sigma_2\rangle,\langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle;\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JM\rangle\\
&\quad=\sum_{\bar\tau_1,\bar\tau_2}\sum_{L}\left\langle\begin{matrix}\langle\bar\sigma_1,\bar\sigma_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\\ (\bar\tau_1,\bar\tau_2)\quad(\frac{1}{2},\frac{1}{2})\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (\tau_1,\tau_2)\end{matrix}\right\rangle\\
&\qquad\times\left\langle\begin{matrix}(\bar\tau_1,\bar\tau_2)(\frac{1}{2},\frac{1}{2})\\ L\qquad\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle\\
&\qquad\times|\langle\bar\sigma_1,\bar\sigma_2\rangle(\bar\tau_1,\bar\tau_2)L,\langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tfrac{1}{2},\tfrac{1}{2})\tfrac{3}{2};JM\rangle. \qquad (4.1a)
\end{aligned}
$$

The inverse transformation reads

$$
\begin{aligned}
&|\langle\bar\sigma_1,\bar\sigma_2\rangle(\bar\tau_1,\bar\tau_2)L,\langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tfrac{1}{2},\tfrac{1}{2})\tfrac{3}{2};JM\rangle\\
&\quad=\sum_{\sigma_1,\sigma_2}\sum_{\tau_1,\tau_2}\left\langle\begin{matrix}\langle\bar\sigma_1,\bar\sigma_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\\ (\bar\tau_1,\bar\tau_2)\quad(\frac{1}{2},\frac{1}{2})\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (\tau_1,\tau_2)\end{matrix}\right\rangle\\
&\qquad\times\left\langle\begin{matrix}(\bar\tau_1,\bar\tau_2)(\frac{1}{2},\frac{1}{2})\\ L\qquad\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle\\
&\quad=|\langle\bar\sigma_1,\bar\sigma_2\rangle,\langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle;\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JM\rangle. \qquad (4.1b)
\end{aligned}
$$

If the matrix elements of the operator

$$
\begin{aligned}
&(T_a^{(2)}+T_b^{(2)})\cdot(T_a^{(2)}+T_b^{(2)})\\
&\quad=C_2(O_{a+b}(6))-C_2(O_{a+b}(5))\\
&\quad=C_2(O_a(6))-C_2(O_a(5))+C_2(O_b(6))\\
&\qquad-C_2(O_b(5))+2T_a^{(2)}\cdot T_b^{(2)} \qquad (4.2)
\end{aligned}
$$

between the states (4.1a) and (4.1b) are calculated, the following relation is derived:

$$
\begin{aligned}
&\left\langle\begin{matrix}\langle\bar\sigma_1,\bar\sigma_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\\ (\bar\tau_1,\bar\tau_2)\quad(\frac{1}{2},\frac{1}{2})\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (\tau_1,\tau_2)\end{matrix}\right\rangle\left\langle\begin{matrix}(\bar\tau_1,\bar\tau_2)(\frac{1}{2},\frac{1}{2})\\ L\qquad\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle\{(\sigma_1(\sigma_1+4)+\sigma_2(\sigma_2+2))\\
&\quad-(\bar\sigma_1(\bar\sigma_1+4)+\bar\sigma_2(\bar\sigma_2+2))+(\bar\tau_1(\bar\tau_1+3)+\bar\tau_2(\bar\tau_2+1))-(\tau_1(\tau_1+3)+\tau_2(\tau_2+1))-1\}\\
&\quad=\sum_{\bar\tau_1',\bar\tau_2'}\sum_{L'}\left\langle\begin{matrix}\langle\bar\sigma_1,\bar\sigma_2\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\\ (\bar\tau_1',\bar\tau_2')\quad(\frac{1}{2},\frac{1}{2})\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (\tau_1,\tau_2)\end{matrix}\right\rangle\left\langle\begin{matrix}(\bar\tau_1',\bar\tau_2')(\frac{1}{2},\frac{1}{2})\\ L'\qquad\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle\\
&\qquad\times2\sqrt{5}(-1)^{J+L+5/2}\begin{Bmatrix}L'&\frac{3}{2}&J\\ \frac{3}{2}&L&2\end{Bmatrix}\langle\langle\bar\sigma_1,\bar\sigma_2\rangle(\bar\tau_1',\bar\tau_2')L'\|T_a^{(2)}\|\langle\bar\sigma_1,\bar\sigma_2\rangle(\bar\tau_1,\bar\tau_2)L\rangle. \qquad (4.3)
\end{aligned}
$$

TABLE II. The $O(5)\supset O(3)$ GCC's $\left\langle \begin{matrix}(1,1)&(\frac{1}{2},\frac{1}{2})\\ L&\frac{3}{2}\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$.

(τ_1,τ_2)	J	$L=1$	$L=3$
$(\frac{1}{2},\frac{1}{2})$	$\frac{3}{2}$	$-\sqrt{\frac{3}{10}}$	$\sqrt{\frac{7}{10}}$
$(\frac{3}{2},\frac{1}{2})$	$\frac{1}{2}$	1	
	$\frac{5}{2}$	$\sqrt{\frac{7}{15}}$	$\sqrt{\frac{8}{15}}$
	$\frac{7}{2}$		1
$(\frac{3}{2},\frac{3}{2})$	$\frac{3}{2}$	$\sqrt{\frac{7}{10}}$	$\sqrt{\frac{3}{10}}$
	$\frac{5}{2}$	$\sqrt{\frac{8}{15}}$	$-\sqrt{\frac{7}{15}}$
	$\frac{9}{2}$		1

Since the $O(5)\supset O(3)$ GCC's are known from Sec. III, the $O(6)\supset O(5)$ GCC's can be obtained from Eq. (4.3) for different choices of $\langle\bar\sigma_1,\bar\sigma_2\rangle$, $\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle$, and (τ_1,τ_2). Consider as an example $\langle\bar\sigma_1,\bar\sigma_2\rangle=\langle\sigma,1\rangle$, $\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle=\langle\sigma+\frac{1}{2},\frac{3}{2},\frac{1}{2}\rangle$, and $(\tau_1,\tau_2)=(\frac{3}{2},\frac{1}{2})$. In that case $(\bar\tau_1,\bar\tau_2)$ and $(\bar\tau_1',\bar\tau_2')$ can take the values (1), (2), (1,1), and (2,1) and Eq. (4.3) reduces to the following set of equations:

$$
\begin{aligned}
X(1)\sigma=&-X(2)(5(\sigma-1)(\sigma+5)/8)^{1/2}\\
&+X(1,1)(3(\sigma+1)(\sigma+3)/8)^{1/2}, \qquad (4.4a)\\
X(2)(\sigma+6)=&-X(1)(5(\sigma-1)(\sigma+5)/8)^{1/2}\\
&-X(2,1)(3(\sigma+1)(\sigma+3)/8)^{1/2}, \qquad (4.4b)
\end{aligned}
$$

TABLE IV. The $O(6)\supset O(5)$ GCC's $\left\langle \begin{matrix} \langle\sigma+1\rangle & \langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle \\ (\bar{\tau}) & (\frac{1}{2},\frac{1}{2}) \end{matrix} \middle| \begin{matrix} \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2) \end{matrix} \right\rangle$.

$\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle$	(τ_1,τ_2)	$\bar{\tau}=\tau+\frac{1}{2}$	$\bar{\tau}=\tau-\frac{1}{2}$
$\langle\sigma+\frac{3}{2},\frac{1}{2},\frac{1}{2}\rangle$	$(\tau,\frac{1}{2})$	$\left(\frac{2\sigma-2\tau+3}{4(\sigma+3)}\right)^{1/2}$	$-\left(\frac{2\sigma+2\tau+9}{4(\sigma+3)}\right)^{1/2}$
$\langle\sigma+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle$	$(\tau,\frac{1}{2})$	$\left(\frac{2\sigma+2\tau+9}{4(\sigma+3)}\right)^{1/2}$	$\left(\frac{2\sigma-2\tau+3}{4(\sigma+3)}\right)^{1/2}$

$$X(1,1)(\sigma+2) = X(1)(3(\sigma+1)(\sigma+3)/8)^{1/2} - X(2,1)(5(\sigma-1)(\sigma+5)/8)^{1/2}, \tag{4.4c}$$

$$X(2,1)(\sigma+3) = -X(2)(3(\sigma+1)(\sigma+3)/8)^{1/2} - X(1,1)(5(\sigma-1)(\sigma+5)/8)^{1/2}, \tag{4.4d}$$

where

$$X(\bar{\tau}_1,\bar{\tau}_2) = \left\langle \begin{matrix} \langle\sigma,1\rangle & \langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle \\ (\bar{\tau}_1,\bar{\tau}_2) & (\frac{1}{2},\frac{1}{2}) \end{matrix} \middle| \begin{matrix} \langle\sigma+\frac{1}{2},\frac{3}{2},\frac{1}{2}\rangle \\ (\frac{3}{2},\frac{1}{2}) \end{matrix} \right\rangle. \tag{4.5}$$

Equations (4.4) can be solved for the unknown $X(\bar{\tau}_1,\bar{\tau}_2)$ up to an overall factor. Hence

$$\frac{X(2)}{X(1)} = -\left(\frac{2(\sigma-1)}{5(\sigma+5)}\right)^{1/2}, \tag{4.6a}$$

$$\frac{X(1,1)}{X(1)} = \left(\frac{2(\sigma+1)}{3(\sigma+5)}\right)^{1/2}, \tag{4.6b}$$

$$\frac{X(2,1)}{X(1)} = -\left(\frac{(\sigma-1)(\sigma+1)}{15(\sigma+3)(\sigma+5)}\right)^{1/2}. \tag{4.6c}$$

The $X(\bar{\tau}_1,\bar{\tau}_2)$ are completely determined after normalization. The $O(6)\supset O(5)$ GCC's of practical interest are summarized in Tables IV and V.

V. THE $O(5)\supset O(3)$ GCC's $\left\langle \begin{matrix} (\tilde{\tau},\frac{1}{2}) & (1) \\ J' & 2 \end{matrix} \middle| \begin{matrix} (\tau_1,\tau_2) \\ J \end{matrix} \right\rangle$

These $O(5)\supset O(3)$ GCC's are defined as a transformation from the coupled basis $O_a(5)\times O_b(5)\supset O_{a+b}(5)\supset O_{a+b}(3)$ to be uncoupled basis $O_a(5)\times O_b(5)\supset O_a(3)\times O_b(3)\supset O_{a+b}(3)$:

$$|(\tilde{\tau},\tfrac{1}{2}),(1);(\tau_1,\tau_2)JM\rangle = \sum_{J'} \left\langle \begin{matrix} (\tilde{\tau},\frac{1}{2}) & (1) \\ J' & 2 \end{matrix} \middle| \begin{matrix} (\tau_1,\tau_2) \\ J \end{matrix} \right\rangle |(\tilde{\tau},\tfrac{1}{2})J',(1)2;JM\rangle. \tag{5.1}$$

As in Sec. III, the transformatin matrix can be found by evaluating the matrix elements of $C_2(O_{a+b}(5))$:

$$\langle(\tilde{\tau},\tfrac{1}{2})J',(1)2;J|C_2(O_{a+b}(5))|(\tilde{\tau},\tfrac{1}{2})J'',(1)2;J\rangle = \left(\tilde{\tau}(\tilde{\tau}+3)+\frac{19}{4}\right)\delta_{J'J''} + 4\sum_{k=1,3}(-1)^{J+J''} \times \begin{Bmatrix} J' & 2 & J \\ 2 & J'' & k \end{Bmatrix} \langle(\tilde{\tau},\tfrac{1}{2})J'\|T_a^{(k)}\|(\tilde{\tau},\tfrac{1}{2})J''\rangle \times \langle(1)2\|T_b^{(k)}\|(1)2\rangle. \tag{5.2}$$

The matrix elements of $T_b^{(k)}$ are obtained from a d-boson realization of $O(5)$:

$$\langle(1)2\|T_b^{(k)}\|(1)2\rangle = \sqrt{2k+1}. \tag{5.3}$$

The matrix elements of $T_a^{(k)}$ in Eq. (5.2) are evaluated by using a boson + fermion realization of $O(5)$, i.e.,

$$|(\tilde{\tau},\tfrac{1}{2})J'\rangle = \sum_{\tau,L} \left\langle \begin{matrix} (\tau) & (\frac{1}{2},\frac{1}{2}) \\ L & \frac{3}{2} \end{matrix} \middle| \begin{matrix} (\tilde{\tau},\frac{1}{2}) \\ J' \end{matrix} \right\rangle \times (B^\dagger((\tau)L)\times a^\dagger(\tfrac{3}{2}))^{(J')}|0\rangle, \tag{5.4}$$

and

$$T_a^{(k)} = (d^\dagger\tilde{d})^{(k)} + (-1)^{(k+1)/2}/\sqrt{2}(a^\dagger(\tfrac{3}{2})\tilde{a}(\tfrac{3}{2}))^{(k)}. \tag{5.5}$$

The $O(5)\supset O(3)$ GCC's in Eq. (5.4) can be taken from Ref. 9. From Eqs. (5.3) to (5.5) the matrix element (5.2) is obtained, which in turn determines the $O(5)\supset O(3)$ GCC's. Some useful cases are given in Table VI. These determine the structure of low-lying states not only in odd–odd nuclei in the $U_\nu^B(6)\times U_\pi^B(6)\times U_\nu^F(12)\times U_\pi^F(4)$ scheme, but also in the $U^B(6)\times U^F(20)$ scheme for odd-A nuclei discussed in Ref. 20.

VI. THE $O(6)\supset O(5)$ GCC's $\left\langle \begin{matrix} \langle\tilde{\sigma},\frac{1}{2},\frac{1}{2}\rangle & \langle 1\rangle \\ (\tilde{\tau},\frac{1}{2}) & (\tau) \end{matrix} \middle| \begin{matrix} \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2) \end{matrix} \right\rangle$

These $O(6)\supset O(5)$ GCC's appear in the transformation from the coupled basis $O_a(6)\times O_b(6)\supset O_{a+b}(6)\supset O_{a+b}(5)\supset O_{a+b}(3)$ to the uncoupled basis $O_a(6)\times O_b(6)\supset O_a(5)\times O_b(5)\supset O_a(3)\times O_b(3)\supset O_{a+b}(3)$:

TABLE V. Some $O(6)\supset O(5)$ GCC's $\left\langle \begin{matrix} \langle\sigma,1\rangle & \langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle \\ (\bar{\tau}_1,\bar{\tau}_2) & (\frac{1}{2},\frac{1}{2}) \end{matrix} \middle| \begin{matrix} \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2) \end{matrix} \right\rangle$.

$\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle$	(τ_1,τ_2)	$(\bar{\tau}_1,\bar{\tau}_2)=(1)$	$(\bar{\tau}_1,\bar{\tau}_2)=(2)$	$(\bar{\tau}_1,\bar{\tau}_2)=(1,1)$	$(\bar{\tau}_1,\bar{\tau}_2)=(2,1)$
$\langle\sigma+\frac{1}{2},\frac{3}{2},\frac{1}{2}\rangle$	$(\frac{3}{2},\frac{1}{2})$	$\left(\frac{15(\sigma+3)(\sigma+5)}{32(\sigma+2)(\sigma+4)}\right)^{1/2}$	$-\left(\frac{3(\sigma-1)(\sigma+3)}{16(\sigma+2)(\sigma+4)}\right)^{1/2}$	$\left(\frac{5(\sigma+1)(\sigma+5)}{16(\sigma+2)(\sigma+4)}\right)^{1/2}$	$-\left(\frac{(\sigma-1)(\sigma+1)}{32(\sigma+2)(\sigma+4)}\right)^{1/2}$
	$(\frac{3}{2},\frac{3}{2})$			$\left(\frac{\sigma+5}{2(\sigma+2)}\right)^{1/2}$	$-\left(\frac{\sigma-1}{2(\sigma+2)}\right)^{1/2}$
$\langle\sigma+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle$	$(\frac{1}{2},\frac{1}{2})$	$\left(\frac{\sigma+3}{2(\sigma+2)}\right)^{1/2}$		$-\left(\frac{\sigma+1}{2(\sigma+2)}\right)^{1/2}$	
	$(\frac{3}{2},\frac{1}{2})$	$\left(\frac{(\sigma+3)(\sigma+5)}{32\sigma(\sigma+2)}\right)^{1/2}$	$-\left(\frac{5(\sigma-1)(\sigma+3)}{16\sigma(\sigma+2)}\right)^{1/2}$	$-\left(\frac{3(\sigma+1)(\sigma+5)}{16\sigma(\sigma+2)}\right)^{1/2}$	$\left(\frac{15(\sigma-1)(\sigma+1)}{16\sigma(\sigma+2)}\right)^{1/2}$

TABLE VI. The $O(5)\supset O(3)$ GCC's $\left\langle \begin{matrix} (\frac{3}{2},\frac{1}{2}) & (1) \\ J' & 2 \end{matrix} \middle| \begin{matrix} (\tau_1,\tau_2) \\ J \end{matrix} \right\rangle$.

(τ_1,τ_2)	J	$J'=\frac{1}{2}$	$J'=\frac{3}{2}$	$J'=\frac{7}{2}$
$(\frac{1}{2},\frac{1}{2})$	$\frac{3}{2}$	$\sqrt{\frac{1}{8}}$	$\sqrt{\frac{7}{8}}$	$\sqrt{\frac{1}{2}}$
$(\frac{3}{2},\frac{1}{2})$	$\frac{1}{2}$		1	
	$\frac{3}{2}$	$\sqrt{\frac{1}{3}}$	$-\sqrt{\frac{2}{147}}$	$\sqrt{\frac{22}{49}}$
	$\frac{7}{2}$		$-\sqrt{\frac{24}{49}}$	$\sqrt{\frac{25}{49}}$
$(\frac{3}{2},\frac{3}{2})$	$\frac{3}{2}$	$-\sqrt{\frac{21}{40}}$	$-\sqrt{\frac{5}{56}}$	$\sqrt{\frac{27}{70}}$
	$\frac{5}{2}$	$-\sqrt{\frac{1}{15}}$	$\sqrt{\frac{5}{8}}$	$\sqrt{\frac{1}{10}}$
	$\frac{9}{2}$		$\sqrt{\frac{3}{14}}$	$\sqrt{\frac{11}{14}}$
$(\frac{5}{2},\frac{1}{2})$	$\frac{3}{2}$	$\sqrt{\frac{7}{20}}$	$-\sqrt{\frac{15}{28}}$	$\sqrt{\frac{4}{33}}$
	$\frac{5}{2}$	$\sqrt{\frac{3}{5}}$	$\sqrt{\frac{15}{98}}$	$-\sqrt{\frac{121}{490}}$
	$\frac{7}{2}$		$\sqrt{\frac{25}{49}}$	$\sqrt{\frac{24}{49}}$
	$\frac{9}{2}$		$\sqrt{\frac{11}{14}}$	$-\sqrt{\frac{3}{14}}$
	$\frac{11}{2}$			1

$$|\langle\tilde{\sigma},\tfrac{1}{2},\tfrac{1}{2}\rangle\langle 1\rangle;\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JM\rangle$$
$$=\sum_{\tilde{\tau},\tau}\sum_{J',L}\left\langle \begin{matrix}\langle\tilde{\sigma},\frac{1}{2},\frac{1}{2}\rangle\langle 1\rangle \\ (\tilde{\tau},\frac{1}{2})\ (\tau)\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2)\end{matrix}\right\rangle\left\langle \begin{matrix}(\tilde{\tau},\frac{1}{2})(\tau) \\ J'\quad L\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$$
$$\times|\langle\tilde{\sigma},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tilde{\tau},\tfrac{1}{2})J',\langle 1\rangle(\tau)L;JM\rangle. \tag{6.1}$$

A derivation analogous to the one in Sec. IV leads to the relation

$$\left\langle \begin{matrix}\langle\tilde{\sigma},\frac{1}{2},\frac{1}{2}\rangle\langle 1\rangle \\ (\tilde{\tau},\frac{1}{2})\ (1)\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2)\end{matrix}\right\rangle\left\langle \begin{matrix}(\tilde{\tau},\frac{1}{2})(1) \\ J'\quad 2\end{matrix}\middle|\begin{matrix}(\tau_1,\tau_2)\\ J\end{matrix}\right\rangle$$
$$\times\{[\sigma_1(\sigma_1+4)+\sigma_2(\sigma_2+2)]-\tilde{\sigma}(\tilde{\sigma}+4)$$
$$+\tilde{\tau}(\tilde{\tau}+3)-[\tau_1(\tau_1+3)+\tau_2(\tau_2+1)]-\tfrac{3}{2}\}$$
$$=\delta_{\tau_2,1/2}\,2(2J+1)^{-1/2}$$
$$\times\langle\langle\tilde{\sigma},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tau_1,\tfrac{1}{2})J\|T_a^{(k)}\|\langle\tilde{\sigma},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tilde{\tau},\tfrac{1}{2})J'\rangle$$
$$\times\left\langle \begin{matrix}\langle\tilde{\sigma},\frac{1}{2},\frac{1}{2}\rangle\langle 1\rangle \\ (\tau_1,\frac{1}{2})\ (0)\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2)\end{matrix}\right\rangle. \tag{6.2}$$

Since the $O(5)\supset O(3)$ GCC's in Eq. (6.2) are known from Sec. V and the reduced matrix elements can be taken from Ref. 9, the $O(6)\supset O(5)$ GCC's can be obtained for different choices of $\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle$ and (τ_1,τ_2). The cases of practical interest are given in Table VII.

VII. APPLICATIONS

The results derived in Secs. III–VI fully determine the wave functions (2.4) and (2.8) of the lowest states of odd–odd nuclei with $U_\nu^B(6)\times U_\nu^F(12)\times U_\pi^B(6)\times U_\pi^F(4)$ symmetry. They can thus be used for deriving properties of these nuclei such as electromagnetic transition rates and particle-transfer strengths. Hence, the symmetry scheme, which was proposed in Ref. 16 on the basis of energy spectra, can now be tested in further detail.

To illustrate this, let us take the example of neutron or proton transfer. For one-proton transfer reactions that conserve the number of bosons, the transfer strength is given—in simplest approximation—by the square of the reduced matrix element

$$\langle[N_1,N_2]\langle\bar{\sigma}_1,\bar{\sigma}_2\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JL\,\|a_\pi^\dagger(\tfrac{3}{2})\|GS(\nu);\tfrac{1}{2}\rangle, \tag{7.1}$$

where the reduced matrix element is defined with a $3j$-symbol in the convention of Ref. 21. In Eq. (7.1), the ket denotes

TABLE VII. Some $O(6)\supset O(5)$ GCC's $\left\langle \begin{matrix}\langle\sigma+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\langle 1\rangle \\ (\tilde{\tau},\frac{1}{2})\quad(\tau)\end{matrix}\middle|\begin{matrix}\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle \\ (\tau_1,\tau_2)\end{matrix}\right\rangle$.

$\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle$	(τ_1,τ_2)	(τ)	$(\tilde{\tau},\frac{1}{2})=(\frac{1}{2},\frac{1}{2})$	$(\tilde{\tau},\frac{1}{2})=(\frac{3}{2},\frac{1}{2})$	$(\tilde{\tau},\frac{1}{2})=(\frac{5}{2},\frac{1}{2})$
$\langle\sigma+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle$	$(\frac{1}{2},\frac{1}{2})$	(0)	$\left(\frac{\sigma+5}{2(\sigma+3)}\right)^{1/2}$		
	$(\frac{1}{2},\frac{1}{2})$	(1)	$-\left(\frac{\sigma+5}{10(\sigma+3)}\right)^{1/2}$	$\left(\frac{2\sigma}{5(\sigma+3)}\right)^{1/2}$	
	$(\frac{3}{2},\frac{1}{2})$	(0)		$\left(\frac{\sigma(\sigma+6)}{2(\sigma+1)(\sigma+3)}\right)^{1/2}$	
	$(\frac{3}{2},\frac{1}{2})$	(1)	$\left(\frac{(\sigma+5)(\sigma+6)}{10(\sigma+1)(\sigma+3)}\right)^{1/2}$	$-\left(\frac{3\sigma(\sigma+6)}{70(\sigma+1)(\sigma+3)}\right)^{1/2}$	$-\left(\frac{5\sigma(\sigma-1)}{14(\sigma+1)(\sigma+3)}\right)^{1/2}$
$\langle\sigma+\frac{1}{2},\frac{3}{2},\frac{1}{2}\rangle$	$(\frac{3}{2},\frac{1}{2})$	(0)		$\left(\frac{15}{4(\sigma+1)(\sigma+4)}\right)^{1/2}$	
	$(\frac{3}{2},\frac{1}{2})$	(1)	$-\left(\frac{3\sigma(\sigma+5)}{4(\sigma+1)(\sigma+4)}\right)^{1/2}$	$-\left(\frac{(2\sigma+5)^2}{28(\sigma+1)(\sigma+4)}\right)^{1/2}$	$-\left(\frac{3(\sigma-1)(\sigma+6)}{28(\sigma+1)(\sigma+4)}\right)^{1/2}$
	$(\frac{3}{2},\frac{1}{2})$	(1)		1	
$\langle\sigma+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle$	$(\frac{1}{2},\frac{1}{2})$	(0)	$\left(\frac{1}{(\sigma+2)(\sigma+3)}\right)^{1/2}$		
	$(\frac{1}{2},\frac{1}{2})$	(1)	$\left(\frac{(2\sigma+5)^2}{5(\sigma+2)(\sigma+3)}\right)^{1/2}$	$\left(\frac{\sigma(\sigma+5)}{5(\sigma+2)(\sigma+3)}\right)^{1/2}$	
	$(\frac{3}{2},\frac{1}{2})$	(0)		$\left(\frac{9}{4(\sigma+2)(\sigma+3)}\right)^{1/2}$	
	$(\frac{3}{2},\frac{1}{2})$	(1)	$-\left(\frac{\sigma(\sigma+5)}{20(\sigma+2)(\sigma+3)}\right)^{1/2}$	$\left(\frac{27(2\sigma+5)^2}{140(\sigma+2)(\sigma+3)}\right)^{1/2}$	$\left(\frac{5(\sigma-1)(\sigma+6)}{28(\sigma+2)(\sigma+3)}\right)^{1/2}$

the ground state of an odd-neutron nucleus which has the O(6) symmetry of $U^B(6)\times U^F(12)$, i.e.,

$$|GS(\nu);\tfrac{1}{2}\rangle = A^\dagger_\nu([N+1]\langle N+1\rangle(0)0\,\tfrac{1}{2})|0\rangle. \quad (7.2)$$

Furthermore, the bra denotes a state of an odd–odd nucleus, characterized according to the chain (2.1). The matrix element (7.1) thus describes the properties of a transfer reaction such as $^{199}_{80}Hg_{119}\rightarrow{}^{198}_{79}Au_{119}$. From Eq. (2.4) one immediately finds the predicted transfer strength:

$$\langle[N+1]\langle N+1\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tfrac{1}{2},\tfrac{1}{2})\tfrac{3}{2}L\,\|a^\dagger_\pi(\tfrac{3}{2})\|GS(\nu);\tfrac{1}{2}\rangle^2$$
$$=(2L+1)\left\langle\begin{matrix}\langle N+1\rangle\langle\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle & \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (0)\quad(\frac{1}{2},\frac{1}{2}) & (\frac{1}{2},\frac{1}{2})\end{matrix}\right\rangle^2. \quad (7.3)$$

Hence, two 1^-–2^- doublets are predicted to be excited, and their respective strengths can be computed from Eq. (7.3) and Table IV. The results are

$$\langle[N+1]\langle N+1\rangle\langle N+\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tfrac{1}{2},\tfrac{1}{2})\tfrac{3}{2}L\,\|a^\dagger_\pi(\tfrac{3}{2})\|GS(\nu);\tfrac{1}{2}\rangle^2$$
$$=(2L+1)[(N+5)/2(N+3)], \quad (7.4a)$$

and

$$\langle[N+1]\langle N+1\rangle\langle N+\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tfrac{1}{2},\tfrac{1}{2})\tfrac{3}{2}L\,\|a^\dagger(\tfrac{3}{2})\|GS(\nu);\tfrac{1}{2}\rangle^2$$
$$=(2L+1)[(N+1)/2(N+3)]. \quad (7.4b)$$

For one-neutron transfer properties the matrix element

$$\langle\langle\sigma_B\rangle\langle\tilde\sigma_1,\tfrac{1}{2},\tfrac{1}{2}\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JL\,\|a^\dagger_\nu(j)\|GS(\pi);\tfrac{3}{2}\rangle \quad (7.5)$$

should be calculated, where the bra denotes a state of an odd–odd nucleus, characterized according to the chain (2.5), and the ket denotes the ground state of an odd-proton nucleus which has the Spin(6) symmetry of $U^B(6)\times U^F(4)$, i.e.,

$$|GS(\pi);\tfrac{3}{2}\rangle = A^\dagger_\pi(\langle N\rangle\langle N+\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tfrac{1}{2},\tfrac{1}{2})\tfrac{3}{2})|0\rangle. \quad (7.6)$$

The matrix element (7.5) thus describes the properties of a transfer reaction such as $^{197}_{79}Au_{118}\rightarrow{}^{196}_{79}Au_{117}$. From Eq. (2.8) one finds the predicted transfer strength:

$$\langle\langle N\rangle\langle N+\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle(\tau_1,\tau_2)JL\,\|a^\dagger_\nu(j)\|GS(\pi);\tfrac{3}{2}\rangle^2$$
$$=(2j+1)(2J+1)(2L+1)\begin{Bmatrix}\frac{3}{2} & L_F & J\\ \frac{1}{2} & L & j\end{Bmatrix}^2$$
$$\times\left\langle\begin{matrix}\langle N+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle\ \langle 1\rangle & \langle\sigma_1,\sigma_2,\frac{1}{2}\rangle\\ (\frac{1}{2},\frac{1}{2})\quad(\tau_F) & (\tau_1,\tau_2)\end{matrix}\right\rangle^2, \quad (7.7)$$

where $(\tau_F)L_F=(0)0$ for $j=\frac{1}{2}$ transfer and $(\tau_F)L_F=(1)2$ for $j=\frac{3}{2},\frac{5}{2}$ transfer. Simple predictions for the neutron transfer can now be obtained from Eq. (7.7) and Table VII.

Finally, it should be noted that the chains (2.1) and (2.5) define a different basis, and hence, in general states in both classification schemes do not coincide. The transformation matrix between the two bases is an O(6) Racah coefficient multiplied by a $U(6)\supset O(6)$ GCC, i.e.,

$$U(\langle\sigma_B\rangle\langle 1\rangle\langle\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle\langle\sigma_1,\sigma_2,\tfrac{1}{2}\rangle;\langle\bar\sigma_1,\bar\sigma_2\rangle\langle\tilde\sigma_1,\tfrac{1}{2},\tfrac{1}{2}\rangle)$$
$$\times\left\langle\begin{matrix}[N][1] & [N_1,N_2]\\ \langle\sigma_B\rangle\langle 1\rangle & \langle\bar\sigma_1,\bar\sigma_2\rangle\end{matrix}\right\rangle. \quad (7.8)$$

For the lowest states of ^{198}Au shown in Ref. 16, this transformation matrix reduces to unity and such states *do* coincide in both schemes. Specifically, this is true for states with $\langle\sigma_1,\sigma_2,\frac{1}{2}\rangle=\langle N+\frac{3}{2},\frac{1}{2},\frac{1}{2}\rangle$ and $\langle N+\frac{1}{2},\frac{1}{2},\frac{1}{2}\rangle$, the correspondence being given by

$$|[N+1]\langle N+1\rangle\langle N+\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle$$
$$\equiv|\langle N\rangle\langle N+\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle\langle N+\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle, \quad (7.9a)$$
$$|[N,1]\langle N,1\rangle\langle N+\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle$$
$$\equiv|\langle N\rangle\langle N+\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\rangle\langle N+\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2}\rangle(\tau_1,\tau_2)JLM\rangle, \quad (7.9b)$$

where the wave functions on the left-hand side are classified according to Eq. (2.2) and the ones on the right-hand side according to Eq. (2.6).

ACKNOWLEDGMENTS

I am indebted to J. Jolie and A. Frank for useful discussions.

This research was supported by the Science and Engineering Research Council, United Kingdom, and by NATO Grant RG.85/0699.

[1] A. Arima and F. Iachello, Phys. Rev. Lett. **35**, 1069 (1975).
[2] A. Arima and F. Iachello, Ann. Phys. (NY) **99**, 253 (1976).
[3] A. Arima and F. Iachello, Ann. Phys. (NY) **111**, 201 (1978).
[4] A. Arima and F. Iachello, Ann. Phys. (NY) **123**, 468 (1979).
[5] O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[6] F. Iachello and O. Scholten, Phys. Rev. Lett. **43**, 679 (1979).
[7] F. Iachello, Phys. Rev. Lett. **44**, 772 (1980).
[8] A. B. Balantekin, I. Bars, R. Bijker, and F. Iachello, Phys. Rev. C **27**, 1761 (1983).
[9] F. Iachello and S. Kuyucak, Ann. Phys. (NY) **136**, 19 (1981).
[10] R. Bijker and V. K. B. Kota, Ann. Phys. (NY) **156**, 110 (1984).
[11] P. Van Isacker, A. Frank, and H.-Z. Sun, Ann. Phys. (NY) **157**, 183 (1984).
[12] R. Bijker and F. Iachello, Ann. Phys. (NY) **161**, 360 (1985).
[13] F. Iachello and R. Bijker, to be published in Ann. Phys. (NY).
[14] P. Van Isacker, K. Heyde, J. Jolie, and A. Sevrin, Ann. Phys. (NY) **171**, 253 (1986).
[15] I. Bars, in *Bosons in Nuclei*, edited by D. H. Feng, S. Pittel, and M. Vallieres (World Scientific, Singapore, 1984), p. 155.
[16] P. Van Isacker, J. Jolie, K. Heyde, and A. Frank, Phys. Rev. Lett. **54**, 653 (1985).
[17] A. B. Balantekin and V. Paar, Phys. Rev. C **34**, 1917 (1986).
[18] B. G. Wybourne, *Classical Groups for Physicists* (Wiley–Interscience, New York, 1974).
[19] J. Van der Jeugt, J. Phys. A **18**, L745 (1985).
[20] J. Jolie, K. Heyde, P. Van Isacker, and A. Frank, to be published in Nucl. Phys. A.
[21] A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).

A gauge theory for the quantum planar three-body problem

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

(Received 28 January 1986; accepted for publication 12 November 1986)

A several-particle system is called a molecule in the Born–Oppenheimer approximation. The nonrigidity of molecules involves difficulty in molecular dynamics. Guichardet [A. Guichardet, Ann. Inst. H. Poincaré **40**, 329 (1984)] showed recently that the vibration motion cannot in general be separated from the rotation motion, by using the connection theory in differential geometry. The point of his theory is the observation that a center-of-mass system is made into a principal fiber bundle with rotation group as the structure group, and is equipped with a connection by the Eckart condition of rotationless constraint. The base manifold of this bundle is called the internal space. The fact that the connection has nonvanishing curvature gives rise to the nonseparability of vibration from rotation. This is a mathematical meaning of nonrigidity of molecules. As an application of the connection theory due to Guichardet, this paper establishes a gauge theory for nonrigid molecules on the basis of the observation that the vector bundle associated with the principal fiber bundle (the center-of-mass system) provides a setting for quantum mechanics of the "internal" molecular motion. The interest, however, centers on planar triatomic molecules in order to put forward the gauge theory in an explicit manner. The conclusion is this: The internal space of a planar triatomic molecule is diffeomorphic with $\mathbf{R}^3 - \{0\}$, and endowed with Dirac's monopole field which may be interpreted as a Coriolis field induced by the rotation. The angular momentum eigenvalues, which are twice the quantized monopole strengths, assign the complex line bundles over the internal space. The internal states of the molecule are described as the cross sections of the complex line bundle, on which the internal Hamiltonian operator acts in minimally coupling with the monopole field.

I. INTRODUCTION

In this paper, a molecule means a system of particles or atomic nuclei in the Born–Oppenheimer approximation. The quantum mechanics of molecules has been studied by the use of the Eckart frame.[1] However, the Eckart Hamiltonian is interpreted as valid in the vicinity of the equilibrium nuclear position. If one wishes to study molecular motions far from the equilibrium nuclear position, one must become involved with the difficulty in separating off rotation from vibration.

It is Guichardet[2] who proved that the rotation motion cannot be separated from the vibration motion on the basis of the connection theory in differential geometry. The point of his proof is the observation that the center-of-mass system is made into a principal fiber bundle with rotation group as the structure group, and is equipped with a connection by the Eckart condition of rotationless constraint. The base manifold of this principal fiber bundle is called the internal space.

The connection theory fits mechanics of many-body systems. Indeed, the connection form due to Guichardet is expressed as a differential one-form which is associated only with the total angular momentum and the inertia tensor. Thus the Eckart condition is expressed in terms of the connection form. The nonintegrability of the Eckart condition is then measured by the curvature two-form defined as the "covariant" derivative of the connection form. Through the connection theory, the rotation, vibration, and internal motion of the molecule are rigorously realized. The vibration is not equal to the internal motion, while it induces the internal motion. Guichardet's conclusion of the nonseparability of rotation and vibration is stated as follows: Performing a purely vibrational motion, a molecule can, at the end of a finite time, come to a final configuration which is deduced from the initial one by an arbitrary pure rotation.

On the basis of Guichardet's observation, the present author[3] showed that a moving frame, called the Eckart frame, exists relative to which the molecule moves without rotation, but it depends on a choice of the molecular motion and is not unique for any molecular configuration. For this reason, the Eckart frame is not suitable for describing motions of nonrigid molecules in quantum as well as classical mechanics.

The purpose of this paper is to establish quantum mechanics of nonrigid molecules as a gauge theory. The point is the idea that the complex vector bundle associated with the center-of-mass system is a quantum-mechanical arena for internal molecular dynamics. The introduction of the complex vector bundle is a geometric consequence of the conservation of the total angular momentum. For planar molecules this idea is explained easily as follows: Since the rotation group SO(2) acts on the center-of-mass system for the planar molecule, wave functions on the center-of-mass system are expanded into Fourier series in the rotation angle, expansions in the eigenfunctions of the angular momentum operator. Each Fourier coefficient then might be considered as a function of internal coordinates, which describes the internal state of the molecule with the assigned angular momentum eigenvalue. However, as the center-of-mass system

 0022-2488/87/040964-11$02.50

does not decompose into a product space of the space of the rotation angles ($\simeq S^1$) and the internal space, the Fourier coefficient is not a globally defined function on the internal space. Since a plural number of product spaces of S^1 and an open subset of the internal space may cover the center-of-mass system, from the Fourier series of the wave function restricted on each product space one can get a collection of locally defined wave functions on open subsets in the internal space. These local functions are pieced together through gauge transformations. Geometrically speaking, the local functions pieced together are identified with a cross section in a complex line bundle over the internal space. The cross sections in the complex line bundle thus describe the internal states of the planar molecule with the prescribed angular momentum eigenvalue. If the eigenvalue is zero, the complex line bundle becomes trivial, that is, the total space of the bundle becomes a product space of the internal space and the complex line $\mathbb{C}$, so that the sections in this case are nothing but the functions on the internal space. For the molecules in the space, complex vector bundles are used in place of the complex line bundle.

The internal Hamiltonian operator for the internal motions of the planar molecule may be derived from that acting on wave functions of the center-of-mass system. In fact, operating each term of the Fourier series with the Hamiltonian on the center-of-mass system, one can get a local expression of the internal Hamiltonian operator acting on local functions on the internal space. In order to get a global expression, one must pass to the complex line bundle. To look into the internal Hamiltonian operator, a certain kind of first-order differential operator is worth mentioning. The derivative of each term of the Fourier series in the direction of a vibrational vector gives rise to a locally defined first-order differential operator on the internal space, which may be interpreted as expressing the minimal coupling of the internal motion with some magnetic field on the internal space. This first-order operator can be defined rigorously as a covariant derivation in the complex line bundle.

The above idea is summed up in terms of the bundle theory as follows: The complex line bundle, which is assigned by the eigenvalue of the angular momentum operator, is endowed with a linear connection (or a covariant derivation) from the connection of Guichardet on the center-of-mass system. The curvature of the linear connection is interpreted as a magnetic field on the internal space. The covariant derivation describes the minimal coupling of internal motions with the magnetic field. The internal Hamiltonian operator can be expressed globally in terms of the covariant derivation. Thus one can picture that the rotation induces on the internal space the magnetic field with which the internal motion is minimally coupled.

The interest of this article, however, centers on the planar triatomic molecules in order to put forward the gauge theory in an explicit manner. For physical intuition to the fiber bundle theory, Nash and Sen[4] and Eguchi, Gilkey, and Hanson[5] are helpful. The organization of this article is outlined in the following way.

Section II is concerned with the center-of-mass system for the planar triatomic molecule. It is shown that the center-of-mass system is made into a principal fiber bundle $\dot{\mathbb{R}}^4 \to \dot{\mathbb{R}}^3$ with the structure group SO(2), where the superscript dot means that the origin O is gotten rid of. This fiber bundle is closely related with the Hopf fibering $S^3 \to S^2$. The connection defined by the Eckard condition is constructed, and its curvature is calculated to be Dirac's monopole field on $\dot{\mathbb{R}}^3$, the internal space.

Section III shows that the complex line bundle associated with the principal fiber bundle $\dot{\mathbb{R}}^4 \to \dot{\mathbb{R}}^3$ is an arena for quantum mechanics for the internal motion of planar triatomic molecules. The eigenvalues of the total angular momentum operator assign these complex line bundles. The linear connection and its curvature induced from those on the center-of-mass system are discussed. The curvature proves to define, on the internal space Dirac's monopole field of a quantized strength, half the eigenvalue of the angular momentum operator. The linear connection (or covariant derivation) shows that the internal motion is minimially coupled with the monopole field. Gauge transformations for local cross sections are discussed with introducing local coordinate systems.

Section IV is devoted to the internal Hamiltonian operator by which the internal motion is governed. This operator, acting on cross sections in the complex line bundle, is shown to be quadratic in the covariant derivation operator and to contain a centrifugal potential. Thus one understands how the rotation affects the internal motion. Local expressions are also given in the coordinates introduced in Sec. III.

The internal Hamiltonian operators for multiatomic molecules in $\mathbb{R}^3$ are given in Ref. 6 only for zero total angular momentum eigenvalue without reference to the geometry of complex vector bundles in the large.

II. THE CENTER-OF-MASS SYSTEM AS A PRINCIPAL FIBER BUNDLE

A. Settings on the center-of-mass system

Let y_1, y_2, y_3 be position vectors of particles in $\mathbb{R}^2$ with masses m_1, m_2, m_3, respectively. The components of y_k will be denoted by (y^j_k), $j = 1,2$. The configuration space of a molecule, the system of particles, is the linear space of all the triples (y_1, y_2, y_3);

$$Q_0 = \{ y = (y_1, y_2, y_3) \; ; \; y_k \in \mathbb{R}^2 \} . \tag{2.1}$$

We equip Q_0 with the inner product

$$K(x, y) = \sum_{k=1}^{3} m_k (x_k | y_k) , \tag{2.2}$$

where the parentheses denote the standard inner product in $\mathbb{R}^2$.

The rotation group SO(2) acting on $\mathbb{R}^2$ acts also on Q_0 in a natural manner;

$$y = (y_1, y_2, y_3) \to gy = (gy_1, gy_2, gy_3) . \tag{2.3}$$

The center-of-mass system Q is the linear subspace of Q_0 defined by

$$Q = \left\{ y \in Q_0 \; ; \; \sum_{k=1}^{3} m_k y_k = 0 \right\} . \tag{2.4}$$

Since the condition $\Sigma m_k y_k = 0$ is invariant under the

SO(2) action, Q admits the SO(2) action, too. The induced inner product on Q will be also denoted by the same letter K.

To express the action of SO(2) on Q_0 in a simple matrix form, we introduce a nice orthonormal system in Q_0 as follows.

Proposition 1: Let e_1 and e_2 be the standard basis in $\mathbf{R}^2$. Then the following triples constitute an orthonormal system in Q_0 with respect to the inner product K:

$$\begin{aligned} c_1 &= N_0(e_1,e_1,e_1)\,, \\ c_2 &= N_0(e_2,e_2,e_2)\,; \end{aligned} \tag{2.5a}$$

$$\begin{aligned} f_0 &= N_1(m_3e_1,0,-m_1e_1)\,, \\ f_1 &= N_1(m_3e_2,0,-m_1e_2)\,, \\ f_2 &= N_2(-m_2e_1,(m_1+m_3)e_1,-m_2e_1)\,, \\ f_3 &= N_2(-m_2e_2,(m_1+m_3)e_2,-m_2e_2)\,; \end{aligned} \tag{2.5b}$$

where N_k, $k=1,2,3$, are the normalization constants given, respectively, by

$$\begin{aligned} N_0 &= (m_1+m_2+m_3)^{-1/2}\,, \\ N_1 &= (m_1m_3(m_1+m_3))^{-1/2}\,, \\ N_2 &= (m_2(m_1+m_3)(m_1+m_2+m_3))^{-1/2}\,. \end{aligned} \tag{2.6}$$

Proof: To verify that (2.5) is an orthonormal system is a matter of calculation. The vectors f_k, $k=0,...,3$, and c_j, $j=1,2$, form an orthonormal basis in Q and $Q^{\perp}$, the orthogonal complement of Q, respectively.

The vectors f_k are special configurations of the molecule in the center of mass system. (See Fig. 1.)

Proposition 2: Let $g\in$SO(2);

$$g(t)=\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}. \tag{2.7}$$

Then the action of SO(2) on Q is expressed with respect to the basis $\{f_k\}$ in the form

$$\begin{pmatrix} g(t) & & \\ & & \\ & & g(t)\end{pmatrix}, \tag{2.8}$$

where the missing matrix entries are all zero. The action on Q is expressed in the same form as (2.7) with respect to the basis $\{c_j\}$.

Proof: Computing $K(gf_j,f_k)$ to get the coefficients of $gf_j=\Sigma a_{kj}f_k$ results in (2.8). This ends the proof.

It is of practical importance to understand what the orthonormal system (2.5) means for the molecular configuration. Let B denote the center-of-mass vector;

$$\begin{aligned} B &= \sum_{k=1}^{3} m_k y_k \left(\sum_{k=1}^{3} m_k\right)^{-1} \\ &= \sum_{j=1}^{2} B^j e_j\,. \end{aligned} \tag{2.9}$$

Then any triple $y=(y_1,y_2,y_3)$ in Q_0 is put into the usual decomposition

$$(y_1,y_2,y_3)=(x_1,x_2,x_3)+(B,B,B)\,, \tag{2.10}$$

where $x=(x_1,x_2,x_3)$ is in Q. The basis vectors c_j, $j=1,2$, are related with (B,B,B) by

$$(B,B,B)=\sum_{j=1}^{2}\frac{B^jc_j}{N_0}, \tag{2.11}$$

where N_0 is the normalization constant given in (2.6). The triple $x=(x_1,x_2,x_3)$ is of course expressed in terms of the basis vectors f_k, $k=0,...,3$:

$$x=\sum_{k=0}^{3}\xi^k f_k\,. \tag{2.12}$$

Put another way, any configuration of the molecule is expressed as a linear combination of specified configurations f_k.

Corollary 3: The sextuplet $(B^j/N_0,\xi^k)$, $j=1,2$, $k=0,...,3$, serve as the Cartesian coordinates in Q_0.

Proof: This is clear from (2.11) and (2.12).

To simplify notations, we introduce in $\mathbf{R}^2$ the complex vector space structure by setting

$$z_k=(x_k|e_1)+i(x_k|e_2)\,,\quad k=1,2,3\,. \tag{2.13}$$

Then we obtain, after a simple calculation,

$$\begin{aligned} \xi^0+i\xi^1 &= \left(\frac{m_1m_3}{m_1+m_3}\right)^{1/2}(z_1-z_3)\,, \\ \xi^2+i\xi^3 &= \left(\frac{m_2(m_1+m_3)}{m_1+m_2+m_3}\right)^{1/2}\left(z_2-\frac{m_1z_1+m_3z_3}{m_1+m_3}\right). \end{aligned} \tag{2.14}$$

Thus any vector in $Q\cong\mathbf{R}^4$ can be treated as a pair of complex variables which are closely related with the Jacobi vectors frequently used in the multiparticle system.[7] Moreover, a straightforward calculation yields

$$\sum_{k=1}^{3} m_k|z_k|^2=\sum_{j=0}^{3}(\xi^j)^2=K(x,x)\,, \tag{2.15}$$

that is, $K(x,x)$ is equal to the moment of inertia of the molecule.

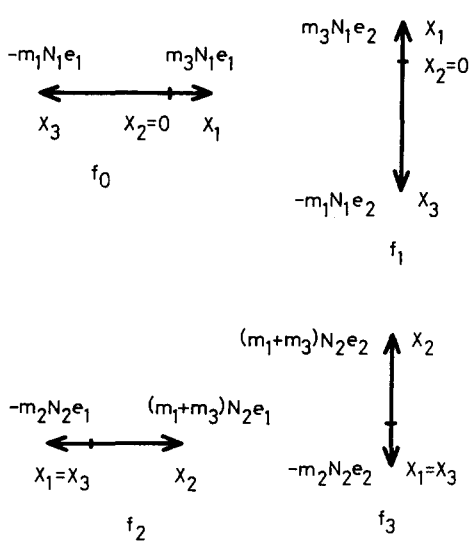


FIG. 1. Configurations corresponding to the orthonormal basis f_k.

B. The principal fiber bundle

We now proceed to a geometry of the center-of-mass system Q. From (2.8) it follows that the SO(2) action on Q is free if the origin is gotten rid of. Further, from (2.14) we see that the origin corresponds to the triple collision of the particles. By $\dot{Q}$ we mean the Q whose origin is excluded.

Then $\dot{Q}$ is diffeomorphic with $\dot{\mathbf{R}}^4 := \mathbf{R}^4 - \{0\}$. We now show the following theorem.

Theorem 4: The center-of-mass system $\dot{Q}$ without the origin is made into a principal fiber bundle with structure group SO(2); $\pi: \dot{Q} \simeq \dot{\mathbf{R}}^4 \to M \simeq \dot{\mathbf{R}}^3$, $M := \dot{Q}/\mathrm{SO}(2)$.

Proof: Following Greub and Petry,[8] we introduce in Q the structure of the quaternion algebra by setting

$$f_1 f_2 = f_3, \quad f_0 f_j = f_j f_0 = f_j, \qquad f_0^2 = -f_j^2 = f_0, \quad j = 1,2,3. \tag{2.16}$$

The conjugate of a quaternion $x = \xi^0 f_0 + \Sigma_{j=1}^3 \xi^j f_j$, $\xi^0, \xi^j \in \mathbf{R}$, is defined by

$$\bar{x} = \xi^0 f_0 - \sum_{j=1}^{3} \xi^j f_j . \tag{2.17}$$

Set

$$\epsilon(t) = f_0 \cos t + f_1 \sin t . \tag{2.18}$$

Then the SO(2) action on Q, given by (2.8), is written as the left action;

$$x \to \epsilon(t) x . \tag{2.19}$$

Since $\overline{\epsilon(t)} f_1 \epsilon(t) = f_1$, invariants under the left action (2.19) are obtained as

$$\bar{x} f_1 x = ((\xi^0)^2 + (\xi^1)^2 - (\xi^2)^2 - (\xi^3)^2) f_1 + 2(-\xi^0 \xi^3 + \xi^1 \xi^2) f_2 + 2(\xi^0 \xi^2 + \xi^1 \xi^3) f_3 . \tag{2.20}$$

On setting

$$\pi(x) = \bar{x} f_1 x = \sum_{k=1}^{3} w^k f_k , \quad x \in \dot{Q}, \tag{2.21}$$

one has the projection $\pi: \dot{Q} \to \dot{\mathbf{R}}^3$ together with

$$\left(\sum_{j=0}^{3} (\xi^j)^2 \right)^2 = \sum_{k=1}^{3} (w^k)^2 . \tag{2.22}$$

Here we have identified $\mathbf{R}^3$ with the linear subspace spanned by $f_k, k = 1,2,3$. Thus if x is restricted on $S^3 \subset \mathbf{R}^4$, π becomes the very Hopf mapping: $S^3 \to S^2$ (Ref. 5). This ends the proof.

The quaternion structure was also used by Satzer, Jr.[9] and Meyer[10] for the planar three-body problem.

The base manifold $M = \dot{Q}/\mathrm{SO}(2)$, diffeomorphic with $\dot{\mathbf{R}}^3$, is called the internal space, which is so to speak the set of all molecule forms independent of their position in $\mathbf{R}^2$. The coordinate system (w^k) is an internal coordinate system; the bond lengths and valence angles are expressible as functions of w^k.

C. The connection and curvature

As was stated in Sec. I, the Eckart condition is closely related with the geometric concept of connection and curvature. For rigorous analysis of the Eckart condition, we have to start with the definition of rotational and vibrational vectors. From (2.19) we have the fundamental vector field[11] F, the infinitesimal generator of the SO(2) action on $\dot{Q}$, as

$$F_x = \dot{\epsilon}(0) x = f_1 x, \quad x \in \dot{Q}, \tag{2.23}$$

where $\dot{\epsilon}(t)$ denotes the derivative of $\epsilon(t)$ with respect to t. In the coordinates (ξ^j) one has

$$F = -\xi^1 \frac{\partial}{\partial \xi^0} + \xi^0 \frac{\partial}{\partial \xi^1} - \xi^3 \frac{\partial}{\partial \xi^2} + \xi^2 \frac{\partial}{\partial \xi^3}, \tag{2.24}$$

where f_k are naturally identified with $\partial/\partial \xi^k$. According to Guichardet,[2] the vector field F is called rotational (or vertical[11]), and the vector fields Y orthogonal to F are called vibrational (or horizontal[8,11]);

$$K_x(Y_x, F_x) = 0, \quad x \in \dot{Q}, \tag{2.25}$$

where K_x is the inner product naturally induced in the tangent space $T_x(\dot{Q})$. We choose the vibrational vector fields $V_k, k = 1,2,3$, as follows:

$$\begin{aligned} V_1(x) &= -f_1 x f_1 = \xi^0 f_0 + \xi^1 f_1 - \xi^2 f_2 - \xi^3 f_3 , \\ V_2(x) &= -f_1 x f_2 = -\xi^3 f_0 + \xi^2 f_1 + \xi^1 f_2 - \xi^0 f_3 , \\ V_3(x) &= -f_1 x f_3 = \xi^2 f_0 + \xi^3 f_1 + \xi^0 f_2 + \xi^1 f_3 . \end{aligned} \tag{2.26}$$

The linear subspace $W_{x,\mathrm{vib}}$ of $T_x(\dot{Q})$ spanned by all the vibrational vectors at x is called the vibrational subspace. The assignment of the vibrational subspace to x gives the connection on $\dot{Q}$ which Guichardet defined,[2] and is equivalent to that analyzed by Greub and Petry.[8]

Theorem 5: The connection form ω is given by

$$\omega = \|\xi\|^{-2}(-\xi^1 d\xi^0 + \xi^0 d\xi^1 - \xi^3 d\xi^2 + \xi^2 d\xi^3), \tag{2.27}$$

where

$$\|\xi\|^2 = \sum_{j=0}^{3} (\xi^j)^2 .$$

Proof: We identify so(2), the Lie algebra of SO(2), with $\mathbf{R}$. A calculation gives

$$\omega(F) = 1, \quad \omega(V_k) = 0, \quad k = 1,2,3 . \tag{2.28}$$

Thus, by definition,[11] ω is the connection form.

It is to be remarked that $\|\xi\|^2$ is equal to the moment of inertia. Thus from (2.27) and (2.28) we see that the ω is a one-form associated only with the moment of inertia and the angular momentum. The Eckart condition of rotationless constraint is then equivalent to the differential equation $\omega = 0$. However, this equation is not completely integrable.[12] That is to say, there do not exist submanifolds of the center-of-mass system on which every motion is vibrational, i.e., the tangent vector of the motion is a vibrational vector. The nonintegrability is closely related with the nonseparability of rotation and vibration, as will be discussed below.

The tangent map π_* of the projection $\pi: \dot{Q} \to M$ provides an isomorphism of $W_{x,\mathrm{vib}}$ onto $T_{\pi(x)}(M)$, so that the vibrational vector fields on $\dot{Q}$ are in one-to-one correspondence with vector fields on M. Let X be a vector field on M. Then a unique horizontal (or vibrational) vector field X^* satisfying $\pi_* X^*_x = X_{\pi(x)}$ is called the horizontal lift of X. Now, using the definition of π_* together with (2.26), we obtain

$$\pi_* V_k = 2 \sum_{j=0}^{3} (\xi^j)^2 \frac{\partial}{\partial w^k}, \quad k = 1,2,3 . \tag{2.29}$$

Put another way, $\frac{1}{2}\|\xi\|^{-2} V_k$ is the horizontal lift of $\partial/\partial w^k$;

$$\left(\frac{\partial}{\partial w^k} \right)^* = \frac{1}{2} \|\xi\|^{-2} V_k, \quad k = 1,2,3 . \tag{2.30}$$

The Lie brackets of these horizontal lifts are calculated in a straightforward manner to give

$$\left[\left(\frac{\partial}{\partial w^1}\right)^*,\left(\frac{\partial}{\partial w^2}\right)^*\right] = -\frac{1}{2}\|\xi\|^{-6}w^3 F \quad \text{(cycl. perm.)}\,. \tag{2.31}$$

These equations are worth noticing; infinitesimal vibrations are coupled to give rise to an infinitesimal rotation. This is a reason why the vibration cannot be separated from the rotation. Equations (2.31) also mean that the equation $\omega = 0$ is not completely integrable (Frobenius' theorem).[12] The non-integrability leads to the curvature form.

Let Ω be the curvature form of the connection ω. Then for vibrational (or horizontal) vector fields U and V one has the formula[11]

$$\Omega(U,V) = -\omega([U,V])\,. \tag{2.32}$$

[In Ref. 11, one finds the formula $\Omega(U,V) = -\frac{1}{2}\omega([U,V])$, but the factor $\frac{1}{2}$ depends on a choice of the definition of exterior products.]

Theorem 6: The curvature form defines a magnetic monopole field on the internal space M;

$$\Omega = \tfrac{1}{2}\|w\|^{-3}(w^3\,dw^1\wedge dw^2 + w^1\,dw^2\wedge dw^3 + w^2\,dw^3\wedge dw^1)\,, \tag{2.33}$$

where (w^k) are given by (2.20) and (2.21), and $\|w\| = (\Sigma_{k=1}^3 (w^k)^2)^{1/2} = \|\xi\|^2$.

Proof: Since Ω is a tensorial two-form[11] (i.e., vanishes for any rotational vector field), and invariant under the SO(2) action [note that SO(2) is Abelian], it defines a two-form on M. We denote it by the same letter Ω. From (2.28), (2.31), and (2.32), we obtain (2.33).

D. The induced metric on *M*

Since the inner product K on Q is invariant under SO(2), it projects to a metric on M through π. For vector fields X and Y on M, one has unique horizontal lifts X^* and Y^*, respectively; $\pi_* X_x^* = X_{\pi(x)}$, etc. The induced metric B on M is then defined by

$$B_{\pi(x)}(X,Y) = K_x(X^*,Y^*)\,. \tag{2.34}$$

Theorem 7: The internal space M is endowed with the metric

$$\tfrac{1}{4}\|w\|^{-1}((dw^1)^2 + (dw^2)^2 + (dw^3)^2)\,. \tag{2.35}$$

Proof: Applying (2.30) and (2.34) to $\partial/\partial w^k$, $k = 1,2,3$, we obtain

$$B_{\pi(x)}\left(\frac{\partial}{\partial w^k},\frac{\partial}{\partial w^j}\right) = \frac{\delta_{kj}}{4\|w\|}\,. \tag{2.36}$$

This ends the proof.

Note that this metric is a generalization of Wilson's G^{-1} matrix.[13] Since $\Sigma_{k=1}^3 (dw^k)^2$ is considered as the standard flat metric on $\mathbf{R}^3 \simeq M$, the internal space is a conformally flat Riemannian manifold.

E. Local description of the connection and curvature

In what follows we give a local description of the connection and curvature we have discussed so far. We introduce the curvilinear coordinates (r,θ,ϕ,ψ) in $\dot{Q}$ by

$$\begin{aligned}\xi^0 + i\xi^1 &= \sqrt{r}\cos(\theta/2)\exp(i[(\psi+\phi)/2])\,,\\ \xi^2 + i\xi^3 &= \sqrt{r}\sin(\theta/2)\exp(i[(\psi-\phi)/2])\,,\end{aligned} \tag{2.37}$$

where

$$r>0,\quad 0\leqslant\theta\leqslant\pi,\quad 0\leqslant(\psi+\phi)/2\leqslant 2\pi,$$
$$-\pi\leqslant(\psi-\phi)/2\leqslant\pi.$$

The action of SO(2) on $\dot{Q}$ is then expressed as

$$(\xi^0 + i\xi^1,\xi^2 + i\xi^3)\to(e^{it}(\xi^0 + i\xi^1),e^{it}(\xi^2 + i\xi^3))\,, \tag{2.38}$$

or

$$\psi\to\psi + 2t \text{ and the others fixed.} \tag{2.39}$$

Then $\psi/2$ is the rotation angle of the "frozen" molecule from an assigned position. (See Fig. 2.)

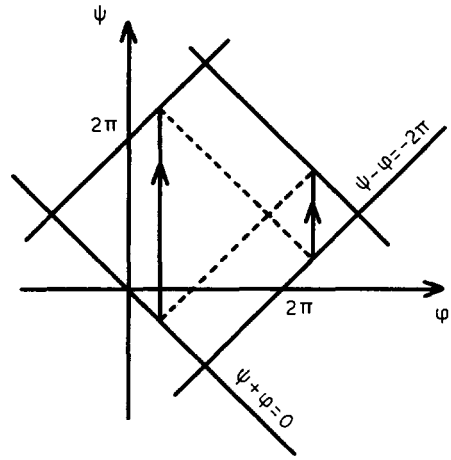


FIG. 2. Angle variables ψ and ϕ. The arrowed lines connected by broken lines represent an SO(2) orbit of a point, a circle.

It is of interest to note how the coordinates other than ψ are related to the Jacobi vectors (2.14). From (2.14) and (2.37), it follows that

$$\begin{aligned}\left(\frac{m_1 m_3}{m_1+m_3}\right)^{1/2}|z_1 - z_3| &= \sqrt{r}\cos\frac{\theta}{2}\,,\\ \left(\frac{m_2(m_1+m_3)}{m_1+m_2+m_3}\right)^{1/2}\left|z_2 - \frac{m_1 z_1 + m_3 z_3}{m_1+m_3}\right| &= \sqrt{r}\sin\frac{\theta}{2}\,.\end{aligned} \tag{2.40}$$

The ϕ is the angle made by two Jacobi vectors. Moreover, the moment of inertia is equal to r from (2.15) and (2.37). (See Fig. 3.)

From the definition of (w^k), we have

$$\begin{aligned}w^1 &= r\cos\theta\,,\\ w^2 &= r\sin\theta\sin\phi\,,\\ w^3 &= r\sin\theta\cos\phi\,,\end{aligned} \tag{2.41}$$

where

$$r>0\,,\quad 0\leqslant\theta\leqslant\pi\,,\quad 0\leqslant\phi\leqslant 2\pi\,.$$

Put another way, the spherical coordinates in $\mathbf{R}^3$ serve as internal coordinates for the planar triatomic molecule.

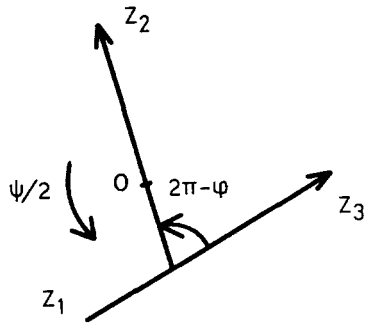


FIG. 3. Jacobi vectors and the angle variables ψ and ϕ.

It is a key to understanding internal motions to get an idea of the topology of the principal fiber bundle $\pi: \dot{Q} \to M$. While $\dot{Q}$ is not a product space $M \times S^1, S^1 \cong SO(2)$, $\dot{Q}$ can be covered by two pieces of product spaces. To see this, we consider the subsets D_+ and D_- of M which consist of points such that $\theta \neq \pi$ and $\theta \neq 0$, respectively. Then $\dot{Q}$ is covered by $\pi^{-1}(D_+)$ and $\pi^{-1}(D_-)$. We define in turn the local cross sections $\gamma_+: D_+ \to \dot{Q}$ and $\gamma_-: D_- \to \dot{Q}$ by

$$\gamma_+: \begin{cases} \xi^0 + i\xi^1 = \sqrt{r}\cos(\theta/2)\,, \\ \xi^2 + i\xi^3 = \sqrt{r}\sin(\theta/2)e^{-i\phi}\,, \end{cases} \quad \text{on } D_+(\theta \neq \pi)\,, \tag{2.42a}$$

$$\gamma_-: \begin{cases} \xi^0 + i\xi^1 = \sqrt{r}\cos(\theta/2)e^{i\phi}\,, \\ \xi^2 + i\xi^3 = \sqrt{r}\sin(\theta/2)\,, \end{cases} \quad \text{on } D_-(\theta \neq 0)\,. \tag{2.42b}$$

By the use of these sections, $\pi^{-1}(D_+)$ and $\pi^{-1}(D_-)$ turn out to be product spaces; $\pi^{-1}(D_+) \cong D_+ \times S^1$ and $\pi^{-1}(D_-) \cong D_- \times S^1$. In fact, every point of $\pi^{-1}(D_\pm)$ is expressed as $e^{it}\gamma_\pm(p), p \in D_\pm$, where the product $e^{it}\gamma_\pm(p)$ denotes a short form of the $SO(2) \cong U(1)$ action (2.38). Setting $t = (\psi + \phi)/2$ in $\pi^{-1}(D_+)$ [resp. $t = (\psi - \phi)/2$ in $\pi^{-1}(D_-)$], we recover the coordinates given by (2.37). Briefly speaking, local sections γ_+ and γ_- are defined locally in $\dot{Q}$ by $\psi + \phi = 0$ and $\psi - \phi = 0$, respectively.

We proceed to the connection and curvature forms. A straightforward calculation using (2.27), (2.33), (2.37), and (2.41) results in

$$\omega = \tfrac{1}{2}(d\psi + \cos\theta\, d\phi)\,, \tag{2.43}$$

$$\Omega = d\omega = -\tfrac{1}{2}\sin\theta\, d\theta \wedge d\phi\,. \tag{2.44}$$

Using (2.42) and (2.43), we obtain the local description of ω as follows:

$$\gamma_+^*\,\omega = \tfrac{1}{2}(-1 + \cos\theta)\, d\phi\,, \quad \text{on } D_+\,, \tag{2.45a}$$

$$\gamma_-^*\,\omega = \tfrac{1}{2}(1 + \cos\theta)\, d\phi\,, \quad \text{on } D_-\,, \tag{2.45b}$$

where the superscript asterisk indicates the pullback. These are gauge potentials of the monopole field Ω. Their gauge transformation is then given by

$$\gamma_+^*\,\omega - \gamma_-^*\,\omega = -d\phi\,, \quad \text{on } D_+ \cap D_-\,. \tag{2.46}$$

This type of gauge transformation was effectively used for the monopole field by Wu and Yang.[14]

F. Remarks on classical mechanics of planar molecules

Classical mechanics for nonrigid molecules was established in Ref. 3 in the Hamiltonian formalism. Following that paper, we make some remarks on classical mechanics for planar triatomic molecules. The phase space, to begin with, is the cotangent bundle $T^*(\dot{Q})$, on which the standard symplectic form is defined. A question as to whether or not the molecular motion is internal, i.e., independent of its attitude in the space $\mathbb{R}^2$, was affirmatively cleared up by using the reduction theory.[15] That is, the reduced phase space by the $SO(2)$ action is diffeomorphic with the cotangent bundle of the internal space, $T^*(M)$. However, this is the case only for planar molecules. If the molecule is in $\mathbb{R}^d$ $(d > 2)$, and if the angular momentum is not zero, the reduced phase space is larger than the cotangent bundle of the internal space. Hence, in this case the molecular motion is not internal.

The reduction of $T^*(\dot{Q})$ to $T^*(M) \simeq T^*(\dot{\mathbb{R}}^3)$ is in principle the same as that performed in Ref. 16 for the conformal Kepler problem of dimension 4. As a result of the reduction, one finds the reduced symplectic form which is the sum of the standard symplectic form on $T^*(M)$ and a constant multiple of the curvature form Ω viewed as a two-form on $T^*(M)$. This fact was pointed out also by Kummer.[17]

III. THE ASSOCIATED COMPLEX LINE BUNDLES

In this section, we show that the appropriate space on which quantum mechanics for the internal motion of planar triatomic molecules should be set up is the Hilbert space of square integrable cross sections in the complex line bundle associated with the principal fiber bundle $\dot{Q} \to M$.

To get an idea of the associated line bundle, we start with wave functions on the subset $\pi^{-1}(D_+)$ of $\dot{Q}$. Since $\pi^{-1}(D_+) \cong D_+ \times S^1$, any function $f(e^{it}\gamma_+(p))$ on $\pi^{-1}(D_+)$ is periodic in t with period 2π. Thus we can get a Fourier series expansion,

$$f(e^{it}\gamma_+(p)) = \sum_{m=-\infty}^{\infty} c_m^{(+)}(p)e^{imt}\,, \tag{3.1}$$

where $c_m^{(+)}(p)$ is the Fourier coefficient defined by

$$c_m^{(+)}(p) = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it}\gamma_+(p))e^{-imt}\,dt\,, \quad p \in D_+\,.$$

These coefficients are indeed functions on D_+. In a similar manner, we may obtain functions $c_m^{(-)}(p)$ on D_-. Since the internal space M is covered by D_+ and D_-, the pair $c_m^{(+)}$ and $c_m^{(-)}(p)$ should piece together an internal state of the angular momentum eigenvalue m. Thus we are led to the notion of cross sections in the complex line bundle.

A. The associated bundles and linear connections

We start with the associated complex line bundle. Though the discussion runs in parallel with the work of Greub and Petry,[8] we reproduce it from the viewpoint of reducing wave functions on Q to wave sections on M by using the $SO(2)$ action. Fix an integer m and let ρ_m denote the representation of $SO(2)$ in $\mathbb{C}$ given by

$$\rho_m(e^{it})\colon z \to e^{imt}z\,, \quad z \in \mathbb{C}\,. \tag{3.2}$$

Define a left action of SO(2) on $\dot{Q}\times\mathbb{C}$ by

$$E_m(\epsilon(t)):\ (x,z)\to(\epsilon(t)x,e^{imt}z)\,. \tag{3.3}$$

This action gives an equivalence relation in $\dot{Q}\times\mathbb{C}$. The quotient manifold, denoted by $\dot{Q}\times_m\mathbb{C}$, is made into the complex line bundle $L_m=(\dot{Q}\times_m\mathbb{C},\pi_m,M)$ via the commutative diagram

$$\begin{array}{ccc} \dot{Q}\times\mathbb{C} & \overset{q}{\to} & \dot{Q}\times_m\mathbb{C} \\ p_1\ \downarrow & & \downarrow\ \pi_m\,, \\ \dot{Q}\simeq\dot{\mathbb{R}}^4 & \overset{\pi}{\to} & M\simeq\dot{\mathbb{R}}^3 \end{array} \tag{3.4}$$

where p_1 denotes the projection onto the first factor, and q is the natural projection.

A complex-valued function Φ on $\dot{Q}$ is called equivariant with respect to ρ_m (ρ_m-equivariant, for short), if it satisfies

$$\Phi(\epsilon(t)x)=e^{imt}\Phi(x)\,,\quad x\in\dot{Q}\,. \tag{3.5}$$

A map σ: $M\to\dot{Q}\times_m\mathbb{C}$ is called a cross section in L_m, if $\pi\circ\sigma=\mathrm{id}_M$. The ρ_m-equivariant functions are in one-to-one correspondence with the cross sections in L_m (see Refs. 8 and 11), as is shown below: For an equivariant function Φ on $\dot{Q}$, the equivalence class $[(x,\Phi(x))]$ by E_m is constant on each SO(2) orbit in $\dot{Q}$, and hence it becomes a cross section in L_m through the diagram (3.4). Conversely, given a cross section σ in L_m, one can define an equivariant function on $\dot{Q}$ by

$$\Phi_\sigma(x)=q_x^{-1}(\sigma(\pi(x)))\,,$$

where $q_x:\mathbb{C}\to\pi_m^{-1}(\pi(x))$ denotes the isomorphism restricted from q on the fibers. Note that $q_{\epsilon(t)x}(z)=q_x(e^{-imt}z)$. We mean by $q_m^{\#}$ the one-to-one correspondence from the cross sections in L_m to the equivariant functions on $\dot{Q}$.

We now have to realize what the $q_m^{\#}$ does mean for molecular dynamics. Recall that the rotation group SO(2) has the infinitesimal generator F given by (2.24). The total angular momentum operator $\hat{F}$ is then defined by $\hat{F}=-iF$. Let Φ be a smooth equivariant function on $\dot{Q}$. Then one has from (3.5)

$$\hat{F}\Phi=m\Phi\,, \tag{3.6}$$

that is, Φ is an eigenfunction of $\hat{F}$ with the eigenvalue m. Conversely, we understand that Eq. (3.5) is the exponentiation of Eq. (3.6). Hence the eigenspace corresponding to the eigenvalue m of the total angular momentum operator $\hat{F}$ is in one-to-one correspondence with the space of cross sections in L_m. Therefore, the correspondence $q_m^{\#\,-1}$ is thought of as a reduction of the space of wave functions on Q to the eigenspace of the total angular momentum operator, and hence the introduction of the complex line bundle is a geometric consequence of the conservation of the total angular momentum. We may consider this reduction process as a quantum version of the reduction theorem[15] applied for classical molecular mechanics. This reduction of the space of wave functions is easy to generalize to the case of multiatomic molecules in $\mathbb{R}^d$. We notice in addition that in a local sense the equivariant function is a term in the Fourier expansion (3.1). Thus $q_m^{\#\,-1}$ is the operation of picking up the Fourier coefficient.

We proceed to the linear connection on L_m induced from the connection on the principal fiber bundle $\dot{Q}\to M$. Let X be a vector field on M and X^* its horizontal lift; $\pi_*X_x^*=X_{\pi(x)}$. Then for a cross section σ in L_m its covariant derivative with respect to X is defined by

$$\nabla_X\sigma=q_m^{\#\,-1}X^*(q_m^{\#}\sigma)\,. \tag{3.7}$$

The operator ∇ is called the linear connection, which is linear in X and σ, and satisfies for arbitrary functions f the conditions

$$\nabla_{fX}\sigma=f\nabla_X\sigma\,, \tag{3.8}$$

$$\nabla_X f\sigma=(Xf)\sigma+f\nabla_X\sigma\,. \tag{3.9}$$

The curvature of ∇ is defined for vector fields X and Y on M by

$$R(X,Y)\sigma=[\nabla_X,\nabla_Y]\sigma-\nabla_{[X,Y]}\sigma\,. \tag{3.10}$$

By (3.7), R is written also as

$$R(X,Y)\sigma=q_m^{\#\,-1}([X^*,Y^*]-[X,Y]^*)(q_m^{\#}\sigma)\,. \tag{3.11}$$

Theorem 7: The curvature R of the linear connection ∇ on L_m takes the form

$$R(X,Y)\sigma=-im\Omega(X,Y)\sigma\,,\quad m\in\mathbb{Z}\,, \tag{3.12}$$

where Ω is the monopole field given by (2.33).

Proof: Applying (3.11) to $\partial/\partial w^k$, $k=1,2,3$, and using (2.31) and (3.6) with $\Phi=q_m^{\#}\sigma$, we have

$$R\left(\frac{\partial}{\partial w^1},\frac{\partial}{\partial w^2}\right)\sigma=-imw^3\sigma/2\|w\|^3\quad\text{(cycl. perm.)}. \tag{3.13}$$

Thus we have (3.12), as is wanted.

We make some remarks on R which defines a two-form on M. Since the de Rham cohomology class $[R]/2\pi i$ is integral,

$$\frac{1}{2\pi i}\int_{S^2}R=m\,, \tag{3.14}$$

and since $M\simeq\dot{\mathbb{R}}^3$ is simply connected, a theorem of Kostant[18] shows that the bundle L_m and the connection ∇ whose curvature coincides with R are unique up to strong isomorphisms. Equation (3.14) also means that the eigenvalues of the total angular momentum operator $\hat{F}$ are twice the quantized monopole strengths. Indeed the monopole strength is given by (3.14) with the denominator replaced by $4\pi i$.

In conclusion we touch upon a bundle metric on L_m. The complex line bundle L_m is naturally endowed with a Hermitian metric. Let σ_k, $k=1,2$, denote cross sections in L_m. Then the Hermitian metric is defined for σ_1 and σ_2 by

$$(\sigma_1|\sigma_2)(\pi(x))=\overline{q_m^{\#}\sigma_1(x)}\;q_m^{\#}\sigma_2(x)\,, \tag{3.15}$$

where the bar indicates the complex conjugate. The right-hand side of (3.15) is of course invariant under the SO(2) action. It is easy to see that the linear connection ∇ is a metric connection; for any vector field X on M, one has

$$X(\sigma_1|\sigma_2)=(\nabla_X\sigma_1|\sigma_2)+(\sigma_1|\nabla_X\sigma_2)\,. \tag{3.16}$$

With this Hermitian metric, we may think of the Hilbert space of square integrable cross sections in L_m. But before doing so, we must deal with the volume element on M.

B. The inner product for cross sections

This section is concerned with the inner product for cross sections in L_m. The volume element on M for integration must be the one which is reduced from the standard volume element on the configuration space Q_0; $dQ_0 = dy_1 \wedge dy_2 \wedge dy_3$, where dy_k, $k = 1,2,3$, stand for $dy_k^1 \wedge dy_k^2$.

The volume element $d\Sigma_0$ on Q_0 defined by the inner product K is related with the standard one by

$$d\Sigma_0 = m_1 m_2 m_3 \, dQ_0 . \tag{3.17}$$

We recall here that any triple y in Q_0 is written, with respect to the orthonormal frame $\{c_j, f_k\}$, $j = 1,2$, $k = 0,...,3$, as

$$y = \sum_{j=1}^{2} \frac{B^j c_j}{N_0} + \sum_{k=0}^{3} \xi^k f_k . \tag{3.18}$$

Accordingly, the volume element $d\Sigma_0$ is expressed in terms of B^j and ξ^k as

$$\begin{aligned} d\Sigma_0 &= N_0^{-2} dB^1 \wedge dB^2 \wedge d\xi^0 \wedge d\xi^1 \wedge d\xi^2 \wedge d\xi^3 \\ &= N_0^{-2} \, dB^1 \wedge dB^2 \wedge dB^2 \wedge d\Sigma , \end{aligned} \tag{3.19}$$

where $d\Sigma = d\xi^0 \wedge \cdots \wedge d\xi^3$ is the volume element on Q defined by K. Therefore, separating off the center-of-mass coordinates, we obtain the volume element on Q;

$$dQ = \mu \, d\Sigma , \quad \mu = \sum_{k=1}^{3} m_k \Big(\prod_{k=1}^{3} m_k \Big)^{-1} . \tag{3.20}$$

To bring out the volume element on M, it is of practical use to express $d\Sigma$ in terms of the curvilinear coordinates introduced in Sec. II E. A straightforward calculation yields

$$d\Sigma = \tfrac{1}{16} r \sin\theta \, d\psi \wedge dr \wedge d\theta \wedge d\phi . \tag{3.21}$$

We are now ready to obtain the following theorem.

Theorem 8: The inner product of cross sections σ_k, $k = 1,2$, is given by

$$\int_M (\sigma_1 | \sigma_2) \, dM = \int_Q \overline{q_m^{\#} \sigma_1} \, q_m^{\#} \sigma_2 \, dQ , \tag{3.22}$$

where dM is defined by

$$\begin{aligned} dM &:= (\mu\pi/4) r \sin\theta \, dr \wedge d\theta \wedge d\phi \\ &= (\mu\pi/4\|w\|) \, dw^3 \wedge dw^2 \wedge dw^1 . \end{aligned} \tag{3.23}$$

Proof: From (3.21) the inner product for ρ_m-equivariant functions $\Phi_k = q_m^{\#} \sigma_k$, $k = 1,2$, is expressed in the form

$$\begin{aligned} \int_Q \overline{\Phi_1} \Phi_2 \, dQ &= \frac{\mu}{16} \int_0^{4\pi} d\psi \int_M \overline{\Phi_1} \Phi_2 \, r \sin\theta \, dr \, d\theta \, d\phi \\ &= \frac{\mu\pi}{4} \int_M \overline{\Phi_1} \Phi_2 \, r \sin\theta \, dr \, d\theta \, d\phi . \end{aligned} \tag{3.24}$$

Here we have used the fact that $\overline{\Phi_1} \Phi_2$ is constant in ψ. Thus the appropriate volume element on M is to be given by (3.23). From (3.15) and (3.24) with $q_m^{\#} \sigma_k = \Phi_k$, $k = 1,2$, we obtain (3.22). We note here that the volume element defined by the Riemannian metric (2.35) is given by $(4\|w\|)^{-3/2} \, dw^3 \wedge dw^2 \wedge dw^1$, which is different from dM.

C. Local description of the linear connections on L_m

In what follows we give a local description of the linear connection ∇ on L_m in order to get an idea of how the "gauge potential" of the monopole field $m\Omega$ is encompassed within the linear connection. Set

$$\begin{aligned} \Psi_+ &= \exp(im[(\psi + \phi)/2]) , \quad \text{on } \pi^{-1}(D_+) , \\ \Psi_- &= \exp(im[(\psi - \phi)/2]) , \quad \text{on } \pi^{-1}(D_-) . \end{aligned} \tag{3.25}$$

These are local ρ_m-equivariant functions on account of (2.39). By σ_+ and σ_- we mean the corresponding local cross sections in L_m; $q_m^{\#} \sigma_\pm = \Psi_\pm$.

We need first to get the horizontal lifts of $\partial/\partial\alpha^k$, $(\alpha^k) = (r,\theta,\phi)$. Making use of (2.43), one readily obtains, from $\omega((\partial/\partial\alpha^k)^*) = 0$,

$$\begin{aligned} \left(\frac{\partial}{\partial r}\right)^* &= \frac{\partial}{\partial r} , \quad \left(\frac{\partial}{\partial\theta}\right)^* = \frac{\partial}{\partial\theta} , \\ \left(\frac{\partial}{\partial\phi}\right)^* &= \frac{\partial}{\partial\phi} - \cos\theta \frac{\partial}{\partial\psi} . \end{aligned} \tag{3.26}$$

We are now in a position to compute the covariant derivatives of $\sigma_\pm$, using the definition (3.7). We have, in fact,

$$\begin{aligned} \nabla_{\partial/\partial r} \sigma_\pm &= \nabla_{\partial/\partial\theta} \sigma_\pm = 0 , \\ \nabla_{\partial/\partial\phi} \sigma_\pm &= -(im/2)(\mp 1 + \cos\theta) \sigma_\pm . \end{aligned} \tag{3.27}$$

For arbitrary local cross sections $f_+ \sigma_+$ (resp. $f_- \sigma_-$) on D_+ (resp. D_-), their covariant derivatives are given as

$$\nabla_{\partial/\partial\phi}(f_\pm \sigma_\pm) = \left(\frac{\partial}{\partial\phi} - \left(\frac{im}{2}\right)(\mp 1 + \cos\theta) \right) f_\pm \sigma_\pm . \tag{3.28}$$

For notational convenience we introduce the operators $\overset{(+)}{\nabla}_\phi$ and $\overset{(-)}{\nabla}_\phi$ by

$$\overset{(\pm)}{\nabla}_\phi f_\pm = \left(\frac{\partial}{\partial\phi} - \left(\frac{im}{2}\right)(\mp 1 + \cos\theta) \right) f_\pm . \tag{3.29}$$

These equations are interpreted as the minimal coupling. In fact, the terms $(m/2)(\mp 1 + \cos\theta)$ are the ϕ component of the gauge potentials of the monopole field

$$m\Omega = -(m/2) \sin\theta \, d\theta \wedge d\phi , \quad \text{in } D_\pm . \tag{3.30}$$

We now discuss gauge transformations on $D_+ \cap D_-$. From (3.25) we have $e^{-im\phi} \Psi_+ = \Psi_-$ on $\pi^{-1}(D_+) \cap \pi^{-1}(D_-)$, and hence $e^{-im\phi} \sigma_+ = \sigma_-$ on $D_+ \cap D_-$. Suppose that a local section on $D_+ \cap D_-$ is expressed as $f_+ \sigma_+ = f_- \sigma_-$. Then we have the usual gauge transformation

$$f_- = e^{im\phi} f_+ , \quad \text{on } D_+ \cap D_- . \tag{3.31}$$

Applying (3.29) for $f_\pm$, we obtain

$$\overset{(-)}{\nabla}_\phi f_- = e^{im\phi} \overset{(+)}{\nabla}_\phi f_+ , \tag{3.32}$$

which means the tensor property of the covariant derivative.

In conclusion we are to link the local section f_+ to the Fourier coefficient $c_m^{(+)}$ of (3.1) and to get the covariant derivative (3.29) in an elementary manner.

As we pointed out in Sec. II E, $\pi^{-1}(D_+)$ is diffeomorphic with a product space $D_+ \times S^1$, so that the local coordinates in $\pi^{-1}(D_+)$ are (r,θ,ϕ,t). If we make a transformation $t = (\psi + \phi)/2$, we get the coordinates given in (2.37).

The connection form ω is then put into

$$\omega = \tfrac{1}{2}(2\,dt + (-1 + \cos\theta) d\phi) . \tag{3.33}$$

Hence the horizontal lift of $\partial/\partial\phi$, a vibrational vector on $\pi^{-1}(D_+)$, is given by

$$\left(\frac{\partial}{\partial\phi}\right)^* = \frac{\partial}{\partial\phi} - \frac{1}{2}(-1+\cos\theta)\frac{\partial}{\partial t}. \tag{3.34}$$

The Fourier term $c_m^{(+)}e^{imt}$ of (3.1) is of course a local equivariant function. The Ψ_+ in (3.25) is nothing but e^{imt} with $t=(\psi+\phi)/2$. Thus $c_m^{(+)}$ may be identified with f_+. Operating $c_m^{(+)}e^{imt}$ with $(\partial/\partial\phi)^*$, we obtain the covariant derivative of $c_m^{(+)}$;

$$\left(\frac{\partial}{\partial\phi} - \frac{im}{2}(-1+\cos\theta)\right)c_m^{(+)}. \tag{3.35}$$

The similar calculation can be done also for $c_m^{(-)}$.

IV. THE INTERNAL HAMILTONIAN OPERATOR

In this section we obtain the internal Hamiltonian operator acting on the cross sections in the complex line bundle L_m. The internal Hamiltonian operator should be reduced from the usual Hamiltonian operator for the three-body system.

A. The Laplacian

We start with the kinetic energy part of the Hamiltonian operator for the planar triatomic molecule. Recall that the configuration space Q_0 is endowed with the inner product (2.2). Then we note that the kinetic energy operator

$$-\frac{1}{2}\sum_{k=1}^{3}\frac{1}{m_k}\left(\frac{\partial}{\partial y_k}\right)^2, \tag{4.1}$$

where $\partial/\partial y_k$, $k=1,2,3$, denote the gradient operators, is $-\frac{1}{2}$ times the Laplacian Δ_0 defined by the metric K. Hence, in terms of the orthogonal coordinates $(B^j/N_0,\xi^k)$, $j=1,2$, $k=0,...,3$, the Laplacian Δ_0 takes the form

$$\Delta_0 = N_0^2\sum_{j=1}^{2}\left(\frac{\partial}{\partial B^j}\right)^2 + \sum_{k=0}^{3}\left(\frac{\partial}{\partial\xi^k}\right)^2. \tag{4.2}$$

Separating off the first term of the right-hand side, we have the Laplacian Δ on the center-of-mass system Q;

$$\Delta = \sum_{k=0}^{3}\left(\frac{\partial}{\partial\xi^k}\right)^2. \tag{4.3}$$

Here we have tacitly assumed that the linear momentum operator has zero eigenvalue.

We now wish to break up the Laplacian Δ into the vibrational and rotational parts. To this end, we first express $\partial/\partial\xi^k$ as a linear combination of F and V_k given in (2.24) and (2.26). As f_k are identified with $\partial/\partial\xi^k$, we obtain, on account of the inner product of f_k and V_j $(V_0=F)$, $j=0,...,3$,

$$\begin{aligned}
\frac{\partial}{\partial\xi^0} &= \|\xi\|^{-2}(-\xi^1F+\xi^0V_1-\xi^3V_2+\xi^2V_3),\\
\frac{\partial}{\partial\xi^1} &= \|\xi\|^{-2}(\xi^0F+\xi^1V_1+\xi^2V_2+\xi^3V_3),\\
\frac{\partial}{\partial\xi^2} &= \|\xi\|^{-2}(-\xi^3F-\xi^2V_1+\xi^1V_2+\xi^0V_3),\\
\frac{\partial}{\partial\xi^3} &= \|\xi\|^{-2}(\xi^2F-\xi^3V_1-\xi^0V_2+\xi^1V_3).
\end{aligned} \tag{4.4}$$

If (4.4) is substituted into (4.3), the Laplacian Δ will be expressed in terms of F and V_k, but this procedure is awkward. We consider, instead, the functional (twice the kinetic energy)

$$\begin{aligned}
\langle\Phi|-\Delta\Phi\rangle_Q &:= -\mu\int_Q\bar{\Phi}\sum_{k=0}^{3}\frac{\partial^2\Phi}{\partial(\xi^k)^2}d\Sigma,\\
&= \sum_{k=0}^{3}\left\langle\frac{\partial\Phi}{\partial\xi^k}\middle|\frac{\partial\Phi}{\partial\xi^k}\right\rangle_Q,
\end{aligned} \tag{4.5}$$

where Φ is a smooth complex-valued function on Q. Inserting (4.4) into (4.5), we obtain

$$\begin{aligned}
&\langle\Phi|-\Delta\Phi\rangle_Q\\
&= \langle\|\xi\|^{-2}F\Phi|F\Phi\rangle_Q + \sum_{k=1}^{3}\langle\|\xi\|^{-2}V_k\Phi|V_k\Phi\rangle_Q.
\end{aligned} \tag{4.6}$$

This equation means that the kinetic energy is separated into the rotational and vibrational energies; the first term in the right-hand side is rotational and the rest vibrational. However, this fact does not imply that rotation and vibration are not coupled as motions. The coupling appears rather in Eq. (2.31).

Theorem 9: The kinetic energy operator $-\Delta/2$ on the center-of-mass system Q is separated into the rotational energy and vibrational energy operators;

$$\Delta = \|\xi\|^{-2}F^2 + \sum_{k=1}^{3}V_k(\|\xi\|^{-2}V_k). \tag{4.7}$$

Proof: Integration of (4.6) by parts provides the Laplacian Δ in terms of F and V_k. It is here a great convenience to have the formulas

$$\langle V_j\Phi|\Phi\rangle_Q = \langle\Phi|-V_j\Phi\rangle_Q, \quad j=0,...,4, \tag{4.8}$$

where $V_0=F$. These formulas are easy to show. Applying (4.8) to (4.6), we obtain (4.7).

B. The Hamiltonian operator in L_m

We are now in a position to derive the internal Hamiltonian operator in L_m from the usual Hamiltonian operator in $L^2(Q)$. From the discussion in the last section, the Hamiltonian operator in $L^2(Q)$ is given by

$$H = -\tfrac{1}{2}\Delta + U, \tag{4.9}$$

where Δ is the Laplacian defined by (4.3) or (4.7), and U is a potential function depending on the internal coordinates only. Let Φ be a ρ_m-equivariant function of compact support on Q, and σ the corresponding cross section in L_m, $q_m^{\#}\sigma=\Phi$. Then the internal Hamiltonian operator H_m is defined through

$$\langle\Phi|H\Phi\rangle_Q = \int_M(\sigma|H_m\sigma)\,dM. \tag{4.10}$$

Theorem 10: For a planar triatomic molecule, the states of internal motions are described as cross sections in the complex line bundle L_m. The internal Hamiltonian operator H_m acting on the cross section is given, with respect to the volume element (3.23), by

$$H_m = -\frac{1}{2}4\|w\|\sum_{k=1}^{3}\nabla_k^2 + \frac{m^2}{2\|w\|} + U. \tag{4.11}$$

Proof: The left-hand side of (4.10) can be written out by using the results in the previous sections. Applying (4.7) to Φ and using (2.30), (3.6), (3.7), and (3.22), we obtain

$$\langle\Phi|\Delta\Phi\rangle_Q = \int_M \left(\sigma\Big|\Big(-\frac{m^2}{\|w\|}+4\|w\|\sum_{k=1}^{3}\nabla_k^2\Big)\sigma\right) dM, \tag{4.12}$$

where ∇_k, $k=1,2,3$, stand for the covariant derivation with respect to $\partial/\partial w^k$;

$$\nabla_k\sigma = q_m^{\#-1}\Big(\frac{\partial}{\partial w^k}\Big)^* q_m^{\#}\sigma$$
$$= q_m^{\#-1}\Big(\frac{1}{2}\|\xi\|^{-2}V_k\Big) q_m^{\#}\sigma. \tag{4.13}$$

For the potential function U, we have

$$\langle\Phi|U\Phi\rangle_Q = \int_M (\sigma|U\sigma)\, dM. \tag{4.14}$$

From (4.12) and (4.14), the internal Hamiltonian operator turns out to be (4.11). This ends the proof.

We make a remark on H_m. If we were allowed to take a volume element $4\|w\|dM = \mu\pi\, dw$ instead of dM, we would obtain an operator

$$-\frac{1}{2}\sum_{k=1}^{3}\nabla_k^2 + \frac{(m/2)^2}{2\|w\|^2} + \frac{U}{4\|w\|}, \tag{4.15}$$

which would describe states of a charged particle moving in the presence of Dirac's monopole field and the potential, the sum of the second and third terms. This operator with $U=-k/4$, $k>0$ a constant, is the Hamiltonian operator for a generalized hydrogen atom.[19]

C. Local description of H_m

We wish to express H_m in the internal coordinates (r,θ,ϕ). We start again with the Laplacian Δ. A straightforward calculation gives a description of the Laplacian (4.3) in the form

$$\Delta = 4r\frac{\partial^2}{\partial r^2} + 8\frac{\partial}{\partial r} + \frac{1}{r}\Delta_3,$$
$$\Delta_3 = 4\Big[\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}$$
$$\times\Big(\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial\psi^2} - 2\cos\theta\frac{\partial^2}{\partial\phi\,\partial\psi}\Big)\Big], \tag{4.16}$$

where Δ_3 is the standard Laplacian on the three-sphere S^3. If we take into account $F = 2\partial/\partial\psi$, the Laplacian (4.16) is broken up into the rotational and vibrational parts;

$$\Delta = \frac{4}{r}\frac{\partial^2}{\partial\psi^2} + 4r\frac{\partial^2}{\partial r^2} + 8\frac{\partial}{\partial r} + \frac{4}{r}\Big[\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta}$$
$$+\frac{1}{\sin^2\theta}\Big(\frac{\partial}{\partial\phi} - \cos\theta\frac{\partial}{\partial\psi}\Big)^2\Big]. \tag{4.17}$$

This expression corresponds to (4.7); the first term in the right-hand side is the rotational part, and the rest is the vibrational part. We note that the vibrational part can be expressed in terms of the horizontal lifts $(\partial/\partial r)^*$, $(\partial/\partial\theta)^*$, and $(\partial/\partial\phi)^*$ on account of (3.26).

We here take the local ρ_m-equivariant function Ψ_+ and the corresponding local cross section σ_+ in L_m, as we considered in Sec. III C. Let $f_+\sigma_+$ be any local section. Applying (4.17) to $q_m^{\#} f_+\sigma_+$, and using (3.27), we obtain

$$\Delta q_m^{\#} f_+\sigma_+$$
$$= \Big[-\frac{m^2}{r} + 4r\frac{\partial^2}{\partial r^2} + 8\frac{\partial}{\partial r} + \frac{4}{r}\Big(\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta}\Big)$$
$$+\frac{4}{r\sin^2\theta}\Big(\frac{\partial}{\partial\phi} - i\frac{m}{2}(-1+\cos\theta)\Big)^2\Big] f_+ q_m^{\#}\sigma_+. \tag{4.18}$$

On supposing that f_+ has its support in D_+, the internal Hamiltonian operator H_m^+, locally defined on D_+, are obtained through (4.10). A straightforward calculation together with (4.18) gives

$$\langle q_m^{\#}(f_+\sigma_+)|Hq_m^{\#}(f_+\sigma_+)\rangle_Q = \int_M \overline{f_+}\, H_m^+ f_+\, dM, \tag{4.19}$$

where

$$H_m^+ = -\frac{1}{2}\Big[4r\frac{\partial^2}{\partial r^2} + 8\frac{\partial}{\partial r} + \frac{4}{r}\Big(\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta}\Big)$$
$$+\frac{4}{r\sin^2\theta}\Big(\frac{\partial}{\partial\phi} - i\frac{m}{2}(-1+\cos\theta)\Big)^2\Big]$$
$$+\frac{m^2}{2r} + U. \tag{4.20}$$

The H_m^+ is a local description of H_m given by (4.11). The first term $-\frac{1}{2}[\cdots]$ is the kinetic energy operator coupled with the monopole field (3.30), and the second term is the centrifugal potential, where r is the moment of inertia. The coupling and the centrifugal potential disappear at the same time when the angular momentum is zero ($m=0$).

The internal Hamiltonian operator H_m^-, locally defined on D_-, is obtained in the same manner. The gauge transformation in the intersection $D_+\cap D_-$ is easy to show. Let $f_+\sigma_+ = f_-\sigma_-$ on $D_+\cap D_-$. Then from the identity

$$\langle q_m^{\#}(f_+\sigma_+)|Hq_m^{\#}(f_+\sigma_+)\rangle_Q$$
$$= \langle q_m^{\#}(f_-\sigma_-)|Hq_m^{\#}(f_-\sigma_-)\rangle_Q,$$

one obtains

$$e^{im\phi}H_m^+ e^{-im\phi} = H_m^-. \tag{4.21}$$

If $m=0$, the complex line bundle becomes trivial, and the operator $(1/4r)H_m^+ = (1/4r)H_m^-$ in $\mathbb{R}^3$ is a standard Hamiltonian operator with a potential $U/4r$. This is a special case of (4.15). The reduction procedure for obtaining H_m with $m=0$ was already used in principle in Ref. 20.

In conclusion, we remark that in his paper[21] Smith discussed the planar three-body problem in the limit of weak interaction ($U=0$) by using the SO(4) action on $Q\simeq\mathbb{R}^4$, but the internal motion was out of his scope.

D. The outlook

Though we have concentrated on the planar triatomic molecule, our method is easy to generalize to the multiatomic molecule. In fact, recall that the center-of-mass system is made into a principal fiber bundle with the rotation group as the structure group, then quantum mechanics of the internal

molecular motion should be set up on the complex vector bundle associated with the principal fiber bundle with respect to a unitary irreducible representation of the rotation group. The internal Hamiltonian operator can be obtained in the same manner as we have used in this section. In Ref. 6, we have already treated the internal Hamiltonian operator for the n-body system in $\mathbb{R}^3$ under the condition that the angular momentum operator has zero eigenvalue, without reference to the geometry of complex line bundles. Our problem is then reduced to working out the topology of the internal space on which the complex vector bundle is constructed. The quantum three-body problem[7] in $\mathbb{R}^3$ is reinvestigated from our viewpoint.[22]

ACKNOWLEDGMENTS

The author would like to thank Dr. A. Tachibana at Kyoto University for helpful discussions.

This work was supported by a Grant-in-Aid for scientific research from the Ministry of Education of Japan.

[1]B. T. Sutcliffe, in *Quantum Dynamics of Molecules*, edited by R. G. Wooley (Plenum, New York, 1980).
[2]A. Guichardet, Ann. Inst. H. Poincaré **40**, 329 (1984).
[3]T. Iwai, "A geometrics setting for classical molecular dynamics," Technical Report No. 85009, Department of Applied Mathematics and Physics, Kyoto University.
[4]C. Nash and S. Sen, *Topology and Geometry for Physicists* (Academic, London, 1983).
[5]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).
[6]A. Tachibana and T. Iwai, Phys. Rev. A **33**, 2262 (1986).
[7]W. Zickendraht, Ann. Phys. (NY) **35**, 18 (1965).
[8]W. Greub and H.-R. Petry, J. Math. Phys. **16**, 1347 (1975).
[9]W. J. Satzer, Jr., Indiana Univ. Math. J. **26**, 951 (1977).
[10]K. R. Meyer, in *Dynamical Systems*, edited by M. M. Peixoto (Academic, New York, 1973).
[11]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[12]Y. Matsushima, *Differentiable Manifolds* (Dekker, New York, 1972).
[13]E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
[14]T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[15]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[16]T. Iwai and Y. Uwano, J. Math. Phys. **27**, 1523 (1986).
[17]M. Kummer, Indiana Univ. Math. J. **30**, 281 (1981).
[18]B. Kostant, in *Lectures in Modern Analysis and Applications III, Lecture Notes in Mathematics*, Vol. 170 (Springer, Berlin, 1970).
[19]T. Iwai and Y. Uwano, J. Math. Phys. **27**, 1523 (1986).
[20]T. Iwai, J. Math. Phys. **23**, 1093 (1982).
[21]F. T. Smith, J. Math. Phys. **3**, 735 (1962).
[22]T. Iwai, "A geometric setting for internal motions of the quantum three-body system," submitted to J. Math. Phys.

Analytical estimation of critical parameter values for bound states of screened Coulomb potentials

Gerald Rosen
Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

(Received 24 June 1986; accepted for publication 10 December 1986)

As a consequence of the virial theorem and Hellmann–Feynman relation, the ratio of the kinetic energy to the derivative of the total energy, $\Lambda(\lambda) \equiv T(\lambda)/E'(\lambda)$, is stationary and equal to the critical parameter value λ_c at $\lambda = \lambda_c$: $E(\lambda_c) = 0 \Rightarrow \Lambda'(\lambda_c) = 0$ and $\Lambda(\lambda_c) = \lambda_c$. The cubic approximation to the latter equation yields positive roots within 3.02% and 0.33% of the exact λ_c values for the exponential and exponential-cosine screening functions, respectively. An alternative estimation formula for λ_c is also presented and shown to give a value within 0.19% of the exact λ_c value for the exponential screening function.

I. INTRODUCTION

Hamiltonians associated with attractive screened Coulomb potentials arise in many areas of quantum physics and have been of considerable recent interest.[1] By employing convenient physical units,[2] such Hamiltonians are expressible as

$$H = -\tfrac{1}{2}\nabla^2 - (1/r)f(\lambda r), \tag{1}$$

where the screening function $f(\lambda r)$ and reciprocal screening length λ are fixed by the normalization conditions at $r \equiv (x^2 + y^2 + z^2)^{1/2} = 0$,

$$f(0) = 1, \quad f'(0) = -1, \tag{2}$$

in which $f'(\xi) \equiv df(\xi)/d\xi$. Particular physical importance is attached to atomic, nuclear, and solid-state systems that feature the exponential screening function (ESF),

$$f(\lambda r) = e^{-\lambda r} \quad \text{(ESF)} \tag{3}$$

and the exponential-cosine screening function (ECSF),

$$f(\lambda r) = e^{-\lambda r}\cos \lambda r \quad \text{(ECSF)}. \tag{4}$$

Generally, the Hamiltonian (1) admits one or more bound (square-integrable, negative energy) eigenstates for $0 \leq \lambda < \lambda_c$, i.e., non-negative values of the parameter λ less than a certain critical value λ_c. The latter critical parameter value has been determined accurately by numerical computation in the case of (3) and (4) as[3,4]

$$\lambda_c = 1.1906 \quad \text{(ESF)}, \tag{5}$$

$$\lambda_c = 0.7205 \quad \text{(ECSF)}. \tag{6}$$

For values of λ greater than (5) or (6), the Hamiltonian (1) has a strictly positive energy spectrum with no bound state for either screening function (3) or (4), respectively.

The analytical estimation of critical parameter values for screened Coulomb potentials has posed a mathematical accuracy problem when approached via traditional methods.[1] For example, a basic (one-parameter) Rayleigh–Ritz variational procedure for the estimation of (5) or (6) fails to produce values with better than 10% accuracy, as discussed in Appendix A. Although more complicated analytical procedures have produced closer estimates for λ_c values,[1] the comparisons with exact numerical values [such as (5) and (6)] shows that the latter analytical methods do not generally yield approximate values with an accuracy better than about 2% or 3%, notwithstanding considerable technical complexity.[5]

The present communication reports a new practical approach to the analytical estimation of critical parameter values for screened Coulomb potentials. By introducing and performing an approximation evaluation of the function $\Lambda(\lambda)$ defined below in (12), the estimates (22) and (23) are obtained for (5) and (6) in the cases of the screening functions (3) and (4). On the basis of the accuracy evident in the latter examples, this logically simple and easily applied procedure can generally be expected to yield estimates for critical parameter values within 3% of the exact values.

II. PROPERTIES OF THE FUNCTION $\Lambda(\lambda)$

Let $E(\lambda) \equiv \langle H \rangle = T(\lambda) - \langle r^{-1} f(\lambda r) \rangle$ denote the ground state energy eigenvalue for the Hamiltonian (1), where $T(\lambda) \equiv \langle -\tfrac{1}{2}\nabla^2 \rangle$ is the ground state kinetic energy expectation value (depending on λ through the energy eigenstate's dependence on the parameter). As a consequence of the virial theorem $\langle [\mathbf{x}\cdot\nabla, H] \rangle = 0$, one has

$$-2T(\lambda) + \langle (r^{-1} f(\lambda r) - \lambda f'(\lambda r)) \rangle = 0, \tag{7}$$

while the Hellmann–Feynman relation takes the form

$$\frac{dE(\lambda)}{d\lambda} \equiv E'(\lambda) = -\langle f'(\lambda r) \rangle. \tag{8}$$

Hence by combining (7), (8), and the definition of $E(\lambda)$ (Ref. 6), one finds that the energy eigenvalue and kinetic energy expectation value are simply related by a Legendre transform:

$$\lambda E'(\lambda) - E(\lambda) = T(\lambda). \tag{9}$$

From the first normalization condition in (2) it follows that the ground state energy eigenvalue of (1) is the Schrödinger hydrogenic value for $\lambda = 0$, i.e., $E(0) = -\tfrac{1}{2}$. Moreover, the second normalization condition in (2), in combination with (8), implies that $E'(0) = 1$. Hence, in general, one has

$$E(\lambda) = -\tfrac{1}{2} + \lambda - O(\lambda^2) \tag{10}$$

and

$$T(\lambda) = \tfrac{1}{2} - O(\lambda^2) \tag{11}$$

by virtue of (9). Since λ_c is defined as the smallest positive

 0022-2488/87/040975-03$02.50

value for which $E(\lambda_c)=0$, it follows from (9) that the ratio

$$\Lambda(\lambda)\equiv T(\lambda)/E'(\lambda)=\lambda-[E(\lambda)/E'(\lambda)] \tag{12}$$

is stationary and equal to λ_c at $\lambda=\lambda_c$:

$$\Lambda'(\lambda_c)=\frac{d}{d\lambda}\left[\lambda-\frac{E(\lambda)}{E'(\lambda)}\right]\Big|_{\lambda=\lambda_c}=\frac{E(\lambda_c)E''(\lambda_c)}{[E'(\lambda_c)]^2}=0, \tag{13}$$

$$\Lambda(\lambda_c)=\lambda_c. \tag{14}$$

In the cases of (3), (4) and similar forms for the screening function, the second derivative of (12) at $\lambda=\lambda_c$, $\Lambda''(\lambda_c)=E''(\lambda_c)/E'(\lambda_c)$, is negative because $E''(\lambda_c)<0$, and thus $\Lambda(\lambda)$ is a maximum at $\lambda=\lambda_c$.

By expanding the final member of (12) in a Maclaurin series about $\lambda=0$, one obtains

$$\begin{aligned}&\Lambda(0)=\tfrac{1}{2},\quad \Lambda'(0)=-\tfrac{1}{2}E''(0),\\ &\Lambda''(0)=E''(0)+[E''(0)]^2-\tfrac{1}{2}E'''(0),\\ &\Lambda'''(0)=2E'''(0)-3[E''(0)]^2[1+E''(0)]\\ &\qquad\qquad+3E''(0)E'''(0)-\tfrac{1}{2}E''''(0),\end{aligned} \tag{15}$$

where use has been made of the conditions $E(0)=-\frac{1}{2}$ and $E'(0)=1$ which derive from (2) and are manifest in (10). The higher-order derivatives of $E(\lambda)$ at $\lambda=0$ appearing in (15) are obtainable directly from (8) or by exact perturbation-theoretic analysis for a prescribed screening function. In the cases of (3) and (4), one readily evaluates $\Lambda'(0)$, $\Lambda''(0)$, and $\Lambda'''(0)$ in (15) and thereby obtains

$$\Lambda(\lambda)=\hat{\Lambda}(\lambda)+O(\lambda^4), \tag{16}$$

where to order λ^3,

$$\hat{\Lambda}(\lambda)=\tfrac{1}{2}+\tfrac{3}{4}\lambda-\tfrac{3}{8}\lambda^2+\tfrac{3}{16}\lambda^3 \quad \text{(ESF)}, \tag{17}$$

$$\hat{\Lambda}(\lambda)=\tfrac{1}{2}+\tfrac{3}{2}\lambda^2-\tfrac{3}{2}\lambda^3 \quad \text{(ECSF)}, \tag{18}$$

for the two cases, respectively.

III. ESTIMATION OF λ_c BY THE CUBIC APPROXIMATION FORMULA $\hat{\Lambda}(\hat{\lambda}_c)=\hat{\lambda}_c$

As expressed by (13) and (14), the stationary character of $\Lambda(\lambda)$ about $\lambda=\lambda_c$ suggests that the cubic approximation to $\Lambda(\lambda)$, introduced as $\hat{\Lambda}(\lambda)$ in (16), may yield an accurate estimate for λ_c as the relevant root of the equation

$$\hat{\Lambda}(\hat{\lambda}_c)=\hat{\lambda}_c. \tag{19}$$

Indeed, for the screening functions (3) and (4), substitution of the cubic approximations (17) and (18) into (19) produces

$$\tfrac{1}{2}-\tfrac{1}{4}\hat{\lambda}_c-\tfrac{3}{8}\hat{\lambda}_c^2+\tfrac{3}{16}\hat{\lambda}_c^3=0 \quad \text{(ESF)}, \tag{20}$$

$$\tfrac{1}{2}-\hat{\lambda}_c+\tfrac{3}{2}\hat{\lambda}_c^2-\tfrac{3}{2}\hat{\lambda}_c^3=0 \quad \text{(ECSF)}, \tag{21}$$

with the relevant (smallest positive) roots

$$\hat{\lambda}_c=1.1547 \quad \text{(ESF)}, \tag{22}$$

$$\hat{\lambda}_c=0.7181 \quad \text{(ECSF)}. \tag{23}$$

The ESF critical parameter estimate (22) is 3.02% below the exact value (5), while the ECSF critical parameter estimate (23) is only 0.33% below the exact value (6). Hence for purposes that do not require high accuracy, the analytical estimation of critical parameter values can be performed by employing formula (19), the cubic approximation to (14). An alternative (more specialized) analytical approximation formula that works very well for estimation of the ESF critical parameter value is derived and applied in Appendix B.

ACKNOWLEDGMENT

The author would like to thank P. Y. Prath for assisting with numerical evaluations.

APPENDIX A: APPROXIMATION FOR λ_c BY A RAYLEIGH–RITZ VARIATIONAL PROCEDURE

If one computes all expectation values with respect to the Rayleigh–Ritz trial form for the normalized ground state

$$\psi=\pi^{-1/2}\kappa^{3/2}e^{-\kappa r}, \tag{A1}$$

in which κ is a disposable parameter, the virial theorem (7), the Hellmann–Feynman equation (8), and the implied equations [(9)–(15)] remain valid as exact relations. In fact, the Rayleigh–Ritz condition for $\kappa=\kappa(\lambda)$,

$$\frac{\partial\tilde{E}}{\partial\kappa}=0, \tag{A2}$$

with

$$\tilde{E}=\tilde{E}(\lambda,\kappa)\equiv\int\psi^*H\psi\,d^3x \tag{A3}$$

is identical to the virial statement (7), because both equations pertain commonly to the invariance of the energy expectation value under the spatial dilatation transformations $r\to(1+\epsilon)r$. With $\tilde{\lambda}_c$ defined as the smallest positive root of the equation

$$\tilde{E}(\tilde{\lambda}_c,\kappa(\tilde{\lambda}_c))=0, \tag{A4}$$

one finds

$$\tilde{\lambda}_c=1.0000 \quad \text{(ESF)}, \tag{A5}$$

$$\tilde{\lambda}_c=0.6437 \quad \text{(ECSF)} \tag{A6}$$

for the screening functions (3) and (4). The crude approximations shown in (A5) and (A6) are below the corresponding exact values (5) and (6) by about 16.0% and 10.7%, respectively. Here the loss of Rayleigh–Ritz accuracy occurs in part when the root of (A4) is extracted, the slope function $\partial\tilde{E}/\partial\lambda\,|_{\lambda=\tilde{\lambda}_c}\equiv\tilde{E}'(\tilde{\lambda}_c)$ being a fraction of unity in either case, viz.

$$\tilde{E}'(\tilde{\lambda}_c)=0.1250 \quad \text{(ESF)}, \tag{A7}$$

$$\tilde{E}'(\tilde{\lambda}_c)=0.4071 \quad \text{(ECSF)}. \tag{A8}$$

APPENDIX B: ALTERNATIVE ESTIMATION FORMULA FOR λ_c

For either screening function (3) or (4), the upper bound defined by

$$M\equiv\max_{0\le\lambda\le\lambda_c}\{\lambda^{-1}\ln([E'(\lambda)]^{-1})\} \tag{B1}$$

is expressible as

$$M=\lambda_c^{-1}\ln([E'(\lambda_c)]^{-1}) \quad (<2) \tag{B2}$$

as shown by straightforward consideration of the forms of

$E'(\lambda)$ in the two cases.[7] The differential inequality associated with (B1),

$$M \geq \lambda^{-1} \ln([E'(\lambda)]^{-1}) \quad (0 \leq \lambda \leq \lambda_c) \tag{B3}$$

can be rearranged and integrated subject to $E(0) = -\frac{1}{2}$ to yield

$$E(\lambda) \geq -\tfrac{1}{2} + M^{-1}(1 - e^{-M\lambda}). \tag{B4}$$

By setting $\lambda = \lambda_c$ in (B4) and recalling that $E(\lambda_c) = 0$, one finds

$$\lambda_c \leq \tfrac{1}{2}[1 - E'(\lambda_c)]^{-1} \ln([E'(\lambda_c)]^{-1}), \tag{B5}$$

where M has been eliminated by employing (B2). If one replaces $E'(\lambda_c)$ by the Rayleigh–Ritz approximation $\tilde{E}'(\tilde{\lambda}_c)$, the right side of (B5) produces the quantity

$$\tilde{\tilde{\lambda}}_c \equiv \tfrac{1}{2}[1 - \tilde{E}'(\tilde{\lambda}_c)]^{-1} \ln([\tilde{E}'(\tilde{\lambda}_c)]^{-1}), \tag{B6}$$

which can be expected to approximate λ_c. Indeed, for the ESF and ECSF Rayleigh–Ritz values (A7) and (A8), formula (B6) yields

$$\tilde{\tilde{\lambda}}_c = 1.1883 \quad \text{(ESF)}, \tag{B7}$$

$$\tilde{\tilde{\lambda}}_c = 0.7579 \quad \text{(ECSF)}. \tag{B8}$$

The ESF critical parameter estimate (B7) is only 0.19% below the exact value (5), while the ECSF critical parameter estimate (B8) is 5.19% above the exact value (6).[7]

[1] J. Piepenbrink, J. Math. Phys. **13**, 1825 (1972); P. C. Hemmer, *ibid.* **14**, 1140 (1973); R. Dutt, Phys. Lett. A **77**, 229 (1980); J. Killingbeck and S. Galicia, J. Phys. A **13**, 3419 (1980); V. L. Eletsky, V. S. Popov, and V. M. Weinberg, Phys. Lett. A **84**, 235 (1981); C. S. Lai, Phys. Rev. A **23**, 455 (1981); **26**, 2245 (1982); A. Ray and P. P. Ray, Phys. Lett. A **83**, 383 (1981); R. Dutt and Y. P. Varshni, J. Math. Phys. **25**, 2563 (1984), and works cited therein.

[2] Here physical units are prescribed such that $\hbar^2 = m$ and $Ze^2 = 1$ for an atom that has nuclear charge $+Ze$; hence the screening length λ^{-1} is measured in units $\hbar^2/mZe^2$ and the energy E in units $mZ^2e^4/\hbar^2$.

[3] R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **19**, 819 (1978).

[4] D. Singh and Y. P. Varshni, Phys. Rev. A **28**, 2606 (1983).

[5] However, by applying the necessary condition on potentials for bound states [G. Rosen, Phys. Rev. Lett. **49**, 1885 (1982)] $\int |V|^{3/2}\, d^3x > (\frac{1}{2})^{3/2}\pi^2/4$ to the ESF form with $V = -r^{-1}e^{-\lambda r}$, one obtains $\lambda < 1.2138$, an immediate upper bound only 1.95% greater than (5).

[6] Energy estimations based on the virial and Hellmann–Feynmann relations have been made previously; see, e.g., G. Rosen, Phys. Rev. A **20**, 1287 (1979), and works cited therein.

[7] Because $E''(0) = 0$ for the ECSF shown in (4), the curly-bracketed quantity in (B1) vanishes at $\lambda = 0$ for the ECSF case. On the other hand, $E''(0) = -\frac{1}{2}$ for the ESF shown in (3), and the curly-bracketed quantity in (B1) remains fairly close to M over the range $0 \leq \lambda \leq \lambda_c$. This is why (B7) is much closer to (5) than (B8) is to (6).

Unique shape of uniformly polarizable dielectrics

K. R. Brownstein
Department of Physics and Astronomy, University of Maine, Orono, Maine 04469

(Received 10 June 1986; accepted for publication 18 November 1986)

A homogeneous isotropic dielectric body, placed in a uniform external electric field, will acquire a polarization $\mathbf{P}(\mathbf{r})$. If the body is an ellipsoid then $\mathbf{P}$ is uniform. A proof is presented showing that only ellipsoidal bodies have this property.

I. INTRODUCTION

Let $\mathscr{R}$ be a region of space filled with a homogeneous isotropic dielectric. When placed in a (previously) uniform external electric field $\mathbf{E}_0$, the region will acquire a polarization $\mathbf{P}(\mathbf{r})$ and an induced electric field $\mathbf{E}_i(\mathbf{r})$. If $\mathscr{R}$ is a sphere then $\mathbf{P}$ is uniform and parallel to $\mathbf{E}_0$. If $\mathscr{R}$ is an ellipsoid then $\mathbf{P}$ is also uniform; however $\mathbf{P}$ is parallel to $\mathbf{E}_0$ only if $\mathbf{E}_0$ lies along one of the principal axes of the ellipsoid.[1]

Suppose that $\mathscr{R}$ has the property that it acquires a uniform polarization $\mathbf{P}$ under the influence of an arbitrary (previously) uniform external electric field $\mathbf{E}_0$. We know that all ellipsoids (with spheres as a special case) have this property. Apparently it is a "folk theorem" that $\mathscr{R}$ must be an ellipsoid.[2] A recent step forward toward proving this theorem is due to Zapolsky.[2] He considers a family of regions $\mathscr{R}$ whose bounding surfaces are given by

$$(x/a)^k + (y/b)^k + (z/c)^k = \text{const},$$

$k = 2,4,6\ldots$. Zapolsky then shows that, for this family, $\mathscr{R}$ has the desired property only for $k = 2$ (i.e., if $\mathscr{R}$ is an ellipsoid).

The purpose of this article is to supply the proof that $\mathscr{R}$ must be an ellipsoid. To accomplish this, we first define a tensor $\boldsymbol{\alpha}$ which essentially relates $\mathbf{P}$ to $\mathbf{E}_i$. We then show that, for a given $\boldsymbol{\alpha}$, the shape of $\mathscr{R}$ is unique. Since an ellipsoid is known to be one possible such shape, it follows that $\mathscr{R}$ must be an ellipsoid.

II. UNIQUENESS PROOF

Consider a dielectric object having a uniform isotropic permittivity and occupying a region of space $\mathscr{R}$. We assume that this object has the property of acquiring a uniform polarization $\mathbf{P}$ under the influence of any externally applied uniform electric field $\mathbf{E}_0$. There exist three mutually perpendicular directions such that $\mathbf{P}$ is parallel to $\mathbf{E}_0$; we take these to be the x, y, and z axes.

The total electric field $\mathbf{E}$ is the sum of the applied field $\mathbf{E}_0$ and the induced electric field $\mathbf{E}_i$,

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_i. \tag{1}$$

The induced electric field $\mathbf{E}_i$, which is uniform within $\mathscr{R}$, is the electrostatic field due to the volume charge distribution

$$\rho_i = -\nabla\cdot\mathbf{P}. \tag{2}$$

Since $\mathbf{P}$ is uniform within $\mathscr{R}$ and vanishes outside $\mathscr{R}$, this volume charge distribution is equivalent to the surface charge distribution

$$\sigma_i = \mathbf{P}\cdot\mathbf{n}, \tag{3}$$

where $\mathbf{n}$ is the unit outward normal to the surface of $\mathscr{R}$.

Suppose $\mathbf{E}_0 = \mathbf{E}_{01}$ is in the positive x direction. Then $\mathbf{P} = \mathbf{P}_1$ is in the positive x direction and $\mathbf{E}_i = \mathbf{E}_{i1}$ is in the negative x direction (within $\mathscr{R}$). Let $V_{i1}(\mathbf{r})$ be the potential associated with this induced electric field

$$\mathbf{E}_{i1} = -\nabla V_{i1}. \tag{4}$$

On the one hand, we can calculate V_{i1} from the induced charge distribution $\sigma_{i1} = \mathbf{P}_1\cdot\mathbf{n} = P_1\mathbf{i}\cdot\mathbf{n}$,

$$V_{i1}(\mathbf{r}) = \int \frac{\sigma_{i1}(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|}\,dA_s = P_1\int \frac{\mathbf{i}\cdot d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|}, \tag{5}$$

where $d\mathbf{A}_s = \mathbf{n}\,dA_s$ is the vectorial surface area element. On the other hand, since $\mathbf{E}_{i1}$ is uniform and in the negative x direction, V_{i1} must be a linear increasing function of x (within $\mathscr{R}$),

$$V_{i1} = P_1(\alpha_1 x + \beta_1), \quad \alpha_1 \text{ and } \beta_1 \text{ constants}. \tag{6}$$

Here the positive factor P_1 has been inserted for convenience and $\alpha_1 > 0$. From Eqs. (5) and (6) we have (dividing by P_1 and multiplying by $\mathbf{i}$)

$$\int \frac{\mathbf{i}\mathbf{i}\cdot d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|} = (\alpha_1 x + \beta_1)\mathbf{i} \quad (\mathbf{r}\in\mathscr{R}). \tag{7}$$

Similar results hold if $\mathbf{E}_0$ is in the positive y or z directions. Adding these three relations and making use of the fact that $\mathbf{ii} + \mathbf{jj} + \mathbf{kk} = \mathbf{I}$ (the unit tensor), one obtains

$$\int \frac{d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|} = \boldsymbol{\alpha}\cdot\mathbf{r} + \boldsymbol{\beta} \quad (\mathbf{r}\in\mathscr{R}). \tag{8}$$

Here

$$\boldsymbol{\alpha} = \alpha_1\mathbf{ii} + \alpha_2\mathbf{jj} + \alpha_3\mathbf{kk}, \tag{9a}$$

$$\boldsymbol{\beta} = \beta_1\mathbf{i} + \beta_2\mathbf{j} + \beta_3\mathbf{k}. \tag{9b}$$

Note that both the positive definite symmetric tensor $\boldsymbol{\alpha}$ and the vector $\boldsymbol{\beta}$ are independent of $\mathbf{r}$. In Eq. (8) the $\boldsymbol{\beta}\cdot\mathbf{r}$ term may be transformed away by shifting the origin: $\mathbf{r}' = \mathbf{r} + \boldsymbol{\alpha}^{-1}\cdot\boldsymbol{\beta}$. Dropping the prime,

$$\int \frac{d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|} = \boldsymbol{\alpha}\cdot\mathbf{r} \quad (\mathbf{r}\in\mathscr{R}). \tag{10}$$

We shall refer to this new origin as the "center C." Using Gauss' theorem the left-hand side of Eq. (10) may be transformed into a volume integral,

$$\int \frac{\mathbf{r}-\mathbf{s}}{|\mathbf{r}-\mathbf{s}|^3}\,d\tau_s = \boldsymbol{\alpha}\cdot\mathbf{r} \quad (\mathbf{r}\in\mathscr{R}), \tag{11}$$

where $d\tau_s$ is the volume element and the integration is over the region $\mathscr{R}$.

 0022-2488/87/040978-03$02.50

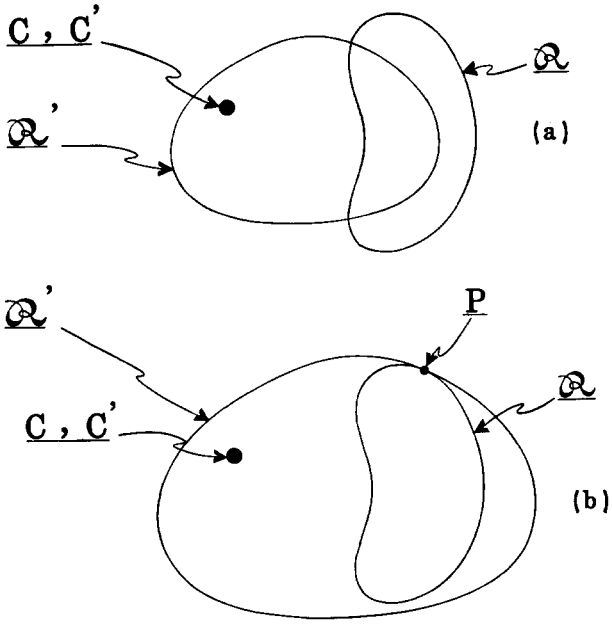


FIG. 1. (a) Region $\mathscr{R}$ is positioned so that its center C coincides with the center C' of region $\mathscr{R}'$. (b) Region $\mathscr{R}'$ is stretched uniformly about its center C until $\mathscr{R}'$ just contains $\mathscr{R}$. Here P is the common tangent point to the surfaces of the two regions.

Two interesting facts may be immediately deduced from Eq. (11).

(i) Integrate both sides $\int(\ )d\tau_r$ over the region $\mathscr{R}$. The left-hand side vanishes by the antisymmetry in $(\mathbf{r},\mathbf{s})$. Thus

$$\int \mathbf{r}\, d\tau_r = 0. \tag{12}$$

This shows that the center C must lie at the centroid of the region $\mathscr{R}$. It does not follow that C must be in $\mathscr{R}$ since $\mathscr{R}$ might be nonconvex or even multiply connected.

(ii) Take the divergence (with respect to $\mathbf{r}$). This gives

$$\int 4\pi\delta(\mathbf{r}-\mathbf{s})d\tau_s = \mathrm{Tr}(\alpha) \quad (\mathbf{r}\in\mathscr{R}).$$

Now the quantities $(\alpha_1,\alpha_2,\alpha_3)$ are simply the "depolarization factors" (N_x,N_y,N_z); see Eq. (6). Thus one recovers the well-known sum rule[3]

$$N_x + N_y + N_z = 4\pi. \tag{13}$$

Equation (11) may be interpreted as the electrostatic field

$$\mathscr{E}(\mathbf{r}) = \alpha\cdot\mathbf{r}, \tag{14}$$

due to the volume charge distribution

$$\rho(\mathbf{r}) = \begin{cases} 1 & (\mathbf{r}\in\mathscr{R}), \\ 0 & (\mathbf{r}\notin\mathscr{R}). \end{cases} \tag{15}$$

The potential $\mathscr{V}(\mathbf{r})$ associated with the electric field $\mathscr{E}(\mathbf{r})$ is, from (14),

$$\mathscr{V}(\mathbf{r}) = -\tfrac{1}{2}\mathbf{r}\cdot\alpha\cdot\mathbf{r} + \mathscr{V}_0 \quad (\mathbf{r}\in\mathscr{R}), \tag{16}$$

where $\mathscr{V}_0$ is a constant chosen so as to ensure that the reference potential is $\mathscr{V}(\infty) = 0$ [note that $\mathscr{V}_0$ has the significance of $\mathscr{V}(\mathbf{r}=0)$ only if C lies within $\mathscr{R}$]. The potential $\mathscr{V}(\mathbf{r})$ may also be expressed in terms of its sources (15),

$$\mathscr{V}(\mathbf{r}) = \int \frac{1}{|\mathbf{r}-\mathbf{s}|}\, d\tau_s. \tag{17}$$

We have shown that the region $\mathscr{R}$ must be of such a shape that if it is uniformly charged [according to Eq. (15)], the resulting electrostatic field $\mathscr{E}(\mathbf{r})$ will be a linear function of $\mathbf{r}$ (for $\mathbf{r}\in\mathscr{R}$) according to Eq. (14). We shall now show that this property is preserved (with the same C and α) under a uniform stretching (or compressing) of the region $\mathscr{R}$ about the center C. It is sufficient to consider an infinitesimal such stretching; a finite one can always be constructed by integration of the infinitesimal case. Under such an infinitesimal uniform stretching the position vector $\mathbf{s}$ of a point on the surface of $\mathscr{R}$ changes to $\mathbf{s}'$:

$$\mathbf{s}' = (1+\lambda)\mathbf{s} \quad (\lambda \text{ infinitesimal}). \tag{18}$$

The new potential $\mathscr{V}'(\mathbf{r})$ is, from Eq. (17),

$$\mathscr{V}'(\mathbf{r}) = \mathscr{V}(\mathbf{r}) + \lambda\int \frac{\mathbf{s}\cdot d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|}. \tag{19}$$

The last term is the contribution due to the additional volume which consists of a "shell" of thickness $\lambda\mathbf{s}\cdot\mathbf{n}$. Using Gauss' theorem and Eqs. (10), (16), and (17) one has

$$\begin{aligned} \mathscr{V}'(\mathbf{r}) &= \mathscr{V}(\mathbf{r}) - \lambda\int \frac{(\mathbf{r}-\mathbf{s})\cdot d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|} + \lambda\mathbf{r}\cdot\int \frac{d\mathbf{A}_s}{|\mathbf{r}-\mathbf{s}|}, \\ \mathscr{V}'(\mathbf{r}) &= \mathscr{V}(\mathbf{r}) + 2\lambda\int \frac{d\tau_s}{|\mathbf{r}-\mathbf{s}|} + \lambda\mathbf{r}\cdot\alpha\cdot\mathbf{r}, \\ \mathscr{V}'(\mathbf{r}) &= \mathscr{V}(\mathbf{r}) + 2\lambda(-\tfrac{1}{2}\mathbf{r}\cdot\alpha\cdot\mathbf{r} + \mathscr{V}_0) + \lambda\mathbf{r}\cdot\alpha\cdot\mathbf{r}, \\ \mathscr{V}'(\mathbf{r}) &= \mathscr{V}(\mathbf{r}) + 2\lambda\mathscr{V}_0. \end{aligned} \tag{20}$$

Since $\mathscr{V}(\mathbf{r})$ changes by only an additive constant (for $\mathbf{r}\in\mathscr{R}$), the new electric field $\mathscr{E}'(\mathbf{r})$ is again given by Eq. (14) with the same C and α.

We shall now show that, given the tensor α, the shape of the region $\mathscr{R}$ is uniquely determined. For a given tensor α we assume the existence of at least one convex region $\mathscr{R}'$ with the property that if it is uniformly charged with charge density $\rho'(\mathbf{r})$ according to Eq. (15), the associated electric field $\mathscr{E}'(\mathbf{r})$ will be given by Eq. (14) for $\mathbf{r}\in\mathscr{R}'$. (The known ellipsoidal solution is an example of such a convex region $\mathscr{R}'$.) Referring to Fig. 1(a), we position the original region $\mathscr{R}$ so that its center C coincides with the center C' of $\mathscr{R}'$ (although C' must lie within $\mathscr{R}'$, we allow the possibility of C not lying within $\mathscr{R}$). Now we stretch (or compress) the region $\mathscr{R}'$ uniformly about C' until $\mathscr{R}'$ just contains $\mathscr{R}$ as shown in Fig. 1(b). This procedure does not alter the shape of $\mathscr{R}'$, the location of C', and the tensor α associated with $\mathscr{R}'$. Denote the common tangent point to the surfaces of $\mathscr{R}$ and $\mathscr{R}'$ by P as shown. Now consider the "difference problem" whose charge density is

$$\rho''(\mathbf{r}) = \rho'(\mathbf{r}) - \rho(\mathbf{r}) = \begin{cases} 1 & \{\mathbf{r}\in(\mathscr{R}'-\mathscr{R})\}, \\ 0 & \{\mathbf{r}\notin(\mathscr{R}'-\mathscr{R})\}, \end{cases} \tag{21}$$

since $\mathscr{R}\subseteq\mathscr{R}'$. By superposition the associated electric field, $\mathscr{E}'' = \mathscr{E}' - \mathscr{E}$, must vanish for all $\mathbf{r}\in\mathscr{R}$. In particular, we have $\mathscr{E}'' = 0$ at P. But this is a contradiction (unless $\rho'' = 0$ everywhere) because all the charge is positive and lies on one side of the tangent plane at P. Therefore $\mathscr{R}' - \mathscr{R} = 0$, i.e., the two regions must coincide. Thus choosing $\mathscr{R}'$ to be the

known ellipsoidal solution, we can conclude that $\mathscr{R}$ must be an ellipsoid.

III. CONCLUSION

We have investigated the possible shape of a region $\mathscr{R}$, filled with a homogeneous isotropic dielectric, which has the property of acquiring a uniform polarization **P** under the influence of an arbitrary (previously) uniform electric field $\mathbf{E}_0$. It is well known that ellipsoids have this property. By demonstrating the uniqueness of the shape of $\mathscr{R}$, we have shown that $\mathscr{R}$ must be an ellipsoid.

[1]A. M. Portis, *Electromagnetic Fields: Sources and Media* (Wiley, New York, 1978), pp. 135 and 136.
[2]H. S. Zapolsky, Am. J. Phys. **55**, 77, 1987.
[3]I am indebted to H. S. Zapolsky for pointing out that the sum rule can be obtained in this manner.

On the excitonic-polaron theory in angular variables

G. Strinati[a)]
Institute of Theoretical Physics, Chalmers University of Technology, Göteborg, Sweden

(Received 24 June 1986; accepted for publication 3 December 1986)

The model of an electron bound to an impurity (or to a core hole) and interacting with a boson field is analyzed in terms of bosonic variables *symmetry adapted* to the spherical symmetry about the impurity center. It is shown that the use of this appropriate set of variables considerably simplifies the solution of the variational problem based on a Lee–Low–Pines-type of ansatz for the ground-state variational wave function.

I. INTRODUCTION

Quasiparticle excitations may sometimes be envisaged as built up by coupling a bare electron (or hole) to a boson field. Well-known examples are the polaron[1] and the plasmaron[2] where the "dressing" of the bare excitation is provided by the coupling to (optical) phonons and to plasmons, respectively. In all cases one starts by reducing the problem to the solution of a model Hamiltonian with a *linear coupling* between the bare particle and the boson field, and then proceeds to determine its ground state and lowest excited states by a variational procedure. Methods of solution differ at this stage, as they are based either on the use of the so-called "coherent states"[3] or on the Feynman path integral approach.[4]

In the problem above one can take full advantage of the conservation of the *total* linear momentum since the medium in which the bare particle is embedded is regarded as being homogeneous and isotropic. The presence of an impurity center (such as a localized core hole) spoils, however, the conservation of linear momentum and makes it apparently more difficult to solve the variational problem. In particular, if the additional interaction between the bare particle and the impurity center can sustain bound states, one is interested in determining the binding energy of the bound quasiparticle. Examples are the bound polaron in ionic insulators and (core) excitons in semiconductors.

Since the presence of an impurity center destroys the homogeneity but not the isotropy, the *total angular momentum* about the impurity center is still a conserved quantity and one may make use of this conservation explicitly to simplify the solution of the variational problem. As previous treatments of this problem based on the use of coherent states[5,6] do not seem to have fully exploited the conservation of angular momentum, we provide in this note the mathematical framework based on this conservation law. Possible applications (such as to the physics of the metallic rare earths) and extensions to systems with lower symmetry remain open.

II. ANGULAR VARIABLES FOR THE BOSON FIELD

For the sake of definiteness, we consider the problem of an electron and a core hole (or impurity) localized at $\mathbf{r} = 0$, which interact among themselves through a *spherically symmetric* bare potential $v_b(r)$. Both particles are also coupled to a plasmon field which provides the screening to $v_b(r)$. The total Hamiltonian can be written as[2,7,8] ($\hbar = 1$)

$$\hat{H} = \frac{\mathbf{p}^2}{2m^*} + v_b(r) + \sum_{\mathbf{k}} \omega_{|\mathbf{k}|} \hat{a}_{\mathbf{k}}^\dagger \hat{a}_{\mathbf{k}} + \sum_{\mathbf{k}} \left(W_{\mathbf{k}}(\mathbf{r}) \hat{a}_{\mathbf{k}} + W_{\mathbf{k}}^*(\mathbf{r}) \hat{a}_{\mathbf{k}}^\dagger \right) . \tag{2.1}$$

Here $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^\dagger$ are annihilation and creation operators for plasmons with wave vector $\mathbf{k}$ that satisfy the usual commutation relation

$$[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta_{\mathbf{k},\mathbf{k}'}; \tag{2.2}$$

$\omega_{|\mathbf{k}|}$ is the plasmon dispersion relation that we may conveniently take of the form[8]

$$\omega_{|\mathbf{k}|} = \omega_p + \mathbf{k}^2/2m, \tag{2.3}$$

where ω_p is the plasma frequency and m is the (bare) electronic mass; $W_{\mathbf{k}}(\mathbf{r})$ is the coupling between the electron (e) and the core hole (h) with the plasmons

$$W_{\mathbf{k}}(\mathbf{r}) = V_{\mathbf{k}}^e(\mathbf{r}) + V_{|\mathbf{k}|}^h, \tag{2.4}$$

where

$$V_{\mathbf{k}}^e(\mathbf{r}) = \overline{V}_{|\mathbf{k}|}^e e^{i\mathbf{k}\cdot\mathbf{r}},$$
$$\overline{V}_{\mathbf{k}}^e = [v(\mathbf{k})\omega_p^2/2\Omega\omega_{|\mathbf{k}|}]^{1/2}, \tag{2.5}$$

and

$$V_{|\mathbf{k}|}^h = -S_c(|\mathbf{k}|)\overline{V}_{|\mathbf{k}|}^e, \tag{2.6}$$

$v(\mathbf{k}) = 4\pi e^2/k^2$ being the Fourier transform of the Coulomb potential, Ω the (quantization) volume, and $S_c(|\mathbf{k}|)$ the structure factor of the core hole. In Eq. (2.1) the effective electronic mass m^* includes possible band-structure effects.

The model Hamiltonian (2.1) treats the excitations of the (metallic) background, wherein the electron and the impurity center are embedded, in terms of "collective variables."[9] Other forms of the coupling (2.5) are possible when the background represents, e.g., electron-hole pair excitations in a semiconductor[10] or optical phonons in an ionic material.[1] In all cases the Hamiltonian can be cast in the form (2.1).

In the absence of the electron, the Hamiltonian that results from (2.1) describes the relaxation about the impurity and can be solved exactly in terms of "displaced" plasmon oscillators.[11] When the sole electron is present, on the other

a) Permanent address: Dipartimento di Fisica, Università degli Studi di Roma "La Sapienza," 00185 Roma, Italy.

 0022-2488/87/040981-05$02.50

hand, the Hamiltonian that results from (2.1) commutes with the total linear momentum operator

$$\hat{\mathscr{P}} = -i\nabla + \sum_{\mathbf{k}} \mathbf{k}\, \hat{a}^{\dagger}_{\mathbf{k}}\, \hat{a}_{\mathbf{k}}, \tag{2.7}$$

because virtual creations or absorptions of plasmons are compensated by the recoil of the electron. This fact is indeed essential for the variational treatment in terms of coherent states[3] in order to obtain the effective mass of the electron due to "dressing" by the boson field, i.e., for obtaining not only the ground state but also the low-lying excited states of the Hamiltonian.

Since the medium about the impurity center is isotropic [cf., Eqs. (2.3) and (2.6)], the Hamiltonian (2.1) should not be affected by rotating the electron *and* the medium simultaneously about the center. To show this invariance explicitly, it is convenient first to rewrite Eq. (2.1) in an alternative form by transforming canonically the boson operators into the angular momentum representation:

$$\begin{aligned} \tilde{\hat{a}}(k\ell m) &= k\left[\frac{\Omega}{(2\pi)^3}\right]^{1/2} \int_{4\pi} d\hat{k}\; Y^{*}_{\ell m}(\hat{k})\hat{a}_{\mathbf{k}}, \\ \tilde{\hat{a}}^{\dagger}(k\ell m) &= k\left[\frac{\Omega}{(2\pi)^3}\right]^{1/2} \int_{4\pi} d\hat{k}\; Y_{\ell m}(\hat{k})\hat{a}^{\dagger}_{\mathbf{k}}, \end{aligned} \tag{2.8}$$

where the tilde signifies that taking the continuum limit is understood before performing the angular integration. One may readily verify that the new operators satisfy the commutation relation

$$[\tilde{\hat{a}}(k\ell m), \tilde{\hat{a}}^{\dagger}(k'\ell' m')] = \delta(k-k')\delta_{\ell\ell'}\delta_{mm'}. \tag{2.9}$$

Expanding the plane wave in Eq. (2.5) into spherical harmonics of $\hat{k}$ and $\hat{r}$ and using the orthonormality of the spherical harmonics, we can then rewrite the Hamiltonian (2.1) as follows:

$$\begin{aligned} \hat{H} = {} & \frac{\mathbf{p}^2}{2m^*} + v_b(r) + \int_0^{\infty} dk\, \omega_k \sum_{\ell m} \tilde{\hat{a}}^{\dagger}(k\ell m)\tilde{\hat{a}}(k\ell m) \\ & + \int_0^{\infty} dk\, \mathscr{V}^{h}_{k0}(\tilde{\hat{a}}(k0) + \tilde{\hat{a}}^{\dagger}(k0)) \\ & + \int_0^{\infty} dk \sum_{\ell m} (\mathscr{V}^{e}_{k\ell m}(\mathbf{r})\tilde{\hat{a}}(k\ell m) \\ & + \mathscr{V}^{e}_{k\ell m}(\mathbf{r})^{*}\tilde{\hat{a}}^{\dagger}(k\ell m)), \end{aligned} \tag{2.10}$$

where we have introduced the notation

$$\begin{aligned} \mathscr{V}^{e}_{k\ell m}(\mathbf{r}) &= \mathscr{V}^{e}_{k\ell}(r)\, Y_{lm}(\hat{r}), \\ \mathscr{V}^{e}_{k\ell}(r) &= 4\pi k\left[\frac{\Omega}{(2\pi)^3}\right]^{1/2} \overline{V}^{e}_{|\mathbf{k}|}\, i^{l} j_l(kr), \end{aligned} \tag{2.11}$$

and

$$\mathscr{V}^{h}_{k0} = 4\pi k\, [\Omega/(2\pi)^3]^{1/2}\, V^{h}_{|\mathbf{k}|}\, Y_{00}(\hat{r}), \tag{2.12}$$

$j_l(kr)$ being the spherical Bessel function of order l and i the imaginary unit. Notice that Eq. (2.10) allows one to analyze the plasmon relaxation into *multipole components*. By the same token, one sees from Eq. (2.10) that only the "monopole" relaxation couples to the spherically symmetric core hole.

Simultaneous rotations of the electron and the plasmon background can now be achieved in terms of the total angular momentum operators

$$\begin{aligned} \hat{L}_{+} &= \ell_{+} + \int_0^{\infty} dk \sum_{lm} [\ell(\ell+1) - m(m+1)]^{1/2} \\ &\quad \times \tilde{\hat{a}}^{\dagger}(k\ell m+1)\tilde{\hat{a}}(k\ell m), \\ \hat{L}_{-} &= \ell_{-} + \int_0^{\infty} dk \sum_{\ell m} [\ell(\ell+1) - m(m-1)]^{1/2} \\ &\quad \times \tilde{\hat{a}}^{\dagger}(k\ell m-1)\tilde{\hat{a}}(k\ell m), \\ \hat{L}_{z} &= \ell_{z} + \int_0^{\infty} dk \sum_{\ell m} m\tilde{\hat{a}}^{\dagger}(k\ell m)\tilde{\hat{a}}(k\ell m), \end{aligned} \tag{2.13}$$

where $\ell = -i\mathbf{r}\times\nabla$ is the angular momentum operator for the electron and $\ell_{\pm} = \ell_x \pm i\ell_y$. It may be verified that the operators (2.13) correctly satisfy the usual commutation relations ($\hbar = 1$)

$$\begin{aligned} &[\hat{L}_z, \hat{L}_{+}] = \hat{L}_{+}, \quad [\hat{L}_z, \hat{L}_{-}] = -\hat{L}_{-}, \\ &[\hat{L}_{+}, \hat{L}_{-}] = 2\hat{L}_z, \end{aligned} \tag{2.14}$$

which justifies identifying them as angular momentum operators in the first place. It may also be verified that *the Hamiltonian (2.10) commutes with the operators (2.13)*, the angular momentum quanta absorbed or released by the plasmon field being exactly compensated by the recoil of the electron. How this conservation law can be exploited to simplify the solution to the variational problem will be considered in the next section.

III. SOLUTION TO THE VARIATIONAL PROBLEM IN ANGULAR VARIABLES

We look for a variational solution to the model Hamiltonian (2.10) based on the *ansatz*

$$|\Phi(\mathbf{r})\rangle_{LM} = \exp\{\tilde{\hat{S}}(\mathbf{r})\}|0\rangle\psi_{LM}(\mathbf{r}). \tag{3.1}$$

Here $|0\rangle$ denotes the ground state of the free-boson field, $\psi_{LM}(\mathbf{r})$ is a single-particle wave function to be specified below, and $\tilde{\hat{S}}(\mathbf{r})$ is taken to be linear in the bosons operators in analogy with the exact solution to the core-hole problem.[11] The symmetry labels LM for the rotation group are common to both sides of Eq. (3.1) since the operator $\tilde{\hat{S}}(\mathbf{r})$ is assumed to be an *invariant* of that group, namely, to satisfy the condition

$$[\hat{L}_{\alpha}, \tilde{\hat{S}}(\mathbf{r})] = 0, \tag{3.2}$$

where the label α distinguishes the three operators (2.13). The ansatz (3.1) is then understood to hold for the lowest state of each symmetry. It may be verified that the (anti-Hermitian) choice

$$\tilde{\hat{S}}(\mathbf{r}) = \int_0^{\infty} dk \sum_{\ell m} (\tilde{f}_{k\ell m}(\mathbf{r})\tilde{\hat{a}}(k\ell m) - \tilde{f}^{*}_{k\ell m}(\mathbf{r})\tilde{\hat{a}}^{\dagger}(k\ell m)) \tag{3.3}$$

satisfies the requirement (3.2) *provided* the functions $\tilde{f}_{k\ell m}(\mathbf{r})$ belong to the symmetry species ℓm of the rotation group:

$$\tilde{f}_{k\ell m}(\mathbf{r}) = \tilde{f}_{k\ell}(r)\, Y_{\ell m}(\hat{r}). \tag{3.4}$$

By minimizing the expectation value of the Hamiltonian (2.10) for the state (3.1) one obtains a set of coupled differ-

ential equations for the unknown functions $f_{k\ell m}(\mathbf{r})$ and $\psi_{LM}(\mathbf{r})$ which need to be solved self-consistently:

$$-\frac{1}{2m^*}\nabla^2 f_{k\ell m}(\mathbf{r})+\omega_k f_{k\ell m}(\mathbf{r})$$
$$-\frac{1}{m^*}\nabla f_{k\ell m}(\mathbf{r})\cdot\frac{\nabla\psi^*_{LM}(\mathbf{r})}{\psi^*_{LM}(\mathbf{r})}+i\mathbf{j}(\mathbf{r})\cdot\nabla f_{k\ell m}(\mathbf{r})$$
$$+\frac{i}{2}f_{k\ell m}(\mathbf{r})\frac{\nabla\cdot(|\psi_{LM}(\mathbf{r})|^2\mathbf{j}(\mathbf{r}))}{|\psi_{LM}(\mathbf{r})|^2}+\frac{1}{4m^*}f_{k\ell m}(\mathbf{r})$$
$$\times\frac{\nabla\cdot(\psi^*_{LM}(\mathbf{r})\nabla\psi_{LM}(\mathbf{r})-\psi_{LM}(\mathbf{r})\nabla\psi^*_{LM}(\mathbf{r}))}{|\psi_{LM}(\mathbf{r})|^2}$$
$$=\mathscr{V}^e_{k\ell m}(\mathbf{r})+\mathscr{V}^h_{k0}\delta_{l,0}, \tag{3.5a}$$

and

$$[(1/2m^*)(-i\nabla+m^*\mathbf{j}(\mathbf{r}))^2+v_{\text{eff}}(r)-E]\psi_{LM}(\mathbf{r})=0. \tag{3.5b}$$

In Eqs. (3.5) $\mathbf{j}(\mathbf{r})$ has the form of a "plasmon current"

$$\mathbf{j}(\mathbf{r})=\frac{1}{2im^*}\int_0^\infty dk\sum_{\ell m}(f_{k\ell m}(\mathbf{r})\nabla f^*_{k\ell m}(\mathbf{r})$$
$$-f^*_{k\ell m}(\mathbf{r})\nabla f_{k\ell m}(\mathbf{r})), \tag{3.6}$$

while $v_{\text{eff}}(r)$ indicates the (spherically symmetric) *effective potential* seen by the electron

$$v_{\text{eff}}(r)=v_b(r)+\frac{1}{2m^*}\int_0^\infty dk\sum_{\ell m}\nabla f^*_{k\ell m}(r)\cdot\nabla f_{k\ell m}(\mathbf{r})$$
$$+\int_0^\infty dk\,\omega_k\sum_{\ell m}f^*_{k\ell m}(\mathbf{r})f_{k\ell m}(\mathbf{r})$$
$$-\int_0^\infty dk\sum_{\ell m}[(\mathscr{V}^e_{k\ell m}(\mathbf{r})+\mathscr{V}^h_{k0}\delta_{l,0})f^*_{k\ell m}(\mathbf{r})$$
$$+(\mathscr{V}^e_{k\ell m}(\mathbf{r})+\mathscr{V}^h_{k0}\delta_{l,0})^*f_{k\ell m}(\mathbf{r})]. \tag{3.7}$$

Notice that Eq. (3.5a) is inhomogeneous owing to the presence of the coupling terms at its right-hand side. We shall specifically be concerned with finding a particular solution to this equation because we require on physical grounds the functions $f_{k\ell m}(\mathbf{r})$ to vanish if the coupling terms are allowed to vanish. Equation (3.5b), on the other hand, has the form of an ordinary single-particle Schrödinger equation where the energy parameter E originates from the subsidiary normalization condition of the trial eigenfunction (3.1).

The structure of Eq. (3.5a) can be simplified as follows. We notice at the outset that the last term at its left-hand side actually vanishes if we take

$$\psi_{LM}(\mathbf{r})=R_L(r)Y_{LM}(\hat{r}) \tag{3.8}$$

with $R_L(r)$ real. The constant hole coupling term at the right-hand side of Eq. (3.5a) may also be eliminated by setting

$$f_{k\ell m}(\mathbf{r})=f^e_{k\ell m}(\mathbf{r})+(\mathscr{V}^h_{k0}/\omega_k)\delta_{l,0}, \tag{3.9}$$

whereby Eq. (3.5a) becomes an equation for the electronic part $f^e_{k\ell m}(\mathbf{r})$ only. The resulting equation is still quite complicated to solve, being apparently nonlinear in the set of functions $f^e_{k\ell m}(\mathbf{r})$ owing to the presence of the current (3.6). However, by exploiting the geometrical as well the dynamical symmetry of the problem at hand we can prove that both the transverse and the radial components of $\mathbf{j}(\mathbf{r})$ vanish identically, i.e.,

$$\mathbf{j}(\mathbf{r})=0. \tag{3.10}$$

The vanishing of the transverse component of $\mathbf{j}(\mathbf{r})$ follows from the identity

$$\sum_{m=-\ell}^{\ell}Y_{\ell m}(\hat{r})\frac{\partial}{\partial\theta}Y^*_{\ell m}(\hat{r})$$
$$=\sum_{m=-\ell}^{\ell}Y_{\ell m}(\hat{r})\frac{\partial}{\partial\varphi}Y^*_{\ell m}(\hat{r})=0, \tag{3.11}$$

where θ and φ are the spherical angles. Equation (3.11), in turn, follows from the property of the spherical harmonics under time reversal and from the addition theorem they satisfy. On the other hand, the radial component of $\mathbf{j}(\mathbf{r})$

$$\hat{r}\cdot\mathbf{j}(\mathbf{r})=\frac{1}{2im^*}\int_0^\infty dk\sum_\ell\frac{(2\ell+1)}{4\pi}\left(f_{k\ell}(r)\frac{d}{dr}f^*_{k\ell}(r)\right.$$
$$\left.-f^*_{k\ell}(r)\frac{d}{dr}f_{k\ell}(r)\right) \tag{3.12}$$

also vanishes *provided* the radial functions $f_{k\ell}(r)$ can be taken to be either real or purely imaginary. Assuming this to be true, we can verify that the solutions of the resulting equation are consistent with this assumption. In fact, making use of Eq. (3.10) and averaging over M we obtain a *radial* equation for $f^e_{k\ell}(r)$ (Ref. 12):

$$-\frac{1}{2m^*}\left[\frac{1}{r}\frac{d^2}{dr^2}(rf^e_{k\ell}(r))-\frac{\ell(\ell+1)}{r^2}f^e_{k\ell}(r)\right]$$
$$+\omega_k f^e_{k\ell}(r)-\frac{1}{m^*}\frac{1}{R_L(r)}\frac{dR_L(r)}{dr}\frac{d}{dr}f^e_{k\ell}(r)$$
$$=\mathscr{V}^e_{k\ell}(r), \tag{3.13}$$

whose solutions are either real or purely imaginary for even or odd values of ℓ, respectively, reflecting a similar property of the source term [cf. Eq. (2.11b)].

The following features can be inferred from Eq. (3.13).

(i) Equation (3.13) is *linear* in $f^e_{k\ell}(r)$ and thus it possesses only one solution for all values of the parameters m^* and $\bar{V}^e_k$ (Ref. 13).

(ii) Solutions for increasing $\ell\,(>0)$ are expected to be progressively suppressed in the region about the core hole and thus they will not contribute appreciably to the relevant portion of the effective potential (3.7) if $R_L(r)$ is sufficiently localized.

(iii) In the limit of large m^* values there is a cancellation between the plasmon relaxations due to the core hole and to the electron.

The properties of the spherical harmonics can also be used to rewrite the effective potential (3.7) in an explicitly invariant form, as shown in the Appendix.

IV. CLOSE-FORM SOLUTION FOR A PARTICULAR CLASS OF SINGLE-PARTICLE ORBITALS

We have shown above that symmetry arguments allow us to reduce the coupled set of equations (3.5) to the simpler radial equations (3.13) and (A4). Solution to these equations has still to be tackled numerically unless the orbital

$R_L(r)$ is restricted, e.g., to a simple hydrogenic form when v_b is a Coulombic potential:

$$R_L(r) = A_L\, r^L e^{-\gamma r}, \tag{4.1}$$

A_L being a normalization constant. In this case a solution to Eq. (3.13) can be obtained in a closed form and it will depend parametrically on γ. To this end, we follow Ref. 6 and express the solution to Eq. (3.13) in terms of the associated Green's function

$$\underset{\sim}{f}^e(r) = \int_0^\infty dr'\, \mathcal{G}(r,r')\, \underset{\sim}{\mathcal{V}}^e(r'). \tag{4.2}$$

(The dependence of these functions on k and ℓ is understood throughout.) The Green's function, in turn, can be expressed in terms of the regular and irregular solutions to the homogeneous equation associated with Eq. (3.13)[14]:

$$\mathcal{G}(r,r') = -\frac{1}{W(r')} \begin{cases} \mathcal{R}(r)\mathcal{J}(r') & (r<r'), \\ \mathcal{J}(r)\mathcal{R}(r') & (r>r'), \end{cases} \tag{4.3}$$

where $W(r)$ is the Wronskian

$$W(r) = \mathcal{J}(r)\frac{d\mathcal{R}(r)}{dr} - \mathcal{R}(r)\frac{d\mathcal{J}(r)}{dr}. \tag{4.4}$$

Suitable boundary conditions require the Green's function to remain bounded everywhere.

The solutions to the homogeneous equation associated with Eq. (3.13) whereby $R_L(r)$ is taken of the form (4.1) can be readily obtained by setting

$$F(r) = r^\eta e^{-\epsilon r/2} w(\zeta r), \tag{4.5}$$

where F stands for either $\mathcal{R}$ or $\mathcal{J}$ and (ϵ,η,ζ) are constants to be determined. The transformed equation for w then reads:

$$\begin{aligned} &\frac{d^2}{dr^2} w(\zeta r) + \left[\frac{2(L+\eta+1)}{r} - 2\gamma - \epsilon\right]\frac{d}{dr} w(\zeta r) \\ &\quad + \{[\eta(\eta-1) + 2\eta(L+1) - \ell(\ell+1)](1/r^2) \\ &\quad - [2\gamma\eta + \eta\epsilon + \epsilon(L+1)](1/r) \\ &\quad + \epsilon^2/4 + \epsilon\gamma - 2m^*\omega_k\} w(\zeta r) = 0, \end{aligned} \tag{4.6}$$

which reduces to the standard Kummer's equation[15]

$$\frac{d^2}{dr^2} w(\zeta r) + \left(\frac{b}{r} - \zeta\right)\frac{d}{dr} w(\zeta r) - \frac{a\zeta}{r} w(\zeta r) = 0, \tag{4.7}$$

provided we identify

$$\begin{aligned} &b = 2(L+\eta+1), \quad \zeta = \epsilon + 2\gamma, \\ &\eta^2 + \eta(2L+1) = \ell(\ell+1), \\ &a\zeta = 2\gamma\eta + \eta\epsilon + \epsilon(L+1), \\ &\epsilon^2 + 4\epsilon\gamma - 8m^*\omega_k = 0. \end{aligned} \tag{4.8}$$

The five unknowns here are thus determined uniquely by solving the quadratic equations for ϵ and η:

$$\begin{aligned} &\epsilon = 2[(\gamma^2 + 2m^*\omega_k)^{1/2} - \gamma], \\ &\eta = \tfrac{1}{2}\,[(2L+1)^2 + 4\ell(\ell+1))^{1/2} - (2L+1)], \end{aligned} \tag{4.9}$$

where the boundary conditions have been used to fix the sign of the square roots. Notice that every unknown turns out to be positive definite.

The Green's function (4.3) can now be written in terms of the regular (Φ) and irregular (Ψ) Kummer's functions:

$$\begin{aligned} &\mathcal{G}(r,r') \\ &\quad = 2m^*[\Gamma(a)/\Gamma(b)]\, \zeta^{b-1} r^\eta r'^{b-\eta} e^{-2\gamma r'} e^{-(\epsilon/2)(r+r')} \\ &\qquad \times \begin{cases} \Phi(a,b,\zeta r)\Psi(a,b,\zeta r') & (r<r'), \\ \Psi(a,b,\zeta r)\Phi(a,b,\zeta r') & (r>r'), \end{cases} \end{aligned} \tag{4.10}$$

where Γ is Euler's gamma function.

The restricted form (4.1) requires us to replace solving the differential equation (A4) for $R_L(r)$ with finding the minimum of the function

$$\begin{aligned} E(\gamma) = A_L^2 \int_0^\infty &dr\, e^{-2\gamma r} r^{2L+1} \\ &\times \left[\frac{2(L+1)\gamma - \gamma^2 r}{2m^*} + v_{\text{eff}}(r) r\right] \end{aligned} \tag{4.11}$$

which determines the parameter γ self-consistently together with the screening functions (4.2).

ACKNOWLEDGMENTS

The author is indebted to G. Wendin for the hospitality at the Institute of Theoretical Physics of the Chalmers University of Technology (Göteborg), as well for his interest in this work. Discussions with W. Bardyszewski during the initial stage of this work are also gratefully acknowledged.

APPENDIX: EFFECTIVE POTENTIAL FOR THE RADIAL EQUATION

The effective potential (3.7) can be rewritten in terms of the radial functions $\underset{\sim}{f}_{k\ell}(r)$ alone. For the last two terms at the right-hand side of Eq. (3.7) this can be simply achieved by using the addition theorem for spherical harmonics. For the second term there, one has to express the gradient operators in spherical coordinates and to make use of the identity

$$\begin{aligned} &\sum_{m=-\ell}^{\ell} \left[\frac{\partial Y_{\ell m}(\hat{r})}{\partial\theta}\frac{\partial Y^*_{\ell m}(\hat{r})}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial Y_{\ell m}(\hat{r})}{\partial\varphi}\frac{\partial Y^*_{\ell m}(\hat{r})}{\partial\varphi}\right] \\ &\quad = [(2\ell+1)/4\pi]\ell(\ell+1), \end{aligned} \tag{A1}$$

which follows from Eq. (3.11) after straightforward manipulations. The result is

$$\begin{aligned} &\sum_{\ell m} \nabla \underset{\sim}{f}^*_{k\ell m}(\mathbf{r})\cdot\nabla \underset{\sim}{f}_{k\ell m}(\mathbf{r}) \\ &\quad = \sum_\ell \frac{(2\ell+1)}{4\pi}\left[\left|\frac{d\underset{\sim}{f}_{k\ell}(r)}{dr}\right|^2 + \frac{\ell(\ell+1)}{r^2}|\underset{\sim}{f}_{k\ell}(r)|^2\right], \end{aligned} \tag{A2}$$

so that:

$$\begin{aligned} v_{\text{eff}}(r) = v_b(r) &- \frac{1}{\pi^{1/2}}\int_0^\infty dk\, \mathcal{V}^h_{k0}\underset{\sim}{f}_{k0}(r) \\ &+ \sum_\ell \frac{(2\ell+1)}{4\pi}\int_0^\infty dk \left\{\frac{1}{2m^*}\left|\frac{d\underset{\sim}{f}_{k\ell}(r)}{dr}\right|^2\right. \\ &+ \left[\frac{\ell(\ell+1)}{2m^* r^2} + \omega_k\right]|\underset{\sim}{f}_{k\ell}(r)|^2 \\ &\left. - 2\,\underset{\sim}{\mathcal{V}}^e_{k\ell}(r)\underset{\sim}{f}^*_{k\ell}(r)\right\}, \end{aligned} \tag{A3}$$

where the contribution of the hole to $f_{k\ell}(r)$ is present only for $\ell = 0$ [cf. Eq. (3.9)]. Taking Eqs. (3.8) and (3.10) into account, the effective Schrödinger equation (3.5b) becomes eventually:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR_L(r)}{dr}\right) + \left\{2m^*[E - v_{\text{eff}}(r)] - \frac{L(L+1)}{r^2}\right\} R_L(r) = 0. \tag{A4}$$

[1]H. Fröhlich, in *Polarons and Excitons*, edited by C. G. Kuper and G. D. Whitfield (Oliver and Boyd, Edinburgh, 1963), p. 1.
[2]B. I. Lundqvist, Phys. Kondens. Mater. **6**, 193 (1967).
[3]T. D. Lee, F. E. Low, and D. Pines, Phys. Rev. **90**, 297 (1953); D. Pines, in Ref. 1, p. 33.
[4]See, e.g., L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981), Chap. 21, and references therein.
[5]A. Quattropani, F. Bassani, G. Margaritondo, and G. Tinivella, Nuovo Cimento B **51**, 335 (1979).
[6]G. Iadonisi and F. Bassani, Nuovo Cimento D **2**, 1541 (1983).
[7]A. W. Overhauser, Phys. Rev. B **3**, 1888 (1971).
[8]C. O. Almbladh and L. Hedin, in *Handbook of Synchrotron Radiation*, edited by E. E. Koch (North-Holland, Amsterdam, 1983), Vol. 1, p. 611.
[9]D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953).
[10]Y. Toyozawa, Prog. Theor. Phys. **12**, 421 (1954).
[11]D. C. Langreth, Phys. Rev. B **1**, 471 (1970).
[12]This averaging process is familiar in the central-field approximation of atomic physics. Compare, e.g., E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U. P., Cambridge, 1970), Sec. 14-8.
[13]This property holds also for the ground state of the free polaron corresponding to zero eigenvalue of the linear momentum (2.7) (cf. Ref. 3).
[14]Compare, e.g., E. A. Kraut, *Fundamentals of Mathematical Physics* (McGraw-Hill, New York, 1967), Sec. 6-22.
[15]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Natl. Bur. Stand., Washington, D. C., 1970), Chap. 13.